

# REVUE DE STATISTIQUE APPLIQUÉE

ALAIN BACCINI

ANTOINE DE FALGUEROLLES

EL MOSTAFA QANNARI

## **Étude de quelques indices de concentration : un essai de présentation unifiée**

*Revue de statistique appliquée*, tome 34, n° 1 (1986), p. 31-44

[http://www.numdam.org/item?id=RSA\\_1986\\_\\_34\\_1\\_31\\_0](http://www.numdam.org/item?id=RSA_1986__34_1_31_0)

© Société française de statistique, 1986, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

# ÉTUDE DE QUELQUES INDICES DE CONCENTRATION : UN ESSAI DE PRÉSENTATION UNIFIÉE

Alain BACCINI, Antoine de FALGUEROLLES, El Mostafa QANNARI

*Laboratoire de Statistique et Probabilités*

*U.A. — C.N.R.S. 745*

*Université Paul Sabatier*

*118, route de Narbonne*

*31062 Toulouse Cedex*

---

## RÉSUMÉ

Dans cet article, on établit que les indices de concentration les plus souvent utilisés, lorsque l'on étudie l'inégalité des revenus, du patrimoine..., peuvent être obtenus à partir de mesures classiques de dissemblance entre lois de probabilités. On obtient ainsi une présentation simple et unifiée de ces indices.

## SUMMARY

In this article, we consider the concentration indices which are commonly used when studying inequalities of income, wealth, ... It is natural, in this context, to examine various measures of dissimilarity between probability distributions. A simple presentation can in this way readily be found, covering most of the concentration indices in a unified manner.

## INTRODUCTION

L'évaluation statistique des inégalités économiques a suscité depuis longtemps de nombreuses études tant descriptives que méthodologiques. Aussi, actuellement, la plupart des manuels présentent des méthodes usuelles d'étude de la concentration de variables telles que revenu, fortune, ... . Courbe de Lorenz et indice de Gini y sont principalement traités.

Il nous a semblé intéressant de donner ici un exposé synthétique de quelques indices usuels de concentration. Motivée par une réflexion pédagogique menée par les auteurs dans le cadre d'enseignements en premier cycle universitaire et en I.U.T., utilisant de façon naturelle la notion de dissemblance entre lois de probabilités, la démarche adoptée permet de dégager des analogies profondes entre ces indices et de rappeler des propriétés rarement rassemblées.

D'autres aspects de la concentration ne sont pas abordés ici (dans le domaine économique : transfert, désagrégation, pauvreté, interprétation en terme d'utilité...; dans le domaine écologique : lien avec les notions de diversité...). Pour une présentation très complète des mesures d'inégalité en

économie, le lecteur se reportera au cours de Ch. GOURIEROUX [5]. Il trouvera un exposé très général de la notion de diversité, de sa mesure et de ses liens avec la notion de concentration dans l'article de G.P. PATIL et C. TAILLIE [9]. Il convient aussi de citer le travail récent de M. BARBUT consacré à l'étude des propriétés des fonctions de concentration de C. GINI [2].

## 1. PRÉSENTATION DE LA NOTION DE CONCENTRATION

$X$  est une variable statistique discrète à valeurs positives (revenu, fortune, ...) observée sur une population  $\Omega$  de taille finie; à chaque unité statistique  $\omega$  est associée la valeur  $X(\omega)$ . On suppose que

$$\sum_{\omega \in \Omega} X(\omega)$$

a un sens concret; cette quantité est appelée « masse » totale des valeurs de la variable  $X$ .

On étudie les observations de  $X$  sous deux systèmes de pondérations des individus :

- à chaque individu  $\omega$  est associé le poids  $1/|\Omega|$ ,
- à chaque individu  $\omega$  est associé un poids égal à sa part de la masse totale, soit

$$X(\omega) / \sum_{\omega \in \Omega} X(\omega).$$

On peut remarquer que cette dernière pondération, comme la précédente, ne dépend pas dans le cas général des unités de mesure du caractère  $X$  (elle est invariante par homothétie).

On notera également que le point de vue adopté ici est aussi celui retenu par P.E. HART [6] qui étudie les pondérations définies par  $X$ ,  $X^2$ , ...

Les deux systèmes de pondérations considérés définissent les deux séries statistiques suivantes :

- la série  $\{(x_k, f_k) | k = 1, 2, \dots, K\}$ , où  $f_k$  désigne la fréquence avec laquelle a été observée la valeur  $x_k$ ,
- la série  $\{(x_k, g_k) | k = 1, 2, \dots, K\}$ , où  $g_k$  désigne la part relative de la masse totale détenue par les individus qui présentent la valeur  $x_k$  :

$$g_k = \frac{|\Omega| f_k x_k}{\sum_{\omega \in \Omega} X(\omega)} = \frac{f_k x_k}{\sum_{k=1}^K f_k x_k}.$$

Ainsi, on associe à la même variable  $X$  deux systèmes de fréquences  $f$  et  $g$ , donc deux fonctions de répartition que l'on notera respectivement  $F$  et  $G$ . L'étude de la concentration consiste alors à comparer ces deux lois empiriques.

*Remarque :*

On a considéré ici l'équipondération des individus par souci de simplification. Il est facile de voir que tout se passe de façon analogue si chaque individu est muni d'un poids  $p(\omega)$  donné a priori.

## 2. NOTATIONS; RAPPEL DE QUELQUES PROPRIÉTÉS

Considérant la variable  $X$  munie de la fonction de répartition  $F$ , on adopte les notations suivantes :

- *moment empirique d'ordre  $h$*  ( $h \in \mathbf{N}$ ) :  $E_F(X^h) = \sum_{k=1}^K f_k(x_k)^h$ ;
- *indices de tendance centrale* :
  - moyenne arithmétique :  $E_F(X)$ ;
  - médiane :  $M_F(X)$ ;
- *indices de dispersion* :
  - variance :  $V_F = E_F[X - E_F(X)]^2$  ;
  - écart-type :  $\sigma_F = \sqrt{V_F}$  ;
  - moyenne des écarts absolus à la moyenne (resp. à la médiane) :  
 $\delta_F = E_F[|X - E_F(X)|]$  (resp.  $E_F[|X - M_F(X)|]$ );
  - différence moyenne de Gini :  
$$\Delta_F = \sum_{k=1}^K \sum_{\ell=1}^K f_k f_\ell |x_k - x_\ell| .$$

### *Remarque*

La différence moyenne de Gini peut être introduite en considérant la variable  $X^*$ , définie sur  $\Omega \times \Omega$ , qui à tout couple  $(\omega, \omega')$  associe l'écart absolu  $|X(\omega) - X(\omega')|$ ;  $X^*$  a pour moyenne  $\Delta_F$  et pour variance  $2 V_F - \Delta_F^2$  (ceci donne la relation bien connue  $\Delta_F \leq \sqrt{2} \sigma_F$ , l'égalité ayant lieu si et seulement si  $X$  est constante).

On adopte des notations analogues pour  $X$  muni de la fonction de répartition  $G$ ; il est bien connu que

$$\forall x \in \mathbf{R}, G(x) \leq F(x).$$

Par suite, la médiane de  $X$  pour  $G$  (connue sous le nom de médiale) est supérieure à la médiane de  $X$  pour  $F$ .

De plus, on peut montrer (voir en 4.2.b) la relation suivante entre les moments empiriques de  $X$  pour  $F$  et pour  $G$  :

$$E_G(X^h) = \frac{E_F(X^{h+1})}{E_F(X)}, h \in \mathbf{N} .$$

## 3. ÉTUDE SYNTHÉTIQUE DE LA CONCENTRATION

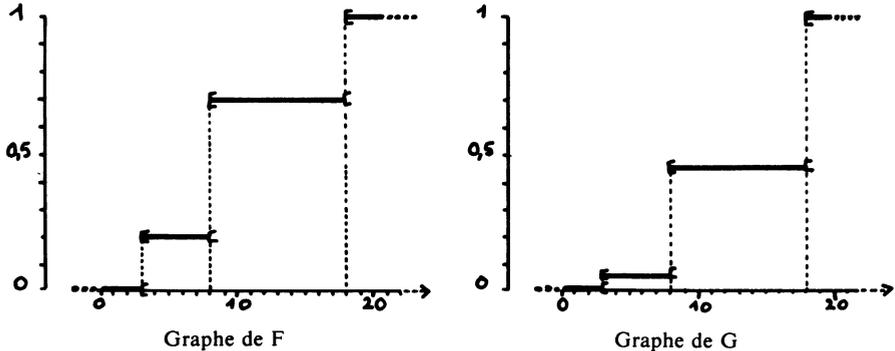
L'étude de la concentration peut être abordée en comparant soit les fonctions de répartition  $F$  et  $G$  (paragraphe 3.1 ci-après) soit les systèmes de fréquences  $f$  et  $g$  (paragraphe 3.2 ci-après).

Les égalités et propriétés formulées se démontrent simplement dans le cas discret considéré dans ce paragraphe; ces démonstrations sont donc omises;

elles seront néanmoins données au paragraphe 4, dans un cadre plus général, notamment à la lumière de la présentation synthétique dégagée en 3.2.e.

### 3.1. Comparaison des fonctions de répartition F et G

Soit donc à comparer les deux fonctions de répartition F et G. A titre d'illustration considérons l'exemple suivant (il sera repris en 3.1.c) :



#### 3.1.a. Indice associé à l'aire comprise entre F et G

$$\text{On considère } A = \int_0^{+\infty} |F(x) - G(x)| dx.$$

On voit que A dépend des unités de mesure de X; il est donc préférable, pour éliminer l'incidence de ces unités, d'étudier la concentration de  $\tilde{X} = X/E_F(X)$ ; on est ainsi amené à définir l'indice :

$$\tilde{A} = \int_0^{+\infty} |\tilde{F}(x) - \tilde{G}(x)| dx.$$

On peut alors démontrer (voir en 4.3.a) que  $\tilde{A}$  est le carré du *coefficient de variation* de X pour F :

$$\tilde{A} = \frac{V_F}{[E_F(X)]^2} = \left[ \frac{\sigma_F}{E_F(X)} \right]^2. \quad (1)$$

#### Remarque

L'indice de HERFINDAHL (voir [5]) n'est autre, au coefficient 1/n près, que le carré du coefficient de variation plus 1, c'est-à-dire le moment empirique d'ordre 2 de  $\tilde{X}$ .

Compte tenu de la relation existant entre les moments empiriques de X pour F et G, il vient encore :

$$\tilde{A} = \frac{E_G(X) - E_F(X)}{E_F(X)}. \quad (2)$$

### 3.1.b. Indice associé à la distance de KOLMOGOROV

On considère  $S = \sup_{x \in \mathbb{R}} |F(x) - G(x)|$  ; on peut démontrer (voir en 4.3.b) que

$$S = \frac{\delta_F}{2E_F(X)}. \quad (3)$$

Cet indice est connu sous le nom de *pourcentage maximum d'égalisation* [8], ou encore d'*écart relatif moyen* [5].

Il vient encore (voir (15)) :

$$S = F[E_F(X)] - G[E_F(X)]. \quad (4)$$

L'interprétation économique de cet indice est immédiate : la concentration est mesurée par la différence entre la proportion d'individus présentant une valeur de la variable inférieure à la moyenne et la proportion de la masse totale qu'ils détiennent.

### 3.1.c. Indice de GINI

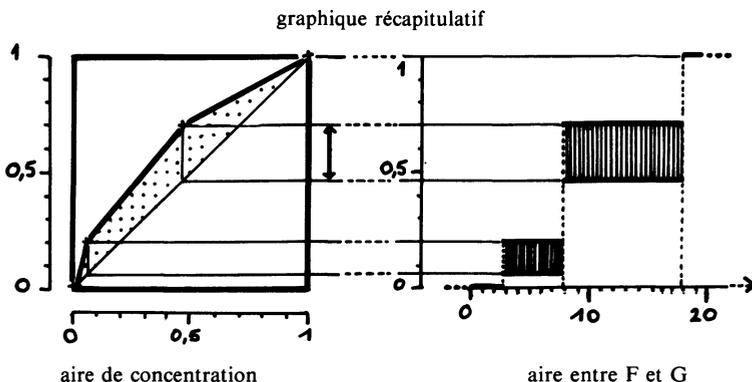
Dans le plan rapporté à un système d'axes orthonormés, on considère les points de coordonnées  $(G(x_k), F(x_k))$ ,  $k = 1, \dots, K$ , autrement dit le « F-F plot » de G et F [4].

Le contour polygônale joignant ces points est appelé courbe de Lorenz ; l'aire comprise entre la première bissectrice et la courbe de Lorenz est appelée aire de concentration ; l'indice de concentration de GINI,  $I_G$ , vaut deux fois cette aire, soit encore :

$$I_G = \frac{\Delta_F}{2E_F(X)}. \quad (5)$$

Outre cette égalité, on montrera en 4.3.c. que  $I_G$  s'obtient encore en considérant la dissemblance entre F et G définie par :

$$\sum_{k=1}^K f_k |F(x_k) - G(x_k)|. \quad (6)$$



### 3.1.d. Indice associé à la différence entre la médiale et la médiane

Par analogie avec l'expression (2) on peut considérer l'indice  $M_G(X) - M_F(X)$ . Là encore divers schémas de standardisation peuvent être envisagés en considérant, par exemple,  $\tilde{X} = X/E_F(X)$  ou  $\tilde{\tilde{X}} = X/M_F(X)$ ; on obtient alors un analogue de l'indice (2).

#### Remarque

On compare ici des quantiles de F et de G (resp. de  $\tilde{F}$  et de  $\tilde{G}$  ou de  $\tilde{\tilde{F}}$  et de  $\tilde{\tilde{G}}$ ). La représentation simultanée de ces quantiles est ancienne et connue sous le nom de « Q-Q plot » de LORENZ.

## 3.2. Comparaison des systèmes de fréquences f et g

L'idée est d'évaluer la dissemblance entre f et g à l'aide d'indices usuels (voir notamment [3]).

### 3.2.a. Indice associé à l'information de KULLBACK

Il est défini par :

$$I_K = \sum_{k=1}^K g_k \text{Log} \frac{g_k}{f_k}.$$

On a donc :

$$I_K = E_F \left[ \frac{X}{E_F(X)} \text{Log} \left( \frac{X}{E_F(X)} \right) \right]. \quad (7)$$

On vérifie simplement que  $I_K$  peut s'écrire aussi sous la forme :

$$I_K = E_G(\text{Log} X) - \text{Log}[E_F(X)]. \quad (8)$$

Sous cette forme, on reconnaît l'indice de concentration de THEIL (voir [5]).

### 3.2.b. Indice associé à la distance du $\chi^2$

La distance du  $\chi^2$  entre f et g, centrée sur f, donne ici l'indice :

$$I_{\chi^2}(f, g) = \left[ \sum_{k=1}^K \frac{(g_k - f_k)^2}{f_k} \right]^{1/2}.$$

On vérifie alors que

$$I_{\chi^2}(f, g) = \frac{\sigma_F}{E_F(X)}. \quad (9)$$

On retrouve ainsi le coefficient de variation de X.

### 3.2.c. Indice associé à la distance de la convergence uniforme

Considérons

$$I_U(f, g) = \frac{1}{2} \sum_{k=1}^K |g_k - f_k|;$$

on vérifie encore simplement que

$$I_U(f, g) = \frac{\delta_F}{2E_F(X)}, \quad (10)$$

*pourcentage maximum d'égalisation* introduit en 3.1.b.

### 3.2.d. Indice associé à la distance d'HELLINGER

Les indices ci-dessus ont été obtenus en choisissant une mesure de dissemblance entre probabilités; dans cette perspective, un indice reposant sur la distance d'HELLINGER peut encore être considéré :

$$I_H = \sum_{k=1}^K (\sqrt{f_k} - \sqrt{g_k})^2.$$

Il vient :

$$I_H = 2 \left( 1 - \frac{E_F(\sqrt{X})}{\sqrt{E_F(X)}} \right). \quad (11)$$

$I_H$  est compris entre 0 et 2, prend la valeur 0 si  $X$  est constante sur  $\Omega$  et approche la valeur 2 si un individu accapare toute la masse des valeurs (et ceci d'autant plus fortement que grande est la population).

#### Remarque

Il convient de mentionner ici une interprétation géométrique de cet indice. Selon P.G. PATIL et C. TAILLIE [9], elle a été introduite pour la première fois en 1946 par A. BATTACHARYYA. Dans  $\mathbf{R}^K$  rapporté à sa base canonique et muni du produit scalaire usuel, les vecteurs  $(\sqrt{f_1}, \dots, \sqrt{f_K})$  et  $(\sqrt{g_1}, \dots, \sqrt{g_K})$ , de norme 1, forment un angle  $\theta$  dont le carré de la mesure peut être approché (lorsque  $\theta$  est petit) par la valeur de l'indice  $I_H$ . En effet :

$$\begin{aligned} I_H &= \sum_{k=1}^K (\sqrt{f_k} - \sqrt{g_k})^2 = 2 \left( 1 - \sum_{k=1}^K \sqrt{f_k} \sqrt{g_k} \right) = 2 (1 - \cos \theta) \\ &= 4 \sin^2 \frac{\theta}{2} \cong \theta^2. \end{aligned}$$

### 3.2.e. Ecriture synthétique de ces indices

En suivant la présentation de la dissemblance entre mesures de probabilités donnée dans [3], on peut retrouver la plupart des indices rencontrés ci-dessus comme cas particuliers d'un indice plus général.

Soit  $\varphi$  une application de  $\mathbf{R}_+ \times \mathbf{R}_+$  dans  $\mathbf{R}$ , convexe, homogène de degré 1 ( $\forall t \in \mathbf{R}_+, \forall (y, z) \in \mathbf{R}_+ \times \mathbf{R}_+ : \varphi(ty, tz) = t\varphi(y, z)$ ) et telle que  $\varphi(1, 1) = 0$ ; un indice synthétique de dissemblance entre  $f$  et  $g$  est donné par :

$$I_\varphi = \sum_{k=1}^K \varphi(f_k, g_k).$$

L'application  $\varphi$  étant homogène, il vient, en explicitant  $g_k$  :

$$I_\varphi = E_F \left[ \varphi \left( 1, \frac{X}{E_F(X)} \right) \right]. \quad (12)$$

Il est facile de voir que les indices de concentration rencontrés au paragraphe 3.2 s'obtiennent en considérant :

$$\varphi(y, z) = z \operatorname{Log} \frac{z}{y} \quad (\text{indice de THEIL}),$$

$$\varphi(y, z) = \frac{(z - y)^2}{y} \quad (\text{carré du coefficient de variation}),$$

$$\varphi(y, z) = \frac{1}{2} |y - z| \quad (\text{pourcentage maximum d'égalisation}),$$

$$\varphi(y, z) = (\sqrt{y} - \sqrt{z})^2 \quad (\text{indice d'HELLINGER}).$$

L'égalité (12) permet alors de retrouver directement les différents résultats énoncés dans ce 3.2.

*Remarque 1*

On retrouverait aussi, par des choix convenables de  $\varphi$ , d'autres indices usuels de concentration (voir [5]) : l'écart moyen des logarithmes et l'écart quadratique des logarithmes par exemple. Cependant certains indices, l'indice de GINI notamment, ne sont pas de cette forme.

*Remarque 2*

L'inégalité de JENSEN permet de montrer que  $I_\varphi \geq 0$ .

#### 4. CAS GÉNÉRAL; DÉMONSTRATIONS

Il est clair que l'étude précédente peut se généraliser au cas d'une variable aléatoire réelle (v.a.r.) positive, définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ .

C'est dans ce contexte que tous les résultats précédemment rencontrés vont être démontrés.

##### 4.1. Définitions

Soit  $X$  une v.a.r. définie sur  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $(\mathbf{R}_+, \mathbf{B}_{\mathbf{R}_+})$ . Soit  $P_X$  la probabilité image de  $P$  par  $X$  et  $F$  sa fonction de répartition. On suppose que  $X$  admet une espérance mathématique,  $E_F(X)$ , non nulle.

On définit la probabilité de masse,  $Q$ , sur  $(\Omega, \mathcal{A})$  par :

$$\forall A \in \mathcal{A}, Q(A) = \int_A \frac{X}{E_F(X)} dP.$$

De même, on définit la probabilité de masse,  $Q_X$ , sur  $(\mathbf{R}_+, \mathbf{B}_{\mathbf{R}_+})$  comme étant la probabilité image de  $Q$  par  $X$ .

$Q$  est, par construction, absolument continue par rapport à  $P$  et a pour densité de RADON-NIKODYM :

$$\frac{dQ}{dP} = \frac{X}{E_F(X)}.$$

Il en résulte que  $Q_x$  est absolument continue par rapport à  $P_x$  et que :

$$\frac{dQ_x}{dP_x}(x) = \frac{x}{E_F(X)}, \text{ pour tout } x \text{ de } \mathbf{R}_+.$$

La fonction de répartition de  $X$  pour  $Q$  est notée  $G$ .

## 4.2. Propriétés générales

### 4.2.a. Propriété 1

$$\forall x \in \mathbf{R}_+, G(x) \leq F(x).$$

En effet,

$$\forall x \in \mathbf{R}_+, G(x) = \int_0^x dQ_x = \int_0^x \frac{t}{E_F(X)} dP_x;$$

1<sup>er</sup> cas :  $x \leq E_F(X)$  ;

alors

$$G(x) = \int_0^x \frac{t}{E_F(X)} dP_x \leq \int_0^x dP_x = F(x);$$

2<sup>e</sup> cas :  $x > E_F(X)$  ;

alors

$$\int_x^{+\infty} \frac{t}{E_F(X)} dP_x \geq \int_x^{+\infty} dP_x,$$

soit

$$1 - \int_0^x \frac{t}{E_F(X)} dP_x \geq 1 - \int_0^x dP_x; \text{ d'où } G(x) \leq F(x).$$

### 4.2.b. Propriété 2

Si  $X$  admet un moment d'ordre  $h+1$  pour  $P$ , alors  $X$  admet un moment d'ordre  $h$  pour  $Q$ , et l'on a la relation :

$$E_G(X^h) = E_F(X^{h+1})/E_F(X).$$

En effet :

$$E_G(X^h) = \int_{\Omega} X^h dQ = \int_{\Omega} X^h \frac{X}{E_F(X)} dP = \int_{\Omega} \frac{X^{h+1}}{E_F(X)} dP = \frac{E_F(X^{h+1})}{E_F(X)}.$$

## 4.3. Concentration par comparaison des fonctions de répartition

### 4.3.a. Indice associé à l'aire comprise entre $F$ et $G$

On procède comme en 3.1.a. en considérant l'aire  $\tilde{A}$  délimitée par les fonctions de répartition  $\tilde{F}$  et  $\tilde{G}$  de  $\tilde{X} = X/E_F(X)$ .

Il vient :

$$\tilde{A} = \int_0^{+\infty} |\tilde{F}(x) - \tilde{G}(x)| dx = \int_0^{+\infty} \int_0^x (1-t) d\tilde{F}(t) dx;$$

une intégration par partie permet d'écrire :

$$\tilde{A} = \left[ x \int_0^x (1-t) d\tilde{F}(t) \right]_0^{+\infty} - \int_0^{+\infty} x [1-x] d\tilde{F}(x). \quad (13)$$

Il suffit alors de supposer que  $\tilde{X}$  admet un moment d'ordre 3 pour  $\tilde{F}$  (donc d'ordre 2 pour  $\tilde{G}$ ) pour montrer, par application de l'inégalité de BIE-NAYMÉ-TCHEBICHEV, que :

$$0 \leq x \int_0^x (1-t) d\tilde{F}(t) = x[\tilde{F}(x) - \tilde{G}(x)] \leq x[1 - \tilde{G}(x)] \leq \frac{E_{\tilde{G}}(\tilde{X}^2)}{x}.$$

On en déduit que le premier terme du second membre de (13) est nul. Compte tenu de ce que  $E_{\tilde{F}}(\tilde{X}) = 1$ , on a :

$$\tilde{A} = V_{\tilde{F}},$$

soit encore

$$\tilde{A} = \frac{V_F}{[E_F(X)]^2}. \quad (14)$$

#### 4.3.b. Indice associé à la distance de KOLMOGOROV

Ayant déjà remarqué que, pour tout  $x$ ,  $G(x) \leq F(x)$ , il vient :

$$\begin{aligned} S &= \sup_{x \in \mathbf{R}_+} |F(x) - G(x)| = \sup_{x \in \mathbf{R}_+} \int_0^x \left[ 1 - \frac{t}{E_F(X)} \right] dP_X \\ &= \frac{1}{E_F(X)} \sup_{x \in \mathbf{R}_+} \int_0^x [E_F(X) - t] dP_X. \end{aligned}$$

La fonction  $\psi$ , définie sur  $\mathbf{R}_+$  par  $\psi(t) = E_F(X) - t$ , est positive sur  $[0, E_F(X)]$  et négative sur  $[E_F(X), +\infty]$ . Il s'ensuit que le maximum est atteint pour  $x = E_F(X)$ ; il vient :

$$S = \frac{1}{E_F(X)} \int_0^{E_F(X)} [E_F(X) - t] dP_X = F[E_F(X)] - G[E_F(X)]. \quad (15)$$

En remarquant maintenant que :

$$\int_0^{+\infty} [E_F(X) - t] dP_X = \int_0^{E_F(X)} [E_F(X) - t] dP_X + \int_{E_F(X)}^{+\infty} [E_F(X) - t] dP_X = 0$$

et que

$$\int_0^{+\infty} |E_F(X) - t| dP_X = \int_0^{E_F(X)} [E_F(X) - t] dP_X + \int_{E_F(X)}^{+\infty} [t - E_F(X)] dP_X$$

il vient

$$\frac{1}{E_F(X)} \int_0^{E_F(X)} |E_F(X) - t| dP_X = \frac{1}{2E_F(X)} \int_0^{+\infty} |E_F(X) - t| dP_X,$$

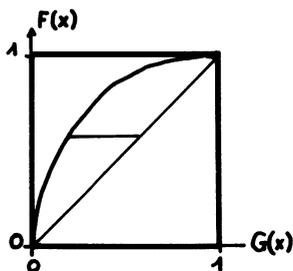
c'est-à-dire

$$S = \frac{\delta_F}{2E_F(X)}. \quad (16)$$

#### 4.3.c. Indice de GINI

Dans le plan rapporté à un repère orthonormé, on considère la courbe de LORENZ définie par les points  $(G(x), F(x))$ ,  $x \in \mathbf{R}_+$ .

Par exemple :



Ainsi que le graphique ci-dessus le suggère, l'aire de concentration est égale à

$$\int_0^{+\infty} [F(x) - G(x)] dF(x),$$

soit encore, puisque  $F(x) \geq G(x)$ ,

$$\int_0^{+\infty} |F(x) - G(x)| dF(x), \quad (17)$$

On retrouve ainsi l'expression d'une distance usuelle entre  $P_X$  et  $Q_X$ . Ceci permet de montrer (voir annexe) que cette aire vaut

$$\frac{\Delta_F}{4E_F(X)}.$$

#### 4.4. Concentration par comparaison des densités

Comme en 3.2.e, considérons une application  $\Phi$  de  $\mathbf{R}_+ \times \mathbf{R}_+$  dans  $\mathbf{R}$ , convexe, homogène de degré 1 et telle que  $\Phi(1, 1) = 0$ . En supposant de plus que

$$\Phi \left( 1, \frac{dQ}{dP} \right)$$

est P-intégrable, un indice synthétique de dissemblance basé sur les densités est alors donné par

$$I_\Phi = \int_{\Omega} \Phi \left( \frac{dP}{dP}, \frac{dQ}{dP} \right) dP,$$

soit :

$$I_\Phi = E_F \left[ \Phi \left( 1, \frac{X}{E_F(X)} \right) \right].$$

Différents exemples de fonction  $\Phi$  conduisant aux indices usuels ont été déjà mentionnés en 3.2.

#### 4.5. Remarque

La concentration s'étudie généralement pour des variables à valeurs positives; toutefois, elle peut s'étendre à des variables prenant des valeurs réelles quelconques mais d'espérance mathématique strictement positive. L'étude présentée ici s'étend alors sans difficulté à ce cas plus général : on définit une mesure de masse (qui n'est plus une mesure de probabilité) de la même façon que dans le cas où  $X$  est positive. Les valeurs négatives interdisent toutefois de considérer des mesures de dissemblance telles que l'information de KULLBACK et la distance d'HELLINGER.

#### ANNEXE

$$\int_0^{+\infty} [F(x) - G(x)] dF(x) = \int_0^{+\infty} F(x) dF(x) - \int_0^{+\infty} G(x) dF(x).$$

La première intégrale du second membre vaut 1/2 puisque  $F(X)$  est distribuée uniformément sur  $[0,1]$  et le second terme vaut :

$$\begin{aligned} \int_0^{+\infty} \int_0^x \frac{t}{E_F(X)} dF(t) dF(x) &= \frac{1}{E_F(X)} \int_0^{+\infty} \int_0^x (t-x) dF(t) dF(x) \\ &+ \int_0^{+\infty} \int_0^x \frac{x}{E_F(X)} dF(t) dF(x); \end{aligned}$$

cette dernière intégrale vaut

$$\int_0^{+\infty} F(x) dG(x)$$

et donne, en intégrant par parties :

$$[F(x) G(x)]_0^{+\infty} - \int_0^{+\infty} G(x) dF(x) = 1 - \int_0^{+\infty} G(x) dF(x).$$

On obtient donc

$$\int_0^{+\infty} G(x) dF(x) = \frac{1}{E_F(X)} \int_0^{+\infty} \int_0^x (t-x) dF(t) dF(x) + 1 - \int_0^{+\infty} G(x) dF(x),$$

d'où l'on déduit

$$\begin{aligned} \int_0^{+\infty} [F(x) - G(x)] dF(x) &= -\frac{1}{2E_F(X)} \int_0^{+\infty} \int_0^x (t-x) dF(t) dF(x) \\ &= \frac{1}{2E_F(X)} \int_0^{+\infty} \int_0^x |t-x| dF(t) dF(x). \end{aligned}$$

En remarquant alors que :

$$\begin{aligned} \int_0^{+\infty} \int_0^{+\infty} (t-x) dF(t) dF(x) &= 0 = -\int_0^{+\infty} \int_0^x |t-x| dF(t) dF(x) \\ &\quad + \int_0^{+\infty} \int_x^{+\infty} |t-x| dF(t) dF(x), \end{aligned}$$

il vient :

$$\begin{aligned} \int_0^{+\infty} [F(x) - G(x)] dF(x) \\ = \frac{1}{4E_F(X)} \int_0^{+\infty} \int_0^{+\infty} |t-x| dF(t) dF(x) = \frac{\Delta_F}{4E_F(X)} \left( = \frac{I_\Gamma}{2} \right). \end{aligned}$$

## BIBLIOGRAPHIE

- [1] A. BACCINI et A. de FALGUEROLLES. — *Statistique descriptive élémentaire*. Cours polycopié, département S.E.E.T.Q.G., I.U.T. B., Université de Toulouse-le-Mirail, 1985.
- [2] M. BARBUT. — Sur quelques propriétés élémentaires des fonctions de concentration de C. GINI. *Mathématiques et Sciences Humaines*, N° 88, p. 5-19, 1984.
- [3] D. DACUNHA-CASTELLE et M. DUFLO. — *Probabilités et Statistiques*, Vol. 1, MASSON, 1982.
- [4] N.I. FISHER. — Graphical methods in nonparametric statistics : a review and annotated bibliography. *International Statistical Review*, Vol. 51, p. 25-58, 1983.
- [5] Ch. GOURIEROUX. — *Mesures d'inégalité, de pauvreté, de concentration*. Cours polycopié, E.N.S.A.E., division S.E.A., 2<sup>e</sup> année, 1981.
- [6] P.E. HART. — Moment distributions in Economics : an exposition. *J.R. Statist. Soc. A.*, Vol. 138, p. 423-434, 1975.
- [7] M.G. KENDALL et A. STUART. — *The advanced theory of statistics*, Vol. 1, GRIFFIN, 1977.
- [8] Y. KONDOR. — An old new measure of income inequality. *Econometrica*, Vol. 39, p. 1041-1042, 1971.
- [9] G.P. PATIL et C. TAILLIE. — Diversity as a concept and its measurement. *J. of the American Statistical Association*, Vol. 77, p. 548-561, 1982.
- [10] E.M. QANNARI. — *Analyses factorielles de mesures, applications*. Thèse de 3<sup>e</sup> cycle, Université Paul Sabatier, Toulouse, 1983.
- [11] M. SAWYER. — La répartition des revenus dans les pays de l'O.C.D.E. *Perspectives économiques de l'O.C.D.E.*, Juillet 1976.