MATHÉMATIQUES ET SCIENCES HUMAINES

MARC BARBUT

# Distributions de type paretien et représentation des inégalités

*Mathématiques et sciences humaines*, tome 106 (1989), p. 53-69
<http://www.numdam.org/item?id=MSH_1989__106__53_0>

*Math. Inf. Sci. hum.* (27e année, n°106, *1989*, pp. 53-69)

# DISTRIBUTIONS DE TYPE PARETIEN ET REPRÉSENTATION DES INÉGALITÉS[1]

Marc BARBUT

Il est constant que l'on ait à comparer entre elles des distributions empiriques (de revenus, de populations urbaines, de patrimoines, etc...) du point de vue de leur *concentration* ou de leur *inégalité*, et de l'évolution de celles-ci dans le temps, dans l'espace, ou dans les deux.

La méthode la plus usitée consiste à calculer, pour chacune des distributions à comparer entre elles, un *indice* ou un *indicateur numérique* de l'inégalité, tel qu'il en existe des quantités, indice de Gini, indices entropiques de Theil, indices d'Atkinson, de Shorroks, de Kolm, etc... ; et bien sûr le coefficient de variation et le $\Phi^2(\chi^2)$.

Dans tous les cas, le procédé consiste à associer à chaque distribution un nombre qui est censé mesurer son écart à l'équirépartition : plus la valeur de l'indice est élevée, plus il y a de l'inégalité.

Bien sûr, on exige de chaque indice proposé qu'il satisfasse à quelques propriétés de bon sens, baptisées axiomes, qui traduisent l'idée que l'on se fait de ce qui doit accroître (ou diminuer) l'inégalité : par exemple, un transfert du bien réparti d'un "riche" à un "pauvre" doit diminuer l'inégalité, donc faire décroître la valeur de l'indice (principe de Pigou et Dalton) ; ou bien, la redistribution en parts égales d'un surplus à partager doit elle aussi diminuer l'inégalité.

L'utilisation de tels indices, très courante chez les économètres notamment, présente deux inconvénients majeurs :

1 - Elle implique une "perte d'information" massive par rapport aux données, puisque chaque distribution est résumée par un seul nombre ; pour un phénomène (l'inégalité) si sensible à la controverse, voir à la passion, de ceux qui en discourent, cette réduction ne va pas sans dommage, et souvent sans graves contresens.

2 - Si les comparaisons fournies par un même indice sont par définition cohérentes entre elles, il n'en va pas de même en général lorsque l'on utilise deux indices ou plus.

---

[1] Ce texte est celui d'une conférence faite au Colloque Interdisciplinaire sur la Modélisation, Universités de Lyon I et Lyon II, 15-16 juin 1989. Il est publié ici avec l'aimable autorisation des organisateurs

La prudence est donc d'avoir recours non à un indice numérique, mais à des expressions *fonctionnelles* de l'inégalité ou de la concentration ; ici deux voies sont possibles :

1 - La voie humble inaugurée par M.O. Lorenz et C. Gini, reprise ensuite dans un contexte différent, par Paul Lévy, qui consiste à associer à chaque distribution étudiée sa *fonction de concentration*, et à comparer entre elles ces fonctions de concentration ; l'ordre entre ces fonctions n'est pas total, mais partiel. Mais c'est souvent lorsque les courbes de concentration se coupent que le spécialiste du phénomène étudié aura le plus de choses à dire : cela met en effet en évidence une ou plusieurs *valeurs critiques* de la variable étudiée pour lesquelles il y a renversement du sens dans la comparaison de l'inégalité (cf. fig.1).

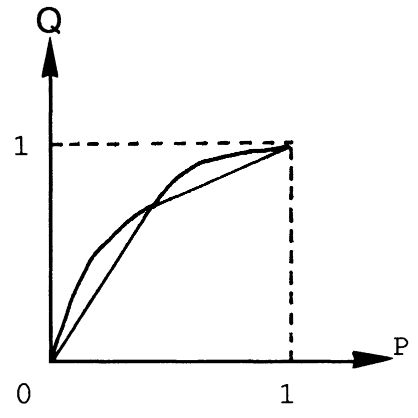


Figure 1
Deux courbes de concentration de Lorentz et Gini se coupant

2 - La voie royale, classique en statistique, qui consiste, chaque fois que les données empiriques s'y prêtent, à *ajuster* à chaque distribution empirique étudiée, une *distribution théorique* d'un *type* donné ; les valeurs des paramètres pour les distributions ajustées fourniront alors tous les éléments utiles à la comparaison du point de vue de la concentration ; cela se voit très bien, par exemple, pour les ajustements à une distribution Log-normale.

A fortiori cette dernière voie doit-elle être préconisée chaque fois que l'on a en tête un "modèle" explicatif raisonnable du phénomène étudié, "modèle" qui conduit à une distribution théorique du type utilisé.

Or il se trouve que pour deux grands *types* de distributions théoriques, les deux voies se rejoignent : elles admettent en effet des *expressions fonctionnelles* de l'inégalité qui sont les plus simples possibles (elles sont *linéaires*) et fournissent une interprétation évidente de leurs paramètres en terme d'inégalité. Ce sont :

- les distributions en fonction puissance, et notamment celles de V. Pareto et P. Lévy, définies par :

$$P(x) = \Pr(X > x) = \begin{cases} 1 & \text{si } x \leq a \\ \left(\dfrac{a+c}{x+c}\right)^{\alpha} & \text{si } x \geq a \end{cases} \qquad \alpha > 0$$

- les distributions exponentielles, définies par :

$$P(x) = \begin{cases} 1 & \text{si } x \leq a \\ e^{-\left(\frac{x-a}{m}\right)} & \text{si } x \geq a \end{cases}$$

Ces dernières sont d'ailleurs, nous le verrons, un cas limite des premières.

En outre, pour ces distributions, l'expression linéaire de leur inégalité fournit des méthodes simples et efficaces d'ajustement de distributions empiriques à des distributions théoriques de leur type.

C'est l'ensemble de ces points que nous allons montrer maintenant.

## 1. MOYENNES CONDITIONNELLES

Soit une distribution définie par sa fonction de répartition F(x) ou, de façon équivalente, par

$$P(x) = 1 - F(x).$$

Supposons que l'espérance de F soit définie.

La *moyenne conditionnelle* en x, notée M(x) est par définition :

$$M(x) = \frac{1}{P(x)} \int_x^{\infty} t dF(t)$$

Dans le cas où F est une distribution de durées de vie, M(x) est l'espérance de vie à l'âge x. Lorsque F est une distribution de revenus, M(x) est le revenu moyen des titulaires de revenus ayant un revenu supérieur à x.

Supposons F de type paretien, avec $\alpha > 1$ (pour que l'espérance existe) ; il vient immédiatement :

$$M(x) = \beta x + \mu \qquad \text{avec :} \quad \beta = \frac{\alpha}{\alpha - 1} > 1$$

$$\mu = c(\beta - 1)$$

. Ainsi, la fonction M(x), qui est l'expression mathématique de la façon la plus courante de *percevoir les inégalités* (si mon revenu est x, quel est le revenu moyen de ceux qui ont un revenu supérieur au mien ?) a ici la forme mathématique *la plus simple*, celle d'une fonction linéaire (ou affine), accessible à un élève de quatrième, par exemple.

En particulier, si c = 0 (distributions paretiennes de premier type $P(x) = \left(\frac{a}{x}\right)^{\alpha}$ , $x \geq a$), on a : $M(x) = \beta x$ .

Le revenu moyen M(x) est un simple multiple de x ; plus le coefficient $\beta$ est grand, plus il y a d'inégalité : $\beta$ est un bon indicateur de l'inégalité.

Dans le cas général, deux moyennes conditionnelles : $M(x) = \beta x + \mu$ et $M'(x) = \beta' x + \mu'$ seront toujours dans le même orde que $\beta$ et $\beta'$ pour les grandes valeurs de x ; mais il pourra arriver qu'elles se coupent et que pour les petites valeurs de x, cet ordre soit inversé : c'est, comme supra, un cas très intéressant puisqu'il met en évidence un seuil critique (de la variable étudiée) de renversement dans le sens de la comparaison (fig. 2).

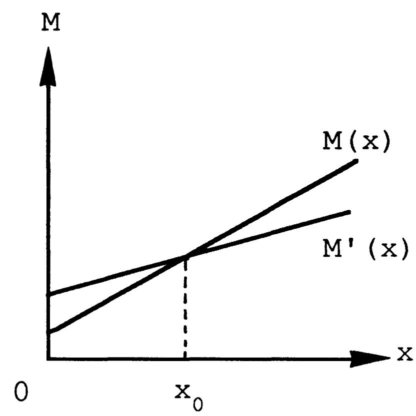


figure 2
Au dessous du seuil $x_0$ (pour $0 < x < x_0$ ) la distribution de moyenne conditionnelle M'(x) est *plus inégalitaire* que celle de la moyenne conditionnelle M(x) ; au-dessus de ce seuil, c'est le contraire.

*Inversement*, demandons-nous quelles sont toutes les distributions absolument continues pour lesquelles la moyenne conditionnelle M(x) est une fonction affine de forme :

$$M(x) = Ax + B .$$

Comme par définition M(x) est une fonction monotone non décroissante de x, et vérifie :

$$\forall x, \; M(x) \geq x$$

il faut : $A > 0$ .

D'autre part, on montre facilement que si pour une valeur $x_0$ de x, on a :

$$M(x_0) = x_0$$

alors la distribution F parente est à support borné par $x_0$ :

$$x \geq x_0 \Rightarrow F(x) = 1 . \quad (P(x) = 0)$$

L'équation fonctionnelle :

$$M(x) = Ax + B = \frac{1}{P(x)} \int_x^{\infty} tdF(t)$$

se transforme par dérivation, compte tenu de $dP(x) = - dF(x)$, en :

$$AP(x)dx = - ((A-1)\, x + B)\, dP(x).$$

*Premier cas.* $A = 1$ , $B > 0$

Il vient :

$$P(x) = K\ e^{-\frac{x}{B}} = e^{-\frac{(x-a)}{B}}$$

le paramètre $a$ étant déterminé par la condition $P(a) = 1$.

Ainsi, les distributions *exponentielles* sont caractérisées par le fait que la moyenne conditionnelle $M(x)$ est de forme :

$$M(x) = x + B$$

où B est l'écart-type de la distribution : le revenu moyen de ceux qui ont un revenu supérieur au mien se déduit du mien par addition de la constante B. C'est cette constante, d'ailleurs égale à l'écart-type de la distribution, qui est le bon indicateur de son inégalité.

*Second cas.* $A > 1$, B quelconque

On a alors :

$$\frac{dP}{P} = - \frac{A}{A-1} \frac{dx}{x + \frac{B}{A-1}} = - \alpha \frac{dx}{x+c}$$

En posant

$$\alpha = \frac{A}{A-1} > 1 \ , \ c = \frac{B}{A-1}$$

D'où :

$$P(x) = \frac{K}{(x+c)^{\alpha}} = \left(\frac{a+c}{x+c}\right)^{\alpha} \quad (P(a) = 1)$$

Ce sont les distributions de Pareto d'exposant $\alpha > 1$ .

*Troisième cas.* $0 < A < 1$ , B quelconque positif.

On n'a $M(x) \geq x$ que pour $x \leq x_0 = \frac{B}{1-A}$ .

L'équation différentielle s'écrit :

$$\frac{dP}{P} = - \frac{A}{1-A} \frac{dx}{(x_0-x)} = \frac{-\gamma dx}{x_0-x} \qquad \gamma = \frac{A}{1-A} > 0$$

D'où :

$$P(x) = K(x_0-x)^{\gamma} \qquad (x \leq x_0)$$

$$P(x) = \left(\frac{x_0-x}{x_0-a}\right)^{\gamma} \qquad P(a) = 1$$

Il s'agit de distributions en fonction puissance sur un intervalle borné $(a,x_0)$ ; en particulier, pour $\gamma = 1$ $(A = \frac{1}{2})$ , c'est la distribution uniforme sur l'intervalle $(a,x_0)$.

Remarquons que la moyenne conditionnelle x a ici l'expression :

$$M(x) = x\,A + x_0\,(1-A)$$

C'est le *barycentre*, avec les poids A et (1-A), entre le revenu x et le revenu le *plus élevé* $x_0$.

Si A est proche de 1, M(x) est proche de x : distribution quasi-égalitaire ; si A est proche de 0, M(x) est proche de $x_0$ : distribution très inégalitaire.

*Dernier cas.* Si A tend vers 1 et B vers 0, on obtient la distribution la *plus égalitaire* dans laquelle P est de type Dirac :

$$P(x) = 1 \quad \text{si} \quad x \leq a \;\; ; \;\; P(x) = 0 \quad \text{si} \quad x > a\,.$$

Tous les revenus sont égaux à un même revenu a.

Le diagramme ci-dessous résume la situation :

$A > 1,\ B \neq 0$
Pareto, cas général
$\alpha > 1$

$A = 1,\ B > 0$
Exponentielles

$A > 1,\ B = 0$
Pareto type 1

$0 < A < 1,\ B > 0$
Distr. Puissance

$A = 1,\ B = 0$
Distr. égalitaire

*Remarque.* On peut également se poser la question de déterminer les distributions pour lesquelles c'est le revenu moyen de ceux qui sont moint riches que moi :

$$m(x) = \frac{1}{F(x)} \int_{-\infty}^{x} t dF(t)$$

qui est une fonction affine de x. La réponse est laissée en exercice au lecteur.

## 2. DEUX GÉNÉRALISATIONS RESTREINTES : LES MOYENNES CONDITIONNELLES D'ORDRE λ.

Pour une variable X à valeurs *positives*, i.e. telle que :

$$1 - F(x) = P(x) = 1 \quad \text{si} \quad x < 0$$

la moyenne conditionnelle d'ordre λ en $x \geq 0$ est définie par :

$$m_\lambda(x) = \left[ \frac{1}{P(x)} \int_{x}^{\infty} t^\lambda dF(t) \right]^{1/\lambda} \qquad \text{pour} \quad \lambda \neq 0$$

$$\text{Log } m_0(x) = \frac{1}{P(x)} \int_{x}^{\infty} (\text{Log } t)\, dF(t) \quad \text{pour} \quad \lambda = 0 \ \text{(moyenne géométrique)}.$$

Ces moyennes sont définies lorsque l'intégrale du second membre converge ; pour $\lambda = 1$, il s'agit de la moyenne ordinaire étudiée au paragraphe précédent ; pour $\lambda = 2$, la moyenne quadratique ; pour $\lambda = -1$, la moyenne harmonique.

On démontre comme on l'a fait (par dérivation) dans le cas $\lambda = 1$ au paragraphe 1 supra, que les distributions pour lesquelles la moyenne conditionnelle d'ordre λ est une fonction linéaire :

$$m_\lambda(x) = Ax \qquad A > 1$$

sont des distributions de Pareto du premier type :

$$P(x) = \left(\frac{a}{x}\right)^\alpha, \quad x \geq a > 0$$

avec :

$$\alpha = \lambda \frac{A^{\lambda}}{A^{\lambda}-1} \quad \text{pour} \quad \lambda \neq 0 \quad \left( A = \left( \frac{\alpha}{\alpha-\lambda} \right)^{1/\lambda} \right)$$

$$\alpha = \frac{1}{\mathrm{Log}\, A} \quad \text{pour} \quad \lambda = 0 \quad (A = e^{1/\alpha}) .$$

Cette propriété ne s'étend pas aux distributions générales de type paretien (avec $c \neq 0$). En effet, pour tout changement d'unité sur la variable, la moyenne d'ordre $\lambda$ subit le même changement d'unité ; par contre, cette moyenne n'est pas définie "aux changements d'origine près". C'est donc bien aux seules distributions paretiennes du premier type (définies au choix près de l'unité, mais non de l'origine) qu'il est naturel qu'elles soient liées de façon simple.

Pour ces distributions, et lorsque l'exposant $\alpha$ est inférieur ou égal à 1 ($0 < \alpha \leq 1$) la moyenne usuelle n'est pas définie ; mais n'importe qu'elle moyenne d'ordre $\lambda < \alpha$ reste définie, est une "valeur centrale" pour les revenus supérieurs ou égaux à x, et est une fonction linéaire de x.

Elle est donc adaptée à une description fonctionnelle simple de l'inégalité de la distribution.

Pour une variable pouvant prendre des valeurs positives ou négatives, la moyenne logarithmique d'ordre $\lambda$ (en x) est définie par :

$$L_{\lambda}(x) = \frac{1}{\lambda} \mathrm{Log} \frac{1}{P(x)} \int_{x}^{\infty} e^{\lambda t}\, dF(t) \quad \text{si} \quad \lambda \neq 0$$

$$L_0(x) = \frac{1}{P(x)} \int_{x}^{\infty} t dF(t) \quad \text{(moyenne ordinaire)}$$

Elle joue, par rapport aux changements d'origine, le même rôle que la moyenne d'ordre $\lambda$ par rapport aux changements d'unité.

On vérifie facilement que :

$$L_{\lambda}(x) = x + A \quad (A > 0)$$

si et seulement si :

$$P(x) = e^{-\frac{(x-a)}{B}} \qquad x \geq a$$

avec :

$$B = \frac{1}{\lambda}(1 - e^{-\lambda A}) \quad \text{pour} \quad \lambda \neq 0 \; ; \; B = A \quad \text{pour} \quad \lambda = 0 .$$

## 3. MÉDIANES, QUANTILES

Comme on l'a rappelé ci-dessus, une distribution de type paretien n'a pas de moyenne lorsque son exposant $\alpha$ vérifie : $0 < \alpha \leq 1$ .

Or ce cas se rencontre parfois dans les applications, notamment en statistique textuelle.

Par contre, la *médiane* $\mu$ est une "valeur centrale" définie quel que soit $\alpha > 0$ ; plus généralement, définissons le quantile conditionnel à q% par $\mu_q(x)$ tel que :

$$P(\mu_q(x)) = q\,P(x) \qquad 0 < q < 1 .$$

La médiane conditionnelle $\mu(x)$ correspond au cas : $q = \frac{1}{2}$ .

Si P est une distribution de type paretien :

$$P(x) = \frac{K}{(x+c)^{\alpha}} \qquad \alpha > 0, \quad x \geq a .$$

On obtient immédiatement :

$$\mu_q(x) = q^{-1/\alpha} x + c(q^{-1/\alpha} - 1) .$$

Et en particulier, pour la *médiane* :

$$\mu(x) = 2^{1/\alpha} x + c(2^{1/\alpha} - 1) .$$

De même, si P est une distribution exponentielle :

$$P(x) = e^{-\frac{(x-a)}{B}}$$

On a :

$$\mu_q(x) = x - B \operatorname{Log} q .$$

Et pour la *médiane* :

$$\mu(x) = x + B \operatorname{Log} 2 .$$

Ainsi, comme dans le cas de la moyenne, *la médiane des valeurs supérieures à x est une fonction affine de x* :

$$\mu(x) = Ax + B$$

avec $A > 1$ pour toutes les distributions paretiennes d'exposant $\alpha > 0$, et $A = 1$ pour toutes les distributions exponentielles.

Mais *attention*, ici, la réciproque n'est plus vraie. Supposons en effet $P(x) = 1$ pour $x \leq a$ , et telle que :

$$\mu(x) = Ax + B \quad \text{pour } x \geq a$$

avec $A \geq 1$ et $B$ donnés.

Donnons-nous $P(x)$ monotone décroissante mais *arbitraire* dans l'intervalle $(a, \mu(a))$ avec :

$$P(a) = 1 \qquad P(\mu(a)) = P\,(Aa + B) = \frac{1}{2} \;.$$

Alors $P(x)$ sera définie pour l'intervalle :

$$a_1 = \mu(a) \leq x \leq \mu(a_1) = a_2.$$

Par :

$$P(x) = \frac{1}{2}\, P\left(\frac{x-B}{A}\right) .$$

On procèdera de même pour définir $P(x)$ dans l'intervalle $a_2 \leq x \leq \mu(a_2) = a_3$ , etc...

Ce qui est par contre vrai en général, c'est que si deux des quantiles conditionnels $\mu_q(x)$ et $\mu_{q'}(x)$ avec $q \neq q'$ sont des fonctions affines de x, alors $P(x)$ est nécessairement paretienne (ou exponentielle).

## 4. CENTRES ET ÉCARTS CONDITIONNELS

Le constat que, pour les distributions paretiennes ou exponentielles les moyennes conditionnelles d'une part, les médianes conditionnelles d'autre part, sont des fonctions affines suggère une généralisation.

Pour tout $\lambda \geq 1$, appelons *centre conditionnel* d'ordre $\lambda$ en x le nombre $C_\lambda(x)$ tel que :

$$(1) \qquad \int_x^{C_\lambda(x)} (C_\lambda(x) - t)^{\lambda-1}\, dF(t) = \int_{C_\lambda(x)}^{\infty} (t - C_\lambda(x))^{\lambda-1}\, dF(t) .$$

Pour $\lambda > 1$, et sous condition que l'intégrale du second membre converge, le centre $C_\lambda(x)$ est, pour une distribution de fonction de répartition F donnée, bien défini ; en particulier pour $\lambda = 2$, c'est la moyenne conditionnelle usuelle, étudiée au paragraphe 1..

Pour $\lambda = 1$, $C_\lambda(x)$ est la médiane, ou éventuellement peut prendre toutes les valeurs d'un intervalle médian.

Pour tout $\lambda \geq 1$, le centre est la valeur de $\theta$ qui, lorsque les intégrales en jeu convergent, minimise :

$$\varphi_\lambda(\theta) = \left[ \frac{1}{P(x)} \int_x^\infty |t-\theta|^\lambda \, dF(t) \right]^{1/\lambda}$$

$\varphi_\lambda(C_\lambda(x))$, valeur du minimum, est l'*écart moyen conditionnel* d'ordre $\lambda$ en x, noté $e_\lambda(x)$ :

$$(2) \quad e_\lambda(x) = \left[ \frac{1}{P(x)} \int_x^\infty |t - C_\lambda(x)|^\lambda \, dF(t) \right]^{1/\lambda} .$$

Supposons F paretienne d'exposant $\alpha > 0$ :

$$1 - F(x) = \left( \frac{a+b}{x+b} \right)^\alpha , \quad x \geq a .$$

Le centre est défini pour tout ordre $\lambda < 1+\alpha$, et l'écart pour $\lambda < \alpha$.

Supposons $\lambda < 1+\alpha$, et dans l'équation (1) définissant le centre, opérons le changement de variable de t en :

$$u = \frac{t+b}{C_\lambda(x)+b}$$

L'équation (1) se ramène à :

$$(3) \quad \int_{Z(x)}^1 (1-u)^{\lambda-1} \frac{du}{u^{\alpha+1}} = \int_1^\infty (u-1)^{\lambda-1} \frac{du}{u^{\alpha+1}}$$

où l'on a posé :

$$Z(x) = \frac{x+b}{C_\lambda(x)+b}$$

Le second membre de (3) est l'intégrale eulérienne :

$$\int_1^\infty (u-1)^{\lambda-1} \frac{du}{u^{\alpha+1}} = \int_0^1 (1-v)^{\lambda-1} \, v^{\alpha-\lambda} \, dv = B(\lambda, \alpha-\lambda+1) .$$

Sa valeur ne dépend que de $\alpha$ et $\lambda$ et est positive. Quant au premier membre, la fonction :

$$\Psi(Z) = \int_Z^1 (1-u)^{\lambda-1} \frac{du}{u^{\alpha+1}} \qquad (0 < Z \leq 1)$$

est une fonction continue, monotone décroissante de Z dans l'intervalle semi-ouvert ]0,1], avec :

$$\Psi(1) = 0 \qquad \text{et} \qquad \lim_{Z\to 0} \Psi(Z) = \infty$$

Donc il existe une valeur $Z_\lambda$ unique telle que :

$$(4) \qquad \Psi(Z) = B(\lambda, \alpha-\lambda+1)$$

et l'on a par définition :

$$0 < Z_\lambda < 1$$

En revenant à la définition de Z(x), il vient :

$$\frac{x+b}{C_\lambda(x)+b} = Z_\lambda$$

c'est-à-dire :

$$C_\lambda(x) = A_\lambda x + (A_\lambda - 1)\, b$$

avec :

$$A_\lambda = \frac{1}{Z_\lambda} > 1 \; .$$

Donc, quel que soit $\lambda \geq 1$, pour toute distribution paretienne d'exposant $\alpha$ tel que $\lambda < 1+\alpha$, le centre d'ordre $\lambda$ est une fonction affine de x.

On vérifie sans peine que, pour $\lambda = 1$ et $\lambda = 2$ respectivement, les valeurs correspondantes fournies par l'équation (4) sont bien :

$$2^{1/\alpha} \qquad \text{et} \qquad \frac{\alpha}{\alpha-1} \; .$$

Supposons maintenant en outre que : $\lambda < \alpha$ . Comme $\lambda \geq 1$, ceci suppose que l'exposant $\alpha$ de la distribution paretienne soit supérieur à 1.

L'écart $e_\lambda(x)$ conditionnel est défini et s'écrit :

$$e_\lambda(x) = \left[ \frac{1}{P(x)} \left( \int_x^{C_\lambda(x)} (C_\lambda - t)^\lambda dF(t) + \int_{C_\lambda(x)}^{\infty} (t - C_\lambda)^\lambda dF(t) \right) \right]^{1/\lambda}$$

En opérant le même changement de variable que supra pour les deux intégrales du second membre, il vient :

$$e_\lambda(x) = (x + b)\, A_\lambda \,.K(\lambda,\alpha)\,.$$

Ainsi, l'écart moyen conditionnel d'ordre $\lambda$, qui, dans les cas où il est défini, mesure la dispersion des valeurs supérieures où égales à x, est lui aussi une fonction affine de x.

Le cas le plus intéressant en pratique est celui où $\lambda = 1$ (écart moyen absolu) ; en effet, il est fréquent dans les applications que l'exposant $\alpha$ soit compris entre 1 et 2 ($1 < \alpha \leq 2$) ; dans ce cas, la distribution n'a pas d'écart-type mais son écart moyen absolu est défini et vaut :

$$e_1(x) = \frac{\alpha}{\alpha - 1} (2^{1/\alpha} - 1)(x + b)\,.$$

De façon tout à fait analogue, on peut voir que si la distribution parente est l'exponentielle $e^{-\frac{(x-a)}{B}}$ le centre conditionnel d'ordre $\lambda$ en x, qui est défini pour tout $\lambda \geq 1$, est égal à :

$$C_\lambda(x) = x + B\, Z_\lambda$$

où $Z_\lambda$ est la solution unique de l'équation :

$$\Psi(Z) = \int_0^Z u^{\lambda-1} e^u\, du = \int_0^\infty u^{\lambda-1} e^{-u}\, du = \Gamma(\lambda)\,.$$

L'écart moyen conditionnel correspondant est constant et vaut :

$$e_\lambda(x) = B \left[ e^{-Z_\lambda} \left( \int_0^{Z_\lambda} u^\lambda e^u\, du + \Gamma(\lambda + 1) \right) \right]^{1/\lambda}\,.$$

## 5. AJUSTEMENTS

Ainsi, pour les variables ayant une distribution de Pareto ou une distribution exponentielle, on dispose, pour *exprimer l'inégalité à chaque niveau x de la variable en fonction de x, d'une infinité de ces procédés qui consistent à comparer à x une "valeur centrale" des valeurs supérieures à x ; cette comparaison est rendue particulièrement intelligible par le fait qu'il s'agit toujours d'une fonction linéaire ou affine de x.*

En pratique, la valeur centrale à utiliser est la moyenne conditionnelle si celle-ci existe, et la médiane conditionnelle sinon.

Mais ces remarques fournissent de surcroît des procédés très simples, et très efficaces, d'ajustement de distributions empiriques à des distributions théoriques de type exponentiel ou paretien.

Soient $x_1, x_2, ..., x_i, ..., x_n$ des valeurs *observées* indiciées de façon que :

$$x_1 > x_2 > ... > x_i > ... > x_n$$

Et soit, pour chaque i, $g(x_i)$ la fréquence *observée* de $x_i$ :

On pose

$$G(x_i) = \sum_{j=1}^{i-1} g(x_j)$$

En particulier, si $\forall_i$, $g(x_i) = \frac{1}{n}$, $G(x_i) = \frac{i-1}{n}$

$$m(x_i) = \frac{1}{G(x_i)} \sum_{j=1}^{i} x_j\, g(x_j) \quad \text{(dans le cas particulier : } m(x_i) = \frac{1}{i-1} \sum_{j=1}^{i-1} x_j \text{)}$$

$G(x_i)$ et $m(x_i)$ sont des estimations de la fonction de répartition $P(x_i)$ et de la moyenne conditionnelle $M(x_i)$ respectivement.

*Ajustement à une exponentielle* $Ke^{-\frac{x}{B}}$

Si la distribution parente est bien exponentielle, les valeurs empiriques doivent être telles que les différences $(m(x_i) - x_i)$ soient à peu près constantes.

On calcule donc , pour chaque i ;

$$y_i = m(x_i) - x_i$$

si les $y_i$ sont à peu près constants, on estime B, avec les précautions d'usage, par

$$B^* = \frac{1}{n} \sum_{1}^{n} y_i$$

*Ajustement à une distribution de Pareto d'exposant $\alpha > 1$*

Un test rapide de la validité de l'hypothèse consiste à regarder si les rapports $\frac{m(x_i)}{x_i}$ sont à peu près constants pour les grandes valeurs de xi.

S'il en est ainsi, on a à réaliser l'ajustement linéaire :

$$y_i = m(x_i) = Ax_i + B + \varepsilon_i. \text{ (où } \varepsilon_i \text{ est le résidu).}$$

Si cet ajustement est bon et si le coefficient A de régression est supérieur à 1, on estime $\alpha$ par :

$$\alpha^* = \frac{1}{A - 1}$$

puis la constante c de l'expression $P(x) = \left(\frac{a+c}{x+c}\right)^{\alpha}$ par :

$$c^* = \frac{B}{A - 1}$$

Et enfin a par :

$$a^* = \frac{m(x_n) - B}{A}$$

*Ajustements à une distribution de Pareto d'exposant* $0 < \alpha < 1$ .

Cette fois-ci, on se servira de la médiane. Pour chaque $x_i$, si :

$$G(x_j) \leq \frac{1}{2} G(x_i) < G(x_{j+1}).$$

On prendra $\mu(x_i) = x_j$.

Et on procèdera à l'ajustement linéaire :

$$\mu(x_i) = y_i = Ax_i + B + \varepsilon_i .$$

D'où, pour les estimations :

$$\alpha^* = \frac{\text{Log } 2}{\text{Log } a} \ , \quad c^* = \frac{B}{A - 1} \ , \quad a^* = \frac{\mu(x_n) - B}{A}$$

*Remarque.* Lorsque l'exposant $\alpha$ est égal à 1 (i.e., a une valeur empirique voisine de 1) la linéarisation de l'ajustement à :

$$P(x) = \frac{a+c}{x+c} \ , \qquad x \geq a$$

est immédiate. En effet, on a dans ce cas :

$$xP(x) = - c \ P(x) + (a+c).$$

On peut donc procéder à l'ajustement linéaire :

$$Y_i = AX_i + B + \varepsilon_i$$

avec

$$Y_i = x_i \, G(x_i), \qquad X_i = G(x_i).$$

D'où

$$C^* = - A \ , \quad a^* = A + B$$

De multiples expériences, sur des "données" économiques (revenus fiscaux français), géographiques (surfaces agricoles en France), démographiques (villes ou agglomérations réparties selon leur population) ou lexicales (fréquences des mots dans des textes) m'ont convaincu de la grande efficacité de ces méthodes si simples d'ajustement.

Resterait à étudier leurs propriétés du point de vue de la statistique mathématique (inférentielle) ; mais ceci est une autre question.

Le point de vue de cette note a été celui de la statistique descriptive : on n'y a pas mentionné les très belles propriétés mathématiques, par rapport à l'addition des variables aléatoires et aux processus stochastiques, des distributions du type Pareto-Lévy, propriétés qui devraient en faire un outil privilégié (et trop souvent négligé de nos jours) de la modélisation mathématique dans les sciences sociales (cf. les travaux de Paul Lévy sur les lois stables, et ceux de Benoit Mandelbrot sur la loi de Pareto-Lévy).

En restant sur l'humble terrain de la description statistique, ces distributions me semblent avoir trois avantages décisifs sur tous les autres types de distributions théoriques utilisées (je pense en particulier à la log-normale) pour rendre compte des répartitions où le phénomène inégalitaire joue un rôle central :

- simplicité et rapidité du "test" permettant de juger si un ajustement paretien (ou exponentiel) est raisonnable,

- *linéarisation naturelle* (i.e., sans le recours à des transformations fonctionnelles, telles que le "passage aux logarithmes", par exemple) de cet ajustement,
- *clarté de l'interprétation*, en termes d'inégalité, des moyennes ou médianes conditionnelles pour lesquelles on réalise cet ajustement.

Ce sont là des remarques de gros "bon sens" ; mais celles-ci ne sont-elles pas les plus fréquemment méconnues ?

## BIBLIOGRAPHIE

- Sur les lois stables et la loi de Pareto-Lévy :
LÉVY P., Théorie de l'addition des variables aléatoires, Gauthier-Villars, Paris, 1937.

MANDELBROT B., The Pareto-Lévy Law and the distribution of income. *International Economic Review*, vol.1, 1960, 79-106.

MANDELBROT B., The stable paretian income distribution, when the apparent exponent is near two. *International Economic Review*, vol.4, janvier 1963.

EYTAN M., Quelques propriétés de la loi de Pareto et leurs incidences en Sciences Humaines. *Mathématiques et Sciences humaines*, n°8, 1964, 21-26.

- Sur les coefficients d'inégalité :
FOSTER James E., Inequality Measurement, in *Proceedings of Symposium in Applied Mathematics*, vol.33, 1985 (Amer. Math. Soc.), 31-68.

- Sur une application de la méthode d'ajustement :
BARBUT M., Des bons et des moins bons usages des distributions paretiennes en analyse des données. *Histoire et Mesure*, CNRS, Paris, T.1, 1988, 111-128.