



Mathématiques et sciences humaines

Mathematics and social sciences

191 | Automne 2010
Variabilité et inégalités

Mesure de la dispersion et des inégalités. Quelques idées

Marc Barbut



Édition électronique

URL : <http://journals.openedition.org/msh/11883>

DOI : 10.4000/msh.11883

ISSN : 1950-6821

Éditeur

Centre d'analyse et de mathématique sociales de l'EHESS

Édition imprimée

Date de publication : 1 septembre 2010

Pagination : 7-32

ISSN : 0987-6936

Référence électronique

Marc Barbut, « Mesure de la dispersion et des inégalités. Quelques idées », *Mathématiques et sciences humaines* [En ligne], 191 | Automne 2010, mis en ligne le 15 février 2011, consulté le 30 avril 2019.

URL : <http://journals.openedition.org/msh/11883> ; DOI : 10.4000/msh.11883

MESURES DE LA DISPERSION ET DES INÉGALITÉS. QUELQUES IDÉES

Marc BARBUT¹

RÉSUMÉ – *On présente sommairement cinq façons bien différentes dans leur esprit et dans leur mode opératoire d’appréhender la notion de dispersion (concentration) d’une distribution statistique ou de probabilité, dans le seul cas univarié. On met l’accent sur des mesures fonctionnelles de la dispersion. Dans un dernier paragraphe, on évoque le cas de la mesure des inégalités.*

MOTS-CLÉS – Concentration, Concentration-dispersion (fonctions de), Diamètre, Dispersion, Écart, Inégalités, Quantiles

SUMMARY – Dispersion and inequality measures: six ideas
This paper is a brief introduction to five very different ways to define the notion of dispersion (concentration) in an empirical or probability distribution, in the univariate case. We emphasize the approach by functional measures of dispersion. The last section is about the functional measures of inequality.

KEYWORDS – Concentration, Concentration (dispersion) functions, Deviation, Diameter, Dispersion, Inequality, Quantiles

INTRODUCTION

En juin 1989, d’éminents chercheurs en sciences sociales avaient organisé à Paris un colloque intitulé « Moyenne, milieu, centre » (cf. Armatte [1991]). C’était, selon la doctrine d’Adolphe Quetelet, témoigner du rôle essentiel, pour l’analyse statistique des données, des valeurs centrales.

Mais pour une distribution, une même valeur centrale peut correspondre à des situations radicalement différentes, notamment en ce qui concerne la dispersion (ou la concentration) autour (attention : ce n’est même pas toujours « autour » !) de ce centre.

Il convenait donc de compléter l’initiative de 1989 par une rencontre consacrée à la dispersion, la variabilité et l’inégalité. Ce fut fait à la Sorbonne en novembre 2009 et c’est un exposé qui y fut présenté en introduction aux notions de dispersion (concentration) en statistique que l’on trouve ci-dessous.

D’autres interventions à ce colloque seront publiées dans *Mathématiques et Sciences humaines* ou encore dans le *Journal Électronique d’Histoire de la Probabilité et de la Statistique* (jehps.net), livraison de juin 2010.

¹ Centre d’analyse et de mathématiques sociales (CAMS), École des hautes études en sciences sociales, 190-198 avenue de France 75244 Paris cedex 13, mbarbut@ehess.fr

Sans qu'il y ait rien d'original, on y insiste sur cinq idées concernant cinq façons différentes d'appréhender la notion de dispersion et de la rendre opératoire. Une sixième section consiste en l'application des plus fécondes des précédentes, celles des mesures *fonctionnelles* de dispersion, à la mesure des inégalités. C'est la sixième idée.

Les deux premiers paragraphes concernent les deux plus usuelles des mesures de la dispersion d'une distribution par un seul nombre, tels que le plus célèbre d'entre eux, la variance.

Mais une même variance (ou une même valeur de tel indicateur numérique de la concentration) peut, comme une même valeur du centre, correspondre à son tour à des situations très différentes.

Pour pallier cet inconvénient, une solution « à la Karl Pearson » consiste à ajouter des moments de la distribution, ceux d'ordre 3 et 4 le plus souvent. Mais où s'arrêter ?

Une autre solution, celle empruntée de 1906 à 1930 environ par les M.O. Lorenz, C. Gini, P. Lévy, M. Fréchet, E. Gumbel, et même, en un certain sens, par G. Sorel dès 1897 : associer à chaque distribution une *fonction* de la variable en jeu, fonction censée représenter au mieux sa dispersion (ou son inégalité).

Bien sûr, il en résulte que, dans la comparaison, on n'aura plus un ordre total (comme c'est le cas pour un indicateur numérique) ; en général, l'ordre ne sera que partiel. Mais n'est-ce pas justement là dans la nature des choses ? D'ailleurs, le calcul de plusieurs moments conduit lui aussi à un ordre partiel.

Deux avertissements pour terminer :

- On n'aborde pas ici, sauf par quelques très sommaires et très partielles indications, la question des indices (ou coefficients) d'inégalité ; la dispersion est le sujet central ;
- Dans ce texte élémentaire on reste dans le cadre le plus simple, celui de la statistique uni-variée. L'extension au cas multi-varié de certaines des notions présentées ici pose de délicats problèmes. Il faudrait y consacrer tout un autre colloque.

1. PREMIÈRE IDÉE : UN CENTRE ET UN ÉCART AU CENTRE

C'est l'idée la plus banale, la plus communément mise en pratique. Le lecteur qui en est familier peut donc sauter ce paragraphe.

1.1 DISTANCES

Soient deux suites d'observations de variables *numériques* portant sur le même ensemble de n individus (ce que l'on appelle deux variables appariées) :

$$x = (x_1, x_2, \dots, x_i, \dots, x_n) \quad y = (y_1, y_2, \dots, y_i, \dots, y_n)$$

En quoi diffèrent-elles ?

Première réponse, la plus immédiate : par leurs différences $(x_i - y_i)$, c'est-à-dire la suite :

$$x - y = (x_1 - y_1, \dots, x_i - y_i, \dots, x_n - y_n)$$

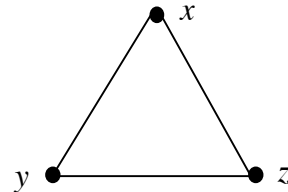
Problème : résumer cette suite de n différences individuelles par un seul nombre.

La solution : associer à chaque couple de suites x et y un nombre *positif* $\Delta(x,y)$, leur *distance*, vérifiant (quels que soient x, y et z) les trois conditions :

$$\Delta(x,y) = 0 \quad \text{ssi} \quad x = y$$

$$\Delta(x,y) = \Delta(y,x) \quad (\text{symétrie})$$

$$\Delta(x,z) \leq \Delta(x,y) + \Delta(y,z) \quad (\text{inégalité triangulaire})$$



1.2 UNE FAMILLE DE DISTANCES

On démontre, on ne le fera pas ici (voir par exemple, Barbut [1986(a)] que, quel que soit l'exposant $\alpha \geq 1$, $\Delta_\alpha(x,y)$ défini par l'équation (1) est une distance :

$$(1) \quad \Delta_\alpha(x,y) = \left(\sum_{i=1}^n |x_i - y_i|^\alpha \right)^{1/\alpha} \quad \alpha \geq 1$$

En pratique, seules trois valeurs de l'exposant α sont utilisées.

$$- \quad \alpha = 1 \quad \Delta_1(x,y) = \sum_{i=1}^n |x_i - y_i|$$

Pour la dimension $n = 2$, on a : $\Delta_1(x,y) = |x_1 - y_1| + |x_2 - y_2|$

c'est la valeur commune de tous les « escaliers » reliant les points x et y (cf. Figure 1), et en particulier du chemin $x \alpha y$. D'où le nom de « distance de Manhattan » qu'on lui donne souvent.

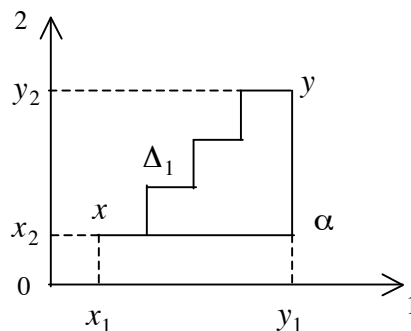


Figure 1

$$- \quad \alpha = 2 \quad \Delta_2(x,y) = \left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{1/2}$$

C'est la classique « distance euclidienne », la longueur de l'hypoténuse du triangle $x a y$ (cf. Figure 2) dans le cas de la dimension $n = 2$.

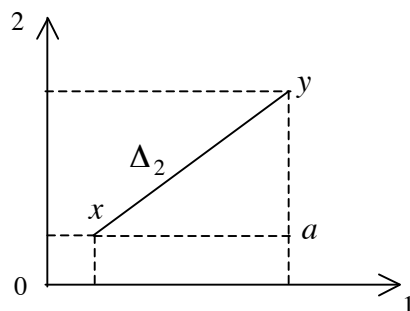


Figure 2

$$- \alpha = \infty \quad \Delta_{\infty}(x, y) = \max_i |x_i - y_i|$$

Pour $n = 2$ (cf. Figure 3), c'est le plus grand côté de l'angle droit dans le triangle $x a y$. Cette distance est appelée *distance de la convergence uniforme*.

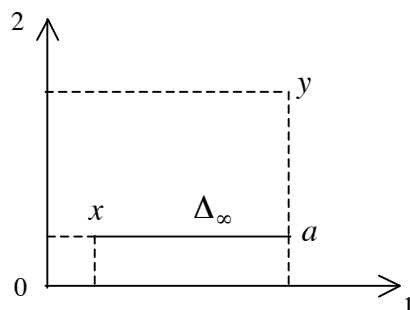


Figure 3

1.3 CENTRES ET ÉCARTS

Soit maintenant une suite x de n observations d'une même variable numérique X (par exemple une même variable X observée sur n individus)

$$x = (x_1, x_2, \dots, x_i, \dots, x_n)$$

Problème : résumer cette suite de n nombres par un seul nombre, qu'on appellera *valeur centrale* (ou typique), ou plus brièvement *centre* de la suite des observations.

Une solution :

- considérons une suite de n observations qui seraient constamment égales à une même valeur λ . On peut la noter λu , où u est la suite constante :

$$u = (1, 1, \dots, 1, \dots, 1)$$

ce qui donne :

$$\lambda u = (\lambda, \lambda, \dots, \lambda, \dots, \lambda)$$

- choisissons une distance Δ . Alors si λ varie, $\Delta(x, \lambda u)$ est une fonction de λ :

$$\varphi(\lambda) = \Delta(x, \lambda u)$$

- faisons varier λ de $-\infty$ à $+\infty$; en général, $\varphi(\lambda)$ aura un minimum pour $\lambda = c$.

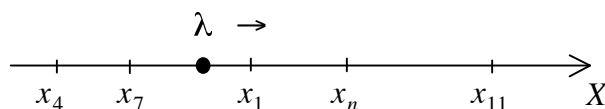


Figure 4

C'est ce minimum c , ou plus précisément $c(x)$, car il dépend évidemment de la suite x des observations, que l'on prendra pour *centre* de celles-ci.

Quant à la distance minimum $\Delta(x, cu)$, on l'appelle *écart* de la suite x pour la distance Δ . C'est une mesure de la *dispersion* (ou dualement de la *concentration*) des observations :

$$E(x) = \Delta(x, cu)$$

Exemples :

– Pour la distance Δ_∞ de la convergence uniforme, on a :

$$\varphi(\lambda) = \Delta_\infty(x, \lambda u) = \max_i |x_i - \lambda|$$

Il est clair (cf. Figure 4) que $\varphi(\lambda)$ est minimum pour :

$$\lambda = \frac{1}{2}(\max_i x_i + \min_i x_i) = c$$

Le centre est le *milieu* de l'intervalle séparant la plus petite des observations de la plus grande (cf. Figure 5).

L'écart correspondant est par suite :

$$E = \frac{1}{2} \left(\max_i x_i - \min_i x_i \right)$$

C'est la moitié de l'intervalle séparant la plus petite des observations de la plus grande.

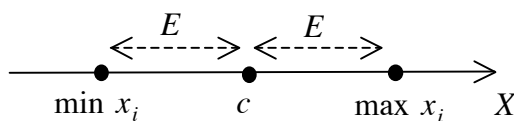


Figure 5

– Pour la distance euclidienne Δ_2 , on a :

$$\varphi(\lambda) = \Delta_2(x, \lambda u) = \left(\sum (x_i - \lambda)^2 \right)^{1/2}$$

Il est clair que $\varphi(\lambda)$ et $\varphi^2(\lambda)$ ont leur minimum pour la même valeur de λ . Or :

$$\varphi^2(\lambda) = \sum (x_i - \lambda)^2 = \sum x_i^2 - 2\lambda \sum x_i + n \lambda^2$$

Ce polynôme du second degré est minimum pour :

$$\lambda = \frac{1}{n} \sum x_i = c$$

Le centre est la *moyenne* arithmétique, notée m dans la suite, des observations.

Quant à l'écart, appelé *écart quadratique*, son carré est :

$$E^2 = \sum (x_i - m)^2 = \left(\sum x_i^2 \right) - nm^2$$

La moyenne $\frac{E^2}{n}$ est ainsi la classique *variance* V :

$$\frac{1}{n} E^2 = V = \frac{1}{n} \left(\sum x_i^2 \right) - m^2$$

Ce qui se dit la variance est « la moyenne des carrés moins le carré de la moyenne ».

Quant à l'écart moyen lui-même, c'est l'écart-type :

$$\sigma = V^{1/2} = \frac{E}{n^{1/2}} = \left(\frac{1}{n} \sum x_i^2 - m^2 \right)^{1/2}$$

– Pour la distance de Manhattan Δ_1 , le centre est la *médiane* μ des valeurs x_i observées, si celles-ci sont en nombre impair. Si leur nombre est pair, on a un intervalle médian, dont n'importe quel point peut être pris pour centre. On a en effet :

$$\varphi(\lambda) = \sum |x_i - \lambda| = \sum_{x_i < \lambda} (\lambda - x_i) + \sum_{x_i > \lambda} (x_i - \lambda)$$

Lorsque λ croît de $-\infty$ à $+\infty$, chacun des deux termes décroît (cf. Figure 4 *supra*) tant qu'il y a moins de x_i inférieurs à λ que de x_i qui lui sont supérieurs ; et croît lorsque c'est l'inverse. Le minimum est donc atteint lorsqu'il y a autant de x_i de part et d'autre de λ .

L'écart correspondant, dit *écart absolu*, est alors :

$$E = \varphi(\mu) = \sum_{x_i > \mu} x_i - \sum_{x_i < \mu} x_i$$

L'interprétation de cette valeur de l'écart est extrêmement claire si les x_i sont par exemple des revenus. Convenons d'appeler « riches » ceux qui ont un revenu supérieur au revenu médian, et « pauvres » ceux dont le revenu est inférieur à μ .

L'écart E se lit alors « revenu global des riches moins revenu global des pauvres ».

Le rapport $\frac{E}{n}$ est appelé *écart moyen absolu*.

NB.

a. Il arrive, chez les économistes notamment, que l'on calcule l'écart absolu à la moyenne m :

$$\varphi(m) = \sum |x_i - m|$$

Pourquoi pas ? mais on n'a plus alors l'interprétation très parlante de l'écart à la médiane.

Supposons en effet n impair, soit : $n = 2k + 1$. Si les x_i sont indicés dans l'ordre croissant :

$$x_1 \leq x_2 \leq \dots \leq x_{k+1} \leq \dots \leq x_{2k+1}$$

et si : $x_h < m \leq x_{h+1}$, on a

$$\sum |x_i - m| = \sum_{x_i > m} x_i - \sum_{x_i < m} x_i + 2(h - k)m$$

Règle d'or. À la métrique euclidienne Δ_2 correspondent la moyenne m et l'écart-type. À la métrique « de Manhattan » Δ_1 , la médiane μ et l'écart moyen absolu *par rapport à μ* .

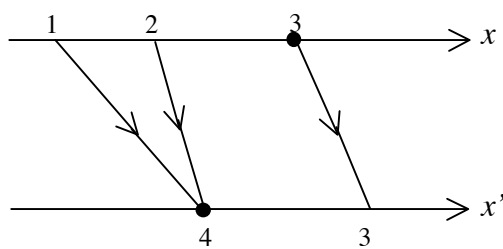
- b. On utilise parfois comme valeur centrale d'une distribution le *mode* M de celle-ci, c'est-à-dire la valeur la plus fréquente, également appelée *dominante*.

C'est, par exemple le cas en démographie, où l'âge auquel la mortalité est maximum est d'ailleurs appelé « vie normale », à bien distinguer de la « vie moyenne » (espérance de vie) et de la « vie probable » (médiane des durées de vie).

L'une des propriétés que l'on doit exiger d'une « valeur centrale », supposée *représentative* de l'ensemble des observations x_i , est que si tous les x_i croissent, la valeur censée les représenter croisse elle aussi.

Or, pour le mode, il y a des cas où c'est le contraire qui se produit : les x_i croissent ($\forall i, x_i > x'_i$) et le mode M diminue ($M < M'$).

Voici un exemple très simple de ce phénomène.



| | | | |
|---------------------------------|-----------------|--------------------------|------------------------|
| $x_1 = x_2 = 0$ | $x_3 = x_4 = 1$ | $x_5 = x_6 = x_7 = 5$ | $M = 5$ (fréquence 3) |
| $x'_1 = x'_2 = x'_3 = x'_4 = 2$ | | $x'_5 = x'_6 = x'_7 = 6$ | $M' = 2$ (fréquence 4) |

Figure 6

1.4 INDICES D'INÉGALITÉ ASSOCIÉS

Les écarts sont tous homogènes à la variable X : ils sont de dimension 1 par rapport à celle-ci, et s'expriment dans la même unité de mesure.

On transforme chacun en un *indice* (ou *indicateur*, ou *coefficient*) d'*inégalité* de la distribution des observations en le divisant par une grandeur le ramenant à la *dimension* 0.

C'est ainsi que l'écart absolu peut être divisé par la somme S des x_i :

$$\frac{E}{S} = \frac{1}{S} \sum_{x_i > \mu} x_i - \frac{1}{S} \sum_{x_i < \mu} x_i, \quad S = \sum x_i$$

Ce qui s'interprète comme la proportion du revenu global attribuée aux « riches » diminuée de celle qui est attribuée aux « pauvres ».

De même, à l'écart-type correspond, en le divisant par la moyenne, le *coefficient de variation* C :

$$C = \frac{\sigma}{m} = \left(\frac{V}{m^2} \right)^{1/2}$$

Dans le cas de la distance Δ_∞ , on peut considérer le rapport qui, à ma connaissance, n'a pas de nom :

$$\frac{\max x_i - \min x_i}{\max x_i + \min x_i}$$

NB. Les trois indices d'inégalité ainsi définis n'ont vraiment de sens que pour les variables *positives*. Sinon, la somme ou la moyenne des x_i peuvent être proches, voire très proches de 0 ; ou négatives : le rapport calculé n'a plus aucun sens en termes d'inégalité.

2. DEUXIÈME IDÉE : UN DIAMÈTRE

Ici, il n'y a plus trace de centre, ni de valeur représentative des x_i . A. Quetelet n'aurait pas aimé.

2.1 DIAMÈTRE D'UNE DISTRIBUTION

Notons toujours $x = (x_1, x_2, \dots, x_n)$ la suite des valeurs observées.

L'idée est de considérer l'ensemble des différences $|x_i - x_j|$. On posera donc, pour tout $\alpha \geq 1$:

$$(2) \quad D_\alpha(x) = \left[\sum_{i=j}^n \sum_{j=1}^n |x_i - x_j|^\alpha \right]^{1/\alpha} = \left[2 \sum_{x_i > x_j} (x_i - x_j)^\alpha \right]^{1/\alpha}$$

D_α est le *diamètre* d'ordre α .

Pour définir un *diamètre moyen*, on peut considérer soit le nombre total n^2 des couples (x_i, x_j) , soit celui des $n(n-1)$ couples tels que $i \neq j$; ce qui conduit aux diamètres moyens :

$$- \quad d_\alpha(x) = \left[\frac{1}{n^2} \sum_1^n \sum_1^n |x_i - x_j|^\alpha \right]^{1/\alpha}$$

$$- \quad \delta_\alpha(x) = \left[\frac{1}{n(n-1)} \sum_i \sum_j |x_i - x_j|^\alpha \right]^{1/\alpha}$$

On a évidemment :

$$\delta_\alpha = \left(\frac{n}{n-1} \right)^{1/\alpha} d_\alpha, \quad \alpha \leq 1$$

Quand le nombre n des observations est assez grand, ces deux diamètres moyens diffèrent peu l'un de l'autre.

Dans le cas $\alpha = 1$:

$$\delta_1 = \frac{1}{n(n-1)} \sum_i \sum_j |x_i - x_j| = \frac{2}{n(n-1)} \sum_{x_i > x_j} (x_i - x_j)$$

En divisant δ_1 par le total $(\sum x_i)$ distribué, nous obtenons l'indice d'inégalité *G de Gini* :

$$(3) \quad G = \frac{\delta_1}{\sum x_i}$$

Dans le cas $\alpha = 2$, on a :

$$\begin{aligned} (D_2(x))^2 &= \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2 = \sum \sum (x_i - 2x_i x_j + x_j)^2 \\ &= n \sum x_i^2 - 2 \sum \sum x_i x_j + n \sum x_j^2 \end{aligned}$$

Or :

$$\sum_i \sum_j x_i x_j = \left(\sum_1^n x_i \right) \left(\sum_1^n x_j \right) = \left(\sum x_i \right)^2$$

D'où finalement :

$$\begin{aligned} (D_2(x))^2 &= 2 \left[n \sum x_i^2 - \left(\sum x_i \right)^2 \right] \\ &= 2n^2 \left[\frac{1}{n} \sum x_i^2 - \left(\frac{\sum x_i}{n} \right)^2 \right] \\ &= 2n^2 V \end{aligned}$$

D'où, pour le diamètre moyen d_2 :

$$d_2(x) = \left[\frac{1}{n^2} \sum \sum (x_i - x_j)^2 \right]^{1/2}$$

$$(4) \quad d_2(x) = (2V)^{1/2} = \sqrt{2} \sigma$$

– Dans le cas où $\alpha = \infty$, on obtient, par passage à la limite quand α tend vers l'infini :

$$D_\infty(x) = 2E_\infty(x) = \max_i x_i - \min_i x_i$$

En fait, écart-moyen et diamètre moyen sont toujours du même ordre de grandeur.

Désignons par $\varepsilon_\alpha(x)$ l'écart-moyen :

$$\varepsilon_\alpha(x) = \left(\frac{1}{n} \sum |x_i - c|^\alpha \right)^{1/\alpha}$$

où c est le centre.

Alors, on peut démontrer (cf. [Barbut, 1986(b)]) la double inégalité :

$$(5) \quad 2 \varepsilon_\alpha(x) \geq \delta_\alpha(x) > \varepsilon_\alpha(x)$$

Ainsi, de même qu'on dit *diamètre*, on pourrait considérer que l'écart est le *rayon* de la distribution.

2.2 L'INÉGALITÉ TRIANGULAIRE

Dans les trois cas ($\alpha = 1$, $\alpha = 2$, $\alpha = \infty$) usuels, on vérifie aisément :

$$(6) \quad D_\alpha(x+y) \geq D_\alpha(x) + D_\alpha(y), \quad \alpha \geq 1$$

– Pour $\alpha = 1$

$$D_1(x+y) = \sum_i \sum_j |(x_i + y_i) - (x_j + y_j)| = \sum_i \sum_j |(x_i - x_j) + (y_i - y_j)|$$

Or, pour chaque couple (i, j) :

$$|(x_i - x_j) + (y_i - y_j)| \geq |x_i - x_j| + |y_i - y_j|$$

D'où l'inégalité (6).

– Pour $\alpha = 2$, on a :

$$\begin{aligned} (D_2(x+y))^2 &= \sum_i \sum_j (x_i + y_i - x_j - y_j)^2 = \sum_i \sum_j (x_i - x_j + y_i - y_j)^2 \\ &= \sum_i \sum_j (x_i - x_j)^2 + \sum_i \sum_j (y_i - y_j)^2 + 2 \sum_i \sum_j (x_i - x_j)(y_i - y_j) \\ &= (D_2(x))^2 + (D_2(y))^2 + 2 \sum_i \sum_j (x_i - x_j)(y_i - y_j) \end{aligned}$$

Mais d'après l'inégalité de Schwarz :

$$\left| \sum_i \sum_j (x_i - x_j)(y_i - y_j) \right| \leq D_2(x) D_2(y)$$

D'où la double inégalité :

$$|D_2(x) - D_2(y)| \leq D_2(x+y) \leq D_2(x) + D_2(y)$$

Pour $\alpha = \infty$, soit i_0 l'indice du maximum des $(x_i + y_i)$, et i_1 celui du minimum. On a :

$$D_\infty(x+y) = x_{i_0} + y_{i_0} - (x_{i_1} + y_{i_1}) = (x_{i_0} - x_{i_1}) + (y_{i_0} - y_{i_1})$$

Or : $x_{i_0} - x_{i_1} \geq \max_i x_i - \min_i x_i = D_\infty(x)$

Et de même pour y . Finalement :

$$D_\infty(x+y) \leq D_\infty(x) + D_\infty(y)$$

Pour démontrer (6) dans le cas général où $\alpha \leq 1$ est réel quelconque, rappelons d'abord que dans un espace de dimension N , la norme $\|z\|_\alpha$ des vecteurs $z = (z_1, \dots, z_v, \dots, z_N)$ définie par :

$$\|z\|_\alpha = \left(\sum_{v=1}^N |z_v|^\alpha \right)^{1/\alpha}$$

vérifie l'inégalité triangulaire (cf. [Hardy, *et alii*, 1934]) ; pour une démonstration géométrique élémentaire, cf. [Barbut, 1986(a)].

Posons : $n_{ij} = x_i - x_j \quad v_{ij} = y_i - y_j$

On a, selon la définition (2) du diamètre :

$$D_\alpha(x) = \left(\sum_{i=1}^n \sum_{j=1}^n |u_{ij}|^\alpha \right)^{1/\alpha} \quad D_\alpha(y) = \left(\sum_{i=1}^n \sum_{j=1}^n |v_{ij}|^\alpha \right)^{1/\alpha}$$

et $D_\alpha(x+y) = \left(\sum_{i=1}^n \sum_{j=1}^n |u_{ij} + v_{ij}|^\alpha \right)^{1/\alpha}$

Posons ensuite : $v(i,j) = n(i-1) + j$. Lorsque i et j prennent chacun toutes les valeurs entières de 1 à n , $v(i,j)$ prend une fois et une seule chacune des valeurs entières de 1 à n^2 .

On a donc, dans la dimension $N = n^2$:

$$D_\alpha(x) = \left(\sum_{v=1}^{n^2} |u_v|^\alpha \right)^{1/\alpha} = \|u\|_\alpha$$

De même : $D_\alpha(y) = \|v\|_\alpha \quad D_\alpha(x+y) = \|u+v\|_\alpha$

L'inégalité triangulaire pour les normes $\|u+v\|_\alpha \geq \|u\|_\alpha + \|v\|_\alpha$ se ramène donc ici l'inégalité (6).

2.3 DEUX RELATIONS POUR LE DIAMÈTRE D_1

Dans le cas de la métrique de Manhattan ($\alpha = 1$), celle qui est associée à la médiane, le diamètre vérifie quelques relations remarquables.

En voici une, dont on trouvera une démonstration dans [Barbut, 1986]. Désignons par x_R l'ensemble des x_i strictement supérieurs à la médiane μ (les « riches ») et par x_P l'ensemble de ceux qui sont inférieurs à μ (les « pauvres »). Alors on a :

$$D_1(x) = D_1(x_R) + D_1(x_P) + 2 \left[\frac{n+1}{2} \right] E_1(x)$$

ou $[a]$ désigne la partie entière de a .

En voici une seconde. Supposons ici $n = 2k + 1$ et que les indices sont tels que :

$$x_1 \leq x_2 \leq \dots \leq x_{k+1} \leq \dots \leq x_{2k+1}$$

Alors, en réarrangeant convenablement les termes de la somme double ci-dessous, on trouve :

$$(7) \quad \frac{1}{2} D_1(x) = \sum_{x_i > x_j} (x_i - x_j) = 2 \sum_{h=1}^k (x_{k+1+h} - x_{k+1-h}) h$$

Dans le dernier membre de (7), que C. Gini connaissait bien, chacune des différences $(x_{k+1+h} - x_{k+1-h})$ est un *intervalle interquantile médian*, ce qui relie l'idée de diamètre à la troisième idée, celle des intervalles interquartiles, dont on va parler maintenant. Dans les notations du paragraphe 3, ci-dessous, on aurait $q = \frac{k+h+1}{2k+1}$ (cf. Figure 7).

3. TROISIÈME IDÉE : LES INTERVALLES INTERQUANTILES MÉDIANS

Dans la graphique de la Figure 7, on a en ordonnées le numéro d'ordre de chaque x_i , ceux-ci étant ordonnés par ordre croissant des indices. Et en abscisses, les x_i correspondants.

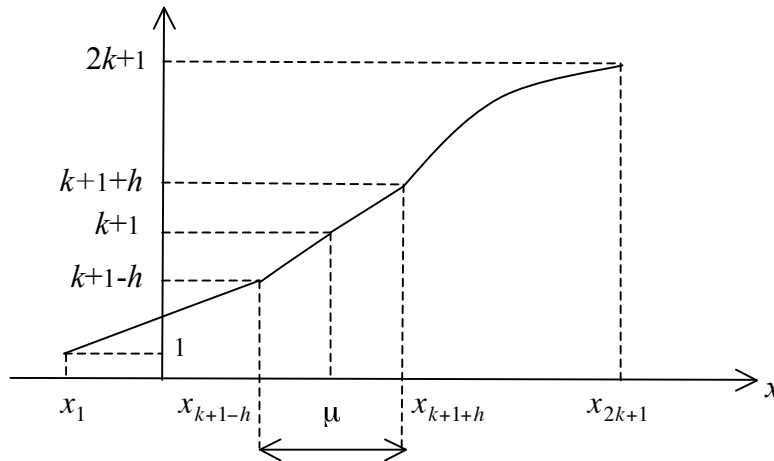


Figure 7

Ramenées à une échelle allant de 0 à 1, les ordonnées seront les proportions successives et croissantes d'un pas constant $\frac{1}{2k+1}$:

$$\frac{1}{2k+1} < \frac{2}{2k+1} < \dots < \frac{k+1-h}{2k+1} < \dots < \frac{k+1+h}{2k+1} < \dots < 1$$

Ici, h varie de 0 à k .

Posons :
$$q = \frac{k+1+h}{2k+1}$$

Alors l'intervalle $(x_{k+1+h} - x_{k+1-h})$ s'écrit $(x_q - x_{1-q})$.

D'une façon générale, si $F(x)$ est une fonction de répartition continue strictement croissante, tout nombre q tel que $1 \leq q \leq 1/2$ correspond par inversion de F un intervalle I_q (cf. Figures 8(a) et 8(b)) :

$$I_q = x_q - x_{1-q}$$

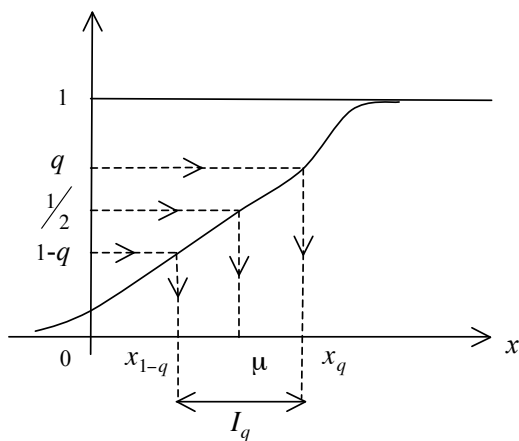


Figure 8(a)

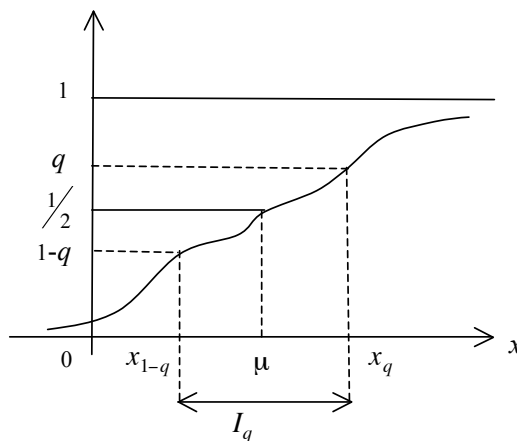


Figure 8(b)

Il est clair (cf. Figure 8(b)) que moins la distribution de fonction de répartition F est concentrée autour de sa médiane (i.e., plus elle est « étalée »), plus, pour une même valeur de q , l'intervalle I_q est grand.

L'intervalle I_q , appelé *intervalle interquantile médian*, est ainsi un indicateur de la dispersion de la distribution.

Les valeurs usuellement considérées sont :

- $q = \frac{3}{4}$ intervalle interquartile
- $q = \frac{h}{10}$, h entier allant de 5 à 10, qui définit les intervalles interdéciles. Mais on utilise aussi les vigintiles, pour lesquels le pas est de 5 %.

Cette façon de « mesurer » la dispersion présente deux avantages considérables par rapport aux précédents (écarts et diamètres).

1. Quelle que soit la variable X , qu'elle ait des moments ou non, et en particulier qu'elle ait un moment d'ordre 1 (moyenne) ou non, *elle est toujours définie*.

Alors que les écarts et les diamètres ne sont définis que si le moment d'ordre 1 existe.

2. Le second avantage, c'est que si, pour une même distribution, on fait varier q , l'intervalle I_q est une *fonction* de q , monotone croissante (ou non-décroissante) lorsque q va de $\frac{1}{2}$ à 1 (cf. Figure 9).

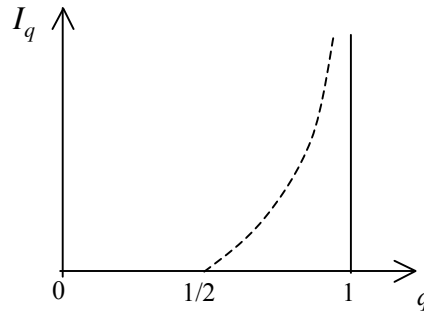


Figure 9

On a ainsi une première définition *fonctionnelle* de la dispersion ; on en verra d'autres.

Exemples.

- Distribution de Pareto :

$$F(x) = 1 - \left(\frac{x_0 + c}{x + c} \right)^\alpha, \quad \alpha > 0, \quad x \geq x_0 > -c$$

Elle n'a de moyenne que pour $\alpha > 1$. Mais, quel que soit α , pour $F(x) = q$, on a :

$$q = 1 - \left(\frac{x_0 + c}{x + c} \right)^\alpha$$

soit : $(1 - q)^{1/\alpha} = \frac{x_0 + c}{x + c}$

soit encore : $x = x_q = -c + (x_0 + c)(1 - q)^{-1/\alpha}$

donc : $x_{1-q} = -c + (x_0 + c)q^{-1/\alpha}$

et : $I_q = x_q - x_{1-q} = [(1 - q)^{-1/\alpha} - q^{-1/\alpha}]$

- Distribution de Cauchy :

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \text{Arc tg} \left(\frac{x - x_0}{a} \right)$$

Pour celle-ci, aucun moment n'est défini, même pas le premier.

En procédant comme ci-dessus, on a : $x_q = x_0 + a \operatorname{tg}\left(q - \frac{1}{2}\right)$

D'où, compte tenu de : $-\operatorname{tg}\theta = \operatorname{tg}(-\theta)$

$$I_q = x_q - x_{1-q} = 2a \operatorname{tg}\left[\Pi(q - 1/2)\right]$$

– Distribution exponentielle :

$$F(x) = 1 - e^{-\frac{(x-x_0)}{\lambda}} \quad \lambda > 0, \quad x \leq x_0$$

On obtient : $x_q = x_0 - \lambda \ell_n(1 - q)$

$$\text{Et : } I_q = \lambda \ell_n\left(\frac{q}{1-q}\right)$$

Ce qui devrait plaire aux amateurs de « log it ».

– Distribution Arc sinus.

Pour $-1 \leq x \leq 1$, on a : $F(x) = \frac{1}{2} + \frac{1}{\Pi} \operatorname{Arc} \sin x$.

D'où (Figure 10(a)) : $f(x) = F'(x) = \frac{1}{\Pi} \cdot \frac{1}{\sqrt{1-x^2}}$.

Ici la masse est concentrée aux extrémités du support, où la densité est infinie.

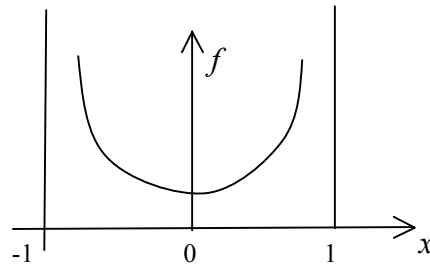


Figure 10(a)

Pour $q > 1/2$, on a : $q - 1/2 = \frac{1}{\Pi} \operatorname{Arc} \sin x_q$

D'où : $x_q = \sin(\Pi(q - 1/2))$

Et (Figure 10(b)) : $I_q = x_q - x_{1-q} = 2 \sin(\Pi(q - 1/2))$

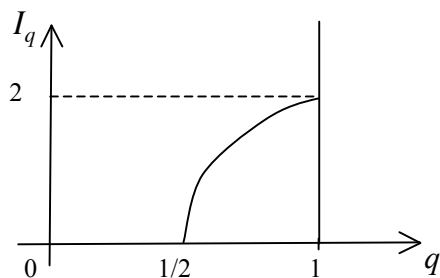


Figure 10(b)

N.B. – On voit sur ces exemples qu'inversement, connaissant l'expression de I_q en fonction de q , on retrouve aisément celle de F (à un changement d'origine près).

– Dans le cas d'observations empiriques : $x = (x_1, x_2, \dots, x_n)$ la fonction de répartition F est une fonction de sauts.

On se ramènera au cas d'une fonction de répartition linéaire par morceaux continue et strictement croissante par interpolation linéaire entre deux sauts successifs.

4. QUATRIÈME IDÉE : LA FONCTION DE DISPERSION DE PAUL LÉVY

L'inconvénient majeur de la mesure de la dispersion par un seul nombre tel qu'un écart ou un diamètre, c'est qu'une même valeur de l'indicateur de concentration utilisé peut correspondre à des réalités très différentes.

Exemple classique : une courbe de densité « en cloche » (cf. Figure 11(a)), où la masse se concentre autour d'une valeur centrale peut avoir même variance qu'une courbe « en U » (cf. Figure 11(b)) où la masse se concentre au contraire aux extrémités de l'intervalle de définition.

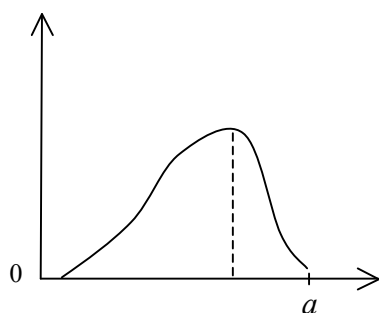


Figure 11(a)

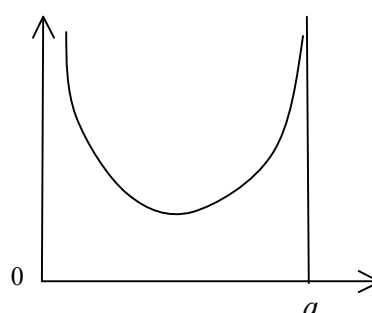


Figure 11(b)

De tels exemples sont faciles à construire. D'où l'intérêt de définitions *fonctionnelles* de la dispersion qui, elles, prennent en compte *toute* la distribution, et pas seulement un paramètre de celle-ci.

On a vu (§ 3) que la considération des quantiles fournit une solution.

L'idée du grand probabiliste Paul Lévy est à première vue assez voisine ; mais on verra qu'en fait elle est notablement différente.

Soit toujours F la fonction de répartition d'une variable X . Supposons-la continue et strictement croissante. Soit alors F^{-1} l'inverse de F .

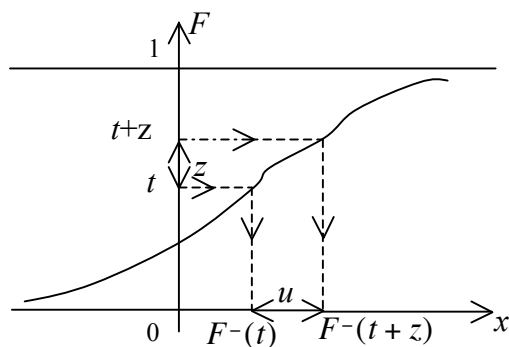


Figure 12

Considérons, sur l'axe des ordonnées, un intervalle $(t, t + z)$ de longueur fixée z , et déplaçons-le en faisant varier t de 0 à $1 - z$.

Pour chaque valeur de t , il y a un intervalle inverse u (cf. Figure 12) :

$$u = F^{-1}(t + z) - F^{-1}(t)$$

Paul Lévy appelle *fonction de dispersion* de la variable X , et note $\omega(z)$, le minimum de u lorsque t varie de 0 à $1 - z$:

$$\omega(z) = \min_t [F^{-1}(t + z) - F^{-1}(t)], \quad 0 \leq z \leq 1$$

Autrement dit, $\omega(z)$ est le plus petit intervalle de variation de X qui contienne la masse z .

Dualement, Paul Lévy définit la *fonction de concentration* de la variable X , soit $Q(y)$ dans ses notations par (cf. Figure 13) :

$$Q(y) = \max_x [F(x + y) - F(x)] \quad y \geq 0$$

C'est le maximum de la masse qui puisse être contenue dans un intervalle de longueur y .

En particulier, $Q(0)$ est le plus grand saut de F ; $Q(0) = 0$ ssi F est continue.

Les fonctions de dispersion et de concentration, dont les définitions sont duales, sont inverses l'une de l'autre :

$$Q = \omega^{-1} \quad \text{et} \quad \omega = Q^{-1}$$

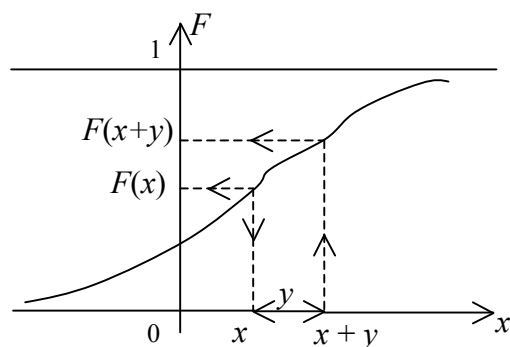


Figure 13

N.B. Les définitions des fonctions ω et Q données ici sont sans problème lorsque la fonction de répartition F est continue et strictement croissante. Dans les autres cas, il y a quelques précautions d'ordre technique à prendre. Nous ne les abordons pas, l'essentiel ici étant que l'idée soit bien comprise.

Revenons à l'exemple d'une distribution « en cloche » comparée à une distribution « en U » ; leurs fonctions de concentration n'ont pas du tout la même allure (cf. Figures 14(a) et 14(b) respectivement).

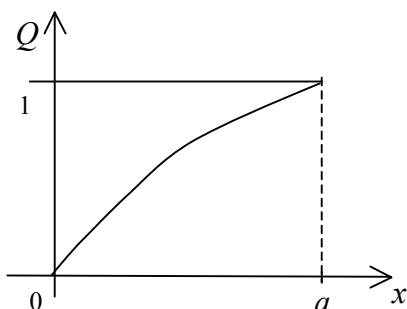


Figure 14(a)

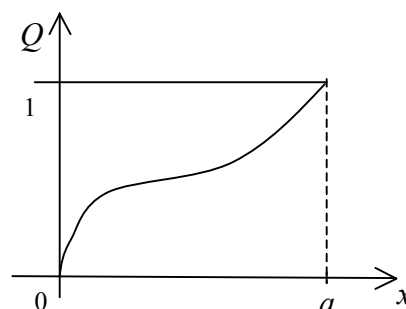


Figure 14(b)

Comme dans le cas des quantiles, on peut se poser la question de l'inversion. Mais là, les choses sont bien différentes.

On a vu dans le paragraphe précédent que la donnée de l'intervalle interquartile médian I_q en fonction de q équivalait, à une translation près, à celle de la fonction de répartition.

Par contre, à une fonction de concentration $Q(y)$ donnée, correspondent :

- dans le cas « en cloche », une infinité, dépendant d'une fonction arbitraire, de fonctions de distributions « en cloche » ayant pu l'engendrer ;
- dans le cas « en U », deux distributions au plus susceptibles de l'avoir engendrée.

Ce qui met bien en évidence la différence profonde entre ces deux types de distributions (unimodale ou bi-modale). Au sujet de cette inversion, on peut voir [Barbut, 2002].

La fonction de concentration de Lévy satisfait à deux inégalités remarquables, du type « inégalité triangulaire ».

- Si X et Y sont deux variables indépendantes en probabilité, désignons par Q_X et Q_Y respectivement leurs fonctions de concentration, et par Q_{X+Y} celle de leur somme.

Alors on a uniformément : $\forall t \geq 0, Q_{X+Y}(t) \leq \min[Q_X(t), Q_Y(t)]$. Le second membre de cette inégalité est évidemment inférieur à $Q_X(t) + Q_Y(t)$.

- Si Q est fonction de concentration d'une variable X donnée, elle satisfait à l'inégalité : $\forall x \geq 0, \forall y \geq 0, Q(x+y) \leq Q(x) + Q(y)$.

Cette inégalité, à bien distinguer de celle qui définit la convexité d'une fonction, est d'ailleurs condition suffisante pour que la fonction de répartition d'une variable positive soit une fonction de concentration.

Sur ces deux points, on pourra consulter [Lévy, 1937] ou [Barbut, 2002].

5. CINQUIÈME IDÉE : LES « SURVIES »

5.1 UNE DISPERSION UNILATÉRALE

En démographie, l'une des variables X beaucoup étudiées est l'âge auquel on meurt, c'est-à-dire la durée de la vie.

Soit un âge x , et $m(x)$ la durée moyenne des vies supérieures à x . La différence :

$$s(x) = m(x) - x$$

est la *survie* moyenne à l'âge x .

On peut d'ailleurs remplacer la durée moyenne par la durée *médiane* $\mu(x)$ des vies supérieures à x . On obtient ainsi comme survie médiane, que les démographes appellent souvent *survie probable* :

$$\sigma(x) = \mu(x) - x$$

Certains statisticiens (W. Lexis, E.J. Gumbel) calculaient les fonctions s et σ à partir d'un âge x particulier, celui où la mortalité est maximum, le mode de la distribution dénommé *vie normale* ; si on le note x_0 , on obtient ainsi :

$$s(x_0) = m(x_0) - x_0 \quad \sigma(x_0) = \mu(x_0) - x_0$$

Si l'on considère (à tort, comme on l'a souligné ci-dessus § 1.3) la vie normale x_0 comme valeur centrale de la distribution des durées de vie, les deux normes $s(x_0)$ et $\sigma(x_0)$ peuvent être considérées comme des *mesures unilatérales de la dispersion* à partir de l'âge modal (« normal ») x_0 .

C'est ainsi que Gumbel [1938] avait relevé qu'aux États-Unis et au début du XX^e siècle, la dispersion ainsi calculée des survies était beaucoup plus grande pour les noirs que pour les blancs, alors que ces derniers avaient une « vie normale » très supérieure à celle des noirs.

Mais on peut plus généralement considérer les fonctions $s(x)$ et $\sigma(x)$ pour chaque valeur possible de la variable x . On obtient ainsi deux expressions *fonctionnelles* de la dispersion pour les valeurs supérieures à x .

Ces expressions ont d'ailleurs un sens quelle que soit la variable X en question. Ce sens est particulièrement clair pour les distributions de revenus ou de richesses (cf. § 6.2 ci-dessous).

5.2 EXEMPLES THÉORIQUES

Dans ce paragraphe, $P(x)$ désigne toujours la proportion $Pr(X > x)$ des valeurs de la variable supérieure à x ; dans certains cas, $P(x)$ peut être interprété comme la probabilité pour que X soit supérieur à x .

$$\text{– Distribution de Pareto : } P(x) = \left(\frac{x_0 + c}{x + c} \right)^\alpha, \quad x \geq x_0 > -c \quad \alpha > 0$$

$$s(x) \text{ n'est définie que pour } \alpha > 1, \text{ et vaut : } s(x) = \frac{x + c}{\alpha - 1}$$

$$\sigma(x) \text{ est défini pour tout } \alpha > 0, \text{ et vaut : } \sigma(x) = (2^{1/\alpha} - 1)(x + c).$$

Dans les deux cas, il s'agit de fonctions linéaires, monotones croissantes de x . La dispersion augmente avec x ; elle tend vers l'infini lorsque x augmente indéfiniment.

$$\text{– Distribution exponentielle : } P(x) = e^{-\frac{(x-x_0)}{\lambda}} \quad x \geq x_0 \quad \lambda > 0$$

$$s \text{ et } \sigma \text{ sont toujours définies, et l'on a : } s(x) = \lambda \quad \sigma(x) = \lambda \ell_n(2)$$

Ici, ℓ_n désigne le logarithme népérien.

La dispersion est constante ; on remarquera d'ailleurs que λ est aussi l'écart-type de la distribution.

$$\text{– Une distribution « sigmoïde » : } P(x) = e^{-\frac{x^2}{2S^2}}, \quad x \geq 0.$$

On obtient : $s(x) = e^{\frac{x^2}{2S^2}} G\left(\frac{x}{S}\right) \sqrt{2\Pi}$, où G est la distribution de Laplace-Gauss :

$$\sigma(x) = (x^2 + 2S^2 \ell(2))^{1/2} - x$$

Chacune d'entre elles est fonction monotone décroissante de x , et tend vers 0 lorsque x augmente indéfiniment. Plus x est grand, moins il y a de dispersion.

– Une loi de mortalité : la loi de Gompertz.

La loi de Gompertz est très fréquemment utilisée pour représenter à partir d'un certain âge (40 ans par exemple) et sauf pour les très grands âges (au-delà de 90-95 ans) les courbes de mortalité des démographes et des actuaires. Elle dépend de deux paramètres λ et ν positifs et a pour expression :

$$P(x) = e^{1/\lambda} e^{-u(x)} \quad \text{avec} \quad u(x) = \frac{1}{\lambda} e^{\nu x}, \quad \lambda, \nu > 0$$

v est le taux (ou « quotient ») de mortalité à l'âge de mortalité maximum (la « vie normale ») et λ le rapport de ce taux au taux de mortalité à l'origine choisie pour x .

On démontre que (cf. par exemple, [Barbut, 2009]) :

$$s(x) = \frac{1}{v} \Phi[u(x)] \quad \text{où} \quad \Phi(z) = e^z \int_z^\infty \frac{e^{-t}}{t} dt, \quad z > 0$$

L'intégrale intervenant dans l'expression de Φ est parfois appelée le « logarithme intégral ».

Quant à la survie probable, son expression est :

$$\sigma(x) = \frac{1}{v} \ell_n(e^{vx} + \lambda \ell_n 2) - x$$

L'une et l'autre sont fonctions monotones décroissantes de x , tendant vers zéro lorsque x augmente indéfiniment.

Mais il semble qu'aux très grands âges, s et σ cessent de décroître et restent supérieurs à une borne strictement positive, comme c'est le cas pour la distribution exponentielle.

5.3 INVERSION DE LA SURVIE MOYENNE

Inversement et sous des conditions assez générales, si l'on connaît l'expression $s(x)$ de la survie moyenne, on peut retrouver $P(x)$; le comportement est donc ici analogue à celui des intervalles interquantiles (cf. § 3) et non à celui de la fonction de concentration de Lévy (cf. § 4).

Supposons que $P(x)$ est dérivable, de dérivée $P' = -f(x)$, où f est la densité de la distribution. Et soit $s(x)$ donnée, continue, strictement positive et dérivable, avec $s' \geq -1$.

$$\text{On a par définition : } m(x) = \frac{1}{P(x)} \int_x^\infty t f(t) dt$$

$$\text{D'où : } s(x) = \left(\frac{1}{P(x)} \int_x^\infty t f(t) dt \right) - x$$

$$\text{D'où : } s(x)P(x) + xP(x) = \int_x^\infty t f(t) dt$$

$$\text{Soit, en dérivant : } s'P + sP' + P + xP' = -x f(x) = xP'$$

Après simplification, il vient donc :

$$(8) \quad \frac{P'}{P} = -\left(\frac{s'}{s} + \frac{1}{s} \right)$$

Comme $s' \geq -1$, on a bien, comme il se doit, $P' \leq 0$.

Désignons par $S(x)$ une primitive de $\frac{1}{s(x)}$: $S'(x) = \frac{1}{s(x)}$.

Par intégration de (8) on a : $\ell_n(P) = -\ell_n(s) - S + c^{te}$

D'où : $P(x) = K \frac{1}{s(x)} e^{-S(x)} = K S'(x) e^{-S(x)} = K(e^{-S(x)})'$

$P(x)$ est donc bien déterminé à la constante multiplicative K près.

6. SIXIÈME IDÉE : MESURER L'INÉGALITÉ PAR DES FONCTIONS

De même que la dispersion, l'inégalité est susceptible de mesures fonctionnelles, bien plus efficaces que les simples indices ou coefficients numériques de l'inégalité, qui ont été évoqués au § 1.4 ci-dessus.

Nous en présentons sommairement deux, l'un classique, l'autre très méconnu.

6.1 LES FONCTIONS DE CONCENTRATION DE LORENZ ET GINI

Elles concernent l'inégalité d'une distribution d'une variable X par rapport à une autre distribution de la même variable (pour une initiation détaillée, cf. [Barbut 1998]).

Prenons l'exemple usuel où X est le revenu distribué dans une population donnée. La proportion de la population ayant un revenu supérieur à x est notée $P(x)$:

$$\Pr(X > x) = P(x)$$

Notons de même $Q(x)$ la proportion du revenu total distribuée à ceux qui ont un revenu supérieur à x .

$P(x)$ est évidemment fonction monotone non croissante de x . Supposons-la, dans le cas de la distribution *théorique* de X continue, strictement décroissante. Alors à chaque valeur de P correspond une valeur x , soit $x(P)$, bien déterminée de X et par conséquent une valeur bien déterminée de Q ; Q est ainsi fonction de la « variable » P , comprise entre 0 et 1 :

$$Q(x(P)) = Q[P] \quad 0 \leq P \leq 1$$

La fonction $Q[P]$ est par définition la *fonction de concentration de la distribution Q par rapport à la distribution P* .

Exemple : X obéit à une loi de Pareto simple :

$$(9) \quad 1 - F(x) = P(x) = \left(\frac{x_0}{x}\right)^\alpha \quad x \geq x_0 \geq 0, \quad \alpha > 1$$

D'où : $dF(x) = -dP(x) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}}$

Le revenu moyen distribué est :

$$m_0 = \int_{x_0}^{\infty} t dF(x) = \alpha x_0^\alpha \int_{x_0}^{\infty} \frac{dt}{t^\alpha} = \frac{\alpha}{\alpha-1} x_0$$

La partie de ce revenu détenue par ceux qui ont un revenu supérieur à x est :

$$R(x) = \int_x^\infty t dF(t) = \alpha x_0^\alpha \int_x^\infty \frac{dt}{t^\alpha} = \frac{\alpha}{\alpha-1} \cdot \frac{x_0^\alpha}{x^{\alpha-1}}$$

D'où pour la proportion $Q(x)$:

$$Q(x) = \frac{R(x)}{m_0} = \left(\frac{x_0}{x}\right)^{\alpha-1}$$

Or, d'après (9) :

$$\frac{x_0}{x} = P^{1/\alpha}$$

Donc, finalement, pour Q :

$$Q[P] = P^{\frac{\alpha-1}{\alpha}}$$

La Figure (15) est celle de la *courbe de concentration* (de Lorenz et Gini) correspondante.

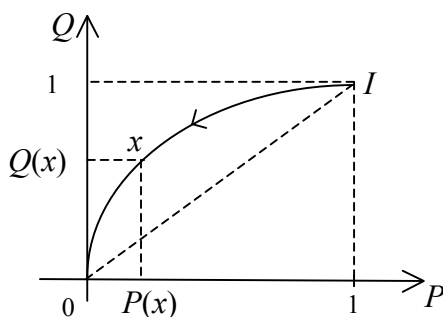


Figure 15

Lorsque x croît de x_0 à l'infini, elle est parcourue, dans le sens indiqué sur la figure, du point I à l'origine 0.

N.B. (1) Dans le cas des observations, P et Q sont des distributions en escalier, avec un nombre fini de sauts. On peut interpoler linéairement entre deux sauts successifs pour se ramener aux conditions ci-dessus.

(2) Dans la suite, on omettra les crochets [] pour simplifier les notations.

(3) Dans le cas particulier donné en exemple, m_0 n'est défini que si l'exposant α est supérieur à 1 ; pour $0 < \alpha \leq 1$, m_0 serait infini et Q n'est plus définie.

D'une façon générale, la fonction de concentration n'existe que si le moment d'ordre 1 (pour P) est défini.

(4) Rappelons enfin que la courbe de concentration de Lorenz et Gini est toujours convexe, et que, pour toute proportion P ($0 < P \leq 1$) donnée de la population, $Q[P]$ est le maximum de revenu que peut détenir cette proportion.

D'une façon générale, et pourvu que le moment d'ordre 1 existe, on a :

$$(10) \quad Q(x) = \frac{1}{m_0} \int_x^\infty t dF(t) = -\frac{1}{m_0} \int_x^\infty t dP(t)$$

D'où, en différentiant :

$$dQ = -\frac{1}{m_0} (-x dP) = \frac{x}{m_0} dP$$

$$\text{Et : } (11) \quad \frac{dQ}{dP} = \frac{x}{m_0}$$

Il en résulte qu'en général la connaissance de la fonction $Q[P]$ entraîne celle de P ; c'est l'*inversibilité* de la fonction de concentration.

En effet, supposons $Q[P]$ convexe, dérivable et de dérivée par rapport à P continue. On a :

$$\frac{dQ}{dP} = h(P)$$

où h est monotone décroissante lorsque P varie de 0 à 1 ; elle est donc inversible, et l'on a, d'après (11) ci-dessus :

$$P = h^{-1}\left(\frac{x}{m_0}\right)$$

Ce qui définit P à un changement d'échelle près sur la variable x .

6.2 INÉGALITÉ AU SENS DE G. SOREL ET M. FRÉCHET

Georges Sorel [1897], commentant la loi de Pareto, faisait remarquer que « le sentiment que l'on a de l'inégalité » était bien traduit par le rapport :

$$\Phi(x) = \frac{m(x)}{x}$$

où, comme dans le §5 *supra*, $m(x)$ est le revenu moyen de ceux qui ont un revenu supérieur à x . En effet, plus ce rapport de $m(x)$ à x est grand, plus celui dont le revenu est x ressentira une infériorité par rapport aux plus riches que lui.

Maurice Fréchet [1925] proposait de prendre la fonction $\Phi(x)$ du revenu x comme mesure de l'inégalité.

Il s'agit-là d'une mesure *fonctionnelle*, d'ailleurs liée à la fonction de survie $s(x) = m(x) - x$ du §5 par la relation :

$$\Phi(x) = \frac{m(x)}{x} = \frac{s(x) + x}{x} = 1 + \frac{s(x)}{x}$$

Il en résulte immédiatement que, comme $s(x)$, la fonction indicatrice de l'inégalité $\Phi(x)$ est *inversible* : sa connaissance entraîne celle de la distribution parente $P(x)$, à un facteur près. On a en effet :

$$s(x) = x(\Phi(x) - 1)$$

Il suffit ensuite d'appliquer le calcul sur $s(x)$ fait au §5.2 ci-dessus.

6.3 RELATION ENTRE CES DEUX MESURES DE L'INÉGALITÉ

On a, par définition de $m(x)$:

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

d'où, d'après (10) :

$$\frac{m(x)}{m_0} = \frac{Q(x)}{P(x)} = \operatorname{tg} \theta \quad (\text{cf. Figure 16})$$

D'autre part, (11) se traduit géométriquement par :

$$\frac{x}{m_0} = \frac{dQ}{dP} = \operatorname{tg} \varphi$$

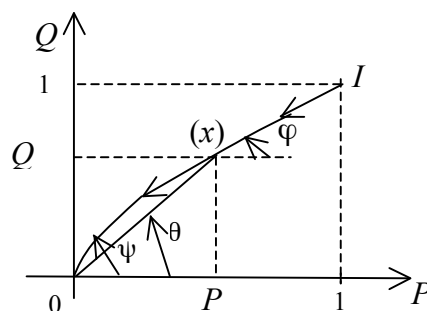


Figure 16

D'où :

$$(12) \quad \Phi(x) = \frac{Q}{P} \cdot \frac{dQ}{dP} = \frac{dP/P}{dQ/Q} = \frac{\operatorname{tg} \theta}{\operatorname{tg} \varphi}$$

La relation (12) fournit donc une expression de l'indicatrice de Fréchet au moyen de la fonction de concentration de Lorenz et Gini.

Elle montre en outre qu'en général, si x tend vers l'infini, le point (x) tend vers 0, et les angles θ et φ tendent tous deux vers Ψ , angle de l'axe des abscisses avec la tangente de 0 à la courbe de concentration. Donc, en général :

$$\lim_{x \rightarrow \infty} \Phi(x) = 1$$

Exceptionnellement, si $\Psi = \frac{\Pi}{2}$, $\Phi(x)$ devient, lorsque x croît indéfiniment, une forme indéterminée de type $\frac{\infty}{\infty}$. Celle-ci peut éventuellement avoir une limite distincte de 1.

C'est notamment le cas pour les distributions de Pareto (la limite est $\frac{\alpha}{\alpha-1}$) et celle de Galton (la limite est 2, pour la forme normalisée).

N.B. Lorsque le moment d'ordre 1 est infini, on peut remplacer la moyenne par la médiane $\mu(x)$, et prendre comme fonction indicatrice de l'inégalité :

$$\Psi(x) = \frac{\mu(x)}{x}$$

Mais ici, l'inversion est moins facile à effectuer.

BIBLIOGRAPHIE

[Il n'est pas usuel qu'un auteur renvoie pour l'essentiel à ses propres articles. Je le fais ici parce qu'il s'agit de textes dont je suis sûr qu'ils sont à la portée de tous].

ARMATTE M. (1991), « Théorie des erreurs, moyenne et loi normale », *Moyenne, milieu, centre*, Paris, éd. de l'EHESS, p. 63-84.

BARBUT M. (1986(a)), « Sur deux familles de valeurs centrales généralisant la moyenne arithmétique : moyennes d'ordre α , centres d'ordre α », *Démographie et sociologie*, Paris, Publications de la Sorbonne, p. 131-145.

BARBUT M. (1986(b)), « Diamètres et écarts », *Mathématiques et Sciences humaines* 93, p. 61-69.

BARBUT M. (1988), « Sur une classe de résumés statistiques : les valeurs centrales », *L'à-peu-près*, Paris, éd. de l'EHESS, p. 109-141.

BARBUT M. (1991), « Moyennes et valeurs centrales », *Moyenne, milieu, centre*, Paris, éd. de l'EHESS, p. 135-151.

BARBUT M. (1998), « Une introduction élémentaire à l'analyse mathématique des inégalités », *Mathématiques et Sciences humaines* 142, p. 27-41. [Repris dans Barbut, 2007, chap. 2].

BARBUT M. (2002), « Une définition fonctionnelle de la dispersion en statistique et en calcul des probabilités : les fonctions de concentration de Paul Lévy », *Mathématiques et Sciences humaines* 158, p. 31-57. Repris dans Barbut, 2007, chap. 5].

BARBUT M. (2007), « Idéologie et modélisation mathématique », *La mesure des inégalités*, Genève, Droz, chap. 10, p. 193-205.

BARBUT M. (2009), « Tablas y curvas de mortalidad en el siglo XVIII », *Historia de la probabilidad y de la estadística* IV, Universidad de Huelva publicaciones, chap. 9, p. 109-151.

FRÉCHET M. (1925), « Une nouvelle représentation analytique de la répartition des revenus », *Actes de la XVIe session de l'Institut international de statistique*, Rome, p. 1-3.

GUMBEL E.J. (1937), *La durée extrême de la vie humaine*, Paris, Hermann.

HARDY G., LITTLEWOOD J., POLYA G. (1934), *Inequalities*, Cambridge, Cambridge University Press.

LÉVY P. (1937), *Théorie de l'addition des variables aléatoires*, Paris, Gauthier-Villars, § 16, p. 43-46, § 29, p. 89-93, § 42, p. 130-133.

SOREL G. (1897), « La loi des revenus », *Le devenir social* 7, p. 578-607.