

MARC BARBUT

**Une famille de distributions : des paretiennes aux « contra-paretiennes ».
Applications à l'étude de la concentration urbaine et de son évolution**

Mathématiques et sciences humaines, tome 141 (1998), p. 43-72

http://www.numdam.org/item?id=MSH_1998__141__43_0

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1998, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) 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/>

UNE FAMILLE DE DISTRIBUTIONS : DES PARETIENNES AUX "CONTRA-PARETIENNES"

Applications à l'étude de la concentration urbaine
et de son évolution

Marc BARBUT

RÉSUMÉ — *Ce texte est consacré à une famille de distributions statistiques — qui comprend les distributions de V. Pareto, celles du type exponentiel et celles que l'on appellera ici "contra-paretiennes" (ou "anti-paretiennes") — dont l'unité tient à ce qu'elles vérifient toutes une même relation fonctionnelle.*

Celle-ci est d'ailleurs interprétable en termes d'inégalité des distributions ; elle fournit en outre une méthode simple et efficace d'ajustement des distributions de la famille à des "données" observées.

Le premier paragraphe constitue un rappel de celles des propriétés mathématiques de cette famille de distributions qui sont à la base de la méthode d'ajustement.

Le paragraphe 2 met la méthode en application, et en détaille toutes les étapes, sur l'exemple de la distribution de la taille des agglomérations françaises selon le recensement de 1901. Le paragraphe 4 fait de même pour la taille, lors des recensements allant de 1806 à 1990, de l'agglomération parisienne : il montre en outre un exemple empirique de distribution "contra-paretienne".

Les paragraphes 3 et 5 donnent quelques aperçus sur l'évolution temporelle de la concentration urbaine et de la croissance de l'agglomération parisienne.

SUMMARY — A family of distributions : from paretian to "contra-paretian"

We define a family of statistical distributions using a common functional equation they satisfy ; this family includes Pareto's distribution, the exponential distribution and a "contra-paretian" distribution.

A very simple method for fitting these theoretical distributions to empirical data is briefly presented, and applied - with many details, and some success - to the rank-size distribution of french cities, during the 19th and 20th centuries, and its temporal evolution.

1. UNE FAMILLE DE DISTRIBUTIONS

1.1. Moyennes conditionnelles linéaires

Une façon intuitive de percevoir l'inégalité d'une distribution, est de se demander, dans le cas de la distribution des revenus par exemple : ceux qui perçoivent plus que mon revenu x , quel revenu perçoivent-ils en moyenne ? Cette moyenne $m(x)$ des revenus supérieurs à x , est-elle très supérieure au mien ? Si oui, j'aurai un fort sentiment d'inégalité. Si par contre elle diffère peu de mon revenu, je jugerai la distribution comme assez égalitaire.

Ce qui vient d'être dit pour les revenus pourrait l'être également si la variable X dont il s'agit est le patrimoine, ou plus généralement s'il s'agit d'une "richesse" quelconque attachée à chaque individu d'une population, au sens que ce dernier terme a en Statistique.

On peut donc baser l'étude de l'inégalité d'une distribution sur la considération de la fonction $m(x)$, moyenne des valeurs de la variable X en jeu qui sont supérieures ou égales à la valeur x .

Dans la suite, on appellera $m(x)$ la *moyenne conditionnelle* en x .

Désignons par $F(x)$ la fonction de répartition de la variable X . On appelle ainsi la fréquence relative des valeurs de X qui sont inférieures à x ; si X est une variable aléatoire, c'est leur probabilité ; sinon, c'est leur proportion, leur pourcentage. On note :

$$F(x) = \Pr(X \leq x).$$

Pour que la moyenne conditionnelle ait un sens, il faut bien entendu que l'intégrale :

$$m_o = \int_{-\infty}^{+\infty} t dF(t)$$

soit définie ; m_o est la *moyenne* de X sur l'ensemble de ses valeurs possibles ; dans le cas où X est une variable aléatoire, m_o est souvent appelée son espérance mathématique.

L'expression "théorique" de la fonction $m(x)$ est alors :

$$(1) \quad m(x) = \frac{1}{P(x)} \int_x^{\infty} t dF(t) \quad (\text{où } P(x) = 1 - F(x) = \Pr(X \geq x)).$$

Remarques :

- Si la variable X était l'âge, et la distribution étudiée celle des âges dans une population (la pyramide des âges), $m(x)$ serait la familière "espérance de vie" à l'âge x .
- La fonction $m(x)$ est liée à la fonction de concentration Q de Lorenz et Gini par la relation :

$$(2) \quad m(x) P(x) = m_o Q(x).$$

Mais elle n'est pas équivalente à Q ; c'est une autre façon de représenter fonctionnellement l'inégalité d'une distribution.

Il est clair, d'après la définition (1) de m , que celle-ci satisfait nécessairement (sur l'ensemble où la variable X est définie) aux deux conditions :

$$(3.1) \quad m(x) \text{ est monotone non décroissante}$$

$$(3.2) \quad m(x) \geq x \quad (\text{et en particulier, } m(o) \geq 0).$$

Le cas le plus simple, le seul qui sera envisagé dans la suite, est celui où $m(x)$ est une fonction linéaire (affine) de x :

$$(4) \quad m(x) = \beta x + h.$$

Remarque : L'expression (4) de la moyenne conditionnelle signifie que, lorsqu'elle a cette expression mathématique, l'inégalité est fonction croissante des deux paramètres β et h .

En particulier, lorsque la variable peut prendre de grandes valeurs, c'est le terme βx qui devient prépondérant.

Si, par exemple, $\beta = 3$, le revenu moyen de ceux qui ont un revenu supérieur à x en est à peu près le triple. Pour $\beta = 2$, ce n'est plus que le double : l'inégalité a diminué.

Quand β est proche de 1, la distribution devient assez égalitaire.

Compte tenu de (3.1) et (3.2), on a nécessairement :

$$\beta \geq 0 \quad \text{et} \quad h \geq 0.$$

- Si $\beta = 0$, on a, quel que soit x , $m(x) = h$, donc $x = h$: la variable étudiée n'a qu'une valeur possible.
- Si $\beta > 0$, la condition (3.2) implique (Figure 1.1) que seules sont possibles les valeurs de x correspondant à la partie de la droite d'équation (4) située au-dessus de la première bissectrice des axes.

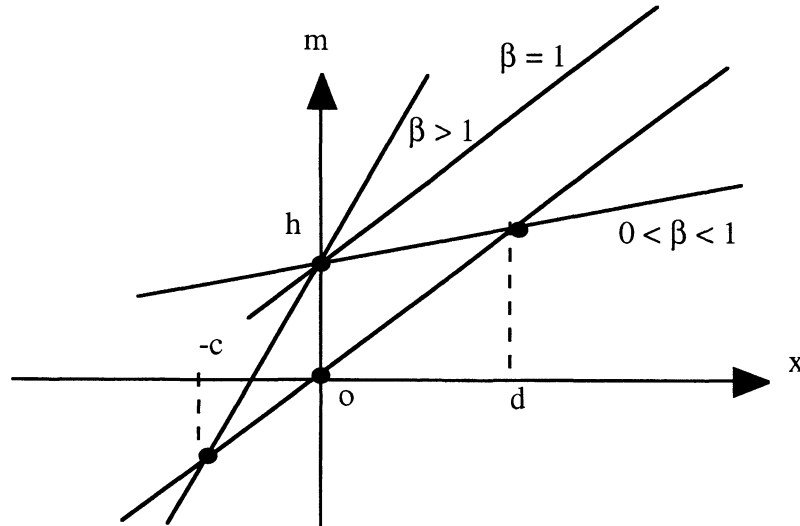


Figure 1.1

Il y a donc trois cas possibles :

1. $\beta > 1$. Il faut : $x \geq -c$ $\left(c = \frac{h}{\beta-1} \geq 0 \right)$. Le *support* de la distribution (c'est-à-dire l'ensemble des valeurs de X qui sont effectivement possibles) est borné inférieurement.
2. $0 < \beta < 1$. Il faut : $x \leq d$ $\left(d = \frac{h}{1-\beta} \geq 0 \right)$. Le support de la distribution est borné supérieurement.
3. $\beta = 1$. C'est le cas intermédiaire où la droite d'équation (4) est parallèle à la première bissectrice. Le support n'est pas borné a priori.

En pratique, pour les variables en jeu dans les phénomènes où l'on s'intéresse à l'inégalité de la distribution (revenus, patrimoines, tailles d'agglomérations mesurées en nombre d'habitants, surfaces agricoles, etc...), il y a un *seuil minimum* x_0 , en général positif, pour le support de la distribution.

On détermine très simplement quelles sont, parmi les distributions absolument continues, c'est-à-dire telles que la fonction de répartition F ait une dérivée :

$$F'(x) = f(x)$$

les distributions ayant une moyenne conditionnelle de la forme linéaire (4) (Cf. [1] de la bibliographie). Voici ce qu'il en est dans les 3 cas énumérés ci-dessus.

1.2. 1^{er} cas $\beta > 1$ — Distributions de Pareto

On a : (5)
$$P(x) = \left(\frac{x_0 + c}{x + c} \right)^\alpha = K(x + c)^{-\alpha} \quad \text{pour } x \geq x_0, (x_0 > -c),$$
 et un 1 sinon.

Le paramètre c est celui défini plus haut : $c = \frac{h}{\beta - 1}$. Quant à l'exposant α , il vaut : $\alpha = \frac{\beta}{\beta - 1} > 1$. La relation entre α et β est d'ailleurs réciproque, et peut s'écrire :

$$\frac{1}{\alpha} + \frac{1}{\beta} = 1.$$

Lorsque l'on aura à ajuster sur une telle distribution de Pareto une distribution empirique (Cf. §.2 infra), portant sur un nombre total N_0 d'observations, on pourra écrire, comme le faisait V. Pareto lui-même :

$$(5') \quad N(x) \approx N_0 \left(\frac{x_0 + c}{x + c} \right)^\alpha \quad \approx : \text{"peu différent de 1"} \\ \text{"presque égal à"}$$

$N(x)$ est le nombre des observations supérieures ou égales à x .

Les distributions de Pareto restent définies lorsque l'exposant α , positif, de (5) et (5') est inférieur ou égal à 1.

Mais en ce cas, elles n'ont plus de moyenne théorique (l'intégrale du second membre de (1) est infinie) ; on dira plus loin un mot de ce cas.

D'une façon générale, le *moment d'ordre* k , m_k :

$$(6) \quad m_k = \int_{x_0}^{\infty} t^k dF(t)$$

n'existe, pour les distributions de Pareto, que si l'entier k est strictement inférieur à α .

En particulier, pour $\alpha \leq 2$ (i.e. $\beta \geq 2$), ce qui est souvent le cas dans les applications, la distribution de Pareto *n'a pas de variance* (l'intégrale (6) est infinie pour $k = 2$), ni par conséquent d'écart-type.

Il s'agit-là d'une propriété la distinguant radicalement de la plupart des distributions présentées dans les cours de statistique, et en particulier de la loi dite "normale", ou de Laplace et Gauss.

Le *comportement asymptotique* de ces deux types de distribution permet de bien voir en quoi d'essentiel elles diffèrent.

La "normale", de densité en e^{-ax^2} ($a > 0$), se rapproche infiniment plus vite de son asymptote que les distributions paretienne dont la densité est, pour x grand, de l'ordre de $\frac{1}{x^{\alpha+1}}$, avec $\alpha > 0$.

Un *critère* commode pour savoir a priori, en présence d'une distribution empirique, dans quel cas l'on se trouve, est le suivant :

- Considérons par exemple la taille des conscrits, dont chacun sait qu'elle est distribuée "normalement" autour de sa moyenne. Posons-nous la question : quelle chance y a-t-il de tomber sur un conscrit dont la taille est plus du double de la taille moyenne ? Réponse : aucune, ou quasiment.
- Prenons maintenant la distribution des revenus fiscaux en France ; la proportion de ceux qui dépassent le double de la moyenne est de l'ordre de 6 à 8%. Ce n'est pas négligeable. Pour la distribution des agglomérations françaises selon leur taille, que l'on étudiera ci-dessous, §.2 et sq. les ordres de grandeurs sont les mêmes. Nous sommes donc en présence d'une toute autre classe de phénomènes que ceux dont rend compte la "loi normale".

Celle-ci est faite pour des variables concentrées autour de leur moyenne, avec des variations de faible amplitude, dont toutes les valeurs pratiquement possibles sont du même ordre de grandeur : celles de "l'homme moyen" cher à A. Quetelet.

Par contre, pour le type paretien, les *grandes valeurs ne sont pas rares*, ont une probabilité non négligeable. Plaçons-nous par exemple dans le cas où $c = 0$; la moyenne m_1 de la distribution de Pareto (5) est égale à :

$$m_1 = \beta x_0 .$$

D'où :

$$\Pr(X > 2m_1) = \left(\frac{x_0}{2m_1}\right)^\alpha = \left(\frac{1}{2\beta}\right)^\alpha .$$

Pour $\beta=3$, valeur fréquente (approximativement) dans certaines applications, on a ainsi :

$$\alpha = \frac{3}{2} .$$

Et :

$$\Pr(X > 2m_1) = \left(\frac{1}{6}\right)^{\frac{3}{2}} = \left(\frac{1}{216}\right)^{\frac{1}{2}} \simeq 7\% .$$

Nous sommes bien dans les ordres de grandeur annoncés.

De telles variables, que l'on rencontre souvent dans les Sciences Sociales ou Humaines, concerneront ce qu'on peut appeler, par opposition à "l'homme moyen", *l'homme extrême* de Vilfredo Pareto.

Remarques.

- Le grand mathématicien du Calcul des Probabilités Paul Lévy a, en 1935, déterminé quelles étaient toutes les distributions de type *stable* par *rappor*t à l'addition des variables aléatoires indépendantes, c'est-à-dire celles que l'on doit s'attendre à trouver, à un changement près d'unité et d'origine, lorsqu'un modèle mathématique linéaire (additif) semble raisonnable en première approximation.

Ce sont, outre la loi de Laplace et Gauss, des distributions qui sont toutes asymptotiquement paretienne, d'exposant α inférieur à 2 (pour un bref exposé synthétique, voir par exemple [2]. Pour la genèse des idées et les démonstrations, lire P. Levy [8]). Ce théorème de P. Levy montre en outre que ces distributions ne correspondent pas *a priori* à des phénomènes "qui ne sont pas au hasard", comme certains le croient naïvement ; il dit même exactement le contraire.

- Les distributions paretiennes ont été retrouvées, dans leur propre domaine, par des spécialistes d'autres disciplines, dans l'ignorance, en général, des travaux antérieurs de Pareto, et de l'identité de la "nouvelle" loi trouvée avec celle qu'il avait découverte.

C'est ainsi que l'usage perdure encore aujourd'hui de parler de la loi de Zipf-Auerbach chez les Géographes, de celle de Zipf (tout court) ou d'Estoup-Zipf chez les Lexicologues.

La bonne dénomination semble être pour toutes les lois approximativement et asymptotiquement paretiennes, celle proposée par B. Mandelbrot : distributions de Pareto-Levy, ce qui rend hommage à la fois à leur inventeur et à leur principal théoricien.

De telles distributions se trouvent d'ailleurs aussi dans les Sciences de la Nature ; elles ne sont en rien spécifiques des Sciences Sociales.

1.3. 2^{ème} cas. $0 < \beta < 1$ — Distributions "contra-paretiennes"

Ici, on a, $x_0 \leq x \leq d$; dans cet intervalle, l'expression de P est :

$$(7) \quad P(x) = \left(\frac{d-x}{d-x_0} \right)^{\gamma} \quad \begin{array}{l} (P(x) = 1 \text{ si } x < x_0 ; \\ P(x) = 0 \text{ si } x > d) \end{array}$$

d est la borne supérieure déterminée supra : $d = \frac{h}{1-\beta}$. Quant à l'exposant γ , positif, il est égal à :

$$\gamma = \frac{\beta}{1-\beta} = -\alpha \quad (\alpha = \frac{\beta}{\beta-1} \text{ est négatif})$$

Il s'agit de distributions en fonction puissance positive, qui sont aussi un cas particulier des distributions dites Beta de K. Pearson.

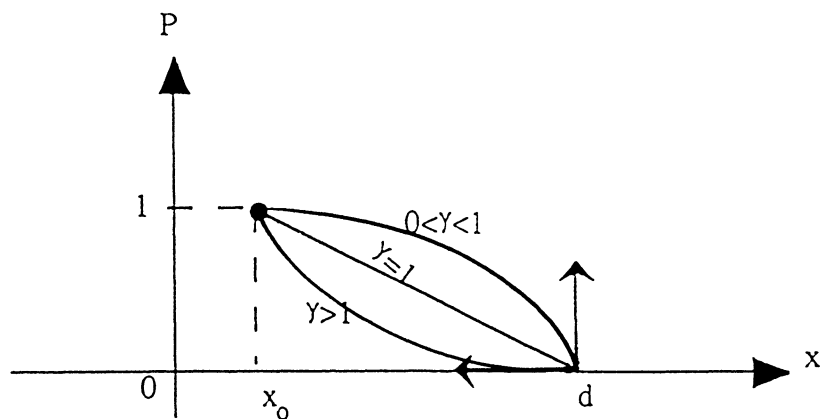


Figure 1.2.

La figure 1.2 montre leur allure, selon que :

- $\gamma > 1$, i.e. $\frac{1}{2} < \beta < 1$
- $\gamma < 1$, i.e. $0 < \beta < \frac{1}{2}$
- $\gamma = 1$, i.e. $\beta = \frac{1}{2}$: distribution uniforme sur l'intervalle (x_0, d) .

Je propose d'appeler ces distributions "contra-paretiennes" (on pourrait aussi dire "anti-paretiennes") parce que dans la famille étudiée ici, tout γ est à l'opposé (et "en dualité") de ce qui se passe pour celles de Pareto :

- la variable est bornée supérieurement, au lieu de l'être inférieurement,
- a fortiori, ses moments existent à tous les ordres,
- l'exposant est positif, au lieu d'être négatif ($\frac{1}{x^\alpha}$ s'écrit $x^{-\alpha}$),
- enfin, les grandes valeurs sont plus que négligeables, elles sont rigoureusement *impossibles*, au lieu de n'être pas rares.

Remarque : Dans ce cas où β est inférieur à 1, et compte-tenu de l'égalité $h = d(1-\beta)$, la moyenne conditionnelle s'écrit :

$$(8) \quad m(x) = \beta x + (1-\beta) d$$

Elle est le barycentre de l'intervalle séparant x de la borne supérieure d .

Si β est voisin de 1, $m(x)$ est voisine de x : situation assez égalitaire.

Si β est voisin de 0, $m(x)$ est proche du maximum d , quelque soit x : situation très inégalitaire.

1.4. 3^{ème} cas. $\beta = 1$ — Distributions exponentielles

Ici, on a :

$$(9) \quad P(x) = e^{-\left(\frac{x-x_0}{h}\right)}, \quad x \geq x_0$$

C'est la distribution exponentielle, qui apparaît ainsi comme intermédiaire entre le cas paretien et le cas "contra-paretien".

Ses moments existent à tous les ordres, et elle converge vers 0, lorsque x augmente indéfiniment, plus vite que toutes les paretiennes ; mais son support n'est pas borné supérieurement.

La moyenne conditionnelle $m(x)$ se déduit ici de x par addition d'une constante :

$$m(x) = x + h.$$

Lorsque x est grand, $m(x)$ devient pratiquement égal à x : les distributions exponentielles ne pourront se rencontrer que dans des situations assez égalitaires.

Remarques.

- La distribution exponentielle, solution de l'équation fonctionnelle (4) lorsque $\beta = 1$, n'est pas pour autant la limite, lorsque β tend vers 1, des distributions parietiennes (ou contra-parietiennes) ; car alors l'exposant α de l'expression (5) (ou γ de l'expression (7)) tend vers l'infini, et la distribution tend vers celle où toutes les valeurs de la variable x (le revenu, par exemple) sont égales au minimum x_0 : c'est l'équirépartition.
- En ce qui concerne la rapidité de convergence de la fonction de répartition vers son asymptote lorsque x augmente indéfiniment, la distribution "normale" de Laplace et Gauss vient s'intercaler entre l'exponentielle et les "anti-parietiennes" ; de même, entre l'exponentielle et les parietiennes, on peut intercaler la distribution Log-normale (ou de Gibrat) : celle d'une variable positive dont le logarithme suit une loi normale.
- La loi Log-normale est parfois mieux adaptée que celle de Pareto pour rendre compte de phénomènes inégalitaires, notamment lorsque l'exposant α est plus grand que 2 (i.e. $1 < \beta < 2$). En ce cas, en effet, la distribution de Pareto a une variance, et nous retombons dans le "domaine d'attraction" de la loi de Laplace-Gauss.

1.5. Médiane conditionnelle, quantile et maximum

Les trois types de distribution pour lesquelles la moyenne conditionnelle est une fonction linéaire de la variable ont d'autres propriétés de linéarité (Cf. [1] de la bibliographie), dues à leur forme analytique. C'est l'une de ces autres propriétés qu'on va maintenant étudier.

Considérons la *médiane* $\mu(x)$ des valeurs supérieures ou égales à x ; il est très facile de voir que c'est une fonction de la forme :

$$(10) \quad \mu(x) = A x + B \quad (A > 0, B \geq 0)$$

Par exemple, dans le cas parietien, on a, dans les notations de (5) :

$$(11) \quad \mu(x) = 2^{1/\alpha} x + c(2^{1/\alpha} - 1) \quad (\text{c'est le cas où } A > 1)$$

Plus généralement, considérons le $k^{\text{ème}}$ quantile de la distribution, défini, pour chaque valeur x de la variable, par :

$$(12) \quad P(\mu_k(x)) = \frac{1}{k} P(x)$$

On peut l'appeler le *quantile conditionnel* d'ordre k . La figure 1.3. représente graphiquement la relation (12), pour $k = 3$.

Alors on a (le calcul est immédiat) :

- 1^{er} cas - Pareto :

$$\mu_k(x) = k^{1/\alpha} x + (k^{1/\alpha} - 1) c$$

- 2^{ème} cas — contra-Pareto :

$$\mu_k(x) = k^{-1/\gamma} x + (1 - k^{-1/\gamma}) d$$

- 3^{ème} cas — exponentielle :

$$\mu_k(x) = x + h \text{ Log } k$$

On a donc dans tous les cas, et pour tout entier k :

$$(13) \quad \mu_k(x) = A_k x + B_k$$

avec :

- $A_k > 1$ dans le premier cas,
- $0 < A_k < 1$ dans le deuxième,
- $A_k = 1$ dans le troisième.

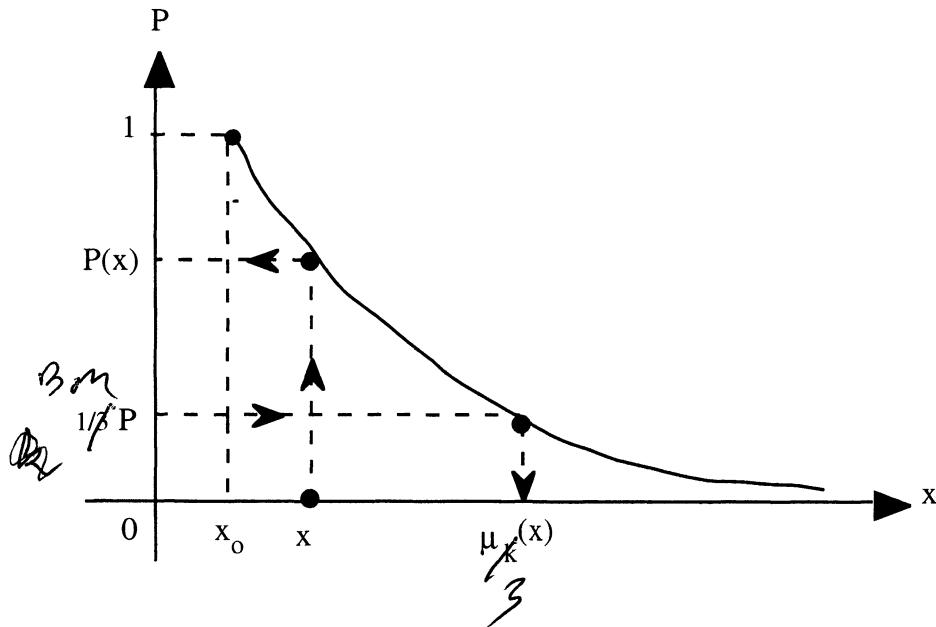


Figure 1.3

Ceci permet de voir le lien entre nos trois types de distributions et les *lois du maximum* trouvées dans les années 1920-1930 par différents auteurs. Expliquons ceci.

D'un échantillon aléatoire d'une même variable X , de fonction de répartition donnée $F(x)$, considérons maintenant, non plus la moyenne, mais la *plus grande valeur* observée, le maximum, soit Y . Si les tirages sont indépendants, on a évidemment :

$$\Pr(Y < x) = [F(x)]^n \quad (n \text{ est la taille de l'échantillon}).$$

Cherchons, à la suite de M. Fréchet et de quelques autres, comment choisir F pour que, à un changement d'origine et d'unité près sur la variable X , le maximum de tout échantillon suive la même loi (une telle distribution sera dite de type *stable pour le maximum*). F doit donc, quelque soit l'entier naturel n , satisfaire à :

$$[F(x)]^n = F(a_n x + b_n) \quad a_n > 0,$$

Soit :

$$n \operatorname{Log} F(x) = \operatorname{Log} F(a_n x + b_n)$$

Posons :

$$G(x) = -\operatorname{Log} F(x).$$

$G(x)$ est positive et monotone décroissante ; elle satisfait à :

$$(14) \quad \forall n \in \mathbb{N} ; \quad n G(x) = G(a_n x + b_n)$$

Dans (14), faisons le changement de variable : $z = a_n x + b_n$; il en résulte :

$$x = A_n z + B_n \quad A_n > 0 ,$$

d'où :

$$(15) \quad \forall n \in \mathbb{N} ; \quad G(A_n z + B_n) = \frac{1}{n} G(z)$$

Aux notations près, c'est l'équation (12) où le $n^{\text{ième}}$ quantile serait de la forme (13).

Nous ne démontrerons pas ici que si tous les quantiles conditionnels sont des fonctions affines de la variable, la distribution est de l'un de nos trois types. Ceci étant admis, on a ainsi trois cas de lois stables pour le maximum :

- 1^{er} cas (loi de Fréchet) : la fonction de répartition est l'exponentielle d'une distribution paretienne de type (5) : $F(x) = \exp\left[-\left(\frac{x_0 + c}{x + c}\right)^\alpha\right] \quad x \geq x_0 > -c, \alpha > 0$.
- 2^{ème} cas (loi de Weibull) : c'est l'exponentielle d'une distribution anti-paretienne de type (7) : $F(x) = \exp\left[-\left(\frac{d - x}{d - x_0}\right)^\gamma\right] \quad x_0 \leq x \leq d, \gamma > 0$.
- 3^{ème} cas (loi de Gumbel) : c'est l'exponentielle d'une exponentielle de type (9).

Pour les grandes valeurs de x (ou celles proches de la borne d dans le 2^{ème} cas), l'exposant de la loi du maximum est proche de zéro ; il en résulte que la fonction de répartition $P = 1 - F$ est alors à peu près égale à cet exposant (dans le langage technique, c'est l'approximation à l'ordre 1).

Ainsi, nos trois types de distribution sont approximativement stables pour le maximum.

C'est, en particulier, le cas pour la distribution paretienne : on retrouve ainsi que "l'homme de Pareto" est bien celui des valeurs extrêmes, au contraire de "l'homme moyen" de Quetelet.

1.6. Application à l'ajustement

Le propriété de linéarité (4) de la moyenne conditionnelle, semble avoir été explicitée pour la première fois en 1925 (deux ans après la mort de Pareto) pour les distributions paretienes, par Maurice Fréchet (Cf. [5]) ; Pareto lui-même ne s'en est probablement jamais aperçu. Elle fournit une méthode simple et efficace d'ajustement de distributions observées à des distributions "théoriques" de l'un de nos trois types.

Il suffit, si les valeurs observées d'une variable donnée sont rangées par valeurs décroissantes :

$$x_1 \geq x_2 \geq \dots \geq x_i \geq \dots \geq x_n$$

de calculer les moyennes empiriques :

$$(16) \quad m_i = \frac{x_1 + x_2 + \dots + x_i}{i} \quad (n \geq i \geq i_0)$$

et d'ajuster linéairement les couples (x_i, m_i) , par telle méthode d'ajustement linéaire que l'on voudra.

Les paragraphes suivants détaillent ce procédé dans le cas de distributions empiriques concernant la Démographie urbaine (et historique).

Préalablement, il convient de faire quatre remarques :

1. Toute la méthode dont on va voir la grande efficacité sur ces exemples "concrets", est basée sur le calcul de moyennes. Pour que celles-ci aient un sens statistique, il faut bien sûr que le dénominateur i des moyennes m_i (formule (16) ci-dessus) soit assez grand ; c'est pourquoi on a indiqué que i doit être au moins égal à un indice minimum i_0 . Comment déterminer approximativement ce minimum, on le verra sur les exemples traités.
2. Il faut être d'autant plus attentif à ce qui vient d'être dit que, pour démontrer la convergence en probabilité de la moyenne d'un échantillon de taille n vers la "moyenne théorique", on utilise le plus souvent la variance de la distribution parente (inégalité de Bienaymé et Tchebicheff, par exemple).

Or, pour les distributions parietiennes d'exposant $1 < \alpha \leq 2$ (on en rencontre souvent dans les applications), il n'y a pas de variance. Heureusement, un théorème, dû à A. Khintchine (1929) montre qu'il suffit, pour que cette convergence soit assurée que la moyenne théorique existe, ce qui est le cas si $\alpha > 1$. Mais, bien sûr, la convergence est ici plus lente que dans le cas de distributions ayant une variance finie.

3. Pour $0 < \alpha \leq 1$, les distributions parietiennes n'ont même plus de moyenne. Mais, nous l'avons vu, (équations (10) et (11) §1.5 supra), elles ont toujours une médiane conditionnelle en x , qui est, elle aussi, fonction affine de x ; on pourra donc appliquer la méthode d'ajustement linéaire, en remplaçant les moyennes observées m_i par les médianes μ_i des i plus grandes observations, pour i supérieur à un seuil i_0 . C'est praticable, mais délicat. On ne traitera pas cette question dans le présent texte (à ce sujet, on peut voir [3]).
4. Une habitude bien ancrée depuis Pareto lui-même, qui procédait ainsi, est de "linéariser" l'ajustement en prenant le logarithme dans l'équation (5'), soit :

$$(17) \quad \text{Log } N(x) = -\alpha \text{Log}(x + c) + C^{\text{te}}$$

Ceci a deux inconvénients majeurs :

- comme c est l'un des paramètres à estimer, elle ne s'applique que pour $c = 0$; ou alors, il ne faut considérer, dans l'application de (17) à un ajustement linéaire, que les *grandes valeurs de* x , (pour lesquelles $x + c \simeq x$), ce qu'on oublie généralement de faire ; Pareto, lui, ne devait pas faire cette erreur, puisqu'il fournit (voir [9] et [10]) de nombreuses valeurs pour c ; mais il n'indique jamais comment il parvient à ses estimations.
- d'autre part, quelque soit la façon dont on procède à un ajustement linéaire :

$$Y = A X + B \quad (\text{dans l'équation (17), } Y = \text{Log } N(x), A = -\alpha, \text{ etc...})$$

on y minimise toujours une fonction des *écarts absolus*, des différences entre valeurs observées y_i , et valeurs estimées $(A x_i + B)$; dans le cas de la formule (17), il s'agira

de *différences de logarithmes*, c'est-à-dire, pour les données elles-mêmes, d'*écarts relatifs*. Comme l'a fort bien dit (et écrit) Pareto lui-même : "ce n'est pas du tout la même chose" (Cf. [11]). Cette difficulté semble l'avoir beaucoup tourmenté.

En particulier, si l'on prend ainsi les écarts relatifs, ce sont justement *les grandes valeurs*, celles qui sont, comme on l'a souligné, les plus importantes, qui voient leur poids très sous-évalué dans le calcul de l'ajustement ; d'où certaines erreurs (et des erreurs certaines).

2. DES DISTRIBUTIONS DE TYPE PARETIEN : LES AGGLOMÉRATIONS FRANÇAISES RANGÉES PAR TAILLE

Les résultats du recensement général de la population française effectué en 1901 sont, d'après les experts, particulièrement fiables.

C'est donc d'abord ceux-ci que l'on va utiliser pour mettre en œuvre la méthode d'ajustement (de distributions paretiennes à des distributions empiriques) qui a été évoquée au §.1.6 supra.

La statistique considérée est celle des *agglomérations* urbaines rangées par taille (population) décroissante. Pour la définition précise de l'agglomération, telle que l'INSEE l'a établie selon des critères de continuité d'habitat et de peuplement autour d'une commune centrale, on pourra consulter [6] de la bibliographie. L'auteur de cet ouvrage, Madame F. Guérin-Pace, chercheur à l'INED, a d'ailleurs eu l'amabilité de me fournir cette statistique non seulement pour l'année 1901, mais pour tous les recensements effectués de 1831 à 1990.

Pour l'étude des données de 1901, on a retenu ici les 233 agglomérations ayant au moins 10.000 habitants. Les colonnes 1, 2 et 6 du tableau 2.1 ci-contre, fournissent un extrait de cette distribution ; l'unité y est le millier d'habitants qu'on peut aussi noter KH. L'agglomération la plus peuplée est, bien sûr, celle de Paris, avec 3.824 KH environ ; la trentième est Dunkerque avec 72,1 milliers d'habitants, etc...

En colonne 3 du tableau, on calcule pour chaque rang la population cumulée :

$$X = x_1 + x_2 + \dots + x_r \quad (x_1 \geq x_2 \geq \dots)$$

et en colonne 4 la moyenne correspondante : $\frac{X}{r} = m(x_r) = m_r$.

C'est ainsi, par exemple, que les 80 premières agglomérations renferment une population cumulée de 10.594,5 milliers d'habitants, et ont une population moyenne de 132,5 KH ; l'ensemble des 233 agglomérations a une population totale de quelque 13.231.500 habitants et une population moyenne m_0 de 56.710 habitants ($m_0 = m_{N_0}$).

Remarque. En colonne 5, on a calculé les rapports $m(x)/x$. C'est parce que, si l'hypothèse d'une distribution paretienne est approximativement vérifiée, on doit avoir (cf. §.1.2. ci-dessus) :

$$m(x) = \beta x + h + \varepsilon(x), \quad \text{avec } \beta > 1 \text{ et } h \geq 0, \varepsilon \text{ est le "résidu".}$$

D'où : $\frac{m(x)}{x} = \beta + \frac{h + \varepsilon}{x} \approx \beta + \frac{h}{x}$ (\approx : approximativement égal à).

La courbe d'équation :

$$y = \beta + \frac{h}{x}$$

est, pour x positif, une branche d'hyperbole, telle que celle représentée sur la Figure 2.1 ci-après.

rang r	population x	population cumulée X	moyenne $m(x) = \frac{X}{r}$	rapport m(x)/x	centre de l'agglomération
1	3.824,35	"	"	1	Paris
2	534,7	4.359	2.179,5	4,1	Lyon
.
10	162	6.435,5	643,5	3,97	Le Havre
.
30	72,1	8.401,25	280	3,83	Dunkerque
.
80	29,4	10.594,5	132,5	4,5	Aix-en Provence
90	27,6	10.880,5	121	4,4	Saint-Brieux
100	25,4	11.144,75	111,5	4,3	Bastia
110	23,1	11.385	103,5	4,5	Lunéville
120	20,9	11.603	96,7	4,6	Cognac
130	19,2	11.781,5	90,6	4,7	Arcachon
140	17,7	11.985,75	85,6	4,8	Hyères
160	15,7	12.321	77	4,9	Autun
180	13,6	12.609,75	70	5,15	Villeneuve/Lot
200	12,1	12.855	64,3	5,3	Givors
216	10,8	13.037	60,4	5,6	Vitré
233	10,1	13.213,5	56,71	5,6	Orange
1	2	3	4	5	6

Tableau 2.1. Recensement de 1901.

Les 233 agglomérations de population supérieure ou égale à 10.000 habitants.
(Dans le tableau, l'unité adoptée est le millier d'habitants, ou KH).

On doit donc constater que ces rapports (dès que les rangs sont assez grands) sont tous supérieurs à une valeur elle-même supérieure à 1, et croissent lorsque la variable x diminue. C'est bien l'indication fournie par la colonne 5 du Tableau 2.1.

Si l'on est dans le cas où $c = 0$ (i.e. $h = 0$), les rapports de la colonne 5 sont à peu près constants.

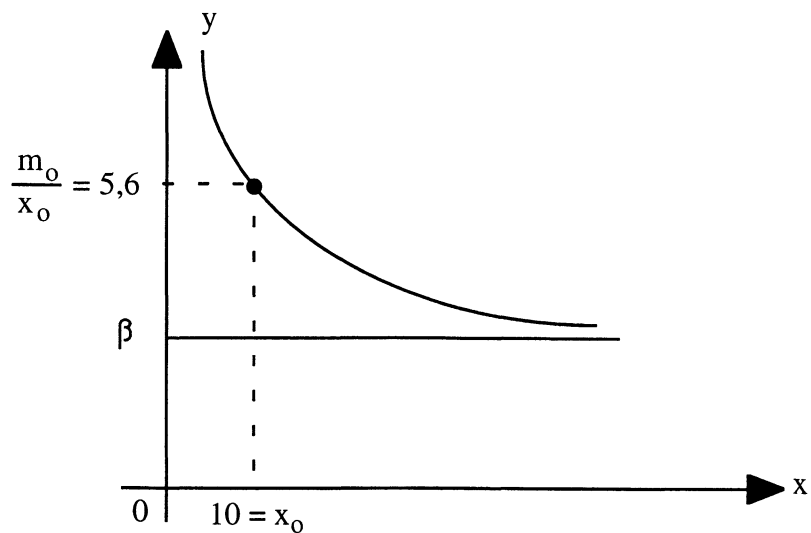


Figure 2.1.

Portons maintenant sur un graphique (Figure 2.2. ci-contre) les douze points d'abscisses x et d'ordonnées correspondantes $m(x)$ (colonnes 2 et 4 du Tableau 2.1), pour les 12 rangs retenus dans le tableau entre 80 et 233 : ils sont assez bien alignés.

Et un simple coup d'œil (ou l'aide d'une règle) fait pressentir, pour une droite d'ajustement d'équation :

$$y = \beta x + h$$

une valeur de β entre 3,5 et 4, et une valeur de h proche de 20.

Avant de vérifier tout cela par le calcul, il faut faire trois remarques :

1. Pour $\beta \geq 2$, ce qui serait le cas ici, l'exposant α de la distribution de Pareto (formules (5) et (5') §.1.2) est compris entre 1 et 2 ; c'est une distribution qui n'a pas de variance (sa variance serait infinie).

Grâce au théorème de Khintchine rappelé au §.1.6, on sait que la convergence en probabilité d'une moyenne d'échantillon de taille n vers la moyenne "théorique", bien que lente, reste assurée lorsque n augmente indéfiniment.

2. Mais bien entendu, une moyenne empirique n'a quelque chance de se rapprocher d'une moyenne théorique que si le dénominateur est assez grand. C'est pour cela que dans l'exemple traité, on n'utilise que des rangs à partir de 80.

Pour les premiers rangs, la moyenne n'a aucun sens statistique. Si l'on avait continué le graphique de la Figure 2.2. vers les petits rangs, on aurait évidemment vu le bel alignement des points disparaître (Figure 2.3). Un procédé pratique pour déterminer à partir de quel rang prendre les moyennes en compte pour l'ajustement, consiste d'ailleurs à construire ce graphique, et voir à partir de quel rang les points s'alignent.

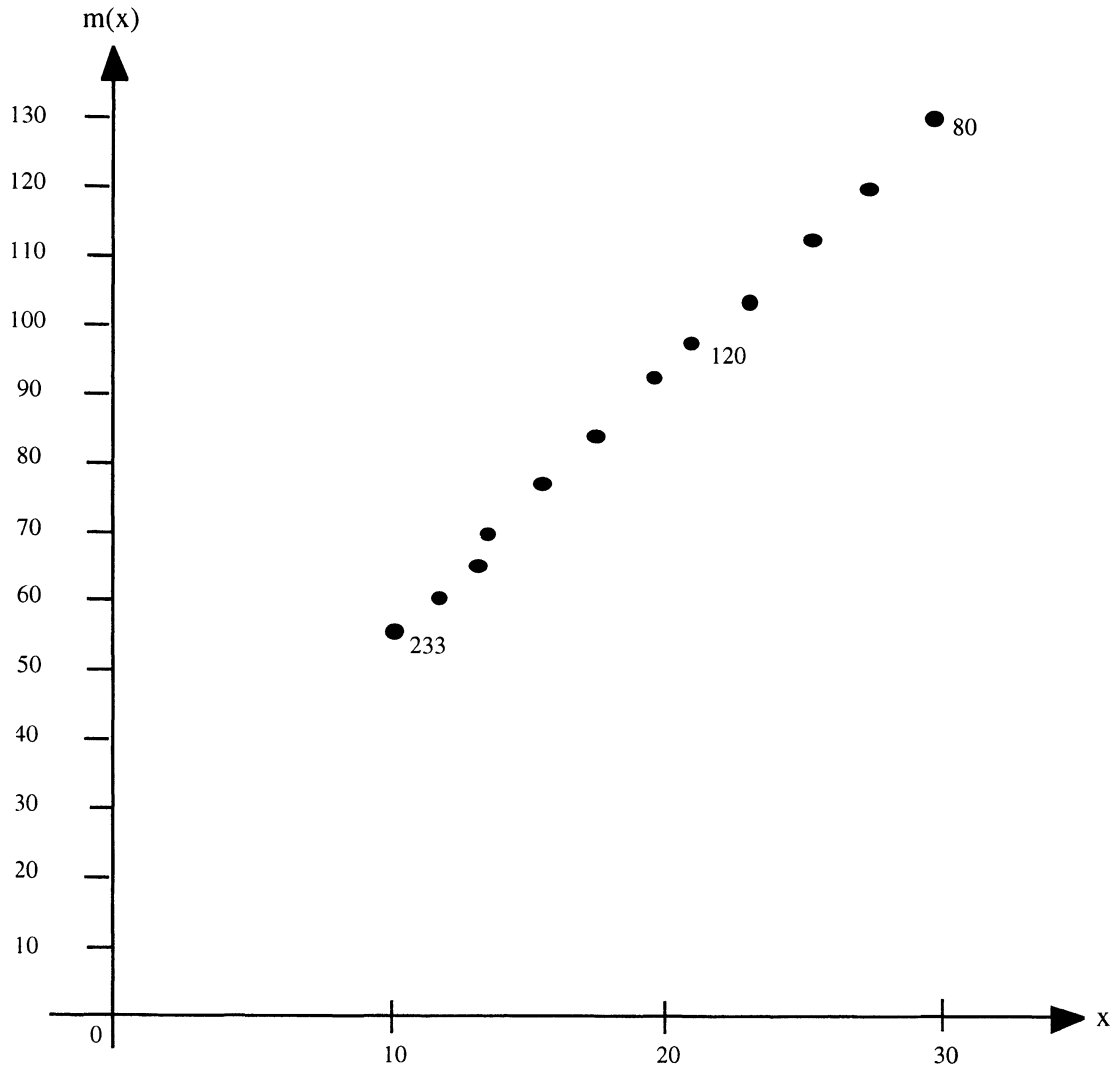


Figure 2.2.

3. Par contre, lorsque l'ajustement linéaire semble devoir être bon, comme l'indique le graphique 2.2., il est tout à fait inutile de le réaliser sur un grand nombre d'observations ; pour qui travaille à la calculette, *ce qui est toujours préférable* pour repérer d'éventuelles erreurs, la seule façon praticable est d'opérer sur peu de points : faire l'ajustement sur toutes les observations du rang 80 au rang 233 (150 observations environ) n'améliorerait guère la qualité de l'ajustement.

D'ailleurs, V. Pareto lui-même (il n'avait pas d'ordinateur) a toujours traité des séries très courtes de 10 à 20 observations au plus (voir dans [9] et [10] celles, nombreuses, qu'il a publiées).

Procédons maintenant à l'ajustement linéaire :

$$m(x) = \beta x + h + \varepsilon(x) = \hat{m}(x) + \varepsilon(x)$$

Compte tenu de la valeur apparemment élevée de la pente β de la droite d'ajustement, il est préférable d'utiliser la droite des moindres carrés qui fait jouer un rôle symétrique aux deux variables x et m .

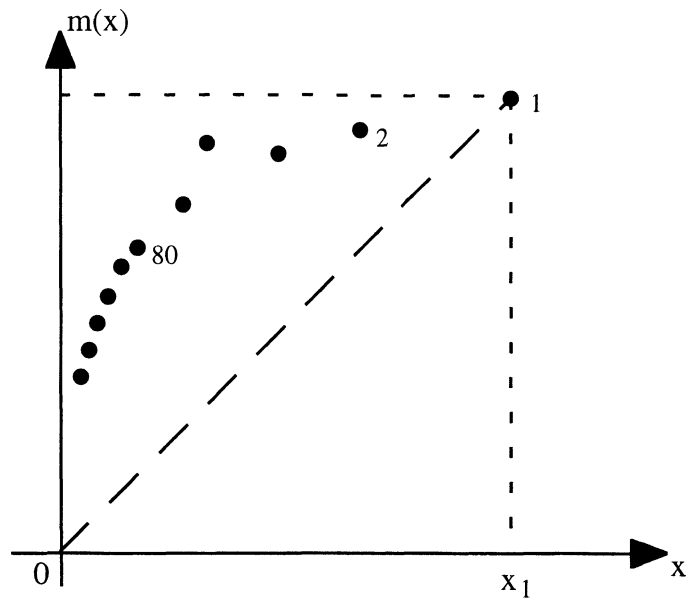


Figure 2.3

La pente de cette droite est donnée par : $\beta = \text{tg } \theta$, où θ (défini à $\frac{\Pi}{2}$ près) est l'angle tel que :

$$\text{tg } 2\theta = \frac{2 \text{Cov}(x, m)}{\text{Var}(x) - \text{Var}(m)} .$$

Comme les droites de régression, elle passe par le "point moyen" du "nuage" des 12 points utilisés ici.

On trouve ainsi les estimations suivantes (le lecteur à tous les éléments lui permettant de les vérifier) :

$$\hat{\beta} = 3,73 \quad \text{et} \quad \hat{h} = 18,9$$

avec un coefficient de corrélation linéaire (σ est l'écart-type) :

$$\rho = \frac{\text{Cov}(x, m)}{\sigma_x \sigma_m} = 0,998 .$$

Comparons les valeurs estimées de $m(x)$ avec les valeurs observées (Tableau 2.2.) pour quelques rangs :

On voit que cet ajustement linéaire est excellent pour les rangs élevés (100 et au-delà) ; la valeur très élevée du coefficient de corrélation ρ , calculé sur les 12 points de la Figure 2.2. le laissait prévoir.

Il est très remarquable qu'ici l'ajustement reste acceptable même pour de petits rangs. Ce n'est pas toujours le cas.

r	x	m(x) observée	$\hat{m}(x)$ estimée	$e(x) = m(x) - \hat{m}(x)$
10	162	643,5	623,1	20,4
20	90,8	376,3	357,5	18,8
30	72,1	280	287,7	- 7,7
40	57	226,5	231,5	- 5
60	36,5	166	155	11
70	31,3	147	135,6	11,4
100	25,4	111,4	113,6	- 2,2
120	20,9	96,7	96,8	- 0,1
160	15,7	77	77,4	- 0,4
200	12,1	64,3	64	0,3
233	10,1	56,71	56,5	0,2

Tableau 2.2. Ajustement linéaire
 $\hat{m}(x) = 3,73 x + 18,9$

N.B. Une valeur très élevée du coefficient de corrélation linéaire doit toujours être considérée de façon critique. Il peut en effet s'agir d'un "artefact", dû à son mode de calcul et à ce que sont les données.

Par exemple, on démontre très facilement que si pour l'un des couples (x, m) les deux coordonnées (ou même une seule d'entre elles) x et m tendent vers l'infini, les autres restant bornés, alors ρ tend vers 1.

Or, il peut se faire que l'un des couples observés (ou quelques-uns d'entre eux) soit très grand par rapport à tous les autres : nous aurons un ρ calculé très proche de 1, mais qui n'aura aucune signification en termes d'ajustement.

Ici, pour nos 12 points, x varie de 10 à 30 (échelle de 1 à 3), et m de 55 à 130 (échelle de 1 à 2,4), les valeurs intermédiaires se répartissant assez uniformément. L'effet factice évoqué ci-dessus ne joue pas.

A partir des estimations de β et de h , calculons les autres paramètres de la distribution de Pareto :

$$N(x) = N_0 \left(\frac{x_0 + c}{x + c} \right)^\alpha$$

$N(x)$ est le rang théorique de l'agglomération de taille x .

On a :

$$\alpha = \frac{\beta}{\beta - 1}, \quad \text{d'où} \quad \hat{\alpha} = \frac{3,73}{2,73} = 1,366 .$$

Et :

$$c = \frac{h}{\beta - 1}, \quad \text{d'où} \quad \hat{c} = \frac{18,9}{2,73} = 6,9$$

N_0 et x_0 sont donnés : N_0 est le nombre 233 des observations ; et $x_0 (= x_{N_0})$ est la plus petite observation, soit 10,1.

D'où :

$$N(x) = 233 \left(\frac{17}{x + 6,9} \right)^{1,366} .$$

En inversant cette relation, on obtient également la taille "théorique" x_r de l'agglomération de rang r :

$$x_r = (x_0 + c) \left(\frac{N_0}{r} \right)^{1/\alpha} - c ,$$

soit ici :

$$\hat{x}_r = \frac{919}{r^{1/\alpha}} - 6,9 = \frac{919}{r^{0,732}} - 6,9 .$$

Le Tableau 2.3 permet de juger la qualité de l'ajustement pour quelques valeurs du rang ; elle est très satisfaisante. On remarquera en particulier (dernière colonne) que l'écart entre distribution théorique et distribution observée, rapporté à une échelle de 0 à 100, a pour maximum de 2% : inutile de tester plus avant l'adéquation de la distribution théorique à la distribution empirique.

On remarquera surtout que *l'ajustement est très bon pour les premiers rangs* (sauf, évidemment, pour Paris), bien que l'ajustement linéaire préalable ait été opéré sur des rangs élevés. On voit bien ainsi que les distributions de Pareto (et plus généralement les asymptotiquement paretiennes de P. Levy) sont, dans l'étude des phénomènes inégaux, particulièrement adaptées pour *rendre compte des grandes valeurs* de la variable en jeu.

On peut enfin, à titre de recouplement, comparer médiane théorique et médiane observée.

La médiane théorique vaut (Cf. §.1.5 supra) :

$$\mu = 2^{1/\alpha} (x_0 + c) - c = 21,3 .$$

Quant à la médiane observée, elle correspond au rang 117, et vaut (agglomération de Moyeuve-Grande) 21,2 ; on a d'ailleurs pour le rang 116 (St-Dié) une population de 21,5 KH. Que demander de plus ?

Pour l'indice d'inégalité de Gini, l'accord est aussi bon. Sa valeur empirique est, pour les 233 agglomérations, de 0,65. Pour la valeur théorique on a :

$$G = \frac{\beta - 1}{\beta + 1} \left(1 + \frac{c}{m_0} \right) \approx 0,647$$

3. UNE INÉGALITE CROISSANTE

J'ai effectué, par la méthode détaillée supra pour 1901, les ajustements paretiens de la distribution des agglomérations de 10 KH et plus pour plus de la moitié des 28 recensements généraux de la population qui ont eu lieu de 1831 à 1990.

Si les résultats ne sont pas toujours aussi bons que pour 1901, ils sont toujours satisfaisants ; le coefficient de corrélation ρ de la phase d'ajustement linéaire ne descend jamais au dessous de 0,99. Et il y a toujours une assez bonne adéquation pour les grandes valeurs de la population.

Ecart $x - \hat{x}$	Valeur théorique estimée \hat{x}	Valeurs observées		Valeur théorique estimée : $N(x) = \hat{r}$	Ecart $r - N$	Ecart % $\frac{ r - N }{233}$
		x	r			
-11,7	546,4	534,7	2	2,06	- 0,06	0
93,5	404,3	497,8	3	2,27	0,73	0,3
-6,7	326,2	319,5	4	4,11	- 0,11	0,05
36,3	276	312,3	5	4,24	0,76	0,3
- 35,8	240,7	204,9	6	7,42	- 1,42	0,6
- 14,3	213,2	197,9	7	7,77	- 0,77	0,3
- 1,06	193,6	192,5	8	8,06	- 0,06	0
12,6	177	189,6	9	8,23	0,77	0,3
- 1,4	163,4	162	10	10,12	- 0,12	0
- 5,9	119,7	113,8	15	16	- 1	0,4
- 4,9	95,6	90,8	20	21,4	- 1,4	0,6
- 2,2	80,2	78	25	25,9	- 0,9	0,4
2,8	69,3	72,1	30	28,6	1,4	0,6
2,15	54,8	57	40	38,2	1,8	0,8
0,56	45,54	46,1	50	49,3	0,7	0,3
- 0,3	42	41,7	55	55,5	- 0,5	0,2
- 2,5	39	36,5	60	64,75	- 4,75	2
- 0,6	31	30,4	75	79,4	- 4,4	2
0,2	28,7	28,9	85	84,25	0,75	0,3
0,73	24,7	25,4	100	96,25	3,75	2
0,3	18,4	18,7	135	133,2	1,8	0,8
0,335	16,56	16,9	150	147,1	2,9	1,3
0,03	13,63	13,6	180	180,4	- 0,4	0,2
0	12,1	12,1	200	200,15	- 0,15	0,6

Tableau 2.3

Remarque. Pour améliorer encore la qualité de l'ajustement, on peut, outre les paramètres β (i.e. $\alpha = \frac{\beta}{\beta - 1}$) et c (i.e. $h = c(\beta - 1)$) de l'expression $N_o \left(\frac{x_o + c}{x + c} \right)^\alpha$, ajuster un troisième paramètre.

Par exemple, prendre pour \hat{x}_o non la plus petite observation, mais la valeur telle que moyenne théorique et moyenne empirique m_o coïncident ; soit : $\hat{x}_o = \frac{m_o - \hat{h}}{\hat{\beta}}$.

Bien sûr, estimer ce troisième paramètre n'a de sens que si les deux moyennes théorique et empirique diffèrent notablement, comme c'est le cas par exemple pour 1990, où leur écart est de quelque 10 milliers

d'habitants. Procéder ainsi améliore l'ajustement du milieu de la distribution, alors que c'est une estimation correcte de β qui importe pour les grandes valeurs.

On peut également, grâce aux données fournies par Bernard Lepetit (dans [7]) sur la population des villes de 10.000 habitants et plus au 18^{ème} siècle et au début du 19^{ème} siècle, remonter plus avant dans le passé.

Certes, ville et agglomération sont deux notions distinctes ; mais elles devaient très peu différer au 18^{ème} siècle et diffèrent encore peu au début du 19^{ème} siècle ; on le voit sur les données de 1836, où je dispose des deux statistiques, celle des villes dans [7] et celle des agglomérations (Cf. Tableau 3.1. ci-dessous).

	Population totale	dont Paris	N_o	m_o	$\hat{\beta}$
Villes	3.800	909	120	31,6	3,1
Agglomérations	4.140	1.013	126	32,8	2,9

Tableau 3.1
Villes et agglomérations de 10KH et plus, en 1836.

Bien entendu, puisque les recensements généraux de la population n'existent que depuis 1801, les données relatives au 18^{ème} siècle résultent d'évaluations faites par les spécialistes de démographie historique, mais sont, selon eux, assez fiables. Les ajustements paretiens sont, en tout cas, aussi bons pour ces données relatives au 18^{ème} siècle (1700, 1750, 1780) que pour celles résultant des recensements.

On n'a pas la place ici de donner le détail de tous ces ajustements. Mais il est intéressant d'observer ce qu'a été l'évolution sur le long terme. Le tableau 3.2 donne la photographie à quatre dates, séparées par des intervalles d'un siècle environ.

	Données				Paramètres estimés				
	Popul. totale	dont Paris	N_o	m_o	β	α	h	c	G (Gini)
1700	1.850	510	60	32	2,73	1,58	0	0	0,47
1806	2.750	581	92	29,9	2,81	1,55	0	0	0,475
1901	13.213,5	3.829,3	233	56,71	3,73	1,366	18,9	6,9	0,65
1990	35.511	9.318,8	435	81,65	4,3	1,3	49,5	15	0,74

Tableau 3.2
Agglomérations de 10 KH et plus. Évolution sur trois siècles.

Ce tableau montre que l'évolution est très lente au cours du 18^{ème} siècle, et elle le reste d'ailleurs pendant le premier tiers du 19^{ème} siècle (cf. sur ce point [12]).

Par contre, la concentration et l'inégalité dans la distribution croissent fortement aux 19^{ème} et 20^{ème} siècles. Cela se lit aussi bien sur les données observées que sur les valeurs estimées des paramètres : β et h augmentent tous les deux.

Certes, ces photographies à quatre dates cachent les accidents de l'évolution, notamment les perturbations consécutives aux guerres de 1870-71, 1914-18 et surtout 1939-45. Mais la tendance sur le long terme est claire.

On pourrait illustrer ce phénomène d'accroissement de l'inégalité par les courbes de concentration de Lorenz et Gini.

Dans le cas de distributions parétiennes, il est plus simple de visualiser l'évolution de l'inégalité en traçant (Figure 3.1) les droites d'équation :

$$m(x) = \beta x + h.$$

En effet, $m(x)$ étant, pour chaque taille d'agglomération, la population moyenne des agglomérations de taille supérieure ou égale à x , plus $m(x)$ est grand par rapport à x , plus l'inégalité dans la distribution est grande, comme on l'a déjà souligné.

En particulier, pour les grandes valeurs de x (les plus grandes agglomérations), c'est le terme βx qui est prépondérant dans le calcul de la moyenne $m(x)$. Aux alentours de 1700 comme de 1800, on peut dire : les villes plus peuplées qu'une ville de population x le sont en moyenne 2,8 fois plus environ. Alors qu'en 1901, elles le sont 3,7 fois plus, et en 1990 entre 4 et 5 fois plus.

Ainsi, l'inégalité croît dans le temps ; mais on peut en outre regarder les écarts de croissance, pour chaque période séculaire. Entre 1806 et 1700, on a, en désignant par D cet écart :

$$D(x) = 2,81x - 2,73x = \frac{8x}{100}.$$

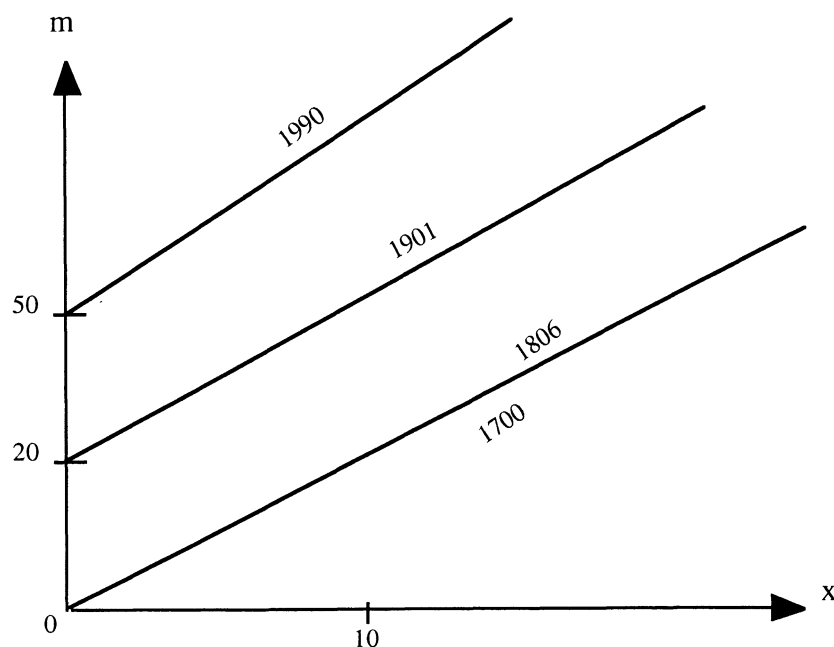


Figure 3.1

Il est très faible, comme prévu ; l'inégalité n'a pratiquement pas augmenté. Pour 1901 et 1806 on a, en chiffres arrondis :

$$D(x) = 1,1x + 16,3 .$$

L'inégalité a fortement augmenté.

De même, entre 1990 et 1901, la différence est :

$$D(x) = 0,6x + 30,6 .$$

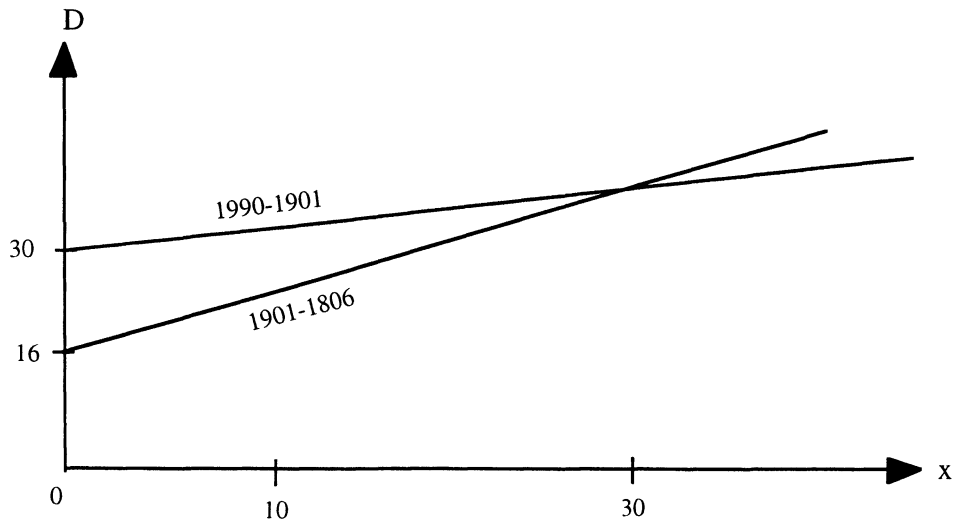


Figure 3.2

L'inégalité augmente encore, mais moins fortement qu'au siècle précédent.

Sur la Figure 3.2, on a tracé les deux droites correspondantes. Elles se coupent en un point représentant une population de 30 milliers d'habitants environ. Cette valeur n'est qu'indicative. Mais on voit qu'au-delà de ce seuil, l'inégalité augmente d'autant moins vite au 20^{ème} siècle (par rapport au 19^{ème}) que le seuil de population x retenu est grand.

Ce phénomène de tassement avec le temps de l'accroissement de l'inégalité, on va le retrouver plus loin sous une autre forme.

4. UNE DISTRIBUTION "CONTRA-PARETIENNE" : LA TAILLE DE LA PLUS GRANDE AGGLOMÉRATION (et de la n-ième ?)

On applique maintenant la méthode qui a été détaillée au §.2 pour la distribution des agglomérations en 1901, non plus à un recensement donné, mais à la distribution de *la plus grande agglomération* (c'est-à-dire celle de Paris) pour chacun des 28 recensements de 1831 à 1990, auxquels on a adjoint ceux de 1821 et 1806 ; ceux-ci concernent les villes, mais on a vu supra qu'à ces dates, il y a peu de différence entre ville et agglomération, au niveau macroscopique.

Nous avons donc 30 tailles de la ville primatale, rangées par valeurs décroissantes (colonne 2 du Tableau 4.1, p.66) ; on remarquera que les rangs et les dates (colonne 1 du Tableau) sont dans le même ordre, sauf pour le recensement de 1946 pour lequel la taille de l'agglomération de Paris se situe entre celles de 1926 et 1931 : perturbation consécutive à la guerre de 1939-45.

Dans la colonne 3 du Tableau 4.1, sont calculées les moyennes conditionnelles $m(x)$. On n'a pas construit ici le graphique cartésien des 30 points (x, m) : le lecteur intéressé pourra le faire. Il constatera qu'une fois de plus les points se répartissent assez bien au voisinage d'une droite, tout au moins pour les 18 derniers d'entre eux (de 1806 à 1906, qui correspond au rang 13), et ceci bien que les dénominateurs soient faibles : ils ne dépassent pas 30 !

L'ajustement de la droite des moindres carrés au modèle :

$$m(x) = \beta x + h + \varepsilon(x)$$

fournit ici les estimations arrondies suivantes : $\hat{\beta} = 0,71$; $\hat{h} = 3.605$ avec un coefficient de corrélation $\rho = 0,994$.

Nous sommes donc dans le deuxième cas de ceux classés au §.1, celui où la distribution est bornée par un maximum d (cf. figure 1.1, p.45 supra) que la variable, ici la taille de la ville primatale, ne peut dépasser et a la forme :

$$N(x) = N_0 \left(\frac{d - x}{d - x_0} \right)^y \quad \text{avec} \quad x_0 \leq x \leq d ; y = \frac{\beta}{1 - \beta} .$$

Ce sont les distributions en "fonction puissance positive" qu'on peut appeler contra-paretiennes ou anti-paretiennes (voir ci-dessus, §.1.3).

Compte-tenu des estimations faites pour β et h , on a ici :

$$\hat{d} = \frac{\hat{h}}{1 - \hat{\beta}} = 12.431 , \quad \text{que l'on arrondit à } 12.450.$$

En prenant pour \hat{x}_0 la valeur 550 (la plus petite observation vaut 581), et en arrondissant β à 0,7 (en fait $0,7 < \beta < 0,71$), on obtient ainsi pour le rang estimé (en fonction de la taille x) :

$$N(x) = \hat{r} = 30 \left(\frac{12.450 - x}{11.900} \right)^{7/3} .$$

La figure 4.1 rappelle l'allure de la courbe "représentative" de cette distribution théorique.

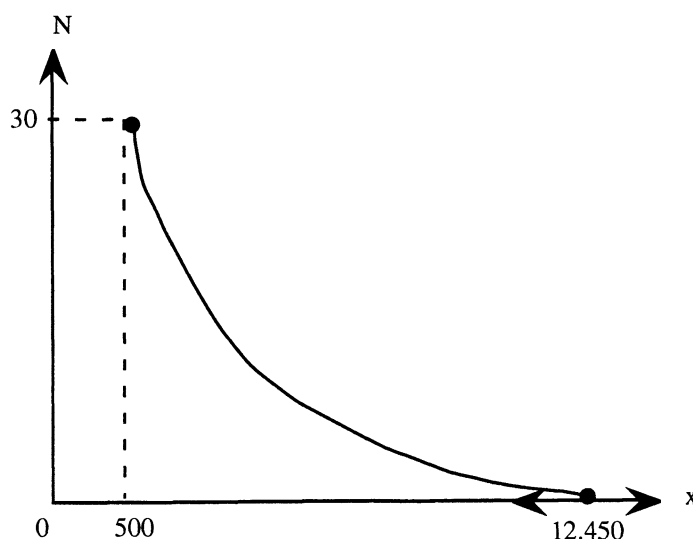


Figure 4.1

Date	Taille x	Moyenne m(x)	Estimation \hat{x} %	$\frac{x - \hat{x}}{x}$ %	Rang r	Estimation \hat{r}	$\frac{ r - \hat{r} }{30}$ %	$\frac{ x - \hat{x} }{12500}$ %
1990	9319		9680	- 4	1	1,33	1,1	3
1982	8673		8722	- 0,6	2	2,06	0,02	0,3
1975	8495,5		8014	+ 5,7	3	2,3	2,5	4
1968	8153,5		7435	+8,9	4	2,8	4	5,6
1962	7339	8396	6926	+5,6	5	4,2	2,7	3,2
1955	6245		6480	- 3,8	6	6,6	1,9	2
1936	5784		6072	- 5	7	7,75	2,5	2,4
1931	5674		5696	- 0,4	8	8,06	0,02	0
1946	5602		5347	+ 4,6	9	8,25	2,5	2,1
1926	5160	7044,5	5019	+ 2,7	10	9,56	1,5	1,1
1921	4726		4709	0	11	10,95	0,02	0
1911	4473		4415	+ 1,3	12	11,8	0,07	0,5
1906	4088,5	6440	4134	- 1,1	13	13,2	0,05	0,4
1901	3824	6254	3866	- 1	14	14,15	0,05	0,4
1896	3787,5	6089,5	3608	+ 4,8	15	14,3	2,5	1,5
1891	3135	5904	3361	- 7,2	16	16,95	3,15	1,6
1886	2983	5732,5	3121	- 4,6	17	17,6	2	1,2
1881	2816	5704	2890	- 2,6	18	18,3	1	0,5
1876	2383	5529	2666	- 12	19	20,3	4,5	2,4
1872	2209	5243,5	2448	- 11	20	21,15	4	2
1866	2116	5080	2237	- 5,7	21	21,6	2	1
1861	1917	4936,5	2031	- 6	22	22,6	2	1
1856	1682	4795	1831	- 8,9	23	23,75	2,5	1,3
1851	1366	4665	1637	- 20	24	25,4	4,5	2,4
1846	1297	4528,5	1445	- 11,4	25	25,8	2,5	1,2
1841	1019	4593,5	1258	- 23,5	26	27,3	4,5	2
1836	1013	4268,5	1075	- 6	27	27,3	1	0,5
1831	835	4146	897	- 7,4	28	28,3	1,2	0,5
1821	714	4027	722	- 1,1	29	29	0	0
1806	581	3912,5	550	+ 5,2	30	29,8	0,07	0,3
1	2	3	4	5	6	7	8	9

Tableau 4.1

Le tableau 4.1 donne en colonnes 4 et 7 les valeurs estimées de la taille et du rang, et en colonnes 8 et 9 les écarts entre observations et estimations rapportées à une échelle de 0 à 100 ; ils ne dépassent pas 6%, ce qui est assez satisfaisant.

Remarques.

1. On a vu que les points (x, m) s'alignaient assez bien (colonnes 2 et 3 du Tableau 4.1) alors que les moyennes portent sur des petits nombres d'observations (de 18 à 30) ; c'est parce qu'ici, toutes les observations étant contenues dans un intervalle borné, la convergence de la moyenne d'un échantillon vers la moyenne théorique est très rapide, (les "grandes valeurs" y sont impossibles), contrairement à ce qui se passe dans le cas paretien, où elle est très lente (les "grandes valeurs" ont un poids non négligeable).
2. Les colonnes 4 et 5 du tableau 4.1 donnent les estimations de la taille x en fonction du rang, et l'écart relatif ; celui-ci est parfois important, pouvant aller jusqu'à près de 25%. Ceci a lieu lorsque les irrégularités dans la croissance de la population urbaine font que deux recensements consécutifs donnent des tailles très voisines comme en entre 1851 et 1846 ou en entre 1841 et 1836 (il peut également y avoir eu des erreurs dans les recensements correspondants). Si l'on avait rapporté les écarts absolus $(x - \hat{x})$ à une échelle de 0 à 12.000 environ, le plus grand écart en pourcentage est inférieur à 6% (colonne 9). On pourra, si on le veut, comparer ceci avec les valeurs fournies par les tables du test de Kolmogorov-Smirnov.

Ce que l'on vient de constater pour la plus grande agglomération, celle de Paris, on le retrouve lorsque l'on remplace la 1^{ère} agglomération par la 2^{ème}, la 10^{ème}, la 50^{ème}, voire la 100^{ème} (qui, d'un recensement à l'autre, ne sont en général pas les mêmes) : les points (x, m) s'alignent assez bien du rang 15 au rang 30. Et les droites d'ajustement correspondantes ont des pentes β du même ordre de grandeur : 0,65 environ pour la deuxième, 0,57 pour la centième.

Même constat pour la population totale des agglomérations d'au moins 10.000 habitants, avec un β légèrement supérieur à celui de Paris (0,72 environ).

Quant au maximum théorique d , son ordre de grandeur serait de 1.600 pour la deuxième agglomération, de 60 pour la 100^{ème}, et de 42.000 pour la population totale. Ces chiffres sont très approximatifs, et donnés sous toute réserve. A titre de comparaison, la taille, lors du dernier recensement (1990) de la 2^{ème} agglomération (qui, selon les années, est soit Lyon, soit Marseille), vaut 1262 ; pour la 100^{ème}, elle est de près de 58, et la population totale est de 35.511 milliers d'habitants.

Remarque. On peut rapprocher le fait que la taille de l'agglomération parisienne (la plus grande ville), ait une distribution "anti-paretienne" de ce qui a été dit au §.1.5 sur les lois du maximum (ici, on aurait approximativement la loi de Weibull). Ce serait une erreur que de faire une telle interprétation. En effet, les lois évoquées au §.1.5 concernent la valeur maximum d'échantillons *extraits d'une même distribution parente*.

Ici, les distributions dont l'agglomération parisienne est le maximum, bien que toutes paretienne, sont toutes *différentes*, et même fortement différentes, on l'a vu, pour des recensements situés à des dates éloignées.

5. CROISSANCE LOGISTIQUE DE L'AGGLOMÉRATION PARISIENNE ?

S'il est exact, le fait que la distribution de la taille de l'agglomération parisienne s'ajuste sur une distribution théorique à support borné implique que dans son évolution temporelle, sa population, notée ici $x(t)$, a une borne supérieure. Le modèle de croissance en fonction du temps le plus simple (et auquel font parfois appel les démographes ; on trouvera des indications dans [6]) assurant l'existence d'une telle borne est le *modèle logistique* pour lequel on aurait :

$$x(t) = d \frac{e^{\lambda(t-t_0)}}{1 + e^{\lambda(t-t_0)}} \quad \lambda > 0 .$$

La figure 5.1 donne l'allure de la courbe représentative. Elle a une asymptote horizontale d'ordonnée d lorsque t tend vers l'infini ; et un centre de symétrie M , d'abscisse t_0 telle que :

$$x(t_0) = \frac{d}{2} .$$

La croissance de l'agglomération parisienne peut-elle être considérée comme du type logistique ? La façon la plus simple de le tester est de remarquer que :

$$y(t) = \frac{x(t)}{d - x(t)} = e^{\lambda(t-t_0)} .$$

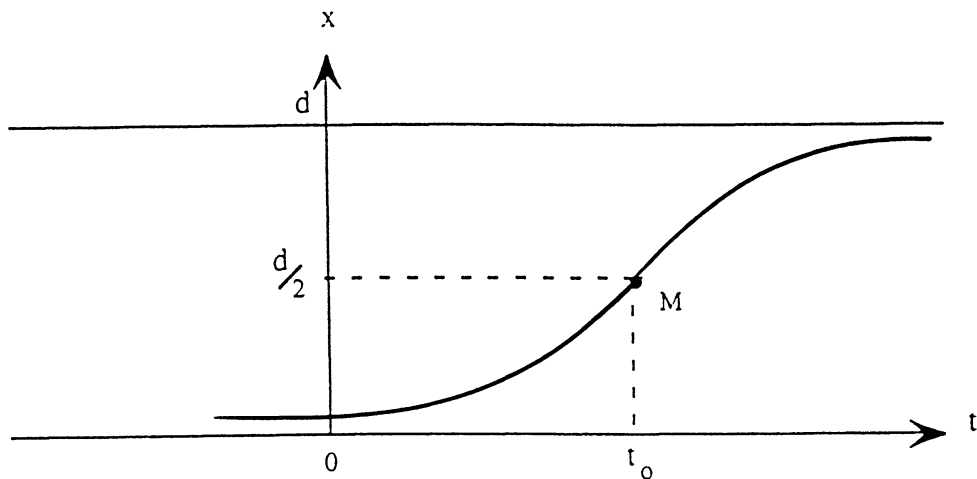


Figure 5.1

D'où : $\text{Log } y(t) = \lambda(t - t_0) = \lambda t - k$.

La variable $z(t) = \text{Log} \frac{x}{d - x}$ (le "logit") est une fonction linéaire du temps.

Si l'on connaît d , ce qui pourrait être le cas ici (on sait que le maximum théorique est aux alentours de 12.000 — 12.500, d'après l'étude faite au paragraphe précédent) on déterminera les 30 couples $(t, z(t))$ d'après les observations et on regardera s'ils sont à peu près alignés.

Dans le Tableau 5.1, colonne 6, on a indiqué quelques valeurs de $z(t)$, pour $d = 12.450$. Pour $d = 12.000$, elle ne sont pas très différentes, au niveau macroscopique où l'on se place ici. Le lecteur pourra compléter la colonne 6 et figurer les 30 points sur un graphique cartésien, qui est omis ici.

En tout cas, le constat est que les 30 points sont bien à peu près alignés, sur une droite de pente, estimée ici à vue :

$$\lambda \approx 0,023.$$

Remarque. L'alignement des points $(t, z(t))$ ne doit pas trop faire illusion : comme on l'a souligné plus haut, minimiser une fonction des écarts à un logarithme, c'est minimiser une fonction des écarts *relatifs*. Ce qui est très différent de minimiser une fonction des écarts absolus ; certains écarts seront sur-évalués, d'autres sous-évalués (Cf. §.1.6 supra).

C'est pourquoi il me semble au moins aussi judicieux d'estimer ici la pente de la droite d'ajustement au "pifomètre" (avec l'aide du papier gradué, cependant) que par un calcul fastidieux, tel celui de la droite des moindres carrés, par exemple.

Reste à estimer la date t_0 correspondant au centre de symétrie de la courbe.

Puisque l'asymptote a une ordonnée de 12.000 KH et quelques, c'est à une population d'environ 6.000 KH que correspond la moitié (Cf. Figure 5.2). Comme le montrent le tableau 5.1., colonnes 1 et 2, et la Figure 5.2, s'il n'y avait pas eu les perturbations apportées par la dernière guerre, c'est aux alentours de 1940 que cette valeur aurait été atteinte.

On prendra donc : $t_0 = 1940$.

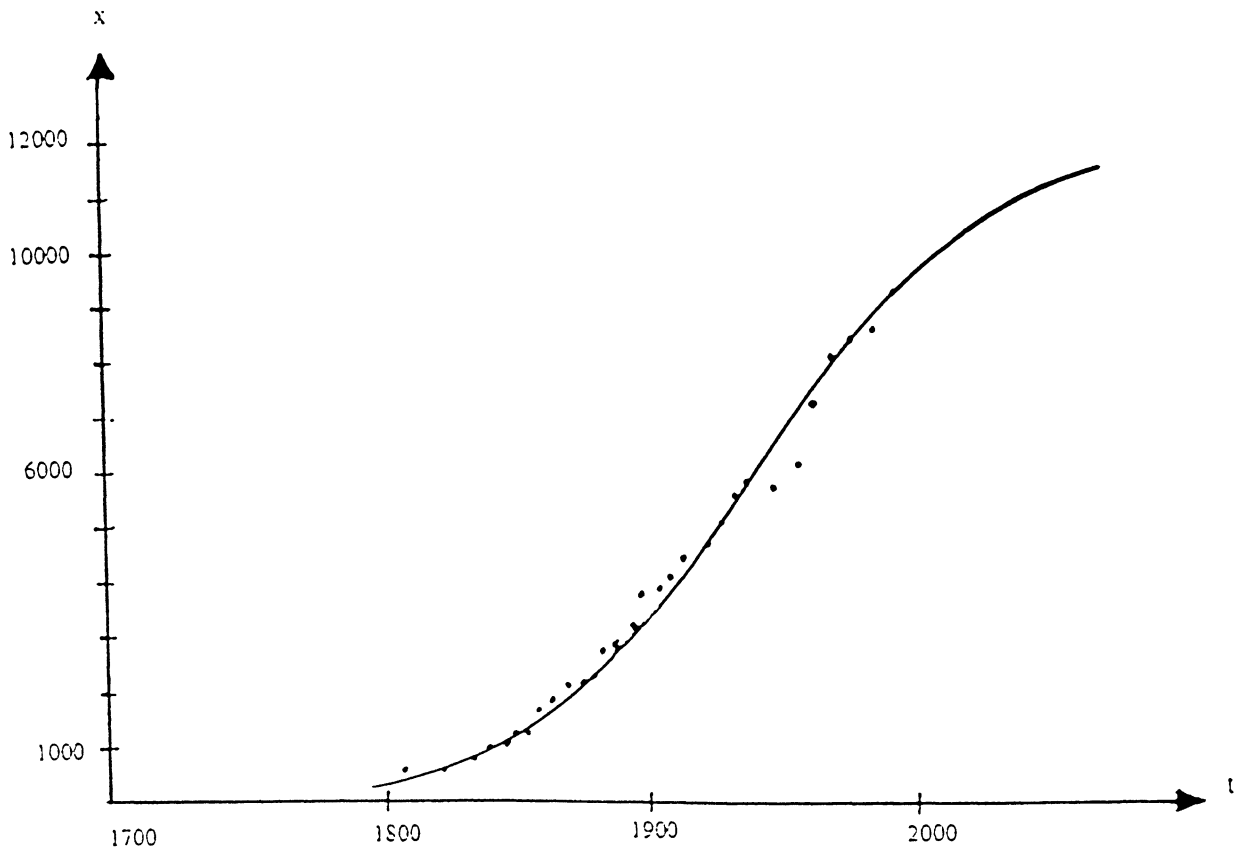


Figure 5.2

Les colonnes 2, 3 et 4 du Tableau 5.1 correspondent ainsi à l'estimation (où t est exprimé en années) :

$$\hat{x}(t) = 12.450 \frac{e^{\lambda(t-1940)}}{1 + e^{\lambda(t-1940)}} \quad (\lambda = 0,023) .$$

La courbe logistique correspondante est tracée en continu sur la Figure 5.2.

Date t	Population x(t)	Estimation \hat{x}	Ecart x - \hat{x}	Ecart % $\frac{x - \hat{x}}{12.500}$	$z(t) = \text{Log} \frac{x(t)}{12.450 - x(t)}$
1990	9319	9467	- 148	- 1,2	1,091
1982	8673	9028	- 355	- 2,9	.
1975	8495	8613	- 118	- 1	.
1968	8153	8171	- 18	0	.
1962	7339	7773,5	- 435	- 3,5	0,362
1955	6245	7293	- 1048	- 8,5	.
1946	5602	6656	- 1046	- 8,5	.
1936	5784	5938	- 153	- 1,2	.
1931	5674	5580	94	+ 0,9	.
1926	5160	5227	- 67	- 0,7	- 0,346
1921	4726	4880	- 154	- 1,2	.
1911	4473	4214	258	+ 2,1	.
1906	4088	3898	189	+ 1,5	.
1901	3824	3596	228	+ 1,8	.
1896	3787	3308	479	+ 4	- 0,828
1891	3135	3035	100	+ 0,9	.
1886	2983	2778	205	+ 1,7	.
1881	2816	2537	279	+ 2,4	.
1876	2383	2312	71	+ 0,8	.
1872	2209	2142	66	+ 0,7	- 1,532
1866	2116	1908	208	+ 1,7	- 1,586
1861	1917	1728	188	+ 1,5	.
1856	1682	1564	118	+ 1	.
1851	1366	1412	- 47	- 0,5	.
1846	1297	1274	23	0	- 2,152
1841	1019	1148	- 129	- 1,1	.
1836	1013	1033	- 42	0	.
1831	835	929	- 94	- 0,8	.
1821	714	748	- 35	- 0,3	- 2,8
1806	581	546	35	0,3	- 3,016
1	2	3	4	5	6

Tableau 5.1. Croissance de la population.
Agglomération parisienne. Ajustement logistique

Les chiffres, comme la figure, mettent bien en évidence l'effet de la guerre 1939-45. D'autres irrégularités apparentes peuvent n'être dues qu'à des hasards ; non ceux de l'histoire, mais d'erreurs de mesure de mon "pifomètre", ou de la mauvaise qualité du recensement. C'est notamment le cas de 1896, d'après les spécialistes (Cf. [4]).

C'est ainsi qu'une estimation légèrement inférieure pour le paramètre λ (0,021 au lieu de 0,023) donne un meilleur ajustement pour les derniers recensements, mais moins bon pour ceux du 19^{ème} siècle.

Quoi qu'il en soit, l'hypothèse d'une croissance logistique pour l'agglomération parisienne semble admissible en première approximation.

Bien entendu, il peut y avoir une rupture dans cette tendance sur le long terme. Seul l'avenir le dira. Selon le présent ajustement, qui n'est sûrement pas le meilleur possible, la population pourrait être autour de 9,9 millions en 2000 (estimation haute) et de 10,4 millions en 2010, si toutefois la tendance reste la même. Attendons les prochains recensements pour savoir ce qu'il en est.

Peut-être y a-t-il des raisons qui expliquent les propriétés mises en évidence empiriquement, et en première approximation, dans cette étude.

Il y a d'ailleurs du choix parmi les modèles mathématiques (tant déterministes que probabilistes) permettant d'engendrer une distribution paretienne ou une croissance logistique.

Ce n'est pas au statisticien qu'il revient d'en proposer un ; seuls les spécialistes (en l'occurrence, ceux de géographie urbaine et de démographie historique) ont les compétences nécessaires pour le faire.

D'autre part, pour compléter cette étude de la distribution des agglomérations de 10 KH ou plus, et de son évolution dans le temps, il faudrait procéder aux ajustements réalisés ci-dessus pour tous les recensements disponibles, et pour tous les rangs ; le phénomène étudié est en effet de dimension 3 (pour les variables retenues) ; la taille x est une fonction $x(t,r)$ de la date du recensement et du rang occupé dans la distribution des tailles. Ici, on n'a procédé qu'à quelques "coupes" de la surface correspondante de l'espace : à t constant, pour quelques valeurs de t (et en particulier pour $t = 1901$) et à r constant, où seule la valeur $r = 1$ a été regardée de près.

Il serait en particulier intéressant de rechercher si pour les autres rangs que le premier, la croissance temporelle de la taille est toujours approximativement logistique ; et si oui, quelle serait la distribution de l'ensemble des bornes supérieures vers lesquelles tendraient asymptotiquement les tailles.

J'y renonce pour l'instant : travailler à la calculatrice, cela évite parfois de dire des bêtises ; mais c'est toujours long et assez fastidieux.

BIBLIOGRAPHIE

- [1] BARBUT, Marc, "Distribution de type paretien et représentation des inégalités", *Mathématiques, Informatique et Sciences humaines*, n°106, 1989, 53-69.
- [2] BARBUT, Marc, "Note sur les moyennes de variables aléatoires", in *Moyenne, Milieu, Centre*, ouvrage collectif, Paris, EHESS, 1991, 31-43.
- [3] BARBUT, Marc, "Une remarque sur l'expression et l'ajustement des distributions Zipf-Mandelbrot en statistique textuelle", in *Mélanges de mathématiques, linguistique, informatique offerts à André Lentin*, ouvrage collectif, Paris, CAMS-EHESS, 1997, 171-177.

- [4] BONNEUIL, Noël, "Cohérence comptable des tableaux de la S.G.F. : recensements de 1851 à 1906, mouvements de la population de 1801 à 1906", *Population*, 4-5, 1989, 809-838.
- [5] FRÉCHET, Maurice, *Une nouvelle représentation analytique de la répartition des revenus*, XVI^{ème} session de l'Institut International de Statistique, Rome, 1925.
- [6] GUÉRIN-PACE, France, "Deux siècles de croissance urbaine", *Anthropos-Economica*, Paris, 1993.
- [7] LEPETIT, Bernard, *Les villes dans la France moderne (1740-1840)*, Paris, Albin Michel, 1988.
- [8] LEVY, Paul, *Théorie de l'addition des variables aléatoires*, Paris, Gauthier-Villars, 1937.
- [9] PARETO, Vilfredo, *Cours d'Économie Politique*, nouvelle édition par G. Busino, Librairie Droz, Genève, 1964.
- [10] PARETO, Vilfredo, *Écrits sur la courbe de répartition de la richesse*, réunis et commentés par G. Busino, Librairie Droz, Genève, 1965.
- [11] PARETO, Vilfredo, "Quelques exemples d'application des méthodes d'interpolation à la statistique" et "Tables pour faciliter l'application de la méthode des moindres carrés", in *Statistique et Économie Mathématique*, textes réunis par G. Busino, Librairie Droz, Genève, 2^{ème} édition 1989, 77-92.
- [12] PUMAIN, Denise, "La dynamique des villes", *Economica*, Paris, 1982.