Figures du Pouvoir
La mesure de l'inégalité

Pierre Arnoux

Cours 4
Mercredi 15 février 2017

Répartition statistique : notions de base

- Histogramme :
- On considère une fonction d'un ensemble E (population) dans $\mathbb{R}$ (grandeur considérée)
- $f : E \to \mathbb{R}$.
- Salaire, richesse, âge, note...
- Histogramme : représentation du nombre d'éléments pour lesquels la grandeur a une valeur donnée.
- Exemple : répartition des notes d'un examen.

Répartition statistique : notions de base
Médiane, étendue, moyenne

- Médiane : valeur telle que la moitié de la population est au-dessus de cette valeur et la moitié au-dessous.
- Minimum : plus petite valeur.
- Maximum : plus grande valeur.
- Etendue : intervalle qui contient toutes les valeurs : c'est l'intervalle [Minimum, Maximum]
- Moyenne : Somme de toutes les valeurs, divisée par le nombre de valeurs.

Répartition statistique : notions de base
Quartile, déciles, centiles

- Quartile : division de l'étendue en quatre intervalle disjoints qui contiennent chacun un quart de la population.
- Premier quartile : L'intervalle correspondant au quart de la population pour lequel la valeur est la plus petite.
- Dernier quartile ; valeur la plus haute.
- La médiane sépare le second et le troisième quartile.
- Décile : division de l'étendue en 10 intervalles
- Centile : division en 100 intervalles ; le dernier centile correspond au 1% du haut.

Répartition statistique : notions de base
Quantiles

- Plus généralement, on parle de quantiles.
- Chaque quantile est défini par un intervalle.
- Il a un seuil inférieur (en anglais : threshold).
- Il a aussi une valeur moyenne, comprise entre les deux seuils : c'est la somme de toutes les valeurs, divisée par le nombre d'individus du quantile.
- Ces seuils et valeurs moyennes sont souvent disponibles.

Répartition statistique : notions de base
Proportion ou probabilité

- Plutôt que de considérer le nombre d'éléments d'une partie de la population, on peut considérer le pourcentage.
- Par exemple, au lieu de considérer le nombre de personnes qui gagnent plus de 50000 euros par an,
- on considère le pourcentage des gens dans la population qui gagnent plus de 50000 euros par an.
- On le notera $P(X > 50000)$ (pourcentage, proportion, ou probabilité)
- Plus généralement, on sera amené à considérer $P(a < X < b)$.
- Avantage : la taille de l'échantillon n'intervient plus directement. C'est plus facile pour faire des comparaisons.

Répartition statistique : notions de base
Moyenne (bis)

- Si la grandeur prend les valeurs $x_1, \ldots, x_d$, où d est le nombre de valeurs
- Et la valeur x_i est prise n_i fois
- Nombre total de personnes : $N = \sum_{i=1}^{i=d} n_i$.
- Valeur moyenne : multiplier x_i par le nombre n_i d'éléments de taille x_i, ajouter,
- et diviser par N, nombre total d'éléments :
- formule : $\frac{\left(\sum_{i=1}^{i=d} x_i n_i\right)}{\left(\sum_{i=1}^{i=d} n_i\right)}$.

Répartition statistique : notions de base
Moyenne (bis)

- Plutôt que de considérer le nombre d'éléments, on peut considérer le pourcentage.
- Cela revient à remplacer la formule :
- $\frac{\left(\sum_{i=1}^{i=d} x_i n_i\right)}{\left(\sum_{i=1}^{i=d} n_i\right)}$
- par $\sum_{i=1}^{i=d} x_i \left(\frac{n_i}{\sum_{i=1}^{i=d} n_i}\right)$
- En raccourci : $\sum_{i=1}^{i=d} x_i p_i$, où $p_i = \frac{n_i}{N}$ est la probabilité de prendre la valeur x_i. La taille de l'échantillon n'intervient plus directement.

Répartition statistique : Fonction de répartition

- Soit X une variable.
- On appelle fonction de répartition de X, et on note F_X, la fonction qui donne le pourcentage de la population pour lequel la grandeur est inférieure ou égale à une valeur donnée.
- $F_X(a) = P(X \leq a)$.
- C'est ce que l'on appelle parfois des données cumulées.

Répartition statistique : Fonction de répartition

- Fonction de répartition et quantité cumulée :
- $F(x)$ pourcentage de la population pour lequel la grandeur est inférieure ou égale à x.
- Propriété caractéristique de la fonction de répartition :
- La fonction de répartition est une fonction croissante qui va de 0 à 1.
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to +\infty} F(x) = 1$

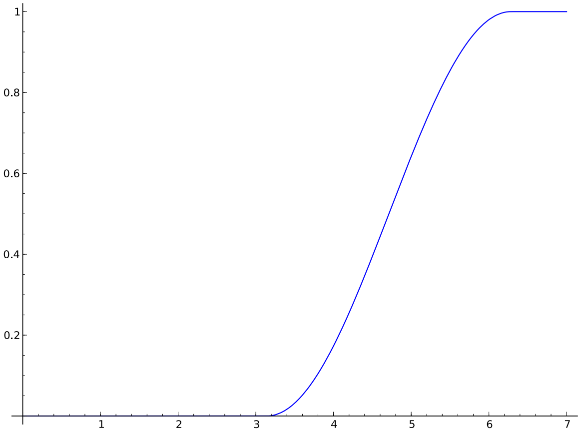
Répartition statistique : complémentaire de la Fonction de répartition

- complémentaire :
- $G(x) = 1 - F(x)$ pourcentage de la population pour lequel la grandeur est supérieure à x.
- Propriété caractéristique de cette fonction :
- c' est une fonction décroissante qui va de 1 à 0.
- $\lim_{x \to -\infty} F(x) = 1$
- $\lim_{x \to +\infty} F(x) = 0$

Répartition statistique : Fonction de répartition

- Propriété caractéristique de la fonction de répartition :
- La fonction de répartition est une fonction croissante qui va de 0 à 1.
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to +\infty} F(x) = 1$

Répartition statistique : Fonction de répartition



Répartition statistique : Fonction de répartition

- On s'intéresse souvent au haut de la répartition.
- On considère alors le complémentaire de la fonction de répartition.
- $F_X(a) = P(X \leq a)$.
- $1 - F_X(a) = P(X > a)$.
- La valeur a telle que $P(X > a) = \frac{1}{100}$, c'est le seuil d'entrée dans le 1% supérieur.
- Seuil du centile supérieur : Upper one percent threshold

Répartition statistique : Fonction de répartition

- La distribution est évidemment discrète.
- Donc la fonction de répartition n'est pas continue.
- Mais si la population est grande, les sauts sont petits :
- On peut *modéliser* la fonction de répartition par une fonction continue.

Densité : motivation

- En pratique, la fonction de répartition fait des "sauts" : distribution "discrète", non continue.
- Pour de très grand nombre, il est plus simple de la remplacer par une fonction continue.
- Un exemple déjà vu : la courbe en cloche de Gauss.
- On peut même souvent supposer que cette fonction de répartition a une pente.

Dérivée

- La dérivée d'une fonction F au point x, si elle existe, est la pente de la tangente.
- Elle représente le taux d'accroissement, ou coefficient de variation de F en x.
- On la note $F'(x)$, et dans ce cours, on la désignera par f.
- A savoir : la dérivée de x^α est $\alpha x^{\alpha-1}$.

Intégrale

- L'intégrale d'une fonction f entre les points a et b est l'aire sous la courbe.
- On la note $\int_a^b f(t)\,dt$.
- Soit f continue ; $F(x) = \int_{-\infty}^x f(t)\,dt$ s'appelle une primitive de f.
- Théorème fondamental du calcul différentiel : On a $F'(x) = f(x)$.
- La dérivation est l'opération inverse de l'intégration.

(R)appel : dérivée et intégrale

- Seule chose à savoir pour nos calculs : La dérivée de x^a est ax^{a-1}
- Une primitive de x^a, si $a \neq -1$, est $\frac{1}{a+1}x^{a+1}$.
- Vraie pour tout a réel.
- On en aura tout le temps besoin.
- Une primitive de $\frac{1}{x}$ est $\ln x$.

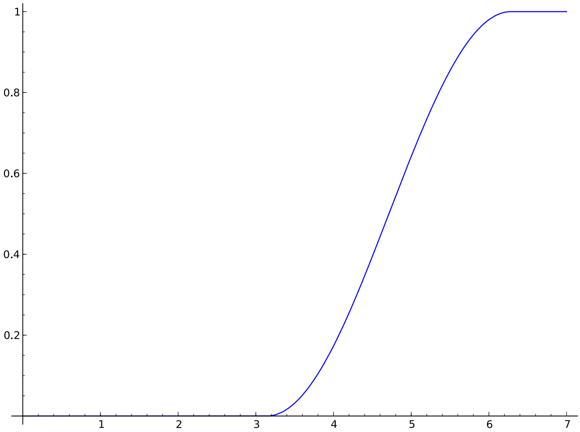
Calcul pratique d'une intégrale

- On veut calculer $\int_a^b f(t)\,dt$.
- Si on connaît une fonction F telle que $F'(x) = f(x)$
- (une primitive de f)
- Alors $\int_a^b f(t)\,dt = F(b) - F(a)$.
- A savoir : si $\gamma \neq -1$, une primitive de de x^γ est $\frac{x^{\gamma+1}}{\gamma+1}$.
- Car la dérivée de $x^{\gamma+1}$ est $(\gamma+1)x^\gamma$.

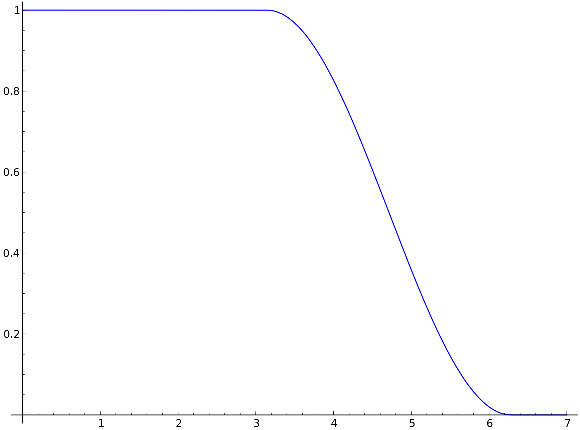
Répartition à densité

- On dit que la répartition de X est à densité f si la fonction de répartition F_X admet f pour dérivée.
- Propriété caractéristique de la densité :
- f est une fonction positive telle que $\int_{-\infty}^{+\infty} f(t)\,dt = 1$.

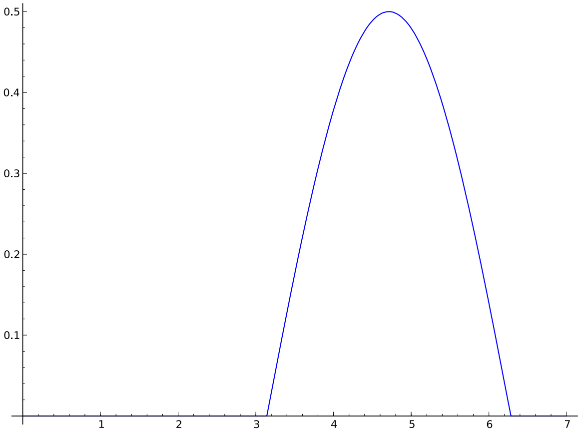
Répartition statistique : Fonction de répartition



Répartition statistique : Complémentaire de la fonction de répartition



Répartition statistique : densité



Répartition à densité : la formule fondamentale

- Si la répartition de X est à densité f on a .
- $P(a < X < b) = \int_a^b f(t)\,dt$.
- Exemple : densité en x^{-2} : $f(x) = 0$ si $x < 1$, $f(x) = x^{-2}$ si $x \geq 1$
- $P(1 < X < 2) = \int_1^2 x^{-2}\,dx$
- $P(1 < X < 2) = [-x^{-1}]_1^2 = \frac{-1}{2} - \frac{-1}{1} = \frac{1}{2}$.
- $P(1 < X < +\infty) = 1$.

Espérance ou moyenne d'une loi finie

- La moyenne est donnée par la somme $\sum_i x_i p_i$.
- On fait la somme de toutes les valeurs obtenues, et on divise par leur nombre.
- C'est ainsi qu'on calcule la moyenne des notes pour un groupe d'étudiants.
- C'est aussi de cette façon que l'on calcule la moyenne d'un examen, avec notes pondérées.

Espérance ou moyenne d'une loi à densité

- La moyenne est donnée par l'intégrale $\int_{-\infty}^{+\infty} t\, f(t)\, dt$.
- C'est une généralisation naturelle de la formule pour la moyenne d'une distribution finie $\sum_i x_i p_i$.
- Elle se calcule bien pour les répartitions les plus simples.
- C'est la même formule que pour le calcul du centre de gravité

La question

- On voudrait étudier la répartition des revenus : deux questions
- Comment représenter cette répartition ?
- C'est forcément une fonction, c'est-à-dire une donnée complexe.
- Peut-on réduire cette fonction à un seul chiffre (indice d'inégalité)
- Sans perdre trop d'information ?

Représentations de la répartition des revenus

- Le revenu est une variable R associée à une population.
- On peut lui attribuer une fonction de répartition F_R
- Ou son complémentaire $G_R = 1 - F_R$ (population ayant plus que le revenu R)
- On peut le faire en population, ou en pourcentage
- On peut le représenter en graphe cartésien
- Ou en coordonnées logarithmiques, ou semi logarithmiques
- Ce n'est pas toujours très lisible.

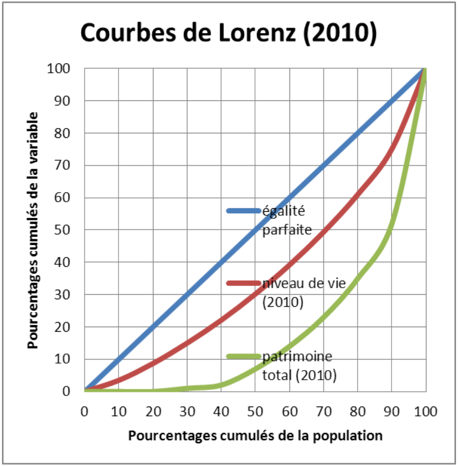
Part du 10%

- Il s'agit de la part que les 10% les plus riches reçoivent du revenu total.
- On peut bien sûr considérer d'autres quantiles (20% les plus riches, 1% les plus riches, 50% les plus pauvres...)
- C'est une mesure facile à obtenir et à comparer.

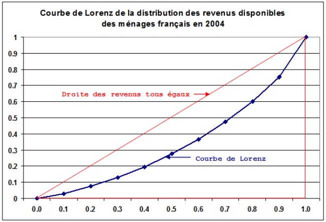
La courbe de Lorenz

- La courbe de Lorenz (1905) associe, aux $x\%$ les plus pauvres de la population, la fraction $y = L(x)$ du revenu qu'ils reçoivent.
- C'est une fonction $L : [0,1] \to [0,1]$
- Si $x = F_R(R)$ est la fonction de répartition du revenu
- Alors $R = F_R^{-1}(x)$ est le revenu maximal des $x\%$ les plus pauvres
- et leur revenu total est $\int_0^{F_R^{-1}(x)} t\, f_R(t)\, dt$
- Donc L se calcule par $L(x) = \frac{\int_0^{F_R^{-1}(x)} t\, f_R(t)\, dt}{\int_0^{+\infty} t\, f_R(t)\, dt} \frac{\int_0^{F_R^{-1}(x)} t\, f_R(t)\, dt}{\mu}$

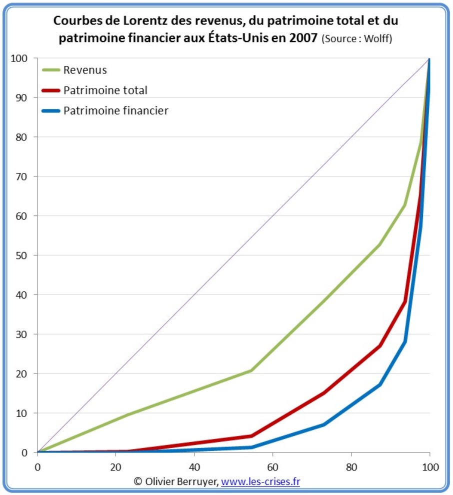
La courbe de Lorenz



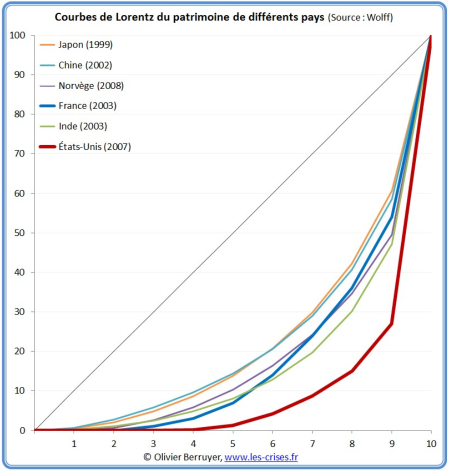
La courbe de Lorenz



La courbe de Lorenz



La courbe de Lorenz



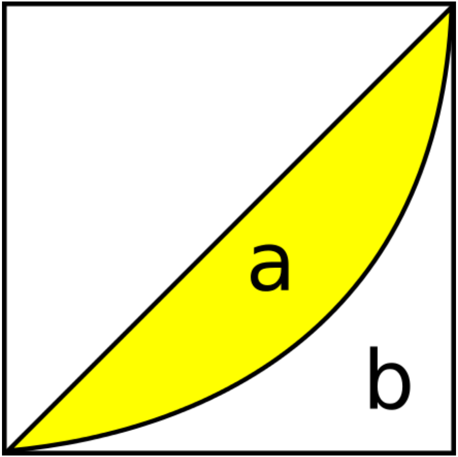
Indices de l'inégalité

- Une courbe ou une distribution, c'est compliqué
- Il est délicat de comparer deux courbes ou deux distributions
- On voudrait pouvoir réduire cette distribution à un seul nombre (indice)
- Pour comparer les inégalités.
- Il existe de nombreux indices
- Il n'y en a pas de parfaits.

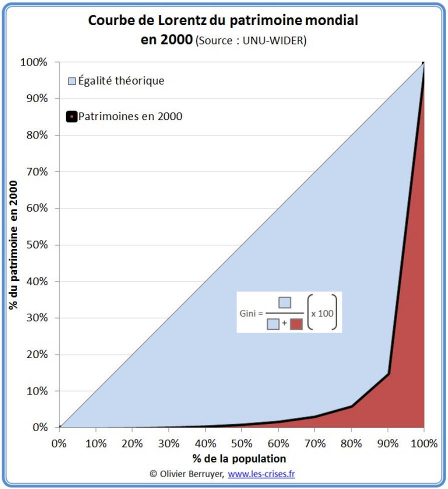
L'indice de Gini

- C'est le rapport entre deux domaines :
- Le domaine entre la courbe de Lorenz et la courbe d'égalité parfaite
- et le triangle sous la courbe d'égalité parfaite.

L'indice de Gini



L'indice de Gini



L'indice de Gini

- L'indice de Gini est le rapport $\frac{a}{a+b}$
- il est compris entre 0 et 1
- $gini = 0$: égalité parfaite.
- $gini = 1$: inégalité parfaite : une seule personne possède tout.
- très utilisé car facile à comprendre et à calculer.

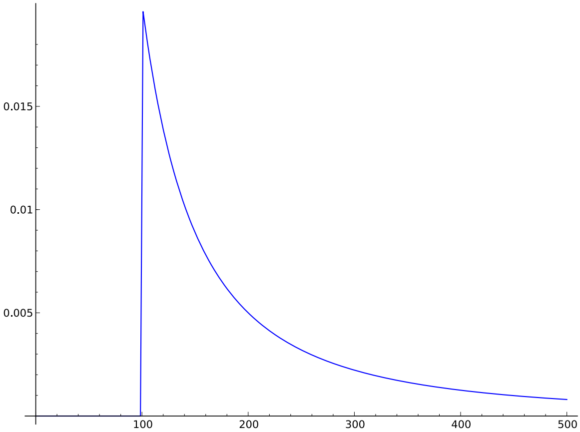
Le coefficient de Pareto-Lorenz

- Il repose sur l'hypothèse que le haut de la courbe des revenus est donné par une loi de Pareto
- Il ne dépend que du haut de la distribution
- Il nécessite quelques calculs de départ.

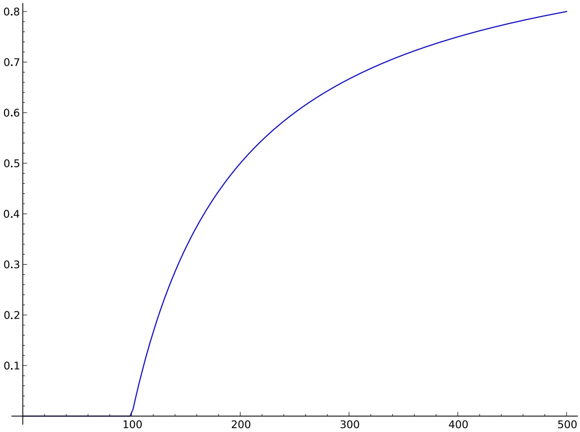
Loi de Pareto

- Définition : loi de Pareto de paramètre x_m, α :
- C'est une loi à densité donnée par $f(t) = 0$ si $t < x_m$
- x_m est le revenu minimum
- $f(t) = \frac{C}{t^{\alpha+1}}$ si $t \geq x_m$.
- On calcule facilement $C = \alpha x_m^\alpha$ (cours précédent)

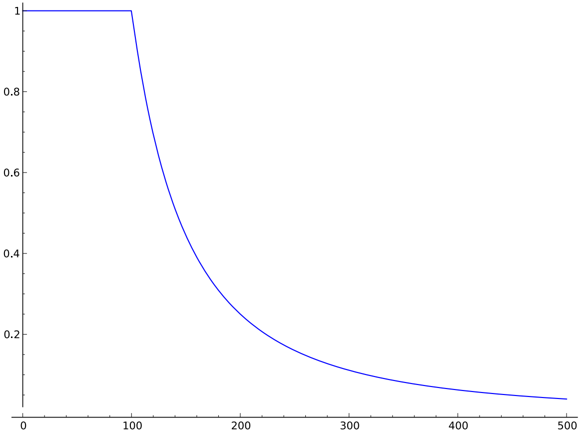
Loi de pareto : densité



Loi de pareto : fonction de répartition



Loi de pareto : probabilité du sommet



Répartition des revenus

- La répartition des revenus et des richesses
- Au moins dans la partie supérieure
- est bien approximée par une loi de Pareto
- Ce qui est important, c'est le coefficient α

Comment reconnaître une loi de Pareto ?

- Que vaut $p(X > a)$?
- $\int_a^{+\infty} \frac{C}{t^{\alpha+1}}\,dt$
- $\int_a^{+\infty} Ct^{-\alpha-1}\,dt = [\frac{C}{\alpha}t^{-\alpha}]_a^{\infty}$
- $\left(\frac{x_m}{a}\right)^{\alpha}$
- Comment reconnaître cette relation et calculer α ?
- En prenant les logarithmes...

Comment reconnaître une loi de Pareto ?

- $P(X > a) = \left(\frac{x_m}{a}\right)^{\alpha}$
- $\log P(X > a) = \log\left(\frac{x_m}{a}\right)^{\alpha}$
- $\log P(X > a) = \alpha \log(x_m) - \alpha \log(a)$
- Si on trace $\log(P(X > a)$ en fonction de $\log(a)$ (repère log-log)
- on obtient une droite !

Comment calculer le coefficient α ?

- Si on connait deux valeurs $P(X > a_1)$ et $P(X > a_2)$
- On a $\log P(X > a_1) - \log P(X > a_2) = -\alpha\,(\log(a_1) - \log(a_2))$
- Donc $\alpha = \frac{\log P(X>a_1) - \log P(X>a_2)}{\log(a_2) - \log(a_1)}$.
- C'est une méthode pratique, qui marche dès que l'on connaît le seuil de deux quantiles.
- En pratique, on cherche le a_1 tel que $P(X > a_1) = 0,1$ et le a_2 tel que $P(X > a_2) = 0,01$ (par exemple)

Une simplification des calculs

- Au lieu de considérer X, on peut considérer $\frac{X}{x_m}$ (choisir le revenu minimum comme unité de compte)
- cette variable suit aussi une distribution de Pareto, avec le même α et $x_m = 1$
- Cela simplifie les calculs !
- On prend désormais comme densité $f(t) = \frac{\alpha}{t^{\alpha+1}}$.

Espérance d'une loi de Pareto

- Le revenu moyen est donné par
- $\int_1^{+\infty} tf(t)\, dt = \int_1^{+\infty} \frac{\alpha}{t^\alpha}\, dt$
- Soit $\frac{\alpha}{\alpha-1}[t^{\alpha-1}]_1^{+\infty}$
- Soit $\frac{\alpha}{\alpha-1}$

La propriété caractéristique

- Le revenu moyen des personnes avec un revenu supérieur à x
- est égal à βx
- où $\beta = \frac{\alpha}{\alpha - 1}$
- Coefficient de Pareto inversé
- en pratique : $\beta = 1,5$: forte égalité
- $\beta = 3$: forte inégalité.

La propriété caractéristique

- On peut montrer que cette propriété caractérise complètement les distributions de Pareto.
- Si une distribution vérifie, pour $x > x_m$, $E(X|X > x) = \beta x$
- Alors c'est une distribution de Pareto
- De coefficient de Pareto $\alpha = \frac{\beta}{\beta-1}$.
- C'est un exercice d'analyse de niveau L2.

Comment calculer le coefficient β ?

- Le coefficient β est très simple à calculer.
- Si on connaît le revenu de seuil R_s d'un certain quantile
- et le revenu moyen μ de ce quantile,
- Alors $\beta = \frac{\mu}{R_s}$
- C'est une méthode indépendante de la précédente.
- Et qui permet aussi de calculer α

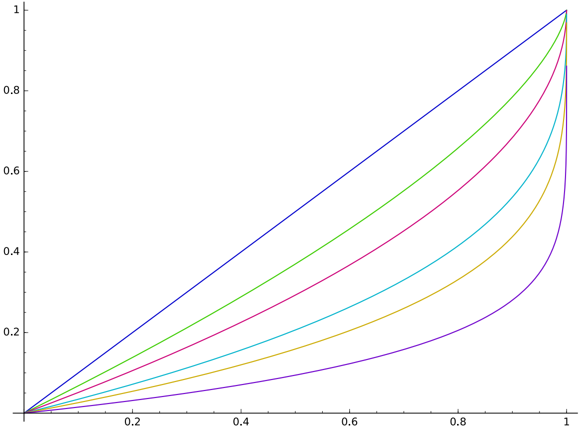
Effet du coefficient β ?

- On peut calculer, en fonction de β, la proportion du revenu total détenu par le quantile supérieur.
- On montre facilement
- (enfin, assez facilement),
- que la fraction p supérieure de la population
- détient une fraction $p^{\frac{1}{\beta}}$ du revenu total.
- Pour le centile supérieur : $0,01^{\frac{1}{\beta}}$.

Courbe de Lorenz d'une distribution de Pareto

- On en déduit la formule pour la courbe de Lorenz d'une distribution de Pareto de paramètre β donné :
- $L(x) = 1 - (1 - x)^{\frac{1}{\beta}}$

Courbes de Lorenz associées à des distributions de Pareto : $\beta = 1.5, 2, 3, 4, 7$



Indice de Gini d'une distribution de Pareto

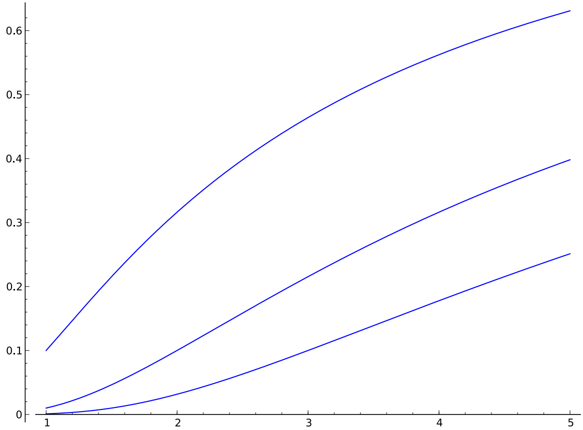
- De la formule $L(x) = 1 - (1-x)^{\frac{1}{\beta}}$ on déduit par un calcul simple l'indice de Gini :
- L'indice de Gini d'une distribution de Pareto de paramètre β est $\frac{\beta-1}{\beta+1}$
- Il vaut 0 pour $\beta = 1$
- Et tend vers 1 quand β tend vers l'infini.

Effet du coefficient β ?

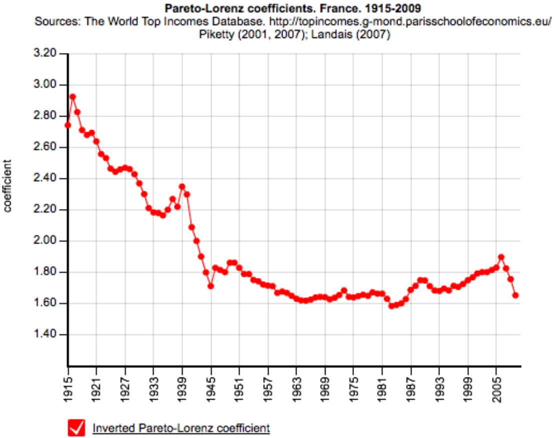
Part des quantiles supérieurs dans le revenu total

β	10%	1%	0,1%
$1,5$	21,5%	4,6%	1%
$1,8$	28%	7,7%	2,2%
2	31,6%	10%	3,1%
$2,5$	39,8%	15,8%	6,3%
3	46,4%	21,5%	10%
4	56,2%	31,6%	17,8%

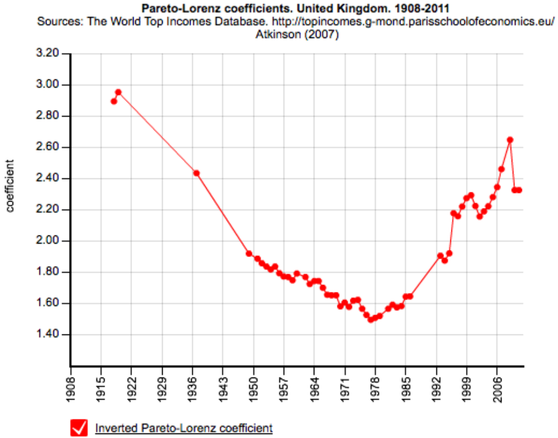
Part du décile, du centile et du millile supérieurs en fonction de β



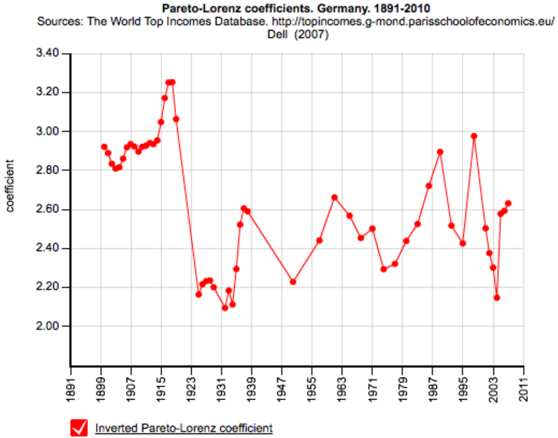
Coefficient β : France



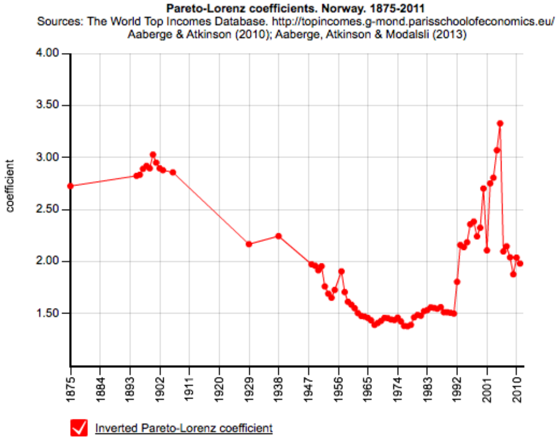
Coefficient β : Grande Bretagne



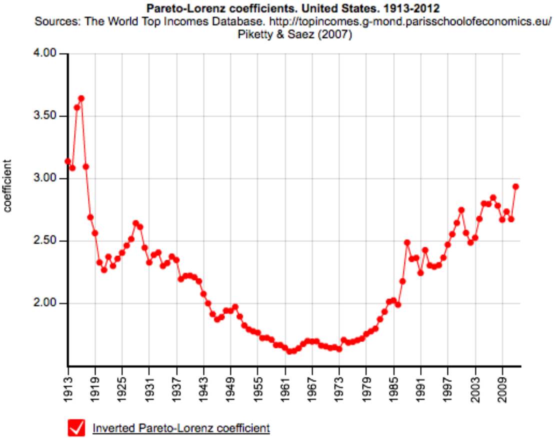
Coefficient β : Allemagne



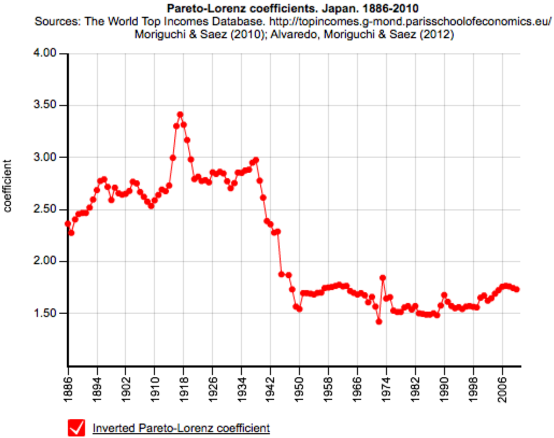
Coefficient β : Norvège



Coefficient β : Etats-Unis



Coefficient β : Japon



Coefficient β : Chine

Coefficient β : Argentine