

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Fisica

Tesi di Laurea

**Disuguaglianza sociale: confronto dell'indice di Gini con
nuovi estimatori di fisica generale**

Relatore

Prof. Luca Stanco

Laureando

Mattia Ceravolo

Anno Accademico 2019/2020

Indice

Introduzione	3
1 Curva di Lorenz	3
1.1 Indice di Gini	4
1.2 Altri indici di disuguaglianza	5
2 Distribuzioni di ricchezza	6
2.1 Società monarchica	6
2.2 Società oligarchica	7
2.3 Distribuzione Gamma	7
2.4 Distribuzione Esponenziale	8
2.5 Distribuzione Lognormale	8
3 Distribuzioni di probabilità della ricchezza in Francia	9
3.1 Verifica della legge di Pareto	9
3.2 Analisi della disuguaglianza sociale	11
3.3 Power-Lax index	15
4 Conclusioni	17

Introduzione

La disuguaglianza sociale, dacché elemento di studio dell'economia, interessò ben presto i fisici quando ci si rese conto che era possibile studiare la ricchezza di una popolazione con modelli di meccanica statistica e ricavare distribuzioni di probabilità che ben la descrivevano empiricamente [1, 2, 3]. Divenne interesse dei fisici anche quantificare la disuguaglianza sociale attraverso l'utilizzo degli indici più conosciuti di Gini, Pietra e Kalkuta e l'introduzione di nuovi estimatori della disuguaglianza sociale [4, 1, 5, 6, 7, 8]. Nel capitolo 1 si andrà a definire la curva di Lorenz con le annesse misure di disuguaglianza più conosciute e si andranno a ridefinire queste misure in funzione della sua densità di probabilità. Nel capitolo 2 si descriveranno le PDF di riferimento nella letteratura scientifica per descrivere la ricchezza delle popolazioni e si andranno a calcolare esplicitamente gli indici di disuguaglianza. Si introdurranno nel capitolo 3 i dati in allegato tecnico al libro "T. Piketty, il capitale nel XXI secolo" [9] e dopo averli usati per verificare la legge di Pareto [10] si studierà nel capitolo 4 la disuguaglianza sociale in Francia mettendo in evidenza gli aspetti critici degli usuali indici di disuguaglianza ed introducendo due nuovi indici l'ultimo dei quali definito nella lettura scientifica dell'econo-fisica [6].

1 Curva di Lorenz

Data una distribuzione di probabilità $P(m)$ sulla variabile casuale "ricchezza" m definita in $[0, \infty]$ la curva di Lorenz esprime la dipendenza tra:

$$Y(r) = \frac{\int_0^r mP(m)dm}{\int_0^\infty mP(m)dm}$$

e la distribuzione cumulativa

$$X(r) = \int_0^r P(m)dm$$

I punti $(X(r), Y(r))$ definiscono quindi la curva di Lorenz, che quindi esprime la percentuale di ricchezza posseduta da una data percentuale di popolazione ordinata per ricchezza. Il fatto che la curva sia sempre compresa nel quadrato di ascisse $[0, 1]$ ed ordinate $[0, 1]$ e che parta sempre nel punto $(0, 0)$ per finire in $(1, 1)$ rende possibile la comparazione di curve di Lorenz ricavate da diverse distribuzioni di probabilità.

Società egualitaria

In una società egualitaria la ricchezza è suddivisa equamente tra i membri della stessa che ne posseggono tutti la stessa quantità. Ne consegue quindi che la percentuale di popolazione ordinata per ricchezza X possiede la percentuale di ricchezza $Y = X$ e la curva di Lorenz sia dunque costituita dai punti del piano (X, X) per $X \in [0, 1]$. Infatti se tutti gli individui posseggono ricchezza m' , $P(m) = \delta(m - m')$ e quindi:

$$X(r) = \int_0^r \delta(m - m')dm = \Theta(r - m')$$

$$Y(r) = \frac{\int_0^r m\delta(m - m')dm}{\int_0^\infty m\delta(m - m')dm} = \frac{m'\Theta(r - m')}{m'} = X(r)$$

dove

$$\Theta(r - m') = \begin{cases} 1 & r > m' \\ \frac{1}{2} & r = m' \\ 0 & r < m' \end{cases}$$

Si noti che la curva di Lorenz che si è intesa per la ricchezza può essere costruita per qualsiasi variabile "posseduta" da agenti diversi. X e Y ad esempio possono essere articoli/citazioni, città/popolazioni, candidati/voti etc.

1.1 Indice di Gini

L'indice di Gini, compreso tra 0 ed 1, vuole misurare la disuguaglianza di un campione intesa come il grado di non uguaglianza dello stesso: lo fa calcolando di quanto la curva di Lorenz del campione differisca dalla curva di uguaglianza vista in precedenza. Sia $Y = L(X)$ la funzione di Lorenz e $L_{pe}(X) = X$ la curva di uguaglianza, allora si definisce l'indice di Gini come:

$$g = 2 \int_0^1 (L_{pe}(x) - L(x))dx$$

ovvero 2 volte l'area compresa tra la curva di uguaglianza e la curva di Lorenz (figura 1). Qualora $L(x)$ sia pari alla curva di uguaglianza l'indice di Gini è pari a 0 ed esprime massima uguaglianza. Un aumento dell'indice di Gini sta a significare un aumento della disuguaglianza fino ad arrivare ad un indice pari ad 1 in una società monarchica nella quale tutta la ricchezza è posseduta da un singolo agente. Si andrà ora a sostituire la curva di Lorenz con la distribuzione di probabilità $P(x)$ da cui è

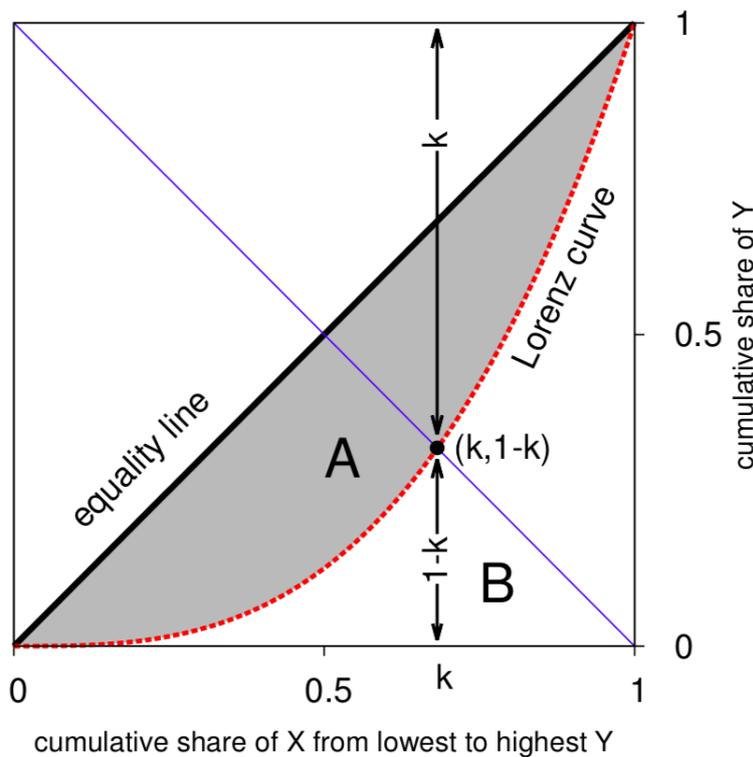


Figura 1: Esempio di curva di Lorenz. La linea nera rappresenta la curva di uguaglianza, quella tratteggiata in rosso è una generica curva di Lorenz. L'indice di Gini è pari a $g = 2A = \frac{A}{A+B}$. Il k-index si trova intercettando la curva $1 - x$ con la curva di Lorenz.

ricavata e si andrà a vedere, attraverso espliciti calcoli¹, il significato dell'indice di Gini riguardo la distribuzione di probabilità della ricchezza. Si osservi che:

$$g = 1 - 2 \int_0^1 L(X)dX = 1 - 2[XL(X)]_0^1 + 2 \int_0^1 XL'(X)dX = -1 + 2 \int_0^1 XL'(X)dX$$

effettuando il cambio di variabili $X = X(m)$, $dX(m) = P(m)dm$ e sfruttando il fatto che $L(X) = \frac{1}{\langle m \rangle} \int_0^X X^{-1}(u)du$ [11] e quindi $L'(X) = \frac{X^{-1}(X(m))}{\langle m \rangle} = \frac{m}{\langle m \rangle}$

$$= \frac{2}{\langle m \rangle} \int_0^\infty mX(m)P(m)dm - 1 = \frac{2}{\langle m \rangle} \left(\int_0^\infty mX(m)P(m)dm - \frac{\langle m \rangle}{2} \right) = \frac{2}{\langle m \rangle} Cov(m, X(m)) \quad (1)$$

¹Sono qui inclusi i calcoli mancanti in [11] ed è derivato in maniera differente il risultato di 4.4 di [11]

Si trova quindi che l'indice di Gini è proporzionale alla covarianza della variabile ricchezza e della sua funzione cumulativa (quante persone posseggono al più quella ricchezza). Se si va ora ad esplicitare la media:

$$\text{Cov}(m, X(m)) = \int_0^\infty mX(m)dP(m)dm - \frac{1}{2} \int_0^\infty mP(m)dm = \int_0^\infty m(X(m) - \frac{1}{2})P(m)dm$$

integrando per parti (separando m da $(X(m) - \frac{1}{2})P(m)$ e derivando m):

$$\text{Cov}(m, X(m)) = [\frac{1}{2}mX(m)(X(m) - 1)]_0^\infty + \frac{1}{2} \int_0^\infty X(m)(1 - X(m))dm = \frac{1}{2} \int_0^\infty X(m)(1 - X(m))dm.$$

Si trova quindi un altro modo di esprimere l'indice di Gini, nota la distribuzione di probabilità $P(m)$ e la sua cumulativa $X(m)$, come

$$g = \frac{1}{\langle m \rangle} \int_0^\infty X(m)(1 - X(m))dm. \quad (2)$$

Si vuole ora mostrare che:

$$g = \frac{1}{\langle m \rangle} E(\max(0, m_1 - m_2)). \quad (3)$$

E' facile dimostrare che $E(\max(0, m_1 - m_2)) = \frac{1}{2}E(|m_1 - m_2|)$. Ora si osservi che $\frac{|x-y|}{2} = \frac{(x+y)}{2} - \min(x, y)$ quindi se le due variabili sono identicamente distribuite con distribuzione $P(m)$ e media $\langle m \rangle$ il valore di aspettazione sarà:

$$\frac{E(|x - y|)}{2} = \langle m \rangle - E(\min(x, y)) \quad (4)$$

Si scrive la distribuzione di probabilità cumulativa di $X(\min(x, y) < m) = 1 - X(\min(x, y) > m) = 1 - X(x > m)X(y > m) = 1 - (1 - X(m))^2$ ed inserendola in (4):

$$\frac{E(|x - y|)}{2} = \langle m \rangle - E(\min(x, y)) = \langle m \rangle + \int_0^\infty md(1 - X(m))^2$$

integrando per parti si ottiene:

$$\begin{aligned} \frac{E(|x - y|)}{2} &= \langle m \rangle + [m(1 - X(m))^2]_0^\infty - \int_0^\infty (1 - X(m))^2 dm \\ &= \langle m \rangle - \int_0^\infty (1 - X(m))dm + \int_0^\infty X(m)(1 - X(m))dm \\ &= \langle m \rangle - [m(1 - X(m))]_0^\infty - \int_0^\infty mP(m)dm + \int_0^\infty X(m)(1 - X(m))dm \\ &= \langle m \rangle g \end{aligned}$$

da cui si trova (3).

1.2 Altri indici di disuguaglianza

Dalla curva di Lorenz si possono ottenere altri indici di disuguaglianza oltre quello di Gini .

k-index

Il Kolkata index o k-index (figura 1) sicuramente più espressivo dell'indice di Gini è rappresentato dalla coordinata X nel piano (X, Y) dell'intersezione della funzione $1 - X$ con la curva di Lorenz. Un k-index pari a k sta a significare che la frazione k di agenti ordinati per ricchezza detiene la frazione

$1 - k$ di ricchezza o che la frazione $1 - k$ di agenti inversamente ordinati per ricchezza possiede la frazione k di ricchezza. Similmente all'indice di Gini il k -index varia tra 0 ed 1 ed è crescente in ordine di disuguaglianza.

Indice di Pietra p

Analogamente all'indice di Gini, l'indice di Pietra p , introdotto dallo statistico italiano Gaetano Pietra, quantifica la distanza della curva di Lorenz dalla curva di uguaglianza.

$$p = \max_{X \in [0,1]} (L(X) - L_{pe}(X))$$

Come l'indice di Gini si vede essere "l'eccedenza" della variabile m_1 sulla variabile m_2 , l'indice di Pietra risulta essere l'eccedenza della variabile m sulla media. Infatti (senza perdersi in ulteriori calcoli) come per l'indice di Gini si trova [5]:

$$p = \frac{1}{\langle m \rangle} \int_{\langle m \rangle}^{\infty} X(m) dm = \frac{1}{\langle m \rangle} E(\max(0, m - \langle m \rangle)).$$

Si può dare un altro significato alla misura dell'indice di Pietra. Questo risulta infatti rappresentare la percentuale di ricchezza che deve essere redistribuita, dai ricchi ai poveri, per ottenere una società perfettamente egualitaria [4]. Come l'indice di Pietra ci sono molti altri indici che possono essere costruiti da distanze geometriche sulla curva di Lorenz ed hanno un significato ben preciso sulla distribuzione di probabilità [6].

2 Distribuzioni di ricchezza

Andiamo ora a vedere le distribuzioni di ricchezza più banali ed alcune presenti in letteratura che sembrano avvicinarsi di più alle distribuzioni empiriche.²

2.1 Società monarchica

In una società monarchica la ricchezza è posseduta da un unico individuo con ricchezza m' ; tutti gli altri hanno ricchezza 0. In questo caso $P(m) = (1 - \epsilon)\delta(m) + \epsilon\delta(m - m')$ con $\epsilon \rightarrow 0$. E quindi integrando:

$$X(r) = \int_0^r P(m) dm = \int_0^r ((1 - \epsilon)\delta(m) + \epsilon\delta(m - m')) dm = (1 - \epsilon) + \epsilon\Theta(r - m') = \begin{cases} 1 - \epsilon & r < m' \\ 1 & r \geq m' \end{cases}$$

Si ha invece per $Y(r)$:

$$Y(r) = \frac{\int_0^r m((1 - \epsilon)\delta(m) + \epsilon\delta(m - m')) dm}{\int_0^{\infty} mP(m) dm} = \frac{m'\epsilon\Theta(r - m')}{m'\epsilon} = \begin{cases} 0 & r < m' \\ 1 & r \geq m' \end{cases}$$

$Y(X)$ sarà sempre pari a 0 tranne quando $r \geq m'$, ovvero quando $X(r)$ è pari ad 1, e quindi $Y(1) = 1$.

Andiamo a calcolare l'indici di Gini come:

$$g = 1 - 2 \int_0^{\infty} Y(m)P(m) dm = 1 - 2 \int_{m'}^{\infty} P(m) dm = 1 - 2 \int_{m'}^{\infty} ((1 - \epsilon)\delta(m) + \epsilon\delta(m - m')) dm = 1 - 2\epsilon = 1$$

e l'indice di Pietra P osservando che $X - Y = (1 - \epsilon)\Theta(r) - (1 - \epsilon)\Theta(r - m')$ è massimo per $0 < r < m'$, vale:

$$P = 1 - 2\epsilon = 1$$

² Andremo a modellizzare con i metodi usati per la curva egualitaria in [8] altre distribuzioni banali. Per le distribuzioni più complesse si verificheranno i risultati sull'indice di Pietra in [6]

2.2 Società oligarchica

In una società oligarchica la ricchezza è posseduta da una frazione p di popolazione equamente suddivisa. Ogni agente della frazione p possiederà la ricchezza m' . Tutti gli altri hanno ricchezza 0. Quindi $P(m) = (1-p)\delta(m) + p\delta(m-m')$:

$$X(r) = \int_0^r ((1-p)\delta(m) + p\delta(m-m'))dm = (1-p) + p\Theta(r-m')$$

$$Y(r) = \frac{\int_0^r m((1-p)\delta(m) + p\delta(m-m'))dm}{\int_0^\infty m((1-p)\delta(m) + p\delta(m-m'))dm} = \frac{pm'\Theta(r-m')}{pm'}$$

ovvero

$$Y(r) = \begin{cases} 0 & r < m' \\ \frac{X(r)}{p} - \frac{(1-p)}{p} & r \geq m' \end{cases}$$

Andiamo a calcolare l'indice di Gini osservando che $Y(X) = 0$ per $X \in [0, 1-p]$ come:

$$\begin{aligned} g &= 1 - 2 \int_{1-p}^1 Y(X)dX = 1 - 2 \int_{1-p}^1 \left(\frac{X}{p} - \frac{1-p}{p} \right) dX = 1 - \left[\frac{X^2}{p} \right]_{1-p}^1 + 2 \left[\frac{X(1-p)}{p} \right]_{1-p}^1 \\ &= 1 - \frac{1}{p} + \frac{(1-p)^2}{p} + 2 \frac{1-p}{p} - 2 \frac{(1-p)^2}{p} = 1-p \end{aligned}$$

e l'indice di Pietra P osservando che $X-Y = (1-p)\Theta(r) - (1-p)\Theta(r-m')$ è massimo per $0 < r < m'$, vale:

$$P = 1-p$$

2.3 Distribuzione Gamma

La distribuzione Gamma è presente nella letteratura come possibile candidata della distribuzione di ricchezza (almeno fino al 97-98-esimo percentile) Calcoliamo la curva di Lorenz per questa distribuzione. Si ha che $P(m) \propto m^\alpha e^{-\beta m}$ con $\alpha, \beta > 0$. Il fattore di normalizzazione vale $X(\infty) = \int_0^\infty m^\alpha e^{-\beta m} dm = \frac{\Gamma(\alpha+1)}{\beta^{1+\alpha}}$. Si procede ora al calcolo di $X(r)$ e $Y(r)$ (dove si ricorda che $\gamma(a, x) = \int_0^x t^a e^{-t} dt$):

$$X(r) = \frac{\int_0^r m^\alpha e^{-\beta m} dm}{\frac{\Gamma(\alpha+1)}{\beta^{1+\alpha}}} = \frac{\gamma(\alpha+1, \beta r)}{\Gamma(\alpha+1)}$$

$$Y(r) = \frac{1}{\Gamma(\alpha+2)} \int_0^{\beta r} m^{\alpha+1} e^{-m} dm = \frac{\gamma(\alpha+2, \beta r)}{\Gamma(\alpha+2)}$$

Da cui si ottiene la curva di Lorenz $L(r) = (X(r), Y(r))$. Si sfrutta ora la proprietà $\gamma(a+1, x) = a\gamma(a, x) - x^a e^{-x}$. Quindi:

$$\gamma(\alpha+2, \beta r) = (\alpha+1)\gamma(\alpha+1, \beta r) - (\beta r)^{\alpha+1} e^{-\beta r}$$

$$\Gamma(\alpha+2)Y = (\alpha+1)\Gamma(\alpha+1)X - (\beta r)^{\alpha+1} e^{-\beta r}$$

$$X - Y = \frac{(\beta r)^{\alpha+1} e^{-\beta r}}{\Gamma(\alpha+2)} \quad (5)$$

Si deriva (5) e si pone uguale a 0 per trovare il massimo e quindi l'indice di Pietra:

$$(X - Y)' = \frac{\beta(\beta r)^\alpha e^{-\beta r}}{\Gamma(\alpha+1)} - \frac{\beta(\beta r)^{\alpha+1}}{\Gamma(\alpha+2)} = 0 \quad (6)$$

da cui si trova:

$$r = \frac{\alpha+1}{\beta} \quad (7)$$

Inserendo (7) in (5) si trova:

$$p = \max_{X \in [0,1]} X - Y(X) = \frac{(\alpha + 1)^{\alpha+1} e^{-(\alpha+1)}}{\Gamma(\alpha + 2)} \quad (8)$$

In modo analogo si trova l'indice di Gini [1]

$$g = \frac{\Gamma(2(\alpha + 1))}{2^{2\alpha+1} \Gamma(\alpha + 1) \Gamma(\alpha + 2)} \quad (9)$$

2.4 Distribuzione Esponenziale

Anche la distribuzione esponenziale è presente in letteratura. E' un particolare caso di una distribuzione gamma con $\alpha = 0$. La distribuzione risulta essere $P(m) = \beta e^{-\beta m}$ definita per $m > 0$ con valore medio $\frac{1}{\beta}$. Sostituendo $\alpha = 0$ in (8) e (9) si trova che $p = \frac{1}{e}$ e $g = \frac{1}{2}$.

Si nota che sia per la distribuzione Gamma che per quella esponenziale il parametro β rappresenta solo un parametro di scala che non cambia la forma della PDF e neanche quella della curva di Lorenz. Per tale motivo non si trova una dipendenza degli indici da β .

2.5 Distribuzione Lognormale

Anche questa presente nella letteratura, la distribuzione Lognormale è la distribuzione di probabilità di una variabile m il cui logaritmo $\log(m)$ segue la distribuzione normale.

$$P(m) = \frac{e^{-\frac{(\log(m)-\mu)^2}{2\sigma^2}}}{m\sqrt{2\pi}\sigma}$$

con $\langle m \rangle = e^{(\mu + \frac{\sigma^2}{2})}$ e $Var(m) = e^{(2\mu + \sigma^2)}(e^{\sigma^2} - 1)$.

Con gli opportuni calcoli vale:

$$X(r) = \frac{1}{2} \operatorname{erfc} \left(\frac{\mu - \log(r)}{\sqrt{2}\sigma} \right)$$

$$Y(r) = \frac{1}{2} \operatorname{erfc} \left(\frac{\mu - \log(r) + \sigma^2}{\sqrt{2}\sigma} \right)$$

dove $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt$.

Seguendo gli stessi calcoli per la distribuzione Gamma, si trova per l'indice di Pietra:

$$X(r) - Y(r) = \frac{1}{2} \operatorname{erfc} \left(\frac{\mu - \log(r)}{\sqrt{2}\sigma} \right) - \frac{1}{2} \operatorname{erfc} \left(\frac{\mu - \log(r) + \sigma^2}{\sqrt{2}\sigma} \right)$$

$$X'(r) - Y'(r) = \frac{e^{-\frac{(\log(r)-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}r\sigma} - \frac{e^{-\frac{(\log(r)-\mu-\sigma^2)^2}{2\sigma^2}}}{\sqrt{2\pi}r\sigma} = 0$$

\iff

$$r = e^{\mu + \frac{\sigma^2}{2}}$$

$$p = \max_{r \in [0, \infty]} (X(r) - Y(r)) = \operatorname{erf} \left(\frac{\sigma}{2\sqrt{2}} \right)$$

3 Distribuzioni di probabilità della ricchezza in Francia

La distribuzione seguita dalla ricchezza di una popolazione è ancora oggetto di studio in diverse discipline. Alla fine del diciannovesimo secolo l'economista italiano Vilfredo Pareto trovò che le distribuzioni di ricchezza seguivano una legge di potenza per i redditi più alti [10]. La probabilità, in questa classe di redditi, che un individuo abbia ricchezza maggiore di m è pari a

$$P_{>}(m) = m_c^\mu \alpha m^{-\mu} \quad (10)$$

dove m_c è la ricchezza dopo la quale comincia a valere la legge di Pareto, μ è una costante maggiore di uno che risulta essere circa 1 per le distribuzioni dei capitali delle aziende e 2 per le distribuzione sui redditi [12] e α è la percentuale di popolazione con ricchezza maggiore di m_c . La distribuzione della classe media rimane tutt'ora oggetto di studio. Le distribuzioni conosciute che sembrano interpolare meglio le distribuzioni di ricchezza empiriche sono le distribuzioni lognormale, esponenziale e gamma [1]. Si può notare che:

$$m^\mu P_{>}(m) = \text{const.} \quad (11)$$

per $m > m_c$. Dalla (11) si trova:

$$m P_{>}(m)^{\frac{1}{\mu}} = \text{const.} = b^{\frac{1}{\mu}}$$

$$L'(X)(1-X)^{\frac{1}{\mu}} = b^{\frac{1}{\mu}} \Rightarrow L'(X) = \frac{b^{\frac{1}{\mu}}}{(1-X)^{\frac{1}{\mu}}} \Rightarrow L(X) = c2 - b^{\frac{1}{\mu}} \frac{(1-X)^{(1-\frac{1}{\mu})}}{(1-\frac{1}{\mu})}$$

e ponendo $L(1) = 1$:

$$L(X) = 1 - b^{\frac{1}{\mu}} \frac{(1-X)^{(1-\frac{1}{\mu})}}{(1-\frac{1}{\mu})} \quad (12)$$

per $X > \alpha$.

3.1 Verifica della legge di Pareto

Si può andare a verificare quanto detto nel capitolo precedente andando ad analizzare i dati del reddito disponibili per la Francia [9]. Il formato dei dati è visibile nella tabella 1. Se si divide la popolazioni in 100 percentili: $x_{0.01}$ esprime la frazione di ricchezza posseduta dalla popolazione con ricchezza compresa tra il 0 percentile ed il primo percentile, $x_{0.02}$ la frazione di ricchezza tra il primo percentile ed il secondo percentile e così via fino a $x_{0.99}$. Qui invece di passare al 100esimo percentile si passa al 99.1esimo percentile e quindi $x_{0.991}$ rappresenta la frazione di ricchezza posseduta dalla popolazione con ricchezza compresa tra il 99esimo percentile ed il 99.1esimo percentile. Si affinano sempre di più i percentili fino ad arrivare a $x_{0.99991}$, $x_{0.99992}$, ..., $x_{0.99993}$, x_1 ; quest'ultimo esprime la frazione di ricchezza posseduta dalla popolazione con ricchezza compresa tra il 99.999esimo percentile ed il 100esimo percentile. Si chiamerà la popolazione con ricchezza compresa tra due percentili "classe". Il formato dei dati, come si vedrà più avanti, permette di costruire con precisione una curva di Lorenz ma per ottenere una curva della distribuzione cumulativa di probabilità ci si vuole riferire non alla percentuale di ricchezza bensì alla ricchezza media \bar{x}_n posseduta dagli agenti della classe n :

$$\bar{x}_n = \frac{x_n \bar{x} N}{N_{classe}} = x_n \bar{x} N_{classe} \quad (13)$$

normalizzata alla media:

$$\bar{x}_n = x_n N_{classe} \quad (14)$$

dove con \bar{x} si intende la ricchezza media dell'anno di riferimento e con N_{classe} l'inverso della frazione di popolazione nella classe di riferimento, quindi 100 per le prime 99 classi, poi 1000 per le altre 9, 10000 e così via... Quindi, ad esempio, per l'anno 2012: $x_{0.99991} = 0.0004678$ e $\bar{x}_{0.99991} = 0.0004678 * 100000 = 46.78$. A questo punto, in riferimento a $P_{>}(x)$, si hanno i dati dell'asse delle ordinate 0.01, 0.02, ... ,

0.99, 0.991, etc. ma mancano i dati dell'asse delle ascisse corrispondenti. Infatti ad esempio $\bar{x}_{0.99991}$ si riferisce alla ricchezza media posseduta dagli agenti con ricchezza compresa tra il 99.99 percentile ed il 99.991 percentile che sarà minore del 99.991 percentile e maggiore del 99.99 percentile. Si assume quindi che, ad esempio, $\bar{x}_{0.99991}$ sia uguale al 99.9905 percentile e si correggono allo stesso modo i dati sulle ordinate per tutte le altre ricchezze. E' facile ora osservare la legge di Pareto in scala logaritmica (figura 2 (A)), infatti $\log_{10} P_{>}(x) = \log_{10} b - \mu \log_{10} x$.

Anno	1950	2013
Reddito Medio [€]	9433.04	34398.7
$x_{0.01}$	0.000467122	0.0000190735
$x_{0.02}$	0.00107986	0.00023073
...
$x_{0.99}$	0.0404241	0.0370219
$x_{0.991}$	0.0048805	0.0043884
...
$x_{0.999}$	0.0121228	0.0099745
$x_{0.9991}$	0.0014992	0.0012151
...
$x_{0.9999}$	0.0035435	0.0030863
$x_{0.99991}$	0.0004292	0.0003871
...
x_1	0.0019292	0.0026922

Tabella 1: Dati disponibili per la Francia dal 1950 al 2013

Come si è visto i dati sulla cumulativa sono soggetti ad errore a causa del formato dei dati. Per verificare la legge di Pareto si utilizzeranno quindi i dati sulla curva di Lorenz dell'equazione (12) con $X = n$ e con $Y = \sum_{i \leq n} x_i$. Si può riscrivere la funzione (12) come:

$$\log(1 - L(X)) = \log\left(\frac{b^{\frac{1}{\mu}}}{1 - \frac{1}{\mu}}\right) + \left(1 - \frac{1}{\mu}\right) \log(1 - X) \quad (15)$$

e quindi trovare i parametri μ e b (figura 3) tramite un'interpolazione lineare (figura 2 (b)) calcolata sugli ultimi 28 dati (1% della popolazione più ricca).

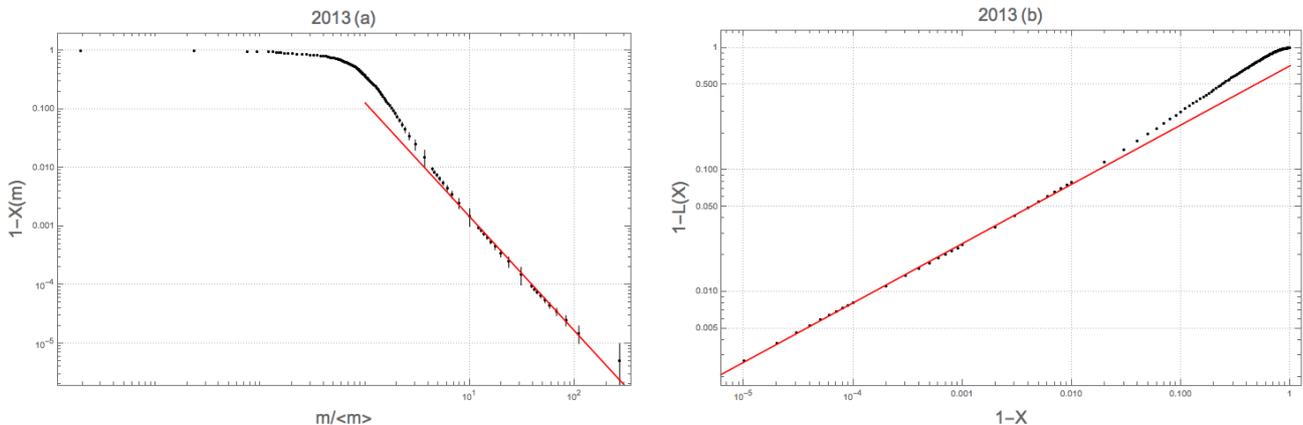


Figura 2: (a) Funzione cumulativa dei dati con fit della retta ricavata dalla distribuzione di Pareto con i parametri della curva in figura 2 (b). (b) Rappresentazione della curva di Lorenz tramite l'equazione (15) per l'anno 2013 con parametri della retta $b = 0.13$ e $\mu = 1.95$. L'indice di correlazione lineare risulta essere 0.99 e rimane lo stesso per tutti gli anni.

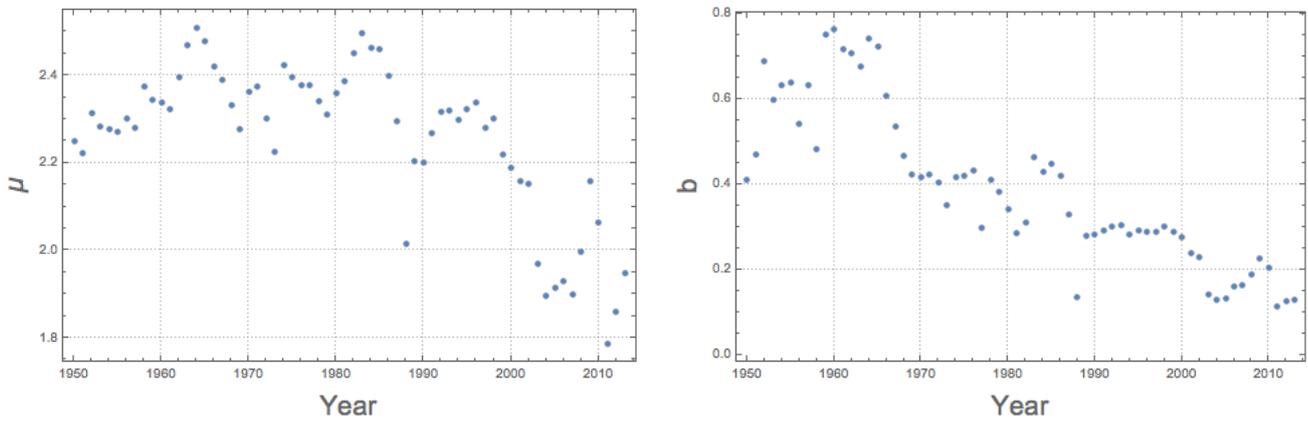


Figura 3: Andamento nel tempo degli indici μ e b calcolati dai parametri dell'interpolazione della retta (15).

3.2 Analisi della disuguaglianza sociale

Andiamo per prima cosa a costruire sui dati della ricchezza posseduta dalla popolazione francese le curve di Lorenz (figura 4) e a calcolare su quest'ultime l'indice di Gini e di Pietra (figura 5). Si può osservare come la ricchezza di ogni classe vari in funzione dell'indice di Gini. Occorre però fare questa analisi in due periodi: dal 1950 al 1984 e dal 1985 al 2013. Infatti le prime classi, subiscono un comportamento anomalo tra il 1984 ed il 1985: il valore della loro ricchezza media diminuisce in modo netto (figura 10). Si può vedere analizzando le curve di Lorenz in figura 4 come dopo il 1984 la ricchezza posseduta dal 50% della popolazione più povera (i.e. la ricchezza delle classi medie) vada ad aumentare (ovvero la curva di Lorenz in quella zona si avvicina alla curva di uguaglianza), contestualmente però è diminuita la ricchezza posseduta dal 5-10% di popolazione più povera ed è aumentata quella posseduta dal 1-2% di popolazione più ricca (la curva di Lorenz agli estremi si è allontanata dalla curva di uguaglianza). I cambiamenti che avvengono quindi nelle distribuzioni di probabilità (figura 9) della ricchezza dopo l'anno 1984 non vengono avvertiti dall'indice di Gini che misura, si ricorda, l'area sottesa alla curva di Lorenz. Calcoliamo quindi per i due periodi di riferimento il coefficiente di correlazione r di Pearson tra la ricchezza media x_n di ogni classe e gli indici di disuguaglianza di Gini, Pietra ed il k-index (figura 6). Gli indici di disuguaglianza sono calcolati dalla curva di Lorenz la cui forma viene approssimata unendo i punti dati dai percentili disponibili da una retta (nel caso dell'indice di Gini, ciò equivale a calcolarlo utilizzando la regola del trapezio con un errore dell'ordine di una parte su 10 000). Come si vede dalla figura 6 fino all'anno 1984 ogni indice di disuguaglianza rappresenta molto bene l'evoluzione delle ricchezze medie dei percentili, infatti quest'ultime oltre un certo percentile sono fortemente correlate e sotto un altro percentile risultano fortemente correlate in modo inverso agli indici di disuguaglianza.

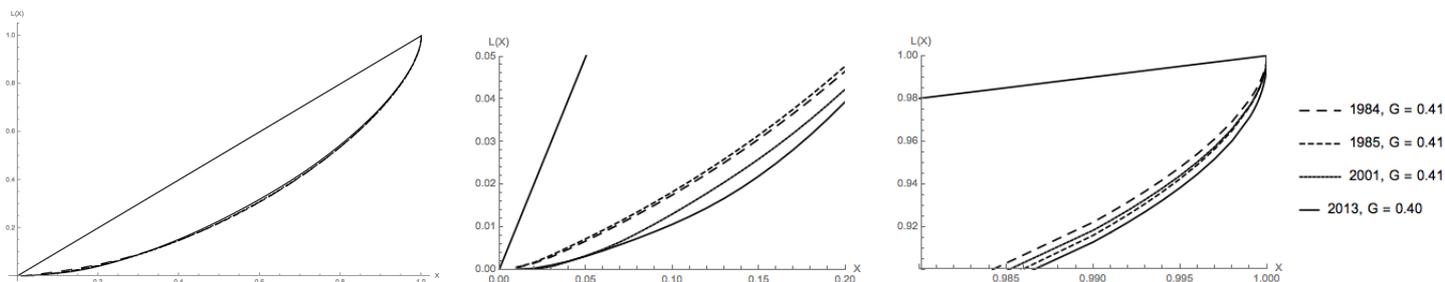


Figura 4: La prima figura mostra le curve di Lorenz per gli anni indicati destra dell'immagine. Le ultime due figure mostrano rispettivamente la parte iniziale e finale delle curve di Lorenz. Si può notare come, nonostante l'indice di Gini sia rimasto invariato tra gli anni 1984 e 2013 e quindi apparentemente non sia cambiata la disuguaglianza sociale, la curva di Lorenz si è allontanata dalla curva di uguaglianza agli estremi per avvicinarsi al centro, ovvero la curva di Lorenz ha cambiato la forma ma ha mantenuto la stessa area sottesa alla curva di uguaglianza.

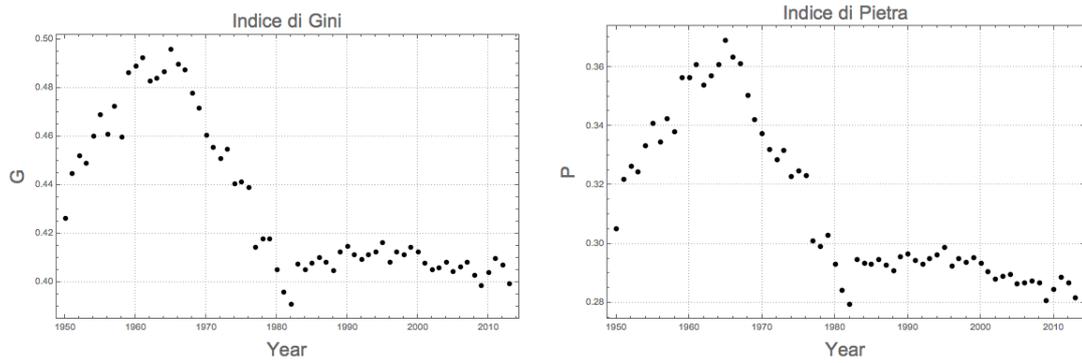


Figura 5: In figura si vede l'indice di Gini e quello di Pietra seguono uno stesso pattern, ripreso anche da altri parametri definiti per misurare la disuguaglianza.

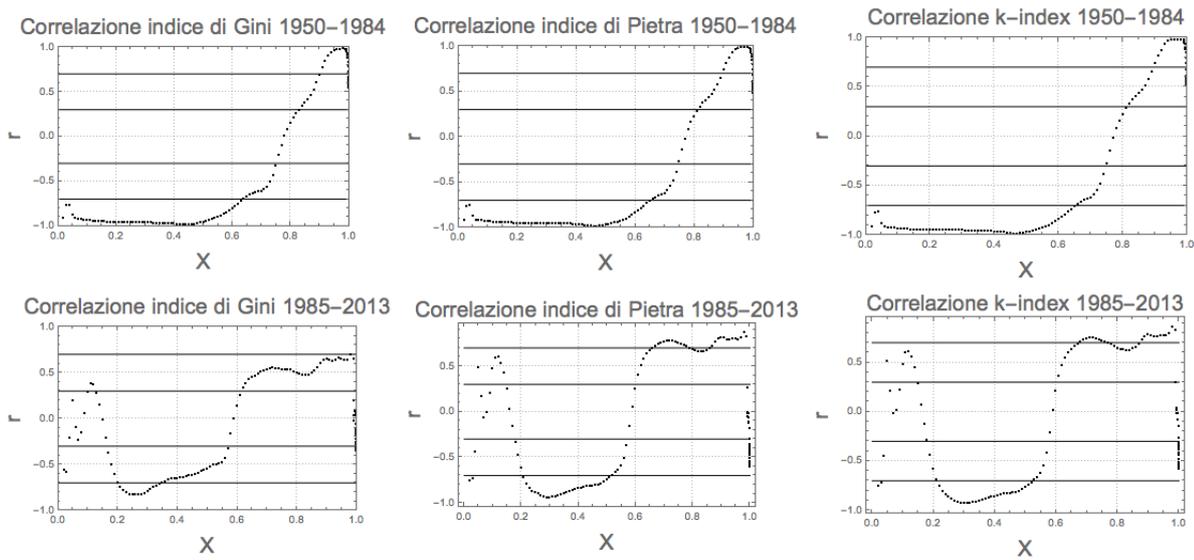


Figura 6: I grafici sono divisi in 5 zone, le x che hanno indice di correlazione nella prima ed ultima zona sono strettamente correlate (nel senso stretto o in modo inverso in basso) con gli indici di disuguaglianza, nella quarta e seconda sono moderatamente correlate e nella terza debolmente correlate.

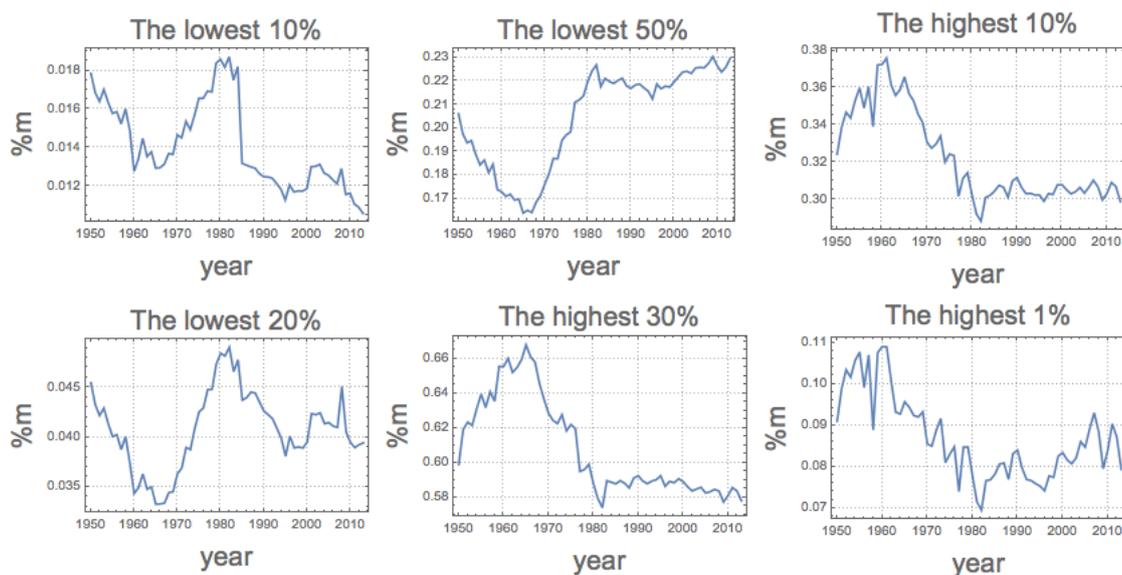


Figura 7: I grafici mostrano la percentuale di ricchezza posseduta dalla popolazione con una ricchezza minore del 10, 20 e 50 percentile e con una ricchezza maggiore del 70, 90 e 99 percentile.

Dal 1985 l'indice di Gini non sembra più descrivere il comportamento della ricchezza delle classi, solo una piccola parte delle classi più povere è fortemente correlata alla disuguaglianza sociale poi si ha solo correlazione moderata o debole da parte delle altre classi. La situazione in questo caso migliora utilizzando il k-index e l'indice di Pietra. Questo comportamento è in parte giustificato dal fatto che gli indici di disuguaglianza rimangono più o meno costanti dal 1984 in poi e trovare le loro correlazioni con la ricchezza delle classi di popolazione non è facile. Tuttavia l'indice di Pietra ed il k-index riescono ad esprimere per una buona parte di popolazione questa correlazione e confermano una debole correlazione nelle classi più povere. A questo scopo si mostra la percentuale di ricchezza posseduta dalla popolazione da un dato percentile in poi o fino ad un dato percentile prendendo come riferimento i percentili 10, 20, 50, 70, 90 e 99 (figura 7). I grafici della figura 7 potrebbero concettualmente essere interpretati come indici di disuguaglianza, infatti dipendono anch'essi dalla distanza della curva di Lorenz dalla curva di uguaglianza, ma questa distanza viene computata per un dato valore dell'asse X (ovvero 0.1, 0.2, 0.5, 0.7, 0.9 e 0.99). Per avere un indice coerente con l'indice di Gini si deve sottrarre ogni indice alla coordinata X di riferimento e poi moltiplicare per 2. Osservando quindi la figura 7 si può affermare che la disuguaglianza, dopo l'anno 1984, è salita se guardiamo i primi due grafici ma è diminuita se ci riferiamo al terzo grafico. Si può invece notare una risalita nell'ultimo grafico (percentuale di ricchezza posseduta dall'1% di popolazione più ricca); ciò conferma le osservazioni fatte sulle curva di Lorenz e le usuali affermazioni che la disuguaglianza sociale stia aumentando [13]. Se si fa una media per ogni percentile si ottiene l'indice di Gini che dal 1984 rimane più o meno costante.

Il problema di usare misure puntuali della curva di Lorenz (percentuali di ricchezza detenute da date percentuali di popolazione) come indice di disuguaglianza è che la loro definizione non è "transfer sensitive" [14], ovvero, non sempre il trasferimento di ricchezza da un agente più povero ad uno più ricco produce un aumento di disuguaglianza e viceversa. Infatti se si prende come indice di disuguaglianza la percentuale di ricchezza posseduta dal 10% di popolazione più povera ogni scambio di ricchezza tra agenti con ricchezza maggiore del 10 percentile farà restare l'indice invariato. Inoltre la definizione di questi indici è arbitraria (il percentile di riferimento viene scelto in modo arbitrario). Nonostante ciò tutti questi indici seguono fino all'anno 1984 lo stesso pattern dell'indice di Gini, mentre dopo l'anno 1984 qualche indice suggerisce un aumento della disuguaglianza sociale, qualcun altro una diminuzione. Presentiamo ora un indice che sia sensibile alla ricchezza delle classi più povere, la cui definizione non includa alcun parametro arbitrario e che soddisfi il principio di "transfer sensitive". Seguendo la definizione degli indici di Pietra e di Gini sulle curve di probabilità, quindi:

$$g = \frac{1}{\langle m \rangle} E(\max(0, m_1 - m_2))$$

$$P = \frac{1}{\langle m \rangle} E(\max(0, m - \langle m \rangle))$$

si definisce l'indice d come:

$$d = E\left(\frac{\max(m_1, m_2)}{\min(m_1, m_2)}\right) = \int_0^\infty \int_0^\infty \frac{\max(m_1, m_2)}{\min(m_1, m_2)} P(m_1)P(m_2) dm_1 dm_2 \quad (16)$$

L'indice d esprime in media, presi casualmente due agenti, il rapporto tra la ricchezza dell'agente più ricco e la ricchezza dell'agente più povero. A differenza dell'indice di Gini è definito per variabili di ricchezza in $(0, \infty)$. Tornando ai dati sulla Francia, come si vede in figura 8 (A), il rapporto tra la ricchezza di due agenti presi casualmente dalla popolazione è compreso tra 3 e 5 fino all'anno 1984, dopo di che passa a 25 l'anno successivo per poi scendere nel tempo. Ciò è dovuto alla ricchezza delle prime classi che influenza in modo molto significativo il valore di d .

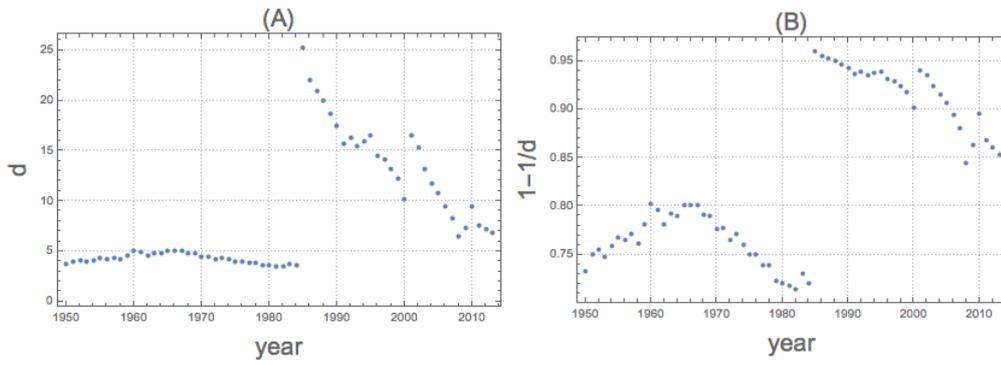


Figura 8: (A) Indice d calcolato per la popolazione francese dagli istogrammi della figura 9. (B) La grandezza $1 - \frac{1}{d}$ è compresa tra 0 ed 1 e segue lo stesso pattern di d .

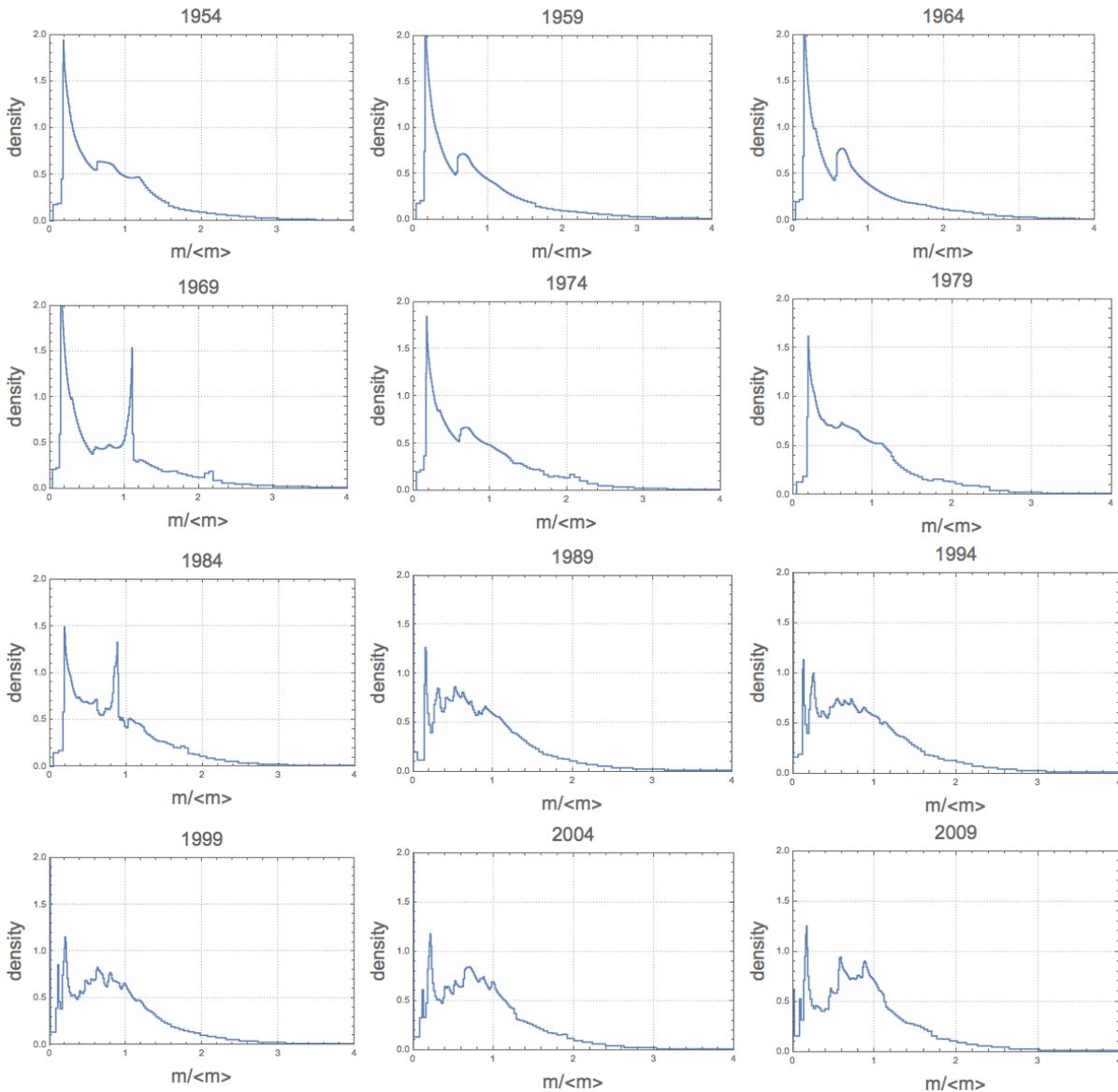


Figura 9: Gli istogrammi sono stati costruiti usando rettangoli di spessore $\bar{x}_{n_i} - \bar{x}_{n_{i-1}}$ ovvero la differenza delle ricchezze medie di due classi successive e altezza $h = \frac{n_i - n_{i-1}}{\bar{x}_{n_i} - \bar{x}_{n_{i-1}}}$. In questo modo si ottiene uno pseudo istogramma con classi di frequenza di diversa ampiezza ma che possiede tutte le caratteristiche di un'usuale distribuzione di probabilità e quindi con un'approssimazione accettabile si possono calcolare tutte le grandezze riguardanti quest'ultima. Come si vede dopo l'anno 1984 il pattern della distribuzione di probabilità cambia radicalmente assumendo una forma molto frastagliata.

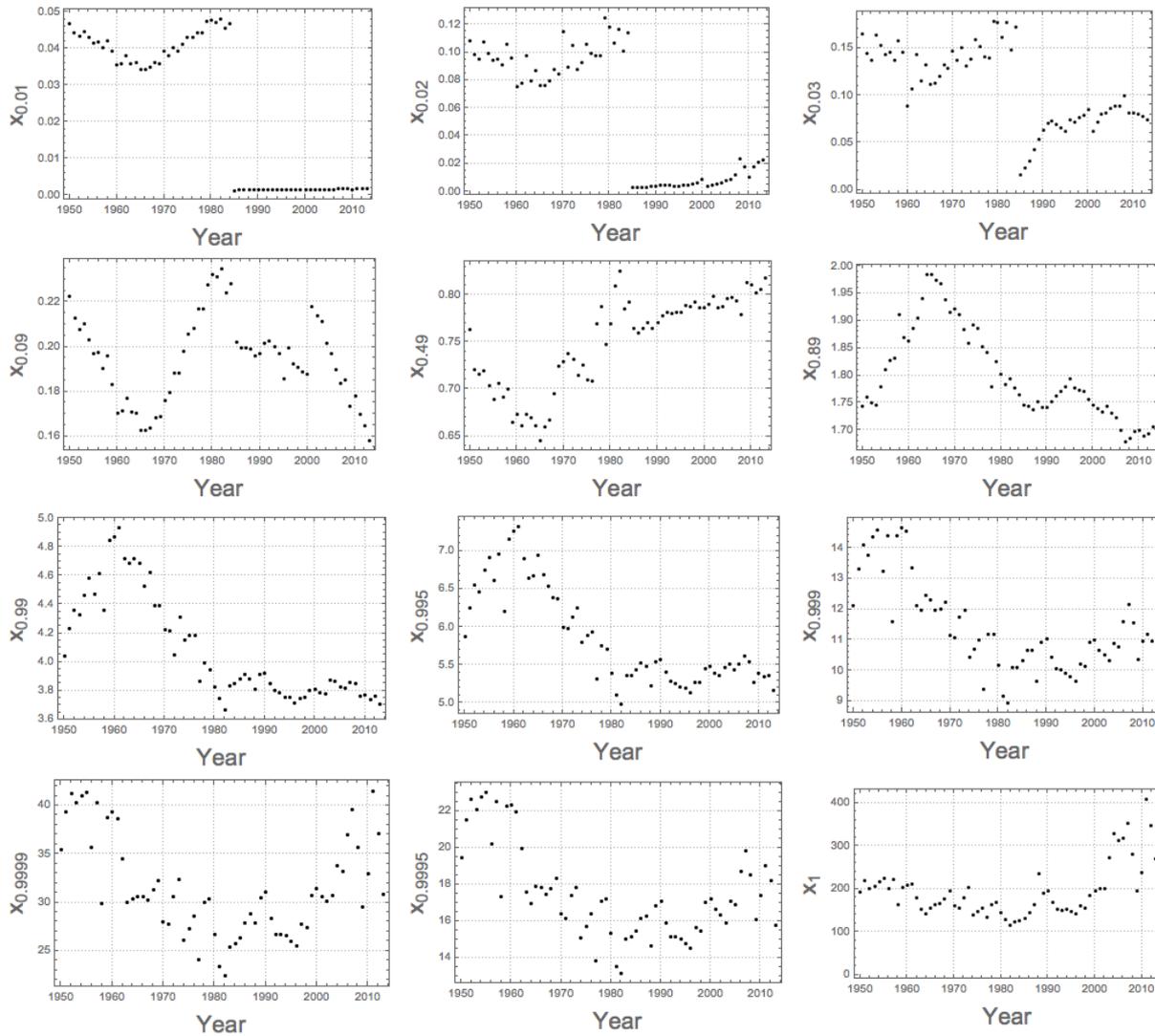


Figura 10: Come si può vedere, mentre fino all'anno 1984, tutte le ricchezze medie di ogni percentile seguono lo stesso pattern, dal 1985 il valore dei primi percentili diminuisce in modo brusco. Il pattern di disuguaglianza è molto chiaro nei percentili di mezzo, meno chiaro nei primi, completamente incorrelato negli ultimi

3.3 Power-Lax index

Si è finora osservato come, dal 1984, i classici indici di disuguaglianza (Gini, Pietra e Kolkata) non seguono più i cambiamenti socio-economici della società francese che avvengono principalmente tra classi più povere e più ricche di popolazione. L'indice d (normalizzato a $1 - \frac{1}{d}$), da noi introdotto, mostra questi cambiamenti. Tuttavia, il cambiamento di ricchezza di un singolo individuo può far variare in modo molto significativo l'indice. Sia, infatti, m_p la ricchezza posseduta dall'agente più povero della popolazione, allora:

$$\lim_{m_p \rightarrow 0} 1 - \frac{1}{d} = 1$$

Si può utilizzare un differente approccio attraverso l'utilizzo del power-law index [4] ϵ per evidenziare i cambiamenti che avvengono nelle classi più povere e più ricche della popolazione. Si trova (Pareto) che anche per le classi di ricchezza più povere vale un'analogia legge di potenza per la distribuzione di ricchezza, tale che:

$$L(x) = x^{\frac{\alpha+1}{\alpha}} \quad (17)$$

con $\alpha > 0$. Per la stima del parametro α non è più possibile usare il metodo dell'interpolazione lineare come fatto in 4.1. Infatti mentre, in quel caso, si avevano altri dati dopo il 99 percentile ora si dispone

solo dei dati del 1 percentile e 2 percentile. Osservando quindi che

$$\log L(x) = \frac{\alpha + 1}{\alpha} \log(x)$$

si stima α dalla pendenza dei primi due percentili, ovvero:

$$\frac{\alpha + 1}{\alpha} = \frac{\log L(x_{0.02}) - \log L(x_{0.01})}{\log(x_{0.02}) - \log(x_{0.01})}$$

Si può notare una correlazione ($r = 0.77$) tra α ed il parametro μ calcolato in 4.1 (Figura 11 (B)).

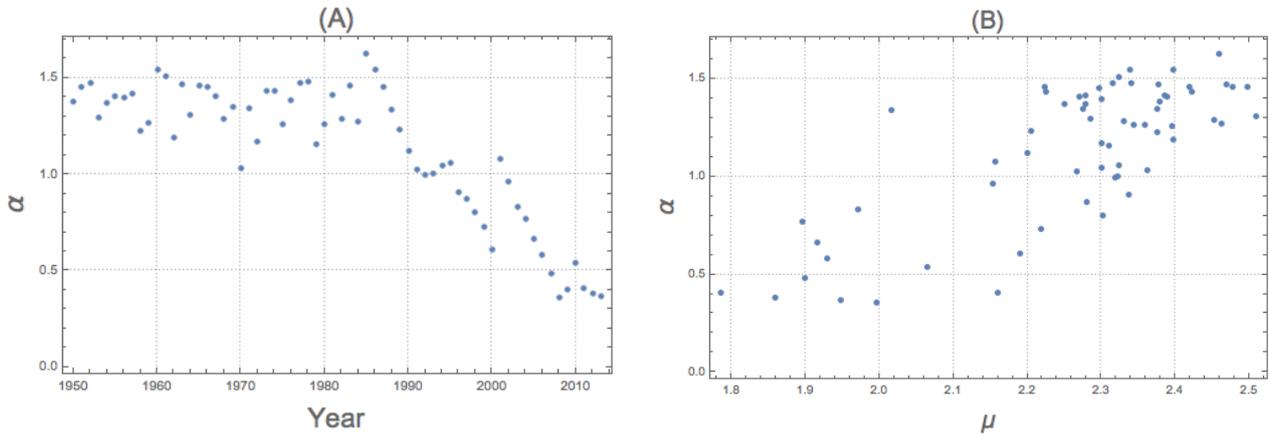


Figura 11: (a) Parametro α dell'equazione 17 in funzione del tempo. Si può vedere una decrescita dopo l'anno 1984. (b) α in funzione β

Si definisce ora la "disparity line" $D(X)$ e la "ratio line" $R(X)$ (figura 12). Sia $L_{>}(X) = 1 - L(1 - X)$ la curva di Lorenz "inversa", ovvero $L_{>}(X)$ sta a significare che la percentuale X di popolazione più ricca possiede la percentuale $L_{>}(X)$ di ricchezza, allora si definisce:

$$D(X) = L^{-1}(L_{>}(X))$$

$$R(X) = \frac{D(X)}{X}$$

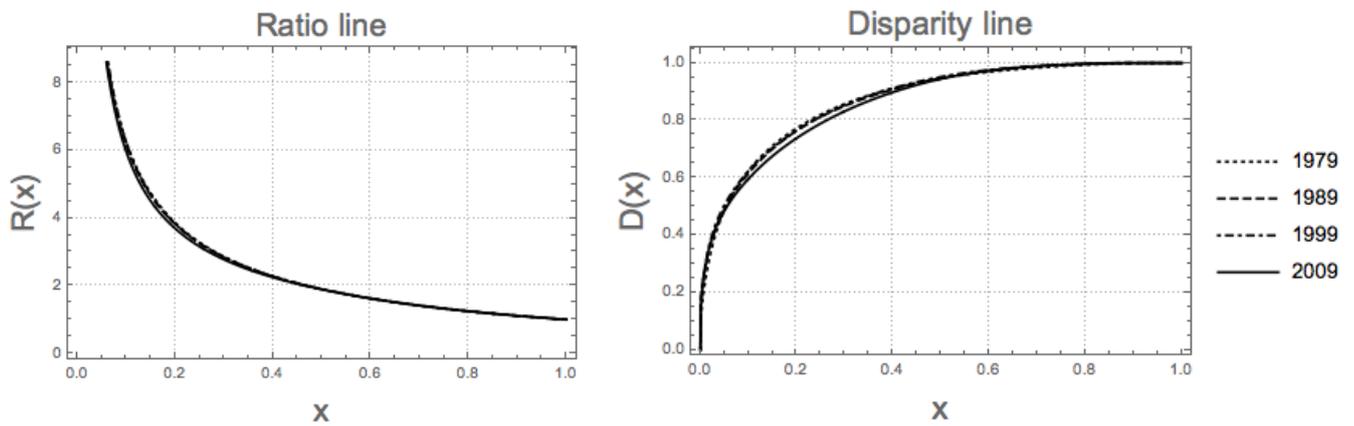


Figura 12: La linea di disparità $D(X)$ sta a significa che la ricchezza posseduta dalla frazione X di popolazione più ricca è la stessa posseduta dalla frazione $D(X)$ di popolazione più povera. Come per la curva di Lorenz la massima uguaglianza si ha per $D(X) = X$. La ratio line $R(X)$ sta a significare la popolazione più povera che possiede la medesima ricchezza della percentuale di popolazione più ricca X è $R(X)$ volte la popolazione X .

Si può facilmente notare [4] come se si estende a tutta la popolazione la distribuzione di Pareto, con parametro di potenza α o μ quest'ultimi due sono decrescenti al crescere dell'indice di Gini ovvero

sono anch'essi misure di disuguaglianza. Si può ora definire il power-index:

$$\epsilon(\alpha, \mu) = \frac{\alpha + \mu}{\mu(\alpha + 1)} \quad (18)$$

Notando che

$$\lim_{X \rightarrow 0} R(X) = X^{-\epsilon(\alpha, \mu)} \quad (19)$$

dalle condizioni $\alpha > 0$ e $\mu > 1$ si ha che $0 < \epsilon < 1$ senza alcuna normalizzazione. Inoltre ϵ è crescente per disuguaglianza. Si può vedere in figura 13 come anche quest'indice evidenzia una crescita di disuguaglianza sociale dopo l'anno 1984. Inoltre l'indice ci mostra come fino all'anno 1984 sostanzialmente i cambiamenti socio-economici non riguardavano le classi di popolazione più ricche e più povere. E' dopo l'anno 1984 che, mentre non ci sono cambiamenti sostanziali nelle distribuzioni di ricchezza delle classi medie (figura 9), si vanno ad impoverire le classi più povere ed arricchire le classi più ricche.

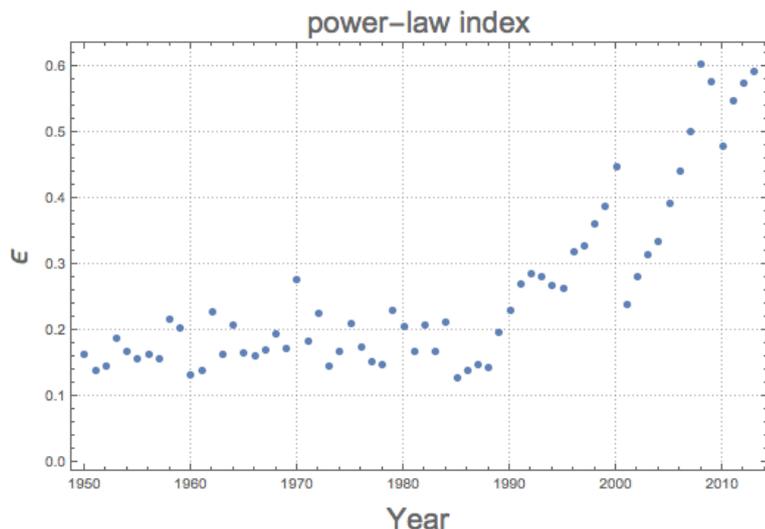


Figura 13: La correlazione r di ϵ con α risulta essere $r = -0.98$, con μ , $r = -0.73$.

4 Conclusioni

Dopo aver visto vari indici e misure di disuguaglianza si sono applicate queste metodologie alla popolazione francese dall'anno 1950 all'anno 2013 e si è andata a vedere la legge di Pareto. I dati di Piketty disponibili per la Francia sono in formato "percentili". Si sono quindi riadattati (calcolando le distribuzioni di probabilità che sono molto più utilizzate in fisica) per poter fare calcoli su di essi. In particolare si è visto come sono particolarmente adatti per eseguire calcoli direttamente sulla curva di Lorenz. Si è notato come fino all'anno 1984 ogni classe di ricchezza era correlata alla disuguaglianza economica. E' dopo l'anno 1984 che si verifica un comportamento anomalo delle distribuzione di probabilità tale per cui i cambiamenti socio-economici riguardavano solo le classi più ricche e più povere di popolazione e gli usuali indici di disuguaglianza non rispondevano a questi cambiamenti. In sostanza la curva di Lorenz cambiava forma ma rimaneva distante della stessa quantità dalla curva di uguaglianza. Abbiamo quindi introdotto un nuovo indice d , che non dipende da alcun parametro arbitrario e mostra i cambiamenti socio-economici che avvengono dopo l'anno 1984. Tuttavia quest'indice risulta essere molto instabile. Un altro indice che risulta tenere conto dei cambiamenti socio-economici successivi all'anno 1984 è il Power Law Index che mentre rimane più o meno stabile fino all'anno 1984 comincia a crescere da quell'anno in poi. Dati analoghi a quelli recuperati per la Francia sono stati insistentemente cercati, anche tramite contatti con la Banca d'Italia e l'Istat, ma non è stato possibile recuperare un campione abbastanza ampio temporalmente.

Riferimenti bibliografici

- [1] Asim Ghosh, Arnab Chatterjee, Jun ichi Inoue, and Bikas K. Chakrabarti. Inequality measures in kinetic exchange models of wealth distributions. *Physica A: Statistical Mechanics and its Applications*, 451:465 – 474, 2016.
- [2] A. Dragulescu and V.M. Yakovenko. Statistical mechanics of money. *The European Physical Journal B*, 17(4):723–729, Oct 2000.
- [3] A. Christian Silva and Victor M Yakovenko. Temporal evolution of the thermal and superthermal income classes in the USA during 1983-2001. *Europhysics Letters (EPL)*, 69(2):304–310, Jan 2005.
- [4] Iddo Eliazar and Morrel H. Cohen. On social inequality: Analyzing the rich–poor disparity. *Physica A: Statistical Mechanics and its Applications*, 401:148 – 158, 2014.
- [5] Iddo I. Eliazar and Igor M. Sokolov. Measuring statistical heterogeneity: The pietra index. *Physica A: Statistical Mechanics and its Applications*, 389(1):117 – 125, 2010.
- [6] Iddo Eliazar. The sociogeometry of inequality: Part i. *Physica A: Statistical Mechanics and its Applications*, 426:93 – 115, 2015.
- [7] Jun ichi Inoue, Asim Ghosh, Arnab Chatterjee, and Bikas K. Chakrabarti. Measuring social inequality with quantitative methodology: Analytical estimates and empirical data analysis by gini and k indices. *Physica A: Statistical Mechanics and its Applications*, 429:184 – 204, 2015.
- [8] Arnab Chatterjee, Asim Ghosh, Jun ichi Inoue, and Bikas K Chakrabarti. Social inequality: from data to statistical physics modeling. *Journal of Physics: Conference Series*, 638:012014, Sep 2015.
- [9] T. Piketty and S. Arecco. *Allegato tecnico al libro Il capitale nel XXI secolo*. Bompiani, 2016.
- [10] H.L. Moore. Cours d'Économie politique. by vilfredo pareto, professeur à l'université de lausanne. vol. i. pp. 430. 1896. vol. ii. pp. 426. 1897. lausanne: F. rouge. *The ANNALS of the American Academy of Political and Social Science*, 9(3):128–131, 1897.
- [11] M. Lubrano. The econometrics of inequality and poverty lecture 4: Lorenz curves, the gini coefficient and parametric distributions. 2014.
- [12] Yoshi Fujiwara. *Pareto-Zipf, Gibrat's Laws, Detailed-Balance and their Breakdown*, pages 24–33. Springer Milan, Milano, 2005.
- [13] Facundo Alvaredo, Thomas Piketty, Emmanuel Saez, Lucas Chancel, and Gabriel Zucman. World Inequality Report 2018. Technical report, 2018.
- [14] Anthony F. Shorrocks and James E. Foster. Transfer Sensitive Inequality Measures. *The Review of Economic Studies*, 54(3):485–497, 07 1987.