Disuguglianza e povertà - Stato dell'arte

Enrico Fornasiero

Aprile 2019

Abstract

Lo scopo di questo elaborato, concepito come preambolo ad un lavoro più approfondito di ricerca sul tema, è quello di mettere in luce e sintetizzare lo sviluppo della ricerca sulla disuguaglianza e sulla povertà e sulla costruzione degli indici per la loro analisi empirica. Prima di iniziare una qualsivoglia analisi è necessario identificare i confini della nostra ricerca, andando a differenziare i concetti di disuguaglianza (in una distribuzione e di conseguenza nella distribuzione delle variabili che andremo a prendere in considerazione) e di povertà. Riferendoci ad una popolazione, il termine disuguaglianza racchiude in sé sia una accezione di tipo economica che di tipo morale. La comprensione del livello di disuguaglianza (o uguaglianza) socialmente accettato in una popolazione trascende da ciò che tratteremo in questo breve elaborato. La discussione non verterà quindi su concetti di tipo morale, anche se ne faremo brevi richiami, ma principalmente sull'evoluzione teorica ed empirica legata alla misurazione del suddetto valore all'interno di una popolazione.

La trattazione degli indici di disuguaglianza verrà svolta in diverse fasi, in una prima parte ci si concentrerà nella definizione dei principi cardine da rispettare nella scelta delle funzioni in grado di rappresentare la distribuzione della variabile campione, nella seconda fase si andrà quindi a delimitare le famiglie di funzioni in grado di esprimere indici di disuguaglianza robusti, mentre nell'ultima fase l'attenzione sarà rivolta alla loro precisione e a tecniche alternative per la misurazione della disuguaglianza.

Per quanto concerne il secondo argomento dell'elaborato, la povertà, ci concentreremo sulla identificazione del segmento della popolazione che si trova al di sotto di una specifica soglia chiamata "poverty line" o "poverty threshold". La trattazione di questo tema segue pressoché gli stessi passaggi di quella precedente; in un primo momento si andranno a definire i principi fondamentali su cui si basa la teoria sottostante, per poi proseguire nel riconoscimento degli indici utilizzati in letteratura concludendo infine con una analisi qualitativa degli stessi.

Dopo avere definito in maniera generale la struttura delle funzioni di disuguaglianza e povertà, ci concentreremo sulle prime specificando alcuni esempi empirici ed alcune scritture in R utilizzabili in una prima analisi sotto determinate circostanze. L' analisi con il software R è facilitata inoltre dalla presenza di pacchetti installabili sulla piattaforma. In questa sezione comprenderemo come l'utilizzo di questi pacchetti non sia sempre possibile a causa della specificità dei dati reperibili dai siti istituzionali; il più delle volte infatti sarà possibile solamente collezionare dati sui redditi suddivisi per gruppi e non singolarmente.

Nella parte finale dell'elaborato andremo inoltre a mostrare alcuni utilizzi di questi indici nella letteratura e ne indagheremo i risultati. L'orizzonte accademico su temi di disuguaglianza e povertà si è via via ampliato negli anni passando da indagini prettamente statistiche, grazie allo sviluppo di indici sempre più precisi, a ricerche di tipo empirico che vanno a mettere in correlazione questi valori a variabili che esulano dalla teoria standard come le aspirazioni individuali o la qualità delle istituzioni pubbliche. Nella parte conclusiva andremo quindi a proporre alcuni di questi risultati augurandoci di dare nuovi spunti di ricerca a chi leggerà questo elaborato.

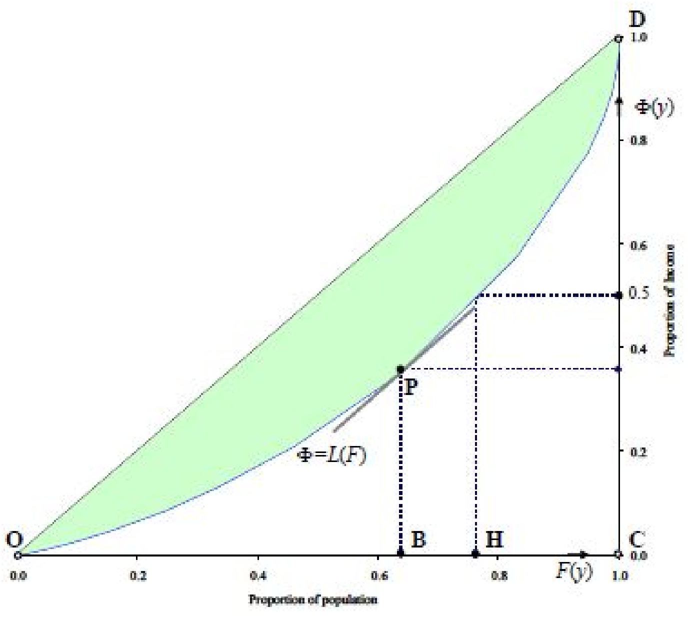
1 Disuguaglianza

L'analisi del livello di disuguaglianza, presente in una determinata popolazione, viene svolta prendendo in considerazione una variabile rappresentativa del fenomeno, una variabile che vada a racchiudere sia elementi di tipo quantitativo che qualitativo. Allo scopo vengono quindi utilizzate, come variabili esplicative, i valori del reddito y_i o dell'utilità, individuale U, sociale W. Prima di iniziare una ricerca su questo argomento è bene chiedersi da chi è rappresentata la singola unità all'interno della popolazione campione; la scelta, oltre che per motivi di ricerca, dipenderà dalla reperibilità del tipo di dato che si vuole prendere in considerazione. Per questo motivo delle volte può essere preferibile fare riferimento al nucleo familiare come unità singola, piuttosto che al singolo individuo. È d'uopo ricordare che nel passaggio dal nucleo familiare ai singoli componenti, qualora fossimo interessati al comportamento individuale partendo da dati raggruppati per unità familiare, va considerato che la variabile reddito non è i.i.d. (Schulter e Trede, 2002 [20]). In questo caso è necessario quindi assumere la possibilità del sorgere di una forte contemporanea dipendenza tra i ricettori di reddito, basti pensare ad esempio alle scelte riguardo l'offerta di lavoro dei genitori all'interno di una famiglia.

Introduciamo ora due caratteristiche molto importanti per la trattazione degli indici di disuguaglianza: l'ordine e la cardinalità. Due indici si dicono ordinariamente equivalenti quando classificano diverse preferenze (nel nostro caso valori di reddito data una distribuzione) nello stesso ordine; due indici si dicono cardinalmente equivalenti, quando oltre a classificare le preferenze nello stesso ordine, mantengono la stessa distanza in proporzione tra essi. La seconda proprietà è quindi molto utile quando si andrà ad analizzare cambiamenti per una data distribuzione a distanza di anni, infatti in questo modo avremo una causalità diretta tra i cambiamenti avvenuti nella distribuzione e quelli nella disuguaglianza senza dovere assumere alcuna unità di misura.

L'implementazione di indici per la misurazione della disuguaglianza ha permesso lo sviluppo di diverse teorie sul tema, alcune prendono piede a partire dalla definizione della Social Welfare Function, altre utilizzano concetti e teoremi presenti in altre discipline (in primis Information Theory e probabilità) ed altre ancora preferiscono costruire i loro indici attraverso un approccio strutturale andando quindi prima a definire le caratteristiche desiderate e da queste, a ritroso, gli indici stessi. Prima di proseguire nella trattazione è importante sottolineare come nell'analisi di disuguaglianza, ma anche di povertà, sia necessario in primis ordinare gli individui in ordine crescente dal soggetto con reddito minore a quello con reddito maggiore; questo procedimento ci permette infatti di ragionare in termini sia di quantili e di rapporto tra i diversi quantili della distribuzione, che di proporzioni.

La rappresentazione grafica maggiormente utilizzata è il diagramma di Lorenz e la Lorenz curve.



In questa rappresentazione grafica troviamo nell'asse verticale la distribuzione cumulata del reddito presente nella popolazione, mentre nell'asse orizzontale la distribuzione cumulata della popolazione stessa. La linea inclinata negativamente che congiunge il valore $F(y) = 0$ e $\Phi(y) = 1$ viene definita come il luogo geometrico dei punti del piano in cui si ha una distribuzione del reddito non diseguale rispetto alla popolazione. La curva di Lorenz misura quindi, per ogni porzione della popolazione, la quantità di reddito cumulata; nella figura precedente notiamo come più del 60 per cento della popolazione (punto B) sia in possesso di meno del 40 per cento della ricchezza totale, indicando quindi la presenza di disuguaglianza nella popolazione stessa. In presenza di più curve di Lorenz diremo quindi che la curva più vicina al segmento $\overline{OD}$ rappresenta la popolazione con una maggiore uguaglianza nella distribuzione della ricchezza, la stessa curva di Lorenz sarà quindi contenuta all'interno dell'area tra la linea di perfetta uguaglianza e la curva con disuguaglianza maggiore.

Tuttavia non tutti gli indici vengono rappresentati attraverso questo grafico, vedremo come una buona parte di essi abbiano rappresentazioni grafiche differenti o non ne abbiano affatto.

Identifichiamo ora alcune proprietà desiderabili e successivamente controlliamo se i più noti ed utilizzati indici le rispettano.

Weak Principle of Transfer

Ipotizziamo di avere due distribuzioni (A e B) in un grafico di Lorenz, supponiamo inoltre che la distribuzione A possa essere raggiunta attraverso una redistribuzione della ricchezza a partire da B in modo che la ricchezza totale rimanga inalterata e la curva di Lorenz di A sia compresa tra B e la linea di perfetta uguaglianza. Fintanto che la misura di disuguaglianza rispetta il Weak Principle of Transfer, questa misura indicherà un minor livello di disuguaglianza in A rispetto che in B. Dato quindi un trasferimento di reddito da un individuo più ricco ad uno più povero, la misura di disuguaglianza diminuirà (in quale misura dipenderà dall'indice utilizzato).

Income Scale Independence

Qualora l'intera popolazione, in ogni sua singola unità, vedesse aumentata la sua ricchezza di uno stesso valore percentuale, l'indice di diseguaglianza deve rimanere invariato.

Principle of Population

Qualora la popolazione in considerazione con n unità, venisse sommata a una identica popolazione in termini di unità e valori di disuguaglianza; il valore della misura della stessa non deve variare.

Decomposability

Questa proprietà si può sintetizzare asserendo che l'indice di disuguaglianza deve rappresentare sia la disuguaglianza della popolazione nella sua interezza che all'interno dei diversi gruppi che la costituiscono (suddivisi per caratteristiche che possono essere di tipo qualitativo o quantitativo). L'obiettivo che ci poniamo è quindi di creare un indice per il quale il valore della disuguaglianza possa essere rappresentato come una qualche media del valore della disuguaglianza presente all'interno dei gruppi che costituiscono la popolazione stessa.

Questa caratteristica viene ampiamente utilizzata nel caso in cui, non conoscendo l'esatta distribuzioni dei redditi all'interno della popolazione, si è costretti a compiere una analisi avendo a disposizione solamente le diverse classi di reddito $(a_1, a_2), (a_2, a_3), ..., (a_k, a_{k+1})$, la media di ciascuna classe μ_i e la corrispondente numerosità n_i.
Grazie a questa proprietà è possibile ricorrere da singole misurazioni di diseguaglianza fatte per singole classi, ad una misura complessiva dell'intera popolazione.

1.1 Misure di Disuguaglianza

In questa sezione andremo ad analizzare alcune misure, non veri e propri indici di diseguaglianza, non utilizzabili per analisi complesse ma che possono indicarci se siamo in presenza o meno di diseguaglianza.

La prima misura è il **Range**, inteso semplicemente come la distanza nel reddito tra l'individuo con il reddito minore e quello con il reddito maggiore:

$$R = y_{max} - y_{min}$$

questa misura non è ovviamente soddisfacente poiché non considera la distribuzione sottostante ai valori estremi.

Un'altra misura è la **Relative Mean Deviation** definita come la distanza assoluta media tra il reddito di ogni singolo individuo e la media della popolazione; introducendo nel calcolo il diagramma di Lorenz possiamo scrivere:

$$M = 2[F(\overline{y}) - \Phi(\overline{y})]$$

Altre misure che si possono derivare dalla curva di Lorenz sono la **Minimal Majority Inequality**, che rappresenta la proporzione della popolazione necessaria per raggiungere la metà della ricchezza cumulata, e l'**Equal Shares Coefficient** che rappresenta la proporzione della popolazione con reddito pari o inferiore al reddito medio.

Un indice molto utilizzato in letteratura che sfrutta le proprietà della curva di Lorenz è il famoso **Coefficiente di Gini**

$$G = \frac{1}{2n^2\overline{y}} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|$$

Sebbene largamente utilizzato in letteratura, l'indice presenta un comportamento particolare riguardo i trasferimenti di reddito tra individui. In sintesi il coefficiente sovrastima i trasferimenti di reddito che avvengono nei pressi del reddito medio nei confronti di trasferimenti di uguale entità tra persone distanti nella distribuzione. Nonostante questo aspetto, questo indice si presta ottimamente per le analisi di disuguaglianza in quanto, come vedremo, rispetta tutte le proprietà elencate in precedenza a differenza di molti altri indici che tratteremo.

Vale la pena citare, come ultimi indici riferiti alla curva di Lorenz, l' **Indice di Amato** (1968) e l'**Indice di Pietra**; (Barry C. Arnold, 2012 [4]). L'indice di Amato consta nel calcolo della lunghezza della curva di Lorenz, è quindi un indice (come altri) che è basato su una proprietà geometrica della curva stessa; mentre l'indice di Pietra corrisponde alla massima distanza verticale tra la curva di Lorenz e la linea di uguaglianza.

Altre misure possono essere derivate da formule puramente statistiche che permettono di analizzare la dispersione di una variabile. Un esempio è la **Varianza**:

$$V = \frac{1}{n} \sum_{i=1}^{n} [y_i - \overline{y}]^2$$

è facilmente riscontrabile come questo indice non rispetti uno dei principi elencati in precedenza, ovvero la "Scale Independence". Osserviamo infatti come, duplicando tutti i redditi degli individui presenti in una ipotetica popolazione, il valore della varianza quadruplicherebbe andando in contrasto con il suddetto principio. Una soluzione proposta dalla letteratura è quella di standardizzare il valore e definire il **Coefficiente di Variazione**:

$$c = \frac{\sqrt{V}}{\overline{y}}$$

Un altro modo per ovviare al problema è applicare una trasformazione logaritmica alla varianza stessa ottenendo la **Varianza Logaritmica**:

$$v = \frac{1}{n} \sum_{i=1}^{n} \left[\log\left(\frac{y_i}{\overline{y}}\right)\right]^2$$

1.2 Indici basati sulla funzione di utilità sociale

Trattandosi di una analisi compiuta su una popolazione, alcuni autori hanno proposta di derivare nuovi indici di disuguaglianza a partire da una generica funzione di utilità sociale. La funzione di utilità permette di ordinare i possibili stati della società seguendo le preferenze dei suoi componenti, o di una figura esterna come ad esempio un dittatore, attraverso delle variabili quali ad esempio il reddito.

La funzione di utilità sociale è *individualistica* e *non decrescente*; nel nostro caso se il reddito in uno stato A è maggiore che in uno stato B per ciascun individuo i, diremo che la funzione corrispondente allo stato A è preferita a quella rappresentativa dello stato B. La funzione è inoltre *simmetrica* ovvero tratta i redditi degli individui in maniera anonima ed il valore W non dipende da indicatori individuali $(w_1, w_2, ...)$. La funzione di utilità sociale è inoltre *additiva, strettamente concava* e con *elasticità costante.*

$$W(y_1, y_2, \ldots, y_n) = \sum_{i=1}^{n} U(y_i) = U(y_1) + U(y_2) + \ldots + U(y_n)$$

$$U(y_i) = \frac{y_i^{1-\varepsilon} - 1}{1-\varepsilon}$$

Queste ultime tre proprietà definiscono alcune interessanti proprietà per l'analisi della variazione della disuguaglianza. L'additività permette ad esempio di asserire che trasferimenti di reddito tra individui diversi, sono indipendenti dalla loro posizione nella distribuzione della ricchezza della popolazione. L'elasticità costante definisce che decrementi del peso di U' per un dato aumento proporzionale nel reddito devono essere costanti a qualsiasi livello di reddito. Se il reddito aumenta dell'1% (con reddito iniziale di 100, 1000, 10000 etc) ciò si traduce in una diminuzione del peso sociale pari a ε%.

Applichiamo ora queste proprietà al calcolo di indici di disuguaglianza. Chiamiamo $U(\overline{y})$ il valore dell'utilità individuale quando la ricchezza nazionale risulta essere equamente distribuita nella popolazione, e y_e il reddito che, se ricevuto da ogni individuo, farebbe risultare il valore totale di benessere nella popolazione uguale al valore attuale; necessariamente $y_e < \overline{y}$ in quanto possiamo sempre non redistribuire l'intera ricchezza nazionale, ma solo una parte di essa, e comunque ottenere lo stesso livello di disuguaglianza.

Definiamo quindi l'**Indice di Dalton** con avversione alla disuguaglianza ε:

$$D_\varepsilon = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}\left[y_i^{1-\varepsilon} - 1\right]}{\overline{y}^{1-\varepsilon} - 1}$$

o

$$D_\varepsilon = 1 - \frac{\overline{U}}{U(\overline{y})}$$

l'indice è 0 quando la ricchezza è perfettamente distribuita $U(\overline{y}) = \overline{U}$.

Atkinson (1970) [5] criticò questo indice asserendo che aggiungendo valori costanti diversi da 0 l'indice variava; propose quindi una variazione che prese il nome di **Indice di Atkinson**:

$$A_\varepsilon = 1 - \frac{U^{-1}(\overline{U})}{\overline{y}}$$

$$A_\varepsilon = 1 - \left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\overline{y}}\right]^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}$$

$$A_\varepsilon = 1 - \frac{y_e}{\overline{y}}$$

1.3 Indici basati su Information Theory

In teoria dell'informazione viene applicata la teoria della probabilità per analizzare il peso delle informazioni collegate alla realizzazione o meno di un determinato evento. Supponiamo di avere tre possibili scenari, ognuno numerato in ordine crescente con abbinate tre diverse probabilità p, $p \neq 0, p < 1$. Osserviamo che l'evento numero 1 è avvenuto, quale valore diamo all'informazione ricevuta $h(p_1)$? Se la realizzazione del fenomeno 1 era altamente probabile, il valore assegnato a questa informazione sarà prossimo allo zero in quanto non aggiunge nulla di nuovo a quello che già sapevamo; viceversa, se la probabilità dell'evento numero 1 era molto bassa, il valore assegnato a $h(p_1)$ dovrà essere molto alto. Definiamo quindi $h(p_i)$ come inversamente proporzionale a p_i. Allo scopo di creare una corrispondenza tra la teoria dell'informazione e le tematiche di disuguaglianza definiamo gli eventi in considerazione tra loro indipendenti in modo che

$$h(p_1 p_2) = h(p_1) + h(p_2)$$

L'unica funzione che rispetta questi vincoli è

$$h = -log(p)$$

In presenza di n stati, con relative probabilità, definiamo una unica funzione in grado di aggregare questi stati in modo che sia possibile descrivere il "grado di disordine" complessivo del sistema. Il numero in questione risulterà basso quando la probabilità dell'evento i è pari a 1 e zero per tutti gli altri eventi (sistema ordinato, informazione non è preziosa), crescente man mano che la probabilità totale viene distribuita diversamente a tutti gli stati possibili. Definiamo quindi il concetto di entropia:

$$\begin{aligned} \text{entropy } &= \sum_{i=1}^{n} p_i h\left(p_i\right) \\ &= -\sum_{i=1}^{n} p_i \log\left(p_i\right) \end{aligned}$$

.

Theil (1967) [23] applicò questa teoria apparentemente estranea allo studio della disuguaglianza, al calcolo degli indici della stessa andando a sostituire agli n possibili eventi le n unità che compongono una popolazione e trasformando p_i in s_i, ovvero nella percentuale di reddito dell'individuo i rispetto al reddito totale.

$$s_i = \frac{y_i}{n\bar{y}}$$

L'**Indice di Theil** viene quindi costruito come differenza tra il valore massimo di entropia presente nella popolazione (massima disuguaglianza) ed il valore attuale di entropia nella distribuzione del reddito:

$$\begin{aligned} T &= \sum_{i=1}^{n} \frac{1}{n} h\left(\frac{1}{n}\right) - \sum_{i=1}^{n} s_i h\left(s_i\right) \\ &= \sum_{i=1}^{n} s_i \left[h\left(\frac{1}{n}\right) - h\left(s_i\right)\right] \\ &= \sum_{i=1}^{n} s_i \left[\log\left(s_i\right) - \log\left(\frac{1}{n}\right)\right] \\ &= \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\bar{y}} \log\left(\frac{y_i}{\bar{y}}\right) \end{aligned}$$

L'indice derivato porta in dote una particolare proprietà: immaginiamo che una persona povera riceva un trasferimento da una persona più ricca per cui s_1 diventa $s_1 + \Delta s$ e s_2, $s_2 - \Delta s$. La variazione dell'indice sarà dunque negativa pari a:

$$\begin{aligned} \Delta T &= \Delta s \left[h\left(s_2\right) - h\left(s_1\right)\right] \\ &= -\Delta s \log\left(\frac{s_2}{s_1}\right) \end{aligned}$$

ed il valore dipenderà esclusivamente dal rapporto tra i due diversi s, e di conseguenza esclusivamente dalla loro distanza nella distribuzione del reddito. Trasferimenti tra individui con redditi simili avranno meno impatto sull'indice T mentre trasferimenti tra individui "lontani" avranno una magnitudine maggiore sulla misura.

La funzione $-log(s)$ appartiene a una famiglia di funzioni più ampia riassumibile nella formula

$$h(s) = \frac{1 - s^\beta}{\beta}$$

dalla quale è possibile estrapolare un indice generico utilizzabile come punto di partenza per la costruzione di nuovi indici:

$$\frac{1}{\beta + \beta^2} \sum_{i=1}^{n} s_i \left[s_i^\beta - n^{-\beta}\right]$$

il valore β è libero di variare, anche se solitamente vengono presi in considerazione valori standard, e rappresenta il peso che viene dato al concetto di distanza tra due valori all'interno della distribuzione.

Per $\beta = -1$ otteniamo ad esempio

$$-\sum_{i=1}^{n} \log(ns_i)$$

che non è altro che n volte la **Deviazione Logaritmica Media**:

$$L = \frac{1}{n}\sum_{i=1}^{n} [\log(1/n) - \log(s_i)]$$

Per $\beta = 0$ otteniamo T, mentre per $\beta = 1$ l'**Indice di Herfindahl**:

$$H = \sum_{i=1}^{n} s_i^2$$

A completare questa argomento richiamiamo i principi precedentemente trattati e definiamo il seguente teorema.

Teorema: Qualsiasi misura di disuguaglianza che soddisfa simultaneamente il principio di Weak Principle of Transfer, Decomposability, Scale Independence and Population Principle deve essere espressa secondo

$$E_\theta = \frac{1}{\theta^2 - \theta}\left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\overline{y}}\right]^\theta - 1\right]$$

oppure con $J(E_\Theta)$, trasformazione di E_Θ.

1.4 Indici e Principi

Dopo avere identificato una larga parte di indici solitamente utilizzati in letteratura cerchiamo ora di capire quali di essi rispettano i principi identificati all'inizio di questa sezione; la violazione di questi principi non ci deve fare rifiutare a priori l'indice ma semplicemente ci deve rendere consapevoli di alcune imprecisioni che potrebbero emergere se utilizziamo un indice non adatto alle nostre ricerche.

Prendiamo in considerazione il *Weak Principle of Transfer*, in questo caso tutti gli indici presentati riflettono un comportamento coerente con quello identificato. Le uniche misure che non rispettano questo principio sono v (la varianza logaritmica), R ed M poiché, in particolar modo per gli ultimi due, è possibile che la curva riferita ad A sia parzialmente contenuta tra B e la linea di uguaglianza e parzialmente no.

Per quanto riguarda invece l'*Independence* dell'indice all'incremento proporzionato di egual misura di tutti i redditi individuali, la varianza e l'indice di Dalton violano questa ipotesi (anche se per l'indice di Dalton ciò dipende anche dalle ipotesi sulla cardinalità della funzione U usata).

Il principio di *Decomposability* illustra un problema che va a colpire il coefficiente di Gini, osserviamo questo esempio

	East	
	A:(60,70,80)	
	B:(60,60,90)	
	A	B
$\bar{y}$	70.00	70.00
G	0.063	0.095
A_1	0.007	0.019
A_2	0.014	0.036
T	0.007	0.020

	West	
	A:(30,30,130)	
	B:(10,60,120)	
	A	B
$\bar{y}$	63.33	63.33
G	0.351	0.386
A_1	0.228	0.343
A_2	0.363	0.621
T	0.256	0.290

	East-West combined	
	A:(60,70,80,30,30,130)	
	B:(60,60,90,10,60,120)	
	A	B
$\bar{y}$	66.67	66.67
G	0.275	0.267
A_1	0.125	0.198
A_2	0.236	0.469
T	0.126	0.149

in cui due diversi programmi A e B vengono proposti a due diversi sottogruppi di una popolazione. L'unione dei due sottogruppi in un unico gruppo mantiene inalterati l'ordine dei valori della disuguaglianza per gli indici di Atkinson (con diversa avversione al rischio) indicando maggiore uguaglianza con il programma A, mentre inverte il risultato per quando riguarda il coefficiente di Gini che, separatamente, indica maggiore disuguaglianza in B, ma considerando un unica popolazione indica A come il programma più disuguale. Questo fenomeno sembra avvenire quando si è in presenza di campionamenti relativamente piccoli, per questo motivo diversi autori, come Barret e Donald (2009) [6], hanno proposto variazioni all'indice per renderlo più efficiente e preciso nella misurazione della disuguaglianza, suggerendo inoltre tecniche di inferenza migliori in caso di campionamenti ridotti (Monte Carlo e metodi bootstrap). In presenza di popolazioni più numerose questo inconveniente tende a scomparire, anzi l'indice di Gini risulta essere facilmente utilizzabile in caso di popolazioni di cui conosciamo solamente le classi di reddito (sottogruppi della popolazione), la media di ciascuna classe e la loro frequenza relativa.

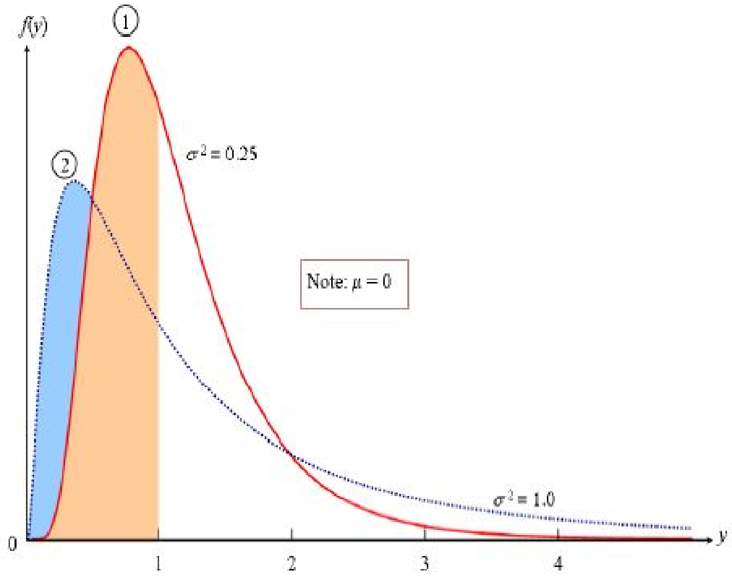
1.5 Distribuzioni

Dopo avere identificato gli indici utilizzabili per la nostra analisi dobbiamo porre l'attenzione sulla distribuzione della variabile in considerazione (nel nostro caso il reddito). Noteremo infatti come il comportamento di determinate funzioni vada ad approssimare la reale distribuzione del reddito in una popolazione, aiutandoci quindi a cogliere le differenze esistenti per i diversi valori di y.

La funzione che dovrà essere presa in considerazione nel nostro caso sarà del tipo:

$$f = \phi(y; a, b)$$

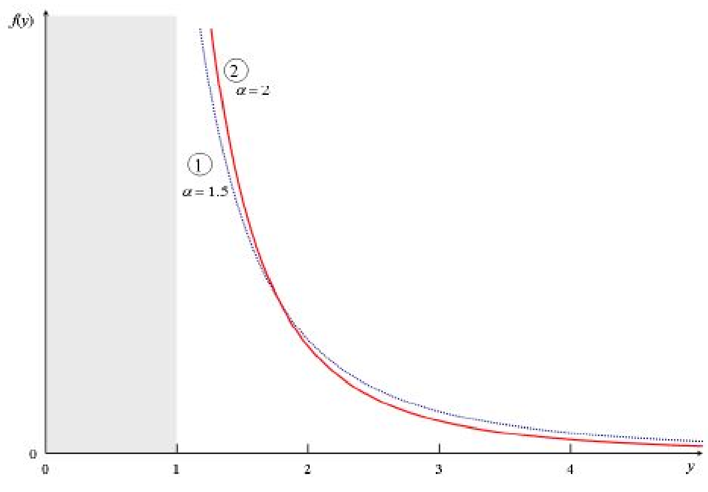
L'approssimazione dipenderà quindi dai valori a, b della funzione; non sempre si riuscirà ad associare una funzione alla distribuzione del reddito del campione, il più delle volte sarà però possibile individuare in certi gruppi di reddito un comportamento molto simile a funzioni note. Osservazioni empiriche mostrano come la distribuzione del reddito in una popolazione sia concentrata a sinistra del valore mediano della stessa, il grafico che solitamente si ottiene è quindi un grafico associabile a una distribuzione log-normal



Notiamo innanzitutto come la distribuzione non tenga conto dei valori di $y < 0$. Questo avviene per facilitare il calcolo in quanto per valori negativi di reddito si rende necessario aggiustare gli indici trovati in precedenza in quanto, a titolo esemplificativo, con redditi negativi la curva di Lorenz non è contenuta solamente tra gli assi ma anche al di sotto dell'asse orizzontale andando ad inficiare il calcolo delle aree necessario per la costituzione dell'indice di Gini. Una distribuzione di questo genere incorpora quindi diverse proprietà utili alla nostra ricerca. Innanzitutto essendo una distribuzione derivata da una distribuzione normale, su di essa possono essere compiuti test statistici sul logaritmo del reddito assumendo la condizione di normalità; inoltre, la distribuzione lognormal assicura che le curve di Lorenz siano simmetriche rispetto alla bisettrice del quadrante con origine in O. Se nella nostra ricerca risulta una curva di Lorenz simmetrica, o pressoché simmetrica, possiamo fare assunzioni su una distribuzione del reddito di questo tipo. Questa distribuzione assicura inoltre che le curve di Lorenz non si intersechino mai, è quindi immediato riconoscere quale curva assicuri maggiore uguaglianza e quale maggiore disuguaglianza.

Come per la normale, anche in questo caso la rappresentazione della curva dipende dai parametri μ e σ^2, dove μ rappresenta la media del logaritmo di y e σ^2 la sua varianza.

Anche se intuitiva, la distribuzione lognormal non sempre riesce ad approssimare il comportamento della variabile y nella coda destra della distribuzione. Per questo motivo, trattando valori di reddito superiori al valore medio, si introduce la distribuzione di Pareto la cui funzione di densità è:

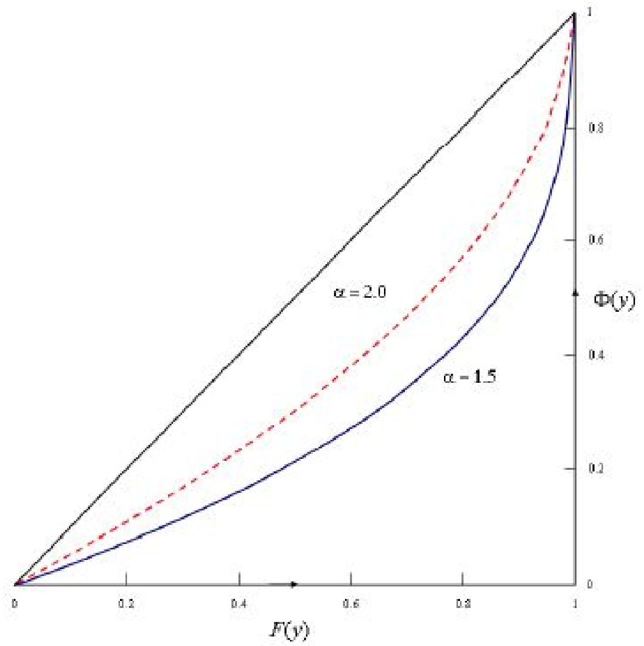


$$f(y) = \frac{\alpha y^{\alpha}}{y^{\alpha+1}}$$

La distribuzione di Pareto riesce a stimare abbastanza bene la distribuzione dei redditi medio-alti, mentre è totalmente inadatta a descrivere il range di redditi bassi-medi. Una proprietà molto utile deriva dalla legge di Van der Wijk's per la quale una volta definito un reddito "base" y, il valore del reddito medio del sottogruppo, che ha un valore di reddito maggiore o uguale alla base, è By con

$$B = \frac{\alpha}{\alpha - 1}$$

Notiamo come all'aumentare di α il valore diminuisca indicando una diminuzione della distanza tra i redditi.



Come nel caso della distribuzione lognormal, anche quella di Pareto assicura curve di Lorenz non secanti. Curve di Lorenz associate con valori di α elevati sono inoltre più vicine alla curva di perfetta uguaglianza.

Non sempre queste forme funzionali riescono a descrivere in maniera accurata la distribuzione del reddito in una popolazione, il più delle volte riescono ad avvicinarsi alla distribuzione di particolari gruppi della popolazione. Nella prossima sezione si andrà quindi ad introdurre un metodo più preciso per l'approssimazione della distribuzione considerando sia la "forma" dei dati a disposizione (il più delle volte raggruppati per categoria di reddito), sia analizzando nel dettaglio la distribuzione all'interno di queste categorie.

1.6 Dati e distribuzioni

La ricerca di dati per lo svolgimento di ricerche sul tema può avvenire principalmente in due modi: attraverso la somministrazione di questionari o ricavando i dati di interesse da dichiarazioni obbligatorie individuali o familiari definite da criteri di legge. In entrambi i casi ci si può trovare di fronte a una raccolta di dati non precisa, è dimostrato infatti come molti soggetti, in particolare i soggetti che si trovano agli estremi nella distribuzione, tendano a mentire riguardo il valore del loro reddito. Nel caso del questionario per non mettere alla mercé la loro condizione di reddito, nel caso della raccolta dalle dichiarazioni legali per motivi di evasione fiscale.

Una constatazione aggiuntiva va inoltre fatta in riferimento ai questionari, il più delle volte l'unità a cui è sottoposto il questionario (individuo o famiglia) non va a indicare il valore preciso del reddito, bensì indica in quale range/categoria si trova. Osserviamo quindi come è possibile ovviare a questo problema.

Definiamo un generico indice di disuguaglianza J come

$$J = \int_0^\infty h(y) f(y) dy$$

con $f(y)$ funzione di densità e $h(y)$ la funzione dell'indice (ad esempio Theil $h(y) = \frac{y}{\bar{y}} \log \left(\frac{y}{\bar{y}} \right)$)

Determiniamo classi arbitrarie di reddito come

$$(a_1, a_2)\,(a_2, a_3)\,(a_3, a_4) \ldots (a_k, a_{k+1})$$

e riscriviamo J come

$$\sum_{i=1}^{k} \left[\int_{\alpha_i}^{a_{i+1}} h(y) f(y) dy \right]$$

In questo modo otteniamo un indice che prende in considerazione le diverse classi ma andiamo ad inficiare la costruzione della distribuzione $F(y)$ in particolar modo non riusciamo ad identificare il tipo di distribuzione presente all'interno di ogni classe.

Una soluzione è quella di costruire il limite superiore e inferiore del valore di J per ogni sotto-classe.

$$J_L \leq J \leq J_U$$

J_L è costruito in modo tale che all'interno di ogni classe (a_i, a_{i+1}) sia presente perfetta uguaglianza (tutti ricevono lo stesso reddito medio).

$$J_{\mathrm{L}} = \sum_{i=1}^{k} \frac{n_i}{n} h(\mu_i)$$

con μ uguale alla media della classe di reddito di riferimento, ed n_i/n la frequenza relativa di popolazione all'interno della fascia di reddito.

J_U è invece costruito in modo che sia presente il valore massimo di disuguaglianza, utilizzando questo approccio assumiamo che un individuo all'interno della classe considerata possa avere reddito massimo a_{i+1} o minimo a_i.

$$\lambda_i = \frac{a_{i+1} - \mu_i}{a_{i+1} - a_i}$$

λ rappresenta lo scostamento dal valore medio che verrà utilizzato nella costruzione del valore massimo di disuguaglianza dell'indice di interesse.

$$J_{\mathrm{U}} = \sum_{i=1}^{k} \frac{n_i}{n} \left[\lambda_i h(a_i) + [1 - \lambda_i] h(a_{i+1}) \right]$$

Utilizzando questa strategia non assumiamo alcuna distribuzione $F(y)$ anche se ovviamente la costruzione di questi valori dipende dai valori a_i scelti. Abbiamo così ottenuto un range di valori in cui J si trova. Per quanto detto in riferimento al principio di **Decomposability** gli unici indici in grado di descrivere il valore di disuguaglianza totale come una qualche media del valore dei sottogruppi che compongono la popolazione sono l'indice di Gini e gli indici derivanti dal concetto di Entropia (per approfondimenti Shorrocks [22]).

Proviamo ad applicare questo concetto alla curva di Lorenz.

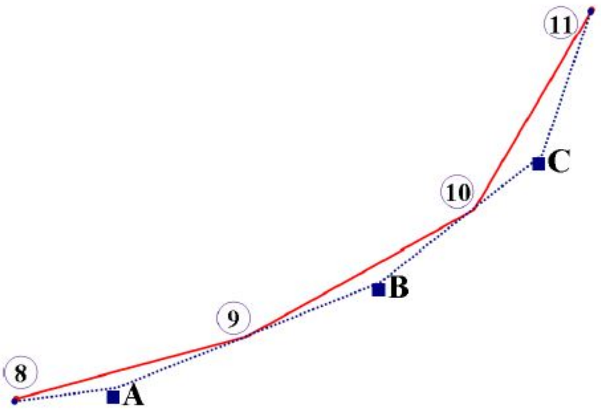


Figure 5.13: Upper and Lower Bound Lorenz Curves

Le linee continue che uniscono i punti nella figura, rappresentano il caso di perfetta uguaglianza; mentre la linea tratteggiata definisce il caso di massima disuguaglianza. Ci accorgiamo quindi che la curva di Lorenz per la popolazione in considerazione dovrà (i) passare per tutti i punti [8,9,10..], (ii) essere convessa e (iii) non passare al di sotto della linea di massima disuguaglianza.

Facendo riferimento a Gastwirth 1972 [12], possiamo esemplificare quanto detto riguardo il limite inferiore e superiore, analizzando l'indice di Gini. Introduciamo il concetto di "mean diffenìrence" Δ

$$\begin{aligned}\Delta &= \int_{-\infty}^{\infty}\int_{\infty}^{\infty}|x-y|dF(x)dF(y)\\ &= 2\int F(x)[1-F(x)]dx\\ &= 4\int x[F(x)-1/2]dF(x)\end{aligned}$$

come misura di dispersione in riferimento alla concentrazione di popolazione e reddito nella curva di Lorenz. Riscriviamo ora l'indice di Gini come rapporto tra questa dispersione ed il valore medio del reddito

$$G = \Delta/2\mu$$

Basandoci sulla formula appena trovata riflettiamo che per ogni gruppo il valore minimo di disuguaglianza si ottiene quando tutti gli individui appartenenti alla categoria di reddito in considerazione, registrano il medesimo valore di reddito; mentre il caso di massima disuguaglianza si ottiene quando gli individui si polarizzano ai due estremi della classe di reddito stessa.
Data quindi una serie ordinata di n numeri raggruppati in $k+1$ sottogruppi, allora il valore empirico della "mean difference" è pari a (Yntema 1933):

$$\begin{aligned}\Delta^* &= \binom{n}{2}^{-1}\sum_{i<j}\sum|x_i-x_j|\\ &= \sum_{i\neq j}\sum_{i}\gamma_i\gamma_j\,|\mu_i-\mu_j|\\ &+ \sum_{i=1}^{k+1}\gamma_i^2\Delta_i^*\end{aligned}$$

dove μ_i rappresenta il valore medio di ogni gruppo, Δ_i^* la mean difference e γ_i il numero di osservazioni presenti in ogni gruppo.
Riscriviamo quindi $G = \Delta^*/2\mu$. Il numeratore dell'indice ci permette di avere una rappresentazione della somma della mean difference tra i diversi gruppi e un termine di correzione che pesa la mean difference all'interno di ogni sottogruppo attraverso il fattore γ_i^2.

Quando tutte le osservazioni all'interno di un gruppo risultano essere identiche, il termine di correzione si riduce a 0 mentre assume un valore positivo quando ipotizziamo che le osservazioni siano polarizzate agli estremi della classe di reddito

$$\begin{aligned} GL = (2\mu)^{-1} \sum_{i \neq j} \sum \gamma_i \gamma_j |\mu_i - \mu_j| \\ \leqslant G \leq GL + \overline{D} = GU \end{aligned}$$

Attraverso l'utilizzo della mean difference siamo quindi in grado di trovare i confini superiori ed inferiori senza compiere assunzioni sulla funzione di densità sottostante con $\overline{D}$:

$$\begin{aligned} \overline{D} = & \mu^{-1} \sum_{i=1}^{k+1} \gamma_i^2 \left(\mu_i - a_{i-1}\right) \times \\ & \left(a_i - \mu_i\right) \left(a_i - a_{i-1}\right)^{-1} \end{aligned}$$

In Cowell [9] questo procedimento viene ulteriormente affinato nell'eventualità in cui il ricercatore è a conoscenza, se non dell'intera distribuzione del reddito, almeno dell'andamento della stessa (ovvero se crescente o decrescente all'interno di una classe di reddito). Consideriamo inizialmente il caso in cui non siamo a conoscenza dell'andamento della distribuzione e riscriviamo i fattori che permettono di identificare il minimo ed il massimo limite di un indice come:

$$\begin{aligned} f_1(y) = \tfrac{n_\theta}{n} \quad & \text{if } y = \mu_\theta, \quad \theta = 1, 2, \ldots, \omega \\ = 0, \quad & \text{otherwise} \end{aligned}$$

$$\begin{aligned} f_2(y) &= \lambda_\theta \frac{n_\theta}{n} && \text{if } y = a_\theta \\ &= \left[1 - \lambda_\theta\right] \frac{n_\theta}{n}, && \text{if } y = a_{\theta+1} \\ &= 0, && \text{otherwise} \end{aligned}$$

Se siamo in possesso dell'ulteriore informazione riguardo l'andamento della distribuzione all'interno di un intervallo (pensiamo alle code sinistre e destre), in particolare se la funzione è decrescente, in questo caso possiamo riscrivere $f_1(y), f_2(y)$ come:

$$\begin{aligned} f_3(y) \quad &= \tfrac{n_\theta}{2n[\mu_\theta - a_\theta]}, \quad && \text{if } a_\theta < y < 2\mu_\theta - a_\theta \\ &= 0, && \text{otherwise} \end{aligned}$$

$$\begin{aligned} f_4(y) &= \frac{n_\theta}{n} \frac{a_\theta + a_{\theta+1} - 2\mu_\theta}{a_{\theta+1} - a_\theta}, && \text{if } y = a_\theta \\ &= \frac{2n_\theta}{n} \frac{\mu_\theta - a_\theta}{\left[a_{\theta+1} - a_\theta\right]^2}, && \text{elsewhere} \end{aligned}$$

Se la funzione non è decrescente $f_3(y), f_4(y)$ risultano identiche a $f_1(y), f_2(y)$. Utilizzando queste 4 funzioni per scrivere un generico indice di disuguaglianza I possiamo definire che

$$I_1 \leqq I_3 \leqq \hat{I} \leqq I_4 \leqq I_2$$

andando a restringere ulteriormente la banda che contiene il vero valore $\hat{I}$.

Per quanto concerne invece la ricerca del limite inferiore e superiore di una misura di entropia possiamo fare riferimento ai lavori di Gastwirth 1975 [13], Cowell e Metha e altri [9], Mussard e altri [17].
Riscriviamo quindi una misura di entropia generica come

$$I^\beta = \frac{1}{\beta[\beta+1]} \left[\int_0^\infty \left[\frac{y}{\mu}\right]^{\beta+1} f(y) dy - 1 \right]$$

data questa espressione possiamo riscrivere l'indice sotto forma dell'unione della misura della disuguaglianza presente all'interno di ogni sottoclasse e del valore derivante dalla disuguaglianza presente tra le diverse sottoclassi, ovvero:

$$I^\beta = \sum_{\theta=1}^{\omega} \frac{n_\theta}{\mu} \left[\frac{\mu_\theta}{\mu}\right]^{\beta+1} I^{\beta\theta} + \frac{1}{\beta^2+\beta} \sum_{\theta=1}^{\omega} \left[\left[\frac{\mu_\theta}{\mu} - 1\right] \frac{n_\theta}{n}\right]$$

nella quale $I^{\beta\theta}$ rappresenta la disuguaglianza all'interno delle classi ed il secondo termine quella tra di esse. Il valore minimo si ottiene considerando nullo il valore *within* ed il limite superiore quando esso è positivo.

Di conseguenza possiamo identificare per l'indice di Theil ($\beta = 0$):

$$T_L = \sum_{\theta=1}^{\omega} \frac{n_\theta \mu_\theta}{n\mu} \log \frac{\mu_\theta}{\mu}$$

$$T_W = \sum_{\theta=1}^{\omega} \frac{n_\theta \mu_\theta}{n\mu} \frac{1}{n_\theta} \sum_{i=1}^{n\theta} \frac{y_{\theta i}}{\mu_\theta} \log \frac{y_{\theta i}}{\mu_\theta}$$

$$T_U = T_L + T_W$$

Similarmente è possibile calcolare altre funzioni per diversi valori di β prendendo in considerazione la funzione generatrice e andando di volta in volta a sostituire a β il valore desiderato.

Grazie a questi espedienti è quindi possibile individuare un range (generalmente piccolo come dimostreremo più avanti) in cui il vero valore dell'indice risiede, senza compiere assunzioni sulla distribuzione del reddito all'interno della popolazione. Sempre grazie al lavoro di Gastwirth 1975, si dimostra come un buon compromesso nell'individuazione del reale valore dell'indice -di Gini o altri-si possa ottenere semplicemente utilizzando i due valori estremi.
Per quanto concerne Gini, si dimostra come una buona stima sia rappresentata da

$$\hat{G} = 1/3G_L + 2/3G_U$$

Per le altre misure di disuguaglianza è necessario invece seguire la seguente regola

$$\hat{I} = 2/3I_L + 1/3I_U$$

Un metodo alternativo è quello di cercare di approssimare direttamente i dati alla forma funzionale di nostro interesse. Il procedimento che ne deriva è il seguente: per prima cosa bisogna individuare i parametri della forma funzionale, poi calcolare la misura di disuguaglianza da quei parametri.

Assumiamo che i dati raccolti y_i siano non raggruppati e provenienti da una distribuzione lognormal. Utilizziamo il metodo dei momenti per definire i parametri $\overline{\mu}, \overline{\sigma}^2$. Cerchiamo di identificare i valori per l'indice di Herfindahl:

$$H = \sum_{i=1}^{n} \left[\frac{y_i}{n\overline{y}}\right]^2$$

da cui deriviamo attraverso il metodo dei momenti

$$\widetilde{\sigma}^2 = \log(nH)$$
$$\widetilde{\mu} = \log(\overline{y}) - \tfrac{1}{2}\widetilde{\sigma}^2$$

Un metodo più efficiente è quello del calcolo dei maximum likelihood estimates $\tilde{\mu}, \tilde{\sigma}^2$. Trasformiamo le osservazioni y_i nei corrispettivi logaritmi x_i.

$$\hat{\mu} = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} x_i$$
$$\hat{\sigma}^2 = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} [x_i - \hat{\mu}]^2$$

Per i data raggruppati in classi il metodo da utilizzare è quello dei MLE.

2 Povertà

La discussione sulla povertà e sul calcolo dei suoi indici prende alcuni elementi fondamentali da ciò che è stato precedentemente detto per la disuguaglianza. I punti distintivi in questo frangente sono due: l'identificazione del valore della soglia di povertà, con la relativa conta del numero di persone al di sotto di essa, e la costruzione di indici con le informazioni a disposizione. Amartya Sen nel corso dei decenni ha contribuito fortemente all'analisi di questa tematica, prenderemo spunto dal suo lavoro (1973) [21] nella definizione di alcune caratteristiche fondamentali nella costruzione della soglia di povertà e degli indici. Uno dei metodi più largamente utilizzati nei primi decenni seguenti al secondo dopoguerra per la conta dei "poveri", era il cosiddetto "Headcount ratio", un semplice rapporto tra coloro con un reddito minore della soglia e la popolazione totale. Sen, muovendo una critica a questo metodo, identifica alcuni assiomi, che questo conteggio viola, che invece devono essere considerati nella trattazione del tema.

Monotonicity axiom: Ceteris paribus, una riduzione di reddito di un individuo al di sotto della soglia di povertà, deve far aumentare il valore della misura di povertà.

Transfer axiom: Ceteris paribus, un trasferimento da un agente povero a uno ricco deve fare aumentare il valore della misura di povertà.

Un'altra misura utilizzata è il "**poverty gap**" ovvero la misura aggregata della differenza di reddito degli individui poveri tra il loro valore di reddito e la soglia di povertà.
Definiamo il gap come $g_i = z - y_i$ con z reddito della soglia di povertà, con g_i positivo per individui poveri e negativo per gli altri.

$$Q(x) = A(z, \underline{y}) \sum_{i \in S(x)} g_i \nu_i(z, \underline{y})$$

$Q(x)$ rappresenta quindi il gap aggregato dell'intera popolazione dove A, v dipendono dalle assunzioni fatte, $\underline{y}$ una qualsiasi configurazione di y e $S(x)$ descrive l'insieme di persone con reddito non maggiore a x.
L'indice di povertà risulta essere quindi funzione di x per il quale

$$P = \max_x Q(x)$$

Il peso individuale v dovrà inoltre dipendere dal valore iniziale di reddito, minore reddito, maggiore peso del gap rispetto al valore z.

Axiom E (Relative equity): Per ciascuna coppia i, j: if $W_i(\underline{y} < W_j(\underline{y})$ allora $v_i(z, \underline{y}) > v_j(z, \underline{y})$. Che corrisponde all'utilità marginale se utilizziamo un approccio di tipo utlilitaristico.

Axiom R (Ordinal Rank Weights): Il peso $v_i(z, \underline{y})$ dato ad una persona i dipende dall'ordinamento del benessere sociale.

Più in basso è una persona nell'ordinamento del benessere sociale, maggiore dovrà essere il suo senso di povertà. È da notare come non stiamo parlando di livelli di reddito ma di benessere, in questo modo incorporiamo la possibilità (seppur remota) che una persona i con con $y_i < y_j$ possa avere un livello di benessere maggiore di j. Le difficoltà nel considerare nella ricerca questa ipotesi, hanno spinto l'autore e la letteratura ad evitare il più delle volte un analisi di questo genere; tutto questo si sintetizza nel prossimo assioma.

Axiom M (Monotonic Welfare): Per ogni i, j: se $y_i > y_j$, allora $W_i(\underline{y}) > W_j(\underline{y})$.

2.1 Indici

Come accennato in precedenza l'headcount ratio è stato largamente utilizzato agli albori ma presenta diverse imperfezioni:

$$H = \frac{q}{n}$$

dove q numero di persone al di sotto della soglia di povertà.

Un'altra misura è il poverty gap che non considera il numero di persone povere ma può essere standardizzato; definiamo quindi l'income gap ratio come

$$I = \sum_{i \in S(z)} g_i / qz$$

indicatore della differenza media del gap tra redditi bassi e la soglia di povertà.

Sen propone quindi un indice in grado di rispettare gli assiomi identificandolo in:

$$P = \frac{1}{(q+1)nz} \left[zq(q+1) + q^2 m \left(G - \frac{q+1}{q} \right) \right]$$

con G indice di Gini. Questa funzione identifica una classe di indici di povertà che viene utilizzata tuttora, con innovazioni che variano da ricercatore a ricercatore, nello studio della materia.
Altri indici sono stati proposti da diversi autori rispettando le indicazioni fornite dal lavoro di Sen.

Atkinson (1970), con la costruzione dell'indice di Atkinson per la disuguaglianza, riuscì fortemente ad influenzare diversi ricecatori, nella creazione di nuove misure di povertà. Dalla misura della disuguaglianza si è ad esempio potuto ricavare

$$\pi(y) = 1 - \tilde{b}/z$$

con $\tilde{b} = h(b)$ e $b_i = min(y_i, z)$. b uguale al reddito se al di sotto della soglia di povertà o alla soglia di povertà stessa se l'individuo ha reddito più alto. Le caratteristiche di questa funzione rispecchiano gli assiomi proposti da Sen, il valore di $\pi(y)$ decresce all'aumentare del reddito della fascia più povera della popolazione, variazioni nel reddito di chi si trova al di sopra della soglia di povertà non influiscono sull'indice e sono redistribuzioni in favore dei più poveri vanno a modificare π.

Clarke (1981), seguendo Atkinson, introdusse

$$\pi(\mathbf{y}) = 1 - \left[(1/n) \sum_{j=1}^{n} (b_j/z)^{\lambda} \right]^{1/\lambda}$$

2.2 Misure multidimensionali

Quanto detto finora si può ricondurre ai cosiddetti metodi unidimensionali, infatti queste misure si rifanno unicamente ad una variabile, il reddito. Nel corso dei decenni si è sviluppata la ricerca di indicatori e metodi multidimensionali che prevedono l'utilizzo di più variabili con lo scopo di aiutare il policymaker nell'individuazione della migliore strategia per contrastare la povertà. Tra le numerose dimensioni che interessano gli individui, è importante selezionare quelle che, per valori bassi delle stesse, determinano un effettivo stato di povertà. Tra questi indicatori troviamo anche, ad esempio, lo Human Developement Index, che prende in considerazione valori di reddito, istruzione ed aspettativa di vita standardizzati considerando la dimensione della popolazione.

Riportiamo Alkire, Foster (2011) [2] come esempio.
Come fatto per le misure unidimensionali, viene identificato un vettore di valori soglia, un valore per ciascuna variabile presa in considerazione $z = (z_1, ..., z_d)$. Ad ogni soglia viene assegnato un peso in base all'importanza data a quella dimensione nel calcolo complessivo (la somma dei pesi del vettore $w = (w_1, ..., w_d)$ deve essere uguale a uno). Un vettore colonna $c = (c_1, ..., c_d)$ viene utilizzato per la rappresentazione dell'ampiezza della deprivazione di ogni singolo individuo. Il valore c_i rappresenta la somma dei valori delle dimensioni per le quali l'individuo si trova al di sotto della soglia. L'ultimo step è relativo all'identificazione del valore minimo $0 < k \leq d$ per il quale possiamo definire un individuo povero in senso generale. Viene infine costruita una funzione indicatore che assegna il valore 1 alle persone ritenute povere, 0 alle altre.

Osserviamo come questo metodo possa essere utilizzato non solo in caso di variabili cardinali ma anche ordinali. Infatti attraverso una standardizzazione delle stesse è possibile sempre dire se quel valore si trova sopra o al di sotto della soglia di riferimento.

Nella fase di aggregazione dei dati si possono utilizzare delle matrici per identificare le dimensioni, i valori e la popolazione in un solo grafico.

$$Y = \begin{bmatrix} 13.1 & 14 & 4 & 1 \\ 15.2 & \underline{7} & 5 & \underline{0} \\ \underline{12.5} & \underline{10} & \underline{1} & \underline{0} \\ 20 & \underline{11} & 3 & 1 \end{bmatrix}$$
$$z = \begin{pmatrix} 13 & 12 & 3 & 1 \end{pmatrix}$$

Nella matrice le colonne identificano le varie dimensioni, le righe la popolazione ed il vettore z le soglie. Da questa matrice è possibile derivarne un'altra in cui viene assegnato il valore 1 se la variabile si trova al di sotto della soglia e 0 se sopra per ogni individuo e dimensione nella popolazione. Nel precedente esempio avremo: la seconda persona deprivata in due dimensioni, la terza in quattro e l'ultima in una. Se identifichiamo come poveri gli individui con $k \geq 2$ definiamo poveri gli individui numero due e quattro. In questo caso abbiamo ipotizzato pesi uguali per ogni dimensione, se volessimo assegnare pesi diversi il metodo rimarrebbe comunque inalterato.

In base alle esigenze di ricerca è possibile creare ulteriori matrici andando a modificare la formula utilizzata nella costruzione della matrice, ad esempio potremmo creare una matrice (che prende in considerazione solo le persone povere) dove al posto dei valori 1 siano riportate le differenze tra il valore e la soglia attraverso la formula $(z_j - y_{ij})/z_j$

$$Y = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.04 & 0 & 1 \\ 0.04 & 0.17 & 0.67 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Quanto fatto pone le basi per il calcolo ad esempio del numero di individui poveri nella popolazione, in questo caso 2 su 4; o per calcoli sul gap tra tra i poveri e le soglie per ogni singola dimensione e per il valore aggregato.

Nel computo multidimensionale della povertà è importante cercare di mantenere la proprietà della decomposibilità che ci permette di analizzare la povertà nei diversi sotto gruppi della popolazione (regioni, gruppi etnici ecc..) andando poi a definire il valore aggregato come media pesata dei sotto gruppi. L'approccio utilizzato rispetta questa importante proprietà, infatti è possibile costruire una matrice in cui invece dei singoli individui vengano utilizzati i diversi sotto gruppi, derivanti a loro volta da matrici composte da individui appartenenti allo stesso sotto gruppo.

3 Esempi applicativi in R

A completamento di quanto descritto in precedenza, viene proposto il lavoro compiuto sul software R, svolto per verificare le assunzioni fatte per i diversi indici di disuguaglianza e per comprendere i passaggi necessari, e le scritture necessarie, per lo svolgimento di lavori di ricerca sul tema. In questa sezione verranno elencati i passaggi effettuati nel software R, attraverso l'applicativo R studio, in modo che possano essere riproposti facilmente. Risulta innanzitutto essenziale esprimere alcune considerazioni riguardo la ricerca dei dati. I dati più facilmente accessibili possono essere trovati attraverso siti istituzionali (italiani e stranieri) e scaricati in formati Excel. Questi dati open source non sono individuali, bensì suddivisi per categorie di reddito di cui si conosce numerosità e media; per dati più affinati è necessario procedere per vie formali attraverso progetti di ricerca che verranno analizzati dalle suddette istituzioni. In caso si avesse la possibilità di procedere per questa via, indichiamo alcuni centri di ricerca ritenuti dalla comunità accademica come maggiormente affidabili e aggiornati: METIS, EUROSTAT ed altri che troverete nell'allegato a questo lavoro.
In questa sezione si farà riferimento alla raccolta dati disponibile nel sito del Ministero Dell'Economia e Della Finanza compiuta nell'anno 2018, in merito al calcolo IRPEF. L'elaborazione viene svolta attraverso diversi step, in un primo momento useremo dati grezzi unicamente per testare se i pacchetti disponibili in R siano affidabili o meno; in seguito inizieremo ad utilizzare le conoscenze acquisite riguardo gli indici di disuguaglianza in caso di suddivisione della popolazione in classi di reddito, provando ad esprimere valutazioni di merito rispetto ai risultati trovati. A conclusione della nostra ricerca proveremo ad applicare formule di interpolazione per affinare ulteriormente i dati e giungere ad un valore maggiormente preciso.

3.1 Indici e pacchetto ineq

Per il calcolo di alcuni indici di disuguaglianza in R è possibile ricorrere al pacchetto *ineq* grazie al quale si può calcolare la disuguaglianza presente in un vettore. Gli indici disponibili sono: Gini, RS, Atkinson, Theil, Kolm, coefficiente di variazione e l'entropia. Per potere utilizzare questo pacchetto dobbiamo quindi essere a conoscenza della reale distribuzione del reddito nella popolazione, non solo delle classi di reddito che la compongono. Utilizzare questo pacchetto unicamente sfruttando la media delle sottoclassi e la numerosità della stessa non fa altro che ipotizzare il caso in cui all'interno della classe tutti abbiano il reddito pari al reddito medio. Non potendo quindi fare affidamento a questo pacchetto per svolgere analisi a partire dai dati a nostra disposizione, ci chiediamo solamente se esso produce risultati compatibili con quanto espresso dalla letteratura sulla costruzione degli indici. Per fare questo controllo, creeremo semplicemente un vettore composto da 7 unità e verificheremo se i nostri script per i diversi indici produrranno gli stessi risultati degli indici proposti dal pacchetto ineq. Le scritture proposte non rappresentano l'unico modo in cui gli indici possono essere rappresentati ma sono uno dei diversi metodi possibili.

Per prima cosa è quindi necessario implementare degli script adeguati per ogni indice. Definiamo il vettore in considerazione ed il parametro che verrà utilizzato per gli indici derivanti dal concetto di entropia, lo impostiamo ad un valore pari a 0.5 in quanto è il valore utilizzato di default dal pacchetto.

```
> x <- c(100, 870, 45, 320, 65, 225, 400)
> parametro <- 0.5
```

Gini

```
> G <- 0
> for (i in 1:7) {
+     for (j in 1:7) {
+         if (i != j) {
+             G <- G + (1/(2*length(x)^2*mean(x)))*(abs(x[i] - x[j]))
+         }
+     }
+ }
```

Confrontiamo ora il risultato per G ed il risultato di Gini(x) dal pacchetto ineq.

```
> G
[1] 0.4747795
> Gini(x)
[1] 0.4747795
```

Il risultato risulta essere il medesimo. Dimostriamo ora come lo stesso avvenga in considerazione dell'indice di Atkinson, Theil e entropia.

Atkinson

```
> atkinson <- 0
> for (i in 1:7) {
+     atkinson <- atkinson + (x[i]/mean(x))^(1 - parametro)
+ }
> A <- 1 - ((1/length(x))*atkinson)^(1/(1 - parametro))
> A
[1] 0.1899722
> Atkinson(x)
[1] 0.1899722
> |
```

Theil

```
> T <- 0
> for (i in 1:7) {
+     T <- T + (1/length(x))*(x[i]/mean(x))*log(x[i]/mean(x))
+ }
> Theil(x)
[1] 0.3833566
> T
[1] 0.3833566
> |
```

Entropia

```
> entro <- 0
> for (i in 1:7) {
+     entro <- entro + (x[i]/mean(x))^parametro
+ }
> E <- (1/(parametro^2 - parametro))*((1/length(x))*entro - 1)
> E
[1] 0.3999382
> entropy(x)
[1] 0.3999382
> |
```

Questo controllo ci permette di affermare che il pacchetto ineq di R permette di calcolare correttamente questi indici quando siamo in presenza di un vettore reddito composto dal reddito individuale di ogni componente della popolazione.

3.2 Lower and Upper bounds

Cerchiamo ora di utilizzare le conoscenze acquisite per compiere una analisi su dati MEF suddivisi in 31 classi di reddito (positive) di cui conosciamo valore medio e numerosità.
Riscriviamo le formule in grado di fornirci il valore minimo e massimo dell'indice di Gini:

$$G_L = (2\mu)^{-1} \sum_{i \neq j} \sum \frac{n_i n_j}{n^2} |\mu_i - \mu_j|$$

$$D = \mu^{-1} \sum_{i=1}^{k+1} (\frac{n_i}{n})^2 \frac{(\mu_i - a_{i-1})(a_i - \mu_i)}{(a_i - a_{i-1})}$$

$$G_U = G_L + D$$

e Theil, appositamente riscritto mantenendo la struttura originaria per facilitare il calcolo:

$$T_L = \sum_{i=1}^{\theta} \frac{n_i}{n} \frac{\mu_i}{\mu} ln(\frac{\mu_i}{\mu})$$

$$T_W = \sum_{i=1}^{\theta} \frac{n_i \mu_i}{n\mu} (\frac{n_{sup}}{n} \frac{a_{i+1}}{\mu_i} ln(\frac{a_{i+1}}{\mu_i}) + \frac{n_{inf}}{n_i} \frac{a_i}{\mu_i} ln(\frac{a_i}{\mu_i})))$$

$$T_U = T_L + T_W$$

n_{inf}, n_{sup} rappresentano le unità necessarie per mantenere il valore medio costante nella sotto classe nel caso in cui gli individui abbiano livelli di reddito polarizzati ai valori estremi della classe stessa.
Procediamo ora nel calcolo dei suddetti indici ai dati MEF suddivisi per sesso. Cerchiamo quindi di individuare se esiste una differenza tra uomini e donne nel livello di disuguaglianza. Per prima cosa, dopo avere importato i dati in R, è necessario creare due diverse matrici, una per gli uomini ed una per le donne. Proponiamo di seguito gli script utilizzati nel calcolo degli indici per gli uomini, gli stessi passaggi dovranno poi essere ripetuti per l'individuazione degli stessi per le donne.

```
> frequenzem
 [1]  982946  273470  227803  203069  187975  168286  165929  332949  329983  778258 1134065
[12]  974048 1464767 3000999 3751562 1375006 1942049  983656 1004501  274032  206938  300066
[23]  116748   98467  149896  104288  128414   97466   70075   44796   32923
> media_maschi
 [1]    437.018   1243.003   1748.075   2254.701   2752.516   3247.709   3752.196   4510.964
 [9]   5508.282   6739.947   8771.808  11007.268  13504.633  17683.524  22824.251  27443.084
[17]  31764.111  37300.253  44281.324  52372.749  57396.438  64684.186  72445.738  77418.660
[25]  84694.022  94711.011 108979.940 133120.066 171130.329 238948.864 617928.986
> totmediomaschi
[1] 25156.97
> nmaschi
[1] 20905430
> 
```

Per prima cosa identifichiamo alcuni valori utili; *frequenzem* rappresenta la numerosità di ogni sottoclasse, *media.maschi* la media di ogni sottoclasse, *totmediamaschi* la media totale della popolazione e *n.maschi* il numero di individui. Successivamente creiamo due vettori contenenti gli estremi di ogni classe di reddito che ci torneranno utili nel calcolo della disuguaglianza all'interno delle classi.

```
> supbound
 [1] 1.0e+03 1.5e+03 2.0e+03 2.5e+03 3.0e+03 3.5e+03 4.0e+03 5.0e+03 6.0e+03 7.5e+03
[11] 1.0e+04 1.2e+04 1.5e+04 2.0e+04 2.6e+04 2.9e+04 3.5e+04 4.0e+04 5.0e+04 5.5e+04
[21] 6.0e+04 7.0e+04 7.5e+04 8.0e+04 9.0e+04 1.0e+05 1.2e+05 1.5e+05 2.0e+05 3.0e+05
[31] 1.0e+08
> infbound
 [1]      0   1000   1500   2000   2500   3000   3500   4000   5000   6000   7500  10000
[13]  12000  15000  20000  26000  29000  35000  40000  50000  55000  60000  70000  75000
[25]  80000  90000 100000 120000 150000 200000 300000
> 
```

In merito all'ultima classe di reddito, i dati del Ministero la classificano come **"oltre 300.000"**; questa rappresentazione è per noi inutilizzabile in quanto non possiamo lasciare la classe aperta con estremo superiore pari a ∞, altrimenti otterremo indici senza un valore economico. Per lo stesso motivo, in riferimento all'indice di Theil, utilizzare un valore di 0 come limite inferiore della classe dei redditi più bassi, avrebbe come risultato un valore non definito, in quanto il logaritmo di 0 tende a $-\infty$. Osservando la figura precedente notiamo come sia stato arbitrariamente assegnato un valore massimo alla classe di reddito "oltre 300.000". Tale classe ha quindi ora estremi $(300.000; 100.000.000)$. Non andiamo invece a modificare subito lo 0 nella prima classe, in quanto almeno per il calcolo di Gini non crea alcun tipo di problema.
Procediamo nel calcolo del valore minimo e massimo dell'indice di Gini:

```
> GL_maschi <- 0
> for (i in 1:31) {
+     for (j in 1:31) {
+         if (i != j) {
+             GL_maschi <- GL_maschi + ((0.5*frequenzem[i]*frequenzem[j])/(nmaschi^2*totmediom
aschi))*(abs(media_maschi[i] - media_maschi[j]))
+         }
+     }
+ }
> D_maschi <- 0
> for (i in 1:31) {
+     D_maschi <- D_maschi + ((frequenzem[i]^2)/(nmaschi^2*totmediomaschi))*(supbound[i] - med
ia_maschi[i])/(supbound[i] - infbound[i])*(media_maschi[i] - infbound[i])
+ }
> GL_maschi
[1] 0.4448852
> D_maschi
[1] 0.004317755
> GU_maschi <- GL_maschi + D_maschi
> GU_maschi
[1] 0.4492029
> |
```

Constatiamo che la differenza tra valore minimo $0,4448852$ e massimo $0,4492029$ è pari a $0,004317755$, un risultato che ci permette di affermare che questa procedura è decisamente accurata e restringe alla terza cifra decimale lo scostamento nel quale si trova il reale valore di G. Volendo utilizzare la regola $G = 1/3GL + 2/3GU$ confermata da diversi autori come buona approssimazione del valore reale otteniamo $G = 0,447763667$.

Per il calcolo dell'indice di Theil, modifichiamo innanzitutto il valore della prima classe di reddito da $(0; 1.000)$ a $(1; 1.000)$ in modo da ovviare ai problemi precedentemente descritti. L'indice di Theil è fortemente influenzato dalle classi di reddito alte, ci aspettiamo quindi un range tra massimi e minimo sicuramente più alto di quello registrato da Gini, in quanto le stime prodotte, in particolar modo l'assunzione sul limite superiore, andranno certamente a penalizzare la precisione del calcolo. Per il calcolo del numero di individui con reddito pari agli estremi di una generica classe di reddito, tali per cui rimane costante il valore medio della classe procediamo in questo modo

```
> infbound[1] <- 1
> infbound
 [1]      1   1000   1500   2000   2500   3000   3500   4000   5000   6000   7500  10000
[13]  12000  15000  20000  26000  29000  35000  40000  50000  55000  60000  70000  75000
[25]  80000  90000 100000 120000 150000 200000 300000
> nsup_maschi <- frequenzem*(media_maschi - infbound)/(supbound - infbound)
> ninf_maschi <- frequenzem - nsup_maschi
> nsup_maschi
 [1]  429011.1451  132908.3040  113024.5160  103443.9220   94933.2740   83371.7540
 [7]   83693.1520  170125.0890  167724.2890  383913.1893  576925.2960  490563.7910
[13]  734645.4373 1610650.7620 1765892.2997  661416.2107  894673.1193  452531.5644
[19]  430059.4227  130041.8536   99182.8060  140556.4820   57106.9978   47631.6446
[25]   70361.5157   49130.1950   57657.5029   42625.3458   29614.1565   17447.5330
[31]     104.9867
> ninf_maschi
 [1]  553934.85  140561.70  114778.48   99625.08   93041.73   84914.25   82235.85
 [8]  162823.91  162258.71  394344.81  557139.70  483484.21  730121.56 1390348.24
[15] 1985669.70  713589.79 1047375.88  531124.44  574441.58  143990.15  107755.19
[22]  159509.52   59641.00   50835.36   79534.48   55157.80   70756.50   54840.65
[29]   40460.84   27348.47   32818.01
> |
```

questo deriva dalle seguenti espressioni:

$$n_\theta \mu_\theta = n_i y_i + n_{i+1} y_{i+1}$$

$$n_i + n_{i+1} = n_\theta$$

tale per cui

$$n_{i+1} = \frac{n_\theta(\mu_\theta - y_i)}{(y_{i+1} - y_i)}$$

Avendo trovato i vettori corrispondenti calcoliamo Theil

```
> TL_maschi <- 0
> for (i in 1:31) {
+     TL_maschi <- TL_maschi + (frequenzem[i]/nmaschi)*(media_maschi[i]/totmediomaschi)*log(me
dia_maschi[i]/totmediomaschi)
+ }
> TW_maschi <- 0
> for (i in 1:31) {
+     TW_maschi <- TW_maschi + ((frequenzem[i]*media_maschi[i])/(nmaschi*totmediomaschi))*((ns
up_maschi[i]/frequenzem[i])*(supbound[i]/media_maschi[i])*log(supbound[i]/media_maschi[i]) + (
ninf_maschi[i]/frequenzem[i])*(infbound[i]/media_maschi[i])*log(infbound[i]/media_maschi[i]))
+ }
> TL_maschi
[1] 0.4125581
> TU_maschi <- TL_maschi + TW_maschi
> TU_maschi
[1] 0.5067546
>
```

Osserviamo come la differenza TW sia decisamente elevata come potevamo aspettarci. Proviamo a calcolare lo stesso indice andando a non considerare la classe di reddito più alta e osserviamo come tale differenza vada notevolmente a ridursi.

```
> TW_maschi <- 0
> for (i in 1:30) {
+     TW_maschi <- TW_maschi + ((frequenzem[i]*media_maschi[i])/(nmaschi*totmediomaschi))*((ns
up_maschi[i]/frequenzem[i])*(supbound[i]/media_maschi[i])*log(supbound[i]/media_maschi[i]) + (
ninf_maschi[i]/frequenzem[i])*(infbound[i]/media_maschi[i])*log(infbound[i]/media_maschi[i]))
+ }
> TW_maschi
[1] 0.00618296
> |
```

La differenza che prima si attestava a 0.0941965 ora si è ridotta a 0.00618296, un valore notevolmente inferiore. La classe di reddito più alta, essendo più estesa delle altre ed essendo stata approssimata fa infatti aumentare il valore della disuguaglianza all'interno della stessa classe andando a fare aumentare notevolmente il valore totale dell'indice di Theil.

Applicando lo stesso procedimento per la componente femminile della popolazione otteniamo i seguenti risultati (per Theil manteniamo tutte le fasce di reddito non escludendo quella massima).

$$Ginimaschi[0.4448852; 0.4492029]$$

$$Ginifemmine[0.4414275; 0.4461150]$$

$$Theilmaschi[0.4125581; 0.5067546]$$

$$Theilfemmine[0.3659769; 0.3918526]$$

Entrambi gli indici mostrano una maggiore uguaglianza nella popolazione femminile rispetto a quella maschile, in particolare nella popolazione femminile la discrepanza tra minimo e massimo per l'indice di Theil è notevolmente inferiore rispetto alla controparte maschile.

La stessa procedura può essere utilizzata per calcolare questi due indici anche per sottoclassi come le regioni italiane (dati disponibili nel sito del MEF) o le fasce di età. Ciò permette di osservare e chiederci quale sia il motivo delle differenze di uguaglianza nella distribuzione di reddito tra aree del paese ed indagare le relazioni con altre variabili di interesse.

3.3 Interpolazione

Tecniche di interpolazione permettono di andare ad assumere il comportamento della distribuzione di reddito all'interno della popolazione, permettendoci di affinare i risultati derivanti dal semplice calcolo del lower e upper bound. La formula da utilizzare, una volta identificata la funzione di densità, è la seguente:

$$\sum_{i=1}^{k}\left[\int_{\alpha_i}^{a_{i+1}} h(y)f(y)dy\right]$$

dove $f(y)$ rappresenta la funzione di densità derivata dall'interpolazione e $h(y)$ definisce la funzione dell'indice di disuguaglianza. Esistono diverse tecniche che fanno al caso nostro, Cowell e Metha [9] ne descrivono alcune, diverse in base alle informazioni a nostra disposizione (conoscenza o meno del valore medio e della numerosità all'interno delle classi di reddito). Proponiamo alcune tecniche nel caso in cui il ricercatore sia a conoscenza sia del valore medio μ_i che della numerosità n_i.

Sono diverse le proprietà che desideriamo vengano rispettate dalla funzione di interpolazione:

1) $f(y) \geq 0$

2) Continuità in ogni intervallo

3) Continuità del primo e secondo ordine

4) $\lim_{y \to a_{i+1}} f(y) = 0$

5) $\lim_{y \to a_{i+1}} f'(y) = 0$

6) Pochi turning points

7) Un margine ristretto di $f(y)$ per ogni intervallo

8) Closed form integration della misura di disuguaglianza

Rispettare contemporaneamente tutte queste proprietà è molto difficile, le seguenti tecniche risultano essere le migliori sebbene non rispettino tutte le proprietà.

Interpolazione Paretiana La funzione di Pareto del primo tipo viene spesso utilizzata per esemplificare il comportamento della distribuzione di reddito per i componenti ad alto reddito della popolazione, è quindi naturale assumere che la funzione di densità paretiana possa essere utilizzata come tecnica di interpolazione.

$$f(y) = A_i y^{-\alpha_i^{-1}}, \qquad y \in [a_i, a_{i+1})$$

dove A_i, α_i sono parametri scelti dal ricercatore.

Interpolazione polinomiale Polynomial spline

$$f(y) = \sum_{k=0}^{K} \gamma_{ik} y^k, \quad y \in [a_i, a_{i+1})$$

questa metodologia non sempre assicura un numero ristretto di turning points quando $K \geq 3$ (valore che permetterebbe di rispettare un gran numero di proprietà descritte); di conseguenza si ricorre spesso alla linea continua o alla quadratica con $K < 0$.

$$f(y) = \gamma_{i0} + \gamma_{i1} y, \quad y \in [a_i, a_{i+1})$$

Split histogram La più semplice delle tecniche di interpolazione, nonché la tecnica che useremo alla fine di questa sezione per un esempio pratico.

$$f(y) = \begin{cases} \frac{n_i/n}{a_{i+1}-a_i}\frac{a_{i+1}-\mu_i}{\mu_i-a_i}, a_i \leq y < \mu_i \\ \frac{n_i/n}{a_{i+1}-a_i}\frac{\mu_i-a_i}{a_{i+1}-\mu_i}, \mu_i \leq y < a_{i+1} \end{cases}$$

L'istogramma risultante sarà quindi diviso in i+1 colonne a loro volta divise in due semicolonne in base al valore medio della classe. La funzione non sarà continua ma risulta essere estremamente robusta per lo scopo.

Utilizziamo i dati del MEF, suddivisi per regioni, per procedere al calcolo dell'indice di Theil attraverso la tecnica di interpolazione dello split histogram. Identifichiamo il valore minimo e massimo dell'indice e verifichiamo conseguentemente se la misura ottenuta con l'interpolazione si trovi all'interno del range e sia congruente con la tecnica 1/3 2/3 suggerita in precedenza come buona approssimazione del valore reale dell'indice. Prendiamo come riferimento i dati della regione Veneto e ne calcoliamo T_L e T_U.

$$T_L = 0.3781034$$

$$T_U = 0.4445465$$

Procediamo ora in R per il calcolo di T; per svolgere l'integrale richiesto è utile istallare il pacchetto *pracma* che permette il calcolo integrale attraverso al funzione integral.

```
> thetainf <- (freqenze_VN/n_VN)*(1/(supbound - infbound))*(supbound - mediaclasse_VN)*(1/(med
iaclasse_VN - infbound))
> thetasup <- (freqenze_VN/n_VN)*(1/(supbound - infbound))*(mediaclasse_VN - infbound)*(1/(sup
bound - mediaclasse_VN))
> |
```

thetainf e thetasup rappresentano le funzioni di interpolazione nel caso in cui y si trovi tra a_i, μ_i e tra μ_i, a_{i+1}.

```
> thinf <- as.vector(thetainf)
> thsup <- as.vector(thetasup)
> inftheil <- 0
> for (i in 1:31) {
+     inftheil <- inftheil + integral(function(y) (y/mediatot_VN)*log(y/mediatot_VN)*thinf[i],
 infbound[i], mediaclasse_VN[i])
+ }
> suptheil <-0
> for (i in 1:31) {
+     suptheil <- suptheil + integral(function(y) (y/mediatot_VN)*log(y/mediatot_VN)*thsup[i],
 mediaclasse_VN[i], supbound[i])
+ }
> THEIL <- inftheil + suptheil
> THEIL
[1] 0.4062692
> |
```

Il valore trovato grazie alla tecnica di interpolazione è pari a

$$T = 0.4062692$$

valore interno al range definito dal minimo e massimo precedentemente individuati.
Verifichiamo se T è prossimo al valore derivante dalla regola

$$T_{/3} = 0.3781034\frac{2}{3} + 0.4445465\frac{1}{3} = 0.4002511$$

L'indice trovato dalla regola 1/3, 2/3 si avvicina al valore identificato dallo split histogram, per una differenza sulla terza cifra decimale.

4 Esempi di ricerca

In questa ultima sezione si cercherà di andare oltre alla mera analisi matematica-statistica, verranno proposte alcune ricerche accademiche come esempi di analisi in campo microeconomico. Non si andrà quindi a calcolare il valore dell'indice della disuguaglianza ma ad indagare le ragioni che muovono questo indicatore.

Fin dall'introduzione di questa tematica in campo accademico, parallelamente alla ricerca statistica, ci si è interrogati sulle cause della disuguaglianza e sulle politiche fiscali o monetarie, socialmente accettabili, in grado di mitigare questo fenomeno promuovendo una maggiore uguaglianza. Uno dei motivi di ricerca maggiormente presenti riguarda l'effetto della crescita economica sul livello di disuguaglianza tra e all'interno dei diversi gruppi di reddito che compongono una popolazione. Tra i diversi autori che hanno esaminato la relazione tra crescita e disuguaglianza riportiamo Ravillon M. (2001) [19], dove l'autore asserisce che la crescita economica porti sia benefici all'interno delle fasce più povere della popolazione che un incremento della disuguaglianza tra i diversi gruppi. L'analisi dei dati che propone mostra chiaramente questo processo, tutte le diverse fasce di reddito beneficiano della crescita economica ma in diverso modo; il risultato finale è quindi di un miglioramento degli standard di vita delle famiglie al di sotto della soglia di povertà ma al contempo il proseguimento di ciò che potremmo chiamare "snowball effect", per il quale chi era in condizione agiate (avendo accesso anche a strumenti finanziari diversi da chi ha meno reddito) riesce a sfruttare meglio il periodo di crescita aumentando in maniera maggiore il proprio livello di reddito. Dall'analisi dell'autore si evince inoltre come un periodo di crescita economica che interessi una popolazione con un elevato livello iniziale di disuguaglianza porti un minore effetto della stessa sia a livello aggregato che in particolar modo nei gruppi più disagiati.

La domanda di redistribuzione di ricchezza appare come un tema di rilievo nella ricerca accademica, in Ghossub E.A. e Reed R.R. (2017) [15] gli autori spostano il focus sulla presenza o meno di istituzioni finanziarie all'interno di un paese e sul loro corretto o meno funzionamento. I due autori dimostrano come le economie con uno sviluppo finanziario più elevato vadano a limitare il crescere delle diseguaglianze in quanto riescono a fornire strumenti di assicurazione sia per le fasce agiate che per quelle più in difficoltà andando anche a promuovere e stimolare la crescita. Nella loro analisi si individua come strumento di politica monetaria, la stabilità dei prezzi (per limitare il fenomeno inflativo) e lo sviluppo dei sistemi finanziari ad ogni livello.

Tornando ai momenti espansivi del ciclo economico, è importante per uno stato riuscire a non dissipare lo slancio di crescita presente nel sistema economico del paese. Un rischio è quello di dissipare questa crescita a causa di un sistema pubblico ingessato e corrotto. Un'analisi su questo tema è stata svolta da Chong A. e Gradstein M. (2007) [7] nella quale gli autori confermano come paesi con istituzioni che presentano una pessima qualità mostrano generalmente una disuguaglianza maggiore rispetto a quei paesi in cui ad esempio il fenomeno corruttivo è marginale.

Muovendoci dal generale al particolare, una domanda che possiamo chiederci è capire se esiste una correlazione tra disuguaglianza di reddito e di consumo. Lo scopo è capire se e come una provata disuguaglianza di reddito vada ad influire sugli standard di vita delle diverse fasce di popolazione. In Aguiar M., Bils M. (2015) [1] e Krueger D., Perri F. (2006) [16] troviamo due analisi che portano a conclusioni differenti, nel primo caso viene riscontrata una relazione tra le due variabili, nel secondo caso no. Questa differenza dipende sia dal tipo di dati utilizzato che dal metodo quantitativo applicato nei diversi casi.

La letteratura accademica lascia quindi ampio spazio di ricerca non essendo riuscita a definire chiaramente le cause e gli effetti della disuguaglianza sulla popolazione, ciò che è certo è che con il passare degli anni l'affinamento del metodo di calcolo degli indici di disuguaglianza e povertà utilizzati in queste ricerche e la mole di dati in continuo aumento porteranno a risultati sempre più precisi e uniformi.

A conclusione di questa breve introduzione sull'argomento povertà e disuguaglianza, riportiamo un articolo di Genicot, Ray (2017) [14] che analizza il fenomeno sotto una lente inusuale. Il paper presentato dai due autori mette in relazione il diverso livello di aspirazioni personali degli individui presenti nei diversi sottogruppi di reddito della popolazione con il livello di disuguaglianza presente nella stessa. Nel modello utilizzato dimostrano come la possibilità di raggiungere i proprio obiettivi sia di fondamentale importanza nella qualità degli investimenti che l'individuo è portato a compiere nella propria vita, e di come, viceversa, il non avere la possibilità di avvicinarsi ad essi provochi frustrazione e malessere. I due autori individuano in questo fenomeno una delle componenti dell'aumento della disuguaglianza: diversità nella qualità degli obiettivi e nella possibilità di raggiungerli dipendenti dal livello di reddito individuale.

References

[1] Mark Aguiar and Mark Bils. Has consumption inequality mirrored income inequality? *American Economic Review*, 105(9):2725–56, 2015.

[2] Sabina Alkire and James Foster. Counting and multidimensional poverty measurement. *Journal of public economics*, 95(7-8):476–487, 2011.

[3] Paul D Allison. Measures of inequality. *American sociological review*, pages 865–880, 1978.

[4] Barry C Arnold. On the amato inequality index. *Statistics & Probability Letters*, 82(8):1504–1506, 2012.

[5] Anthony B Atkinson et al. On the measurement of inequality. *Journal of economic theory*, 2(3):244–263, 1970.

[6] Garry F Barrett and Stephen G Donald. Statistical inference with generalized gini indices of inequality, poverty, and welfare. *Journal of Business & Economic Statistics*, 27(1):1–17, 2009.

[7] Alberto Chong and Mark Gradstein. Inequality and institutions. *The Review of Economics and Statistics*, 89(3):454–465, 2007.

[8] Frank Cowell. *Measuring inequality.* Oxford University Press, 2011.

[9] Frank A Cowell and Fatemeh Mehta. The estimation and interpolation of inequality measures. *The Review of Economic Studies*, 49(2):273–290, 1982.

[10] Chris Elbers, Jean O Lanjouw, and Peter Lanjouw. Micro–level estimation of poverty and inequality. *Econometrica*, 71(1):355–364, 2003.

[11] James Foster, Joel Greer, and Erik Thorbecke. A class of decomposable poverty measures. *Econometrica: journal of the econometric society*, pages 761–766, 1984.

[12] Joseph L Gastwirth. The estimation of the lorenz curve and gini index. *The review of economics and statistics*, pages 306–316, 1972.

[13] Joseph L Gastwirth. The estimation of a family of measures of economic inequality. *Journal of Econometrics*, 3(1):61–70, 1975.

[14] Garance Genicot and Debraj Ray. Aspirations and inequality. *Econometrica*, 85(2):489–519, 2017.

[15] Edgar A Ghossoub and Robert R Reed. Financial development, income inequality, and the redistributive effects of monetary policy. *Journal of Development Economics*, 126:167–189, 2017.

[16] Dirk Krueger and Fabrizio Perri. Does income inequality lead to consumption inequality? evidence and theory. *The Review of Economic Studies*, 73(1):163–193, 2006.

[17] Stéphane Mussard, Françoise Seyte, and Michel Terraza. Decomposition of gini and the generalized entropy inequality measures. *Economics Bulletin*, 4(7):1–6, 2003.

[18] Graham Pyatt. Measuring welfare, poverty and inequality. *The Economic Journal*, 97(386):459–467, 1987.

[19] Martin Ravallion. Growth, inequality and poverty: looking beyond averages. *World development*, 29(11):1803–1815, 2001.

[20] Christian Schluter and Mark Trede. Statistical inference for inequality and poverty measurement with dependent data. *International Economic Review*, pages 493–508, 2002.

[21] Amartya Sen. Poverty: an ordinal approach to measurement. *Econometrica: Journal of the Econometric Society*, pages 219–231, 1976.

[22] Anthony F Shorrocks. The class of additively decomposable inequality measures. *Econometrica: Journal of the Econometric Society*, pages 613–625, 1980.

[23] Henri Theil. World income inequality and its components. *Economics Letters*, 2(1):99–102, 1979.