

Gli indici di asimmetria della curva di Lorenz per dati raggruppati

The asymmetry indices of the Lorenz Curve for Grouped Data

Tarsitano Agostino

Dipartimento di Economia Politica - Università della Calabria

Abstract: the asymmetry of the Lorenz curve has proven useful to illustrate the trend in polarization of the earnings distribution and to the empirical assessment of the Kuznet's curve. The scope of the present work is to review the indices which have been proposed to quantify the asymmetry of the Lorenz curve and evaluate their performances with grouped data.

Parole chiave: piecewise interpolation.

1. Introduzione

Le misure della concentrazione si configurano come indicatori sintetici della situazione distributiva quantificandone lo scarto dalla equidistribuzione. Tuttavia, come per tanti altri indici statistici, la loro natura di "medie" fa sì che uno stesso valore sia associato a situazioni distributive molto diverse. L'esigenza di affiancare le misure di concentrazione con altri indici che ne aumentino le capacità di caratterizzazione ha portato alla proposta di alcune misure sintetiche della asimmetria della curva di Lorenz. Scopo del presente lavoro è di riassumere la problematica della misura della asimmetria della curva di Lorenz riprendendo gli indici proposti in letteratura (secondo paragrafo); passeremo poi, nel secondo paragrafo, ad analizzare varie forme di interpolazione concatenata (*piecewise interpolation*) della curva di Lorenz al fine di determinare quale sia la tecnica più efficiente per il calcolo approssimato degli indici di asimmetria; nel terzo paragrafo confronteremo le tecniche di interpolazione ai fini del calcolo degli indici in riferimento ad una nota funzione di concentrazione.

2. L'asimmetria della curva di Lorenz. Definizione e misura

Per studiare l'asimmetria della curva di Lorenz è necessario considerare tratti uguali della curva in corrispondenza di intervalli uguali sulla retta di equidistribuzione. La curva è simmetrica se la distanza misurata in parallelo alla diagonale $q=1-p$ (detta polo di simmetria) dalla retta di equidistribuzione è uguale per punti collocati a distanza uguale dal punto divisorio "D" di coordinate (p_D, q_D) in cui il polo di simmetria interseca la funzione di concentrazione $q=L(p)$:

$$L(p_D + p) - p = L(p_D - p) + p \quad \text{per } 0 \leq p \leq \min\{p_D, 1 - p_D\}$$

Zanardi (1965) definisce statisticamente simmetrica una curva di concentrazione per la quale si abbia in riferimento alla figura 1, $A_L = A_U$ od anche $T_L = T_U$. Questo non implica la simmetria dato che T_L e T_U possono provenire da curve e distribuzioni diverse (Giurovich, 1960) e non implica nemmeno il centramento che è una condizione sulla distanza $p-L(p)$ che è necessaria, ma non sufficiente per la simmetria. Gli indici proposti per misurare la asimmetria della curva di Lorenz sono i seguenti:

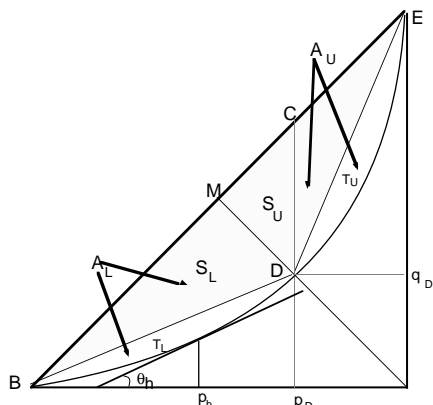


Figura 1: elementi per la misura della asimmetria

$$\begin{aligned}
 \text{Panizzon} \quad A_R &= \text{Log}_3 \left(\frac{S_U + T_U}{S_L + T_L} \right) \\
 \text{Giurovich} \quad G_d &= \frac{T_U}{T_L}; \quad G_s = \frac{T_L}{T_U}; \\
 \text{Patimo} \quad A'_D &= 2 \left(\frac{T_U - T_L}{p_D q_D} \right) \\
 \text{Zanardi} \quad Z &= \frac{(T_U - T_L)}{R/4} \\
 \text{Taguchi} \quad \tau &= -\cos(2\theta_h)
 \end{aligned}$$

Gli indici di Panizzon, Patimo e Zanardi sono nulli in caso di simmetria (il contrario non è sempre vero); assumono valori negativi se la curva ha la gobba in un punto più prossimo all'origine che non al punto (1,1) e positivi se la gobba è più vicina al punto (1,1). I quozienti di Panizzon hanno valore unitario per curve simmetriche (ma non solo) e aumentano all'infinito con il ridursi dell'area al denominatore. L'indice τ varia tra -1 ed 1. È negativo quando il punto di saturazione si trova in corrispondenza di intensità inferiori alla media aritmetica cioè in posizioni più basse rispetto al punto di massima distanza (p_μ, q_μ) ed è positivo se $p_h \geq p_\mu$; è zero se la massima curvatura è nel punto di distanza massima ($x_h = \mu$). Il calcolo degli indici di asimmetria richiede la determinazione del punto divisorio che di solito comporta la soluzione un calcolo iterativo. Noto "D" si possono calcolare le aree T_L e T_U :

$$T_L = \frac{p_D q_D}{2} - \int_0^{p_D} L(p) dp; \quad T_U = \frac{q_D(1 + q_D)}{2} - \int_{p_D}^1 L(p) dp$$

per l'indice di Taguchi occorre individuare il punto (p_h, q_h) in cui la curva di Lorenz raggiunge la massima curvatura (il punto però potrebbe non essere unico).

3. Approssimazione delle aree

Il difetto di approssimazione nelle aree è un inconveniente ben noto della spezzata di Lorenz. Gastwirth e Glauberger (1976), ipotizzando la densità lineare nelle classi, hanno adottato polinomi di Hermite di 3° grado; Kakwani (1976) è arrivato fino ai polinomi di 4° grado, ma senza utilizzare il formalismo delle *spline*. Sia $L(p)$ la funzione di concentrazione sottostante la seriazione osservata. "L" è incognita, ma si ritiene di poterla approssimare con una sequenza concatenata di polinomi di grado "m":

$$L_i(p) = \sum_{j=0}^m b_{ij}^m (p - p_{i-1})^j \quad p_{i-1} \leq p \leq p_i; \quad i = 1, 2, \dots, k$$

I polinomi contigui coincidono nei punti di transizione: $L_i(p_i) = L_{i+1}(p_i)$ assicurando così la continuità della curva interpolante. Se $L(p)$ è dotata di derivate continue fino alla n-esima i coefficienti delle *spline* sono scelti in modo che:

$$L_i^{(r)}(p_{i-1}) = \sum_{j=r}^m \prod_{s=1}^r (j - s + 1) b_{ij}^m (p - p_{i-1})^{j-r} = L_{i-1}^{(r)} \quad i = 2, \dots, k + 1$$

Le derivate prime e seconde della curva di Lorenz nei punti di svolta sono approssimate da:

$$L_i^{(1)}(p_i) = \frac{\mu_i}{\mu}, \quad L_i^{(1)}(p_{i-1}) = \frac{\mu_{i-1}}{\mu}; \quad L_i^{(2)}(p_i) = \frac{1}{\mu f_i}; \quad L_i^{(2)}(p_{i-1}) = \frac{1}{\mu f_{i-1}}; \quad i = 2, \dots, k; \quad \mu = \sum_{i=1}^k \mu_i f_i$$

Tabella 1: coefficienti delle *spline* da utilizzare per l'interpolazione della curva di Lorenz

m	Parametri
2	$b_{i0}^2 = q_{i-1}; \quad b_{i1}^2 = L_{i-1}^{(1)}; \quad b_{i2}^2 = (L_i^{(1)} - L_{i-1}^{(1)})/f_i;$
3	$b_{i0}^3 = q_{i-1}; \quad b_{i1}^3 = L_{i-1}^{(1)}; \quad b_{i2}^3 = 2(L_i^{(1)} - L_{i-1}^{(1)})/f_i, \quad b_{i3}^3 = -(L_i^{(1)} - L_{i-1}^{(1)})/f_i^2;$ $b_{i0}^4 = q_{i-1}; \quad b_{i1}^4 = L_{i-1}^{(1)}; \quad b_{i2}^4 = L_{i-1}^{(2)}/2, \quad b_{i3}^4 = 3(L_i^{(1)} - L_{i-1}^{(1)})/f_i^2 - L_{i-1}^{(2)}/f_i$
4	$b_{i4}^4 = -2(L_i^{(1)} - L_{i-1}^{(1)})/f_i^3 + L_{i-1}^{(2)}/2f_i^2;$
5	$b_{i0}^5 = q_{i-1}; \quad b_{i1}^5 = L_{i-1}^{(1)}; \quad b_{i2}^5 = L_{i-1}^{(2)}/2, \quad b_{i3}^5 = 6(L_i^{(1)} - L_{i-1}^{(1)})/f_i^2 + (L_i^{(2)} - L_{i-1}^{(2)})/2f_i - L_{i-1}^{(2)}/f_i$ $b_{i4}^5 = -8(L_i^{(1)} - L_{i-1}^{(1)})/f_i^3 - (L_i^{(2)} - L_{i-1}^{(2)})/f_i^2 + L_{i-1}^{(2)}/2f_i^2; \quad b_{i5}^5 = 3(L_i^{(1)} - L_{i-1}^{(1)})/f_i^4 + (L_i^{(2)} - L_{i-1}^{(2)})/2f_i^3;$

Il metodo dell'interpolazione concatenata incontra ostacoli negli intervalli terminali $[0, p_1]$ e $[p_k, 1]$ perché qui non è univocamente stabilita l'inclinazione della curva negli estremi zero ed uno. A questo fine ipotizziamo che nel primo intervallo la curva di Lorenz decresca proporzionalmente:

$$\frac{L_0^{(1)}}{L_1^{(1)}} = \gamma \frac{L_1^{(1)}}{L_2^{(1)}} \Rightarrow L_0^{(1)} = \gamma \frac{[L_1^{(1)}]^2}{L_2^{(1)}}; \quad \frac{L_0^{(2)}}{L_1^{(2)}} = \frac{L_1^{(2)}}{L_2^{(2)}} \Rightarrow L_0^{(2)} = \frac{[L_1^{(2)}]^2}{L_2^{(1)}}$$

dove γ è un fattore posto inizialmente pari ad uno e successivamente dimezzato finché $L'_0 < L'$. Per la derivata seconda facciamo una ipotesi analoga perché fornisce valori vicini al vero per molti modelli di curva (ad esempio la curva di Pareto e l'esponenziale). Nell'ultima classe stimeremo regolarmente i parametri legati al punto (p_k, q_k) e porremo pari a zero quelli successivi tranne l'ultimo chiamato ad assicurare che $L_{k+1}(1)=1$.

4. Applicazione

Al fine di valutare l'accuratezza delle interpolazioni proposte scegliamo una curva di Lorenz note delle quale usare i valori teorici: $L(p_i), L'(p_i), L''(p_i)$ per $i=1,2,\dots,k$ per diversi valori di "k".

$$L(p) = p^\alpha e^{-\beta(1-p)}; \quad \alpha > 1, \quad \beta > 0$$

con $\alpha=2.5, \beta=0.5$. Il numero di classi ruota in $\{5, 10, 15, 20\}$. I valori teorici sono riportati nella prima riga della tabella. La colonna "crit." è un criterio ottenuto sommando gli scarti in valore assoluto dal valore teorico del Gini e dello Zanardi. I risultati rispettano le attese: il miglioramento all'aumentare del numero di punti è tra questi; c'è però da precisare che quando i punti sono acquisiti in modo sbilanciato per i vari tratti della curva (ad esempio più rilevazioni per i livelli elevati che non per quelli

inferiori) le misure di asimmetria potrebbero risultarne distorte. Anche il miglioramento al crescere del grado del polinomio non desta sorprese, sebbene il fatto che la quartica sia talvolta meno fallace della quintica suggerisce un affinamento dell'interpolazione nelle classi terminali. Comunque, dalle 15 classi in poi non c'è differenza apprezzabile tra le *spline* di 4° grado e di 5°. Tra gli indici di asimmetria si può notare l'andamento parallelo dell'indice di Zanardi e di Patimo che sembrano più resistenti alle fluttuazioni dovute al grado del polinomio ed al numero di classi, rispetto agli indici di Panizzon e di Giurovich. La stima del Taguchi ha prodotto risultati eterogenei e questo, unito alle critiche già svolte, non contribuisce alla sua adozione come indice specificativo del rapporto di concentrazione. L'esito della sperimentazione può così essere riassunto: il calcolo del Gini con i polinomi di Hermite di 3° grado è perfettibile utilizzando polinomi e di grado superiore: 4° o 5° grado. Tali polinomi possono anche essere invocati per calcolare gli indici di asimmetria della funzione di concentrazione ad esclusione dell'indice di Taguchi i cui calcoli sono spesso contraddittori. Le indicazioni delle misure di asimmetria sono però affidabili come segno e valore solo a partire dalle 15 classi in poi.

Tabella 2: esito della sperimentazione.

Classi	m	Gini	p_b	T_u	T_L	Z	Pat	Pan	Giur	τ	Crit.
		0.48671	0.67791	0.01423	0.05121	-0.30391	-0.33871	-0.27897	3.59782	-0.65793	0.00000
k=5	1	0.47282	0.67527	0.01139	0.04947	-0.32214	-0.34731	-0.29616	4.34206	-0.76177	0.03213
	2	0.46541	0.67778	0.01057	0.04436	-0.29043	-0.30946	-0.26625	4.19773	-0.75283	0.03478
	3	0.45825	0.67757	0.01003	0.04152	-0.27489	-0.28830	-0.25181	4.13948	-0.74393	0.05748
	4	0.47015	0.67787	0.01126	0.04594	-0.29508	-0.31767	-0.27057	4.07985	-0.57183	0.02539
	5	0.45554	0.67782	0.00864	0.04130	-0.28675	-0.29908	-0.26283	4.77820	-0.57274	0.04833
k=10	1	0.48257	0.67335	0.01520	0.05243	-0.30863	-0.33857	-0.28356	3.44975	-0.69059	0.00885
	2	0.46661	0.67648	0.01319	0.05010	-0.31642	-0.33732	-0.28249	3.79804	-0.68549	0.03261
	3	0.47658	0.67513	0.01324	0.04991	-0.30777	-0.33438	-0.28239	3.76869	-0.68044	0.01399
	4	0.48177	0.67648	0.01373	0.05067	-0.30671	-0.33758	-0.28140	3.69016	-0.57850	0.00774
	5	0.47603	0.67526	0.01290	0.04985	-0.31044	-0.33696	-0.28488	3.86337	-0.55542	0.01721
k=15	1	0.48475	0.67485	0.01520	0.05214	-0.30478	-0.33666	-0.27982	3.43000	-0.47855	0.00283
	2	0.47172	0.67579	0.01441	0.05139	-0.31362	-0.33761	-0.28091	3.56732	-0.65776	0.02469
	3	0.48160	0.67505	0.01443	0.05133	-0.30647	-0.33643	-0.28117	3.55776	-0.65425	0.00766
	4	0.48440	0.67658	0.01434	0.05129	-0.30513	-0.33773	-0.27993	3.57751	-0.57935	0.00353
	5	0.48140	0.67490	0.01441	0.05139	-0.30729	-0.33711	-0.28194	3.56681	-0.72803	0.00869
k=20	1	0.48557	0.67715	0.01428	0.05133	-0.30525	-0.33900	-0.28007	3.59505	-0.50659	0.00248
	2	0.47597	0.67780	0.01375	0.05075	-0.31097	-0.33887	-0.28020	3.69077	-0.64271	0.01780
	3	0.48364	0.67764	0.01359	0.05058	-0.30592	-0.33867	-0.28067	3.72094	-0.64003	0.00508
	4	0.48538	0.67782	0.01393	0.05093	-0.30491	-0.33886	-0.27973	3.65547	-0.71030	0.00233
	5	0.48355	0.67773	0.01351	0.05053	-0.30631	-0.33908	-0.28103	3.74183	-0.69876	0.00556

Riferimenti bibliografici

- Gastwirth J.L. Glauberman M. (1976) The interpolation of the Lorenz curve and Gini index from grouped data. *Econometrica*. 44, 479-483.
- Kakwani N.C. (1983) On the estimation of income inequality measures from grouped observations. *Review of Economic Studies*. 483-492.
- Panizzon G. (1955) La concentrazione della proprietà fondiaria in Italia. *Ricerche Economiche*. Vol. 9, 251-278.
- Patimo A. (1977) Su un indice specificativo del rapporto di concentrazione. *Rivista Italiana di Economia, Demografia e Statistica*. 31, 333-350.
- Patimo A. (1978) Dall'indice specificativo del rapporto di concentrazione ad un indice di asimmetria della curva di concentrazione., Atti della Riunione Scientifica della SIS. 29,177-187.
- Taguchi (1968) Concentration-curve methods and structures of skew populations. *Annals of the Institute of Statistical Mathematics*. 20, 107-141.
- Zanardi G. (1965) Lasimmetria statistica delle curve di concentrazione. *Ricerche Economiche*. 19, 355-396.