Coefficiente di Gini composito

La domanda che ci si pone è la seguente: dati n paesi di cui si conoscano i redditi totali R, le popolazioni N, quindi i redditi pro capite medii R_M =R/N, nonché i coefficienti di Gini G, qual'è il coefficiente di Gini del paese immaginario composto dall'insieme degli n paesi?
Si può cominciare dal caso di 2 soli paesi, nella speranza, che rimane peraltro da verificare, che si possa poi procedere per iterazioni successive.
Si tratta dunque di costruire la curva di Lorenz y = y(x) del sistema composito costituito da due paesi aventi curve di Lorenz $y_1 = y_1(x_1)$, $y_2 = y_2(x_2)$, nonché redditi totali R_1, R_2, popolazioni N_1, N_2, redditi pro capite medii $R_{M1} = R_1/N_1$, $R_{M2} = R_2/N_2$ e coefficienti di Gini G_1, G_2.
Poiché, a questo fine, la popolazione deve essere ordinata per redditi pro capite crescenti, ad ogni x1 occorrerà associare una x2 tale per cui i redditi pro capite dei membri più ricchi delle due frazioni siano gli stessi; si dovrà poi dedurne la corrispondente frazione x della popolazione del sistema composito ed infine, a partire da $y_1 = y_1(x_1)$, $y_2 = y_2(x_2)$, si dovrà calcolare la y corrispondente a tale x, che esprime la curva di Lorenz del sistema composito; l'area B da essa sottesa ci permetterà infine di calcolare il coefficiente di Gini secondo la nota formula: G = 1 – 2B.
Ora poiché in generale, per una data frazione x della popolazione, il reddito pro capite del suo membro più ricco è dato da: $r = R_M.dy/dx$, è evidente che il processo suddetto presuppone che le curve di Lorenz originarie siano date in forma matematica e che una soluzione del problema assolutamente generale, indipendente cioè da tale forma, non è quindi possibile.
Nel seguito si farà l'ipotesi che entrambe le curve di Lorenz di partenza abbiano la forma matematica corrispondente alla distribuzione di Pareto, una distribuzione, comunque, che fornisce un grado di approssimazion soddisfacente in molti casi ed è quindi largamente usata.
In tal caso la curva di Lorenz è data da:

$$y = 1 - (1 - x)^{\beta} \qquad (1)$$

dove, per ragioni che saranno subito chiare, abbiamo preferito usare, in luogo della costante α di Pareto, la costante $\beta = (\alpha -1)/\alpha$. Derivando la (1) si ottiene:

$$dy/dx = \beta(1 - x)^{\beta-1} \qquad (2)$$

Per x = 0 la (2) dà: dy/dx = β da cui consegue che β rappresenta il rapporto fra il reddito pro capite dell'individuo più povero della società ed il reddito medio della stessa, rapporto che, evidentemente, non può superare l'unità ($0 < \beta < 1$).
Il coefficiente di Gini è dato da:

$$G = (1 - \beta)/(1 + \beta) \quad \text{cui corrisponde l'inversa:} \quad \beta = (1 - G)/(1 + G) \qquad (3)$$

Le x_1, x_2 dovranno dunque essere associate in modo che sia:

$$R_{M1}\beta_1(1 - x_1)^{\beta 1-1} = R_{M2}\beta_2(1 - x_2)^{\beta 2-1} \qquad (4)$$

Qui di seguito abbiamo sviluppato l'analisi per il solo caso particolare $G_1 = G_2$, $\beta_1 = \beta_2 =\beta$ che è, dal punto di vista matematico, notevolmente meno laborioso ed è comunque istruttivo in quanto permette di isolare l'effetto dovuto alla sola disparità dei redditi medii pro capite.
In tal caso la (4) si semplifica nella:

$$R_{M1}(1 - x_1)^{\beta-1} = R_{M2}(1 - x_2)^{\beta-1}$$

da cui:

$1 - x_2 = k(1 - x_1)$ con $k = (R_{M1} / R_{M2})^{1/(\beta-1)} = 1/[(R_{M1} / R_{M2})^{1/(1-\beta)}]$ (4')

dove, ponendo $R_{M1} / R_{M2} = \rho$ e tenendo conto che $\beta < 1$, $1 - \beta > 0$, possiamo anche scrivere:

$k = 1/\rho^{1/(1-\beta)}$

Sviluppando i calcoli, per il cui dettaglio si rimanda all'Appendice, si constata che la curva di Lorenz complessiva non corrisponde ad una nuova distribuzione di Pareto, anzi non è neanche rappresentata da un'unica funzione matematica, ma bensì da due, ognuna valida in un diverso intervallo di valori della x. Su questa base è comunque possibile calcolare il coefficiente di Gini, dato dalla relazione (12) che qui ripetiamo:

$$G = 1 - 2B =$$
$$= 1 - 2R_2N_2/RN\,[1 - k - (1 - k^{1+\beta})/(1 + \beta)] - [1 - (R_1 + R_2\, k^{\beta})/R(1 + \beta)]\; 2(N_1 + kN_2)/N \qquad (12)$$

Se ad esempio i due paesi hanno popolazioni eguali $N_1 = N_2 = 0{,}5N$, per cui:
$\rho = R_{M1} / R_{M2} = (R_1/N_1)/(R_2/N_2) = R_1/R_2$, la (12) diventa:

$$G = 1 - [1 - k - (1 - k^{1+\beta})/(1 + \beta)]/(1 + \rho) - [1 - (\rho + k^{\beta})/(1 + \rho)(1 + \beta)](1 + k)$$

da cui, ponendo, sempre a titolo di esempio, G = 0,333 (β = 0,5) si ricava il coefficiente di Gini in funzione di ρ (vedi grafico di **Fig.1**).

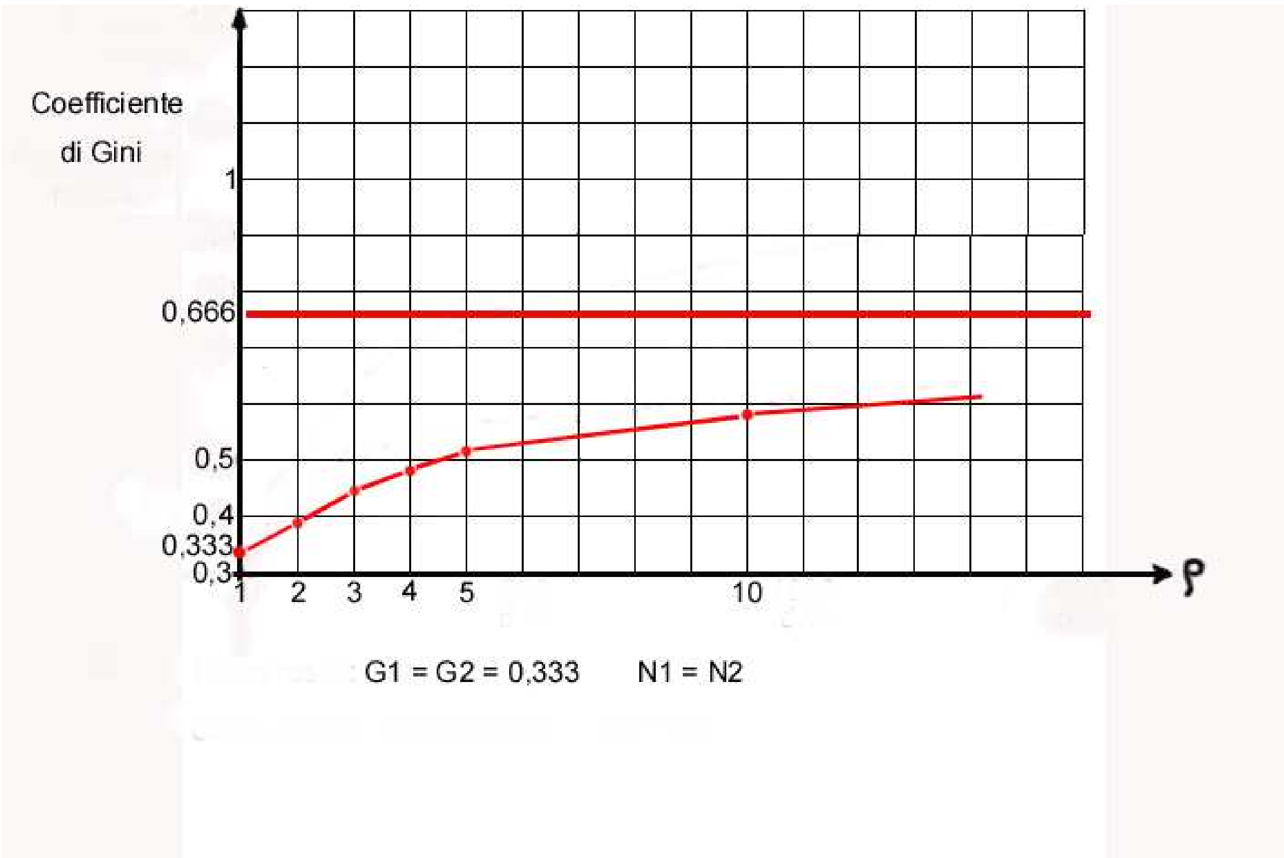

Fig.1: Coefficiente di Gini composito in funzione di ρ

Può essere interessante domandarsi, a titolo di esempio, quale può essere il coefficiente di Gini dell'Unione Europea (27 paesi) presa come un tutto; i dati in mio possesso non mi permettono un calcolo molto preciso, ma la schematizzazione che segue può comunque avere un valore indicativo.
I 27 paesi possono essere divisi in due gruppi come segue:
- Gruppo A di 14 paesi con una popolazione complessiva di circa 360 milioni di abitanti, che comprende tutte le maggiori economie, con redditi medii pro capite un po' superiori ai 20.000€ (la sola eccezione significativa è la Spagna che non è comunque molto al di sotto di tale valore).
- Gruppo B di 13 paesi con una popolazione complessiva di 115 milioni di abitanti, con redditi medii pro capite variabili da 5.000 a 15.000€ (1), fra cui quelli demograficamente più pesanti sono Polonia e Romania
Assumeremo per il primo gruppo un unico reddito medio pro capite di 21.000€ e per il secondo gruppo un valore baricentrico di 10.000€; i coefficienti di Gini del primo gruppo sono più o meno noti e sembrano essere abbastanza ben raggruppati attorno ad un valore di 0,33 (2), quelli del secondo gruppo invece non mi sono noti e quindi, in mancanza di meglio, assumerò per essi lo stesso valore.
In definitiva la nostra schematizzazione comporta quindi:
- Un paese 1 con N_1 = 360 milioni, R_{M1} = 21.000€, G_1 = 0,33
- Un paese 2 con N_2 = 115 milioni, R_{M2} = 10.000€, G_2 = 0,33
Svolgendo i calcoli in base alla formula (12) si trova per il sistema composito un valore di G pari a 0,358; le disparità di reddito fra i due gruppi di paesi provocano quindi un aumento del coefficiente di Gini di circa 2,5 punti percentuali.
Questo fatto può giustificare, almeno in parte, la differenza che di solito si riscontra fra i coefficienti di Gini di uno qualsiasi dei maggiori paesi europei e quello degli Stati Uniti; per la verità per il G di questi ultimi trovo, in numeri diversi dello stesso giornale (l'Economist), per uno stesso periodo (2005 circa), valori notevolmente diversi, da 0,38 a 0,47, il ché mi fa pensare che, sotto questo riguardo, le statistiche in circolazione non siano troppo affidabili; è evidente che l'effetto che abbiamo considerato potrebbe spiegare almeno parte della differenza nel primo caso, ma sarebbe pressoché trascurabile nel secondo.

Note:

1. I dati relativi alle popolazioni ad ai redditi pro capite sono desunti dall'atlante Zanichelli del 2005.
2. Questi dati provengono dall'Economist e dovrebbero essere validi per l'anno 2005.

Possibilità di iterazione

Per quanto riguarda la possibilità, auspicata all'inizio, di procedere iterativamente nel caso di più di 2 paesi, i risultati trovati non sono però troppo favorevoli, poiché, come già sottolineato, la curva di Lorenz del sistema composito non corrisponde più ad una distribuzione di Pareto.
Questo risultato negativo è anzi più generale; se esaminiamo da vicino il procedimento sviluppato in Appendice, ci rendiamo infatti conto del fatto che, anche partendo, per i due paesi originari, da una qualsiasi altra legge matematica, ci troveremmo infine ad avere una curva di Lorenz composita formata da due spezzoni, con leggi matematiche diverse fra loro e da quella di partenza.
Ritornando alla distribuzione di Pareto viene comunque spontaneo domandarsi di quanto sbaglieremmo se prendessimo per il sistema composito, anziché la curva in due spezzoni trovata, la distribuzione di Pareto "equivalente", laddove il criterio di equivalenza, evidentemente, non può essere dato che dall'uguaglianza del coefficiente di Gini; se trovassimo che lo scostamento è modesto, potremmo, quanto meno in via approssimata, usare tale distribuzione equivalente per procedere alle desiderate iterazioni successive.
Prendiamo il caso particolare già considerato, caratterizzato da $G_1 = G_2 = 0{,}333$ ($\beta = 0{,}5$), $N_1 = N_2 = 0{,}5N$; si tratta di un caso particolarmente favorevole perché è possibile dimostrare che se e solo se le popolazioni sono uguali, le derivate dei due spezzoni della curva composita nel punto di giunzione assumono lo stesso valore; in caso contrario, quindi, la curva composita presenterebbe una discontinuità a livello di derivata prima.
Se prendiamo $\rho = R_{M1} / R_{M2} = 5$ ($k = 0{,}04$), dai calcoli precedentemente fatti risulta, per il sistema composito, $G = 0{,}51$ e segue quindi dalla (3):

$$\beta_e = (1 - G)/(1 + G) = 0{,}324$$

dove β_e è evidentemente il valore di β che dà luogo alla distribuzione di Pareto equivalente. La relativa curva di Lorenz è data quindi da:

$$y = 1 - (1 - x)^{0{,}324}$$

il cui andamento, calcolato per punti, risulta dalla curva rossa di **Fig.2**.
La curva di Lorenz reale del sistema composito, calcolata per punti in base alle (5) e (9) (vedi Appendice), è invece rappresentata in blu.
Come si vede le due curve, pur essendo caratterizzate dallo stesso coefficiente di Gini, indicano distribuzioni del reddito notevolmente differenziate; si potrebbe dire che la curva reale manifesta un maggior grado di diseguaglianza dal momento che, ad esempio, il quintile più povero gode di una frazione del reddito totale che è circa la metà di quella della distribuzione equivalente.
E' chiaro che, in questo caso, procedere ad iterazioni successive sulla base della distribuzione equivalente darebbe luogo ad approssimazioni molto grossolane; d'altra parte quando ρ si riduce, le due curve tendono ad avvicinarsi fra loro ed alla curva nera, corrispondente alla distribuzione di Pareto ipotizzata per entrambe le società originarie, con cui vanno ovviamente a coincidere per $\rho = 1$.
Sembra quindi ragionevole pensare che la distribuzione di Pareto equivalente possa dare luogo a gradi di approssimazione accettabili quando il rapporto ρ fra i redditi medi pro capite delle due società originarie non è troppo elevato, per esempio, indicativamente, quando non supera 3; una valutazione più puntuale di questa possibilità resta però rimandata ad ulteriori approfondimenti.
Da un punto di vista generale il fatto che le leggi matematiche di distribuzione del reddito, di Pareto o di altri, non diano luogo a leggi dello stesso tipo nella composizione, ci dice forse qualcosa sulla loro natura; la loro eventuale validità per una o più società sembra infatti da attribuire all'effetto che l'esistenza più o meno lunga di uno stato, di un'unità politica, ha avuto sulla distribuzione del reddito nella sua popolazione.

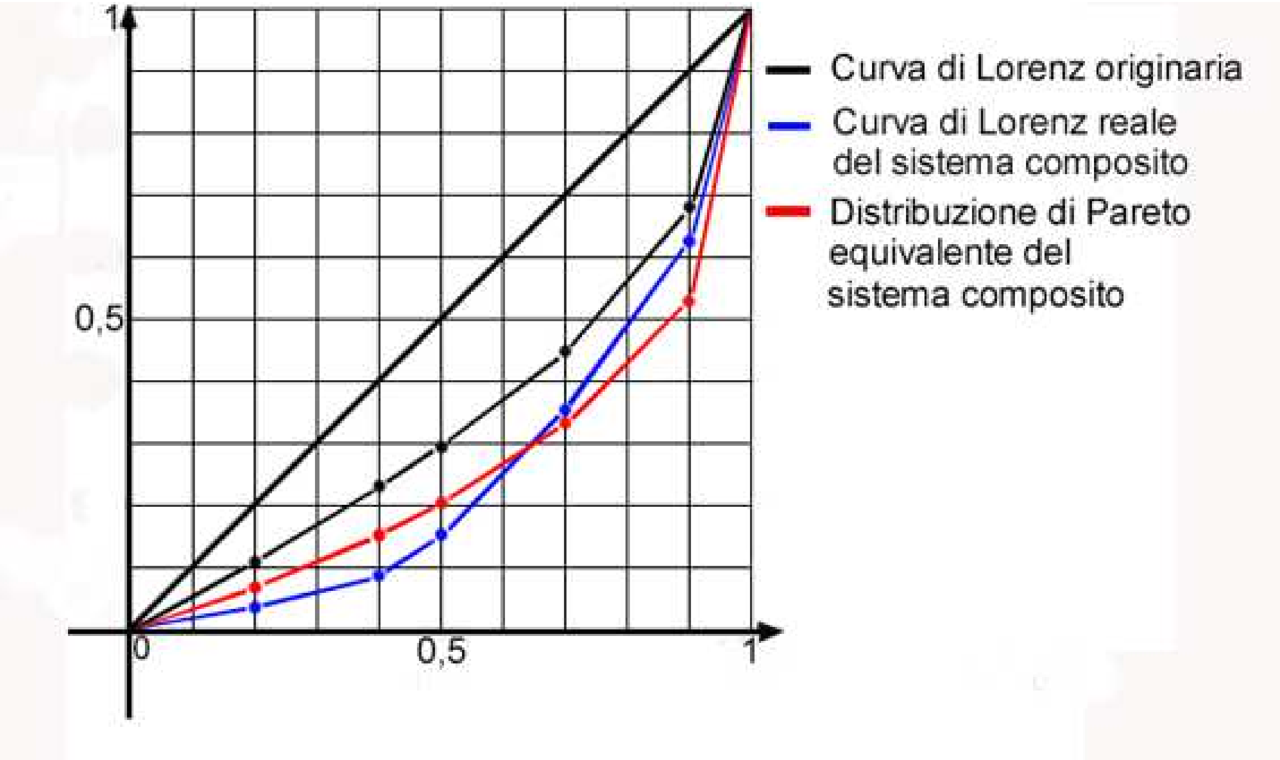

Fig.2: Curve di Lorenz

Appendice

Come abbiamo visto le frazioni di popolazione x_1,x_2, che devono essere associate per determinare la x del sistema complessivo, sono legate dalla relazione (4') (valida per $G_1 = G_2$), che può essere riscritta nella forma: $x_2 = kx_1 + 1 – k$ oppure: $x_1 = x_2/k – (1 – k)/k$ (4'')
Stabiliamo per convenzione di assegnare sempre l'indice 1 al paese contraddistinto dal reddito medio pro capite maggiore: sarà allora sempre: $R_{M1} /R_{M2} = \rho > 1$, $k < 1$
con k che tende a 0 per ρ che tende a infinito.
E' evidente che esisterà sempre una certa frazione x_2 della popolazione del paese più povero i cui redditi pro capite saranno così bassi che neanche il reddito pro capite minimo del paese 1 potrà soddisfare la (4''); questo valore di x_2 corrisponde a $x_1 = 0$ ed è quindi dato da: $x_2 = 1 - k$
La corrispondente frazione della popolazione totale sarà chiaramente: $x = (1 – k) N_2/N$ e, per X compresa fra 0 e tale valore la y(x) sarà determinata dalla sola $y_2(x_2)$ e sarà quindi data da:

$$y' = R_2/R[1 – (1 – x_2)^\beta] = R_2/R[1 – (1 – xN/N_2)^\beta] \qquad (5)$$

Per determinare G del sistema composito dobbiamo calcolare l'area B sottesa dalla curva di Lorenz complessiva y = y(x) di cui una prima parte B_1 sarà quindi ottenuta prendendo l'integrale della (5):

$$\int y'dx = R_2/R[x + (1 – xN/N_2)^{1+\beta} N_2/N(1 + \beta)]$$

fra x = 0 e $x = (1 – k)N_2/N$, per cui:

$$B_1 = R_2N_2/RN [1 – k – (1 – k^{1+\beta})/(1 + \beta)] \qquad (6)$$

Per $x > (1 – k)N_2/N$ le x1,x2 e la x corrispondente sono invece legate dalla relazione:

$$Nx = N_1x_1 + N_2x_2 \qquad x = x_1N_1/N + x_2N_2/N \qquad (7)$$

Combinando questa relazione con (4'') otteniamo:

$$x = x_1(N_1 + kN_2)/N + (1 – k)N_2/N \qquad (8)$$
$$x = x_2(N_1 + kN_2)/kN + (1 – k)N_1/kN$$

e inversamente:

$$x_1 = xN/(N_1 + kN_2) - (1 – k)N_2/(N_1 + kN_2) \qquad (8')$$
$$x_2 = xkN/(N_1 + kN_2) - (1 – k)N_1/(N_1 + kN_2)$$

da cui constatiamo che $x_1 = 0$ per $x = (1 – k) N_2/N$, mentre per $x_1 = 1$ anche x_2 ed x sono uguali ad 1.
Nell'intervallo $(1 – k)N_2/N < x < 1$ la y(x) sarà quindi determinata da entrambi i paesi secondo la relazione:

$$y'' = [1 – (1 –x_1)^\beta] R_1/R + [1 – (1 –x_2)^\beta] R_2/R$$

da cui, sostituendo per x_2 in base alla (4') otteniamo:

$$y'' = [1 – (1 –x_1)^\beta] R_1/R + [1 – k^\beta(1 –x_1)^\beta] R_2/R$$

e poiché $R_1 + R_2 = R$:

$$y'' = 1 – (1 –x_1)^\beta(R_1 + R_2 k^\beta)/R$$

Infine, ponendo per comodità $x_1 = ax + b$, dove a e b hanno le espressioni che risultano dalla (8') e sostituendo, abbiamo:

$$y'' = 1 – (1 – ax - b)^\beta(R_1 + R_2 k^\beta)/R \qquad (9)$$

Si noti che, per $x = (1 – k)N_2/N$, si ha $y' = y'' = (1 – k^\beta)R_2/R$, il chè garantisce che la curva di Lorenz del sistema composito ha andamento continuo.
La seconda parte B_2 dell'area B che cerchiamo sarà quindi ottenuta prendendo, fra $x = (1 – k) N_2/N$ e $x = 1$, l'integrale della (9) che è:

$$x + (1 – ax - b)^{1+\beta}(R_1 + R_2 k^\beta)/Ra(1 + \beta)$$

Quindi, sostituendo ad a,b le loro espressioni, otteniamo:

$$B_2 = [1 - (R_1 + R_2 k^\beta)/R(1 + \beta)] (N_1 + kN_2)/N \qquad (10)$$

e infine, combinando la la (6) e la (10):

$$B = B_1 + B_2 = R_2N_2/RN [1 – k – (1 – k^{1+\beta})/(1 + \beta)] + [1 - (R_1 + R_2 k^\beta)/R(1 + \beta)] (N_1 + kN_2)/N \qquad (11)$$

Da cui:

$$G = 1 – 2B =$$
$$= 1 - 2R_2N_2/RN [1 – k – (1 – k^{1+\beta})/(1 + \beta)] - [1 - (R_1 + R_2 k^\beta)/R(1 + \beta)] 2(N_1 + kN_2)/N \qquad (12)$$

Avvertenza: l'analisi che precede non vale per $\beta = 1$ (reddito pro capite uguale per tutta la popolazione), dato che la (4) perde allora di significato, in corrispondenza del fatto che le due popolazioni non si mescolano mai: si tratta comunque di un caso banale, oltre che poco realistico, poiché la curva di Lorenz composita si riduce allora ad una spezzata con due segmenti di retta, con pendenze proporzionali ai rispettivi redditi (medii o pro capite, in questo caso sono uguali) e punto di giunzione corrispondente a $x = N_2/N$.

Piero Zattoni,   Forlì 2010