Rostra

rubrica a cura di Benito V. Frosini

IMPIEGO DELLE TABELLE DI COGRADUAZIONE PER LA DETERMINAZIONE DELL'INDICE PUNTUALE DI CONCENTRAZIONE Z(p)

Michele Zenga

Università degli Studi Milano.

1. INTRODUZIONE

Per il calcolo dell'indice puntuale di concentrazione Z(p) sono state proposte diverse procedure basate sull'interpolazione della funzione di ripartizione, ovvero del primo momento incompleto (Zenga 1984, 1985; Salvaterra 1987, 1990). Il criterio introduce però un elemento di arbitrarietà dovuto alla scelta del tipo di funzione interpolante. Tale arbitrarietà viene a cadere se si fa riferimento ad una definizione di "inversa" della funzione di ripartizione e del primo momento incompleto ampiamente utilizzata in questi studi. Tuttavia l'impiego diretto di tali "inverse" ai fini del calcolo di Z(p) risulta alquanto complesso per cui il loro utilizzo sarebbe risultato limitato se non si fosse trovato il modo per "snellire" i calcoli. Si mostrerà come sia possibile, ricorrendo alle tabelle di cograduazione, ricavare agevolmente i valori di Z(p).

2. SIMBOLOGIA

Si abbiano n redditi ordinati $0 \leq x_1 \leq x_2 \leq \ldots \leq x_i \leq x \leq \ldots \leq x_n > 0$. Si indichi con $\bar{x}$ la media aritmetica dei redditi. La frequenza relativa e la quota di reddito spettante ad x_i sono rispettivamente date da $f(x_i) = 1/n$ e $q(x_i) = x_i/n\,\bar{x}$; i = 1,2,...,n.

La funzione di ripartizione F(x) ed il primo momento incompleto Q(x) sono dati da

$$F(x) = (\text{numero di } x_i \leq x)/n; \tag{2.1}$$

$$Q(x) = \Big(\sum_{(x_i \leq x)} x_i\Big) / n\,\bar{x} \tag{2.2}$$

Le inverse di F(x) e di Q(x) sono definite come segue:

$$x(p) = F^{-1}(p) = \begin{cases} \inf \{x_i : F(x_i) \geq p\} & \text{per } 0 < p \leq 1 \\ \inf \{x_i : F(x_i) > 0\} & \text{per } p = 0; \end{cases} \tag{2.3}$$

$$x^*(p) = Q^{-1}(p) = \begin{cases} \inf \{x_i : Q(x_i) \geq p\} & \text{per } 0 < p \leq 1 \\ \inf \{x_i : Q(x_i) > 0\} & \text{per } p = 0 \end{cases} \tag{2.4}$$

Come misura puntuale di concentrazione è stato proposto (Zenga 1984) il rapporto

$$z(p) = 1 - x\,(p)/x^*(p) \tag{2.5}$$

che in forza della (2.3) e della (2.4) è possibile definire per qualsiasi p ε [0,1]. La funzione z(p) è a scalini con salti per quei valori di p in corrispondenza dei quali le funzioni x(p) e/o x* (p) cambiano valore, il che può avvenire rispettivamente per p=1/n, 2/n,...,n/n e per p=Q(x_1), Q(x_2),...Q(x_n).

3. LA TABELLA DI COGRADUAZIONE FRA X E X*

Sia X la variabile statistica che assume i valori $x_1,x_2,...,x_n$ con pesi relativi rispettivamente pari a 1/n, 1/n,..., 1/n. Sia X* la variabile statistica che assume i valori $x_1,x_2,...,x_n$ con pesi relativi rispettivamente pari a q(x_1), q(x_2) , ..., q(x_n).

Per ricavare i valori che assume Z(p) è possibile avvalersi della tabella di cograduazione fra X e X*(Leti 1983, pag. 539; Zenga 1988, pag. 236).

Si mostrerà la procedura con un esempio numerico.

Esempio 1. Si abbiano i valori x_1= 5, x_2= 8, x_3= 17, x_4=30 e x_5=40 e se ne ricavi la tabella di cograduazione fra X e X*.

I pesi di X sono uniformemente pari a 0,20, mentre quelli di X* sono: q(x_1)= 0,05; q(x_2)=0,08; q(x_3)=0,17; q(x_4)=0,30 e q(x_5)=0,40.

Per costruire la tabella di cograduazione (Tab.I) si predispone innanzi tutto una tabella a doppia entrata in cui nella riga e nella colonna madre si riportano i valori di X e X* e nella riga e nella colonna marginale rispettivamente i pesi di X e di X* come è indicato nel prospetto che segue. La generica casella sarà indicata con(i,j): l'indice i fa riferimento ad un valore di X e l'indice j ad un valore di X*; ovviamente i,j = 1, 2,...,n. Per riempire le caselle si procede come segue. Si parte dalla casella (1,1) e si scrive al suo interno il minore fra 1/n e q (x_1). Ovviamente esso coincide con q(x_1). Tale valore è pari a 0,05. Nelle altre caselle della riga intestata con x_1 non bisogna riportare (niente) altro in quanto il totale marginale è già stato raggiunto con il peso della prima casella. Si passa allora alla casella (1,2) ed in essa si può scrivere il peso 0,08 corrispondente a q(x_2). In questo modo non si è ancora saturato il peso della colonna intestata con x_1 che è pari a 0,20. Si passa quindi alla casella (1,3) ove si scrive il peso 0,07 che è il valore ancora necessario per saturare la colonna x_1. Si passa ora alla casella (2,3) in cui si scrive il peso 0,10 necessario per saturare $q(x_3)$. Ecc.

Tab. I: Tabella di cograduazione fra X e X*.

	$f(x_i)$	0,20	0,20	0,20	0,20	0,20	1.00
	40	–	–	–	0,20	0,20	0,40
	30	–	0,10	0,20	–	–	0,30
	17	0,07	0,10	–	–	–	0,17
x_j	8	0,08	–	–	–	–	0,08
	5	0,05	–	–	–	–	0,05
	X* / X	5	8	17	30	40	$q(x_j)$
			x_i				

I pesi (positivi) ricavati nelle caselle della tabella di cograduazione saranno indicati, d'ora in avanti, con $v_r(r=1,2,...,k)$. L'indice 1 spetta, nell'esempio considerato, alla casella (1,1), l'indice 2 spetta, alla casella (1,2), ecc. L'indice k spetta alla casella (n,n). È immediato rendersi conto che $n \leq k \leq 2n-1$. È possibile allora indicizzare le caselle della tabella di cograduazione come indicato nel prospetto sotto riportato (Tab. II) in cui: $p_r = \sum_{i=1}^{r} v_i$; $x(p_r)$ e $x^*(p_r)$ indicano rispettivamente i valori della variabile X e della variabile X* corrispondenti alla casella r $^{\underline{ma}}$ di cograduazione.

Tab. II: Prospetto per il calcolo di Z(p) e di x.

r	Vr	Pr	$x(p_r)$	$x^*(p_r)$	$Z(p_r)$	$Z(p_r).Vr$
1	0,05	0,05	5	5	0,000	0,000
2	0,08	0,13	5	8	0,375	0,030
3	0,07	0,20	5	17	0,706	0,049
4	0,10	0,30	8	17	0,529	0,053
5	0,10	0,40	8	30	0,733	0,073
6	0,20	0,60	17	30	0,433	0,087
7	0,20	0,80	30	40	0,250	0,050
8	0,20	1,00	40	40	0,000	0,000
						0,342

È immediato rendersi conto, che per l'esempio in esame:

1) per $0 \leq p \leq 0{,}05 = Q(x_1) = p_1$ si ha $x(p) = 5$ e $x^*(p) = 5$;

2) per $0{,}05 < p \leq 0{,}13 = Q(x_2) = p_2$ si ha $x(p)=5$ e $x^*(p) = 8$;

3) per $0{,}13 < p \leq 0{,}20 = F(x_1) = p_3$ si ha $x(p) = 5$ e $x^*(p) = 17$;

:

8) per $0{,}80 < p \leq 1 = F(x_5) = Q(x_5) = p_8$ si ha $x(p) = 40$ e $x^*(p) = 40$.

Si è così constatato che con la procedura indicata è possibile ricavare Z(p) per ogni $0 \leq p \leq 1$.

Per ricavare il valore dell'indice sintetico ξ basta sommare i prodotti $Z(p_r)\cdot V_r$. Nel caso considerato si ottiene $\xi = 0,342$. I valori $Z(p_r)$ ricavati nel prospetto precedente possono essere riportati in un grafico (fig 1) tenendo presente che

$$Z(p) = \begin{cases} Z(p_1) & \text{per} \quad 0 \le p \le p_1 \\ Z(p_2) & \text{per} \quad P_1 < p \le p_2 \\ \vdots & \\ Z(p_k) & \text{per} \quad p_{k-1} < p \le 1. \end{cases}$$

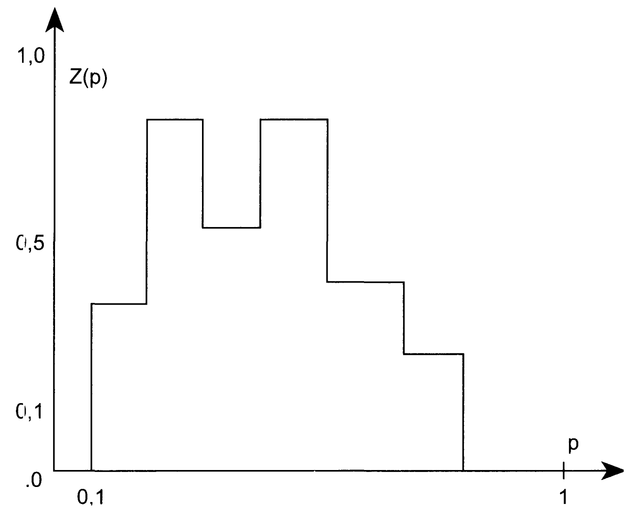

Fig. 1: Diagramma Z(p) per la distribuzione dell'esempio 1.

Ecco ora alcune particolarità sul grafico di Z(p).

Innanzi tutto si rileva che nel caso di una distribuzione di una variabile discreta, vi è sempre un intervallo di p per p prossimo ad 1 per il quale $Z(p) = 0$; trattasi dell'intervallo $(n - 1)/n < p \le 1$. Inoltre, se per p prossimo allo zero risulta $Z(p)= 0$, ciò significa che $x_1 > 0$. In questo caso $Z(p) = 0$ per $0 \le p \le Q(x_1)$.

Conviene ora esaminare il caso di una distribuzione di frequenze

$$X \equiv \begin{cases} x_1, x_2, \ldots, x_j, \ldots, x_s \\ n_1, n_2, \ldots, n_j, \ldots, n_s. \end{cases}$$

Da essa si ricavano le frequenze relative $f(x_j) = n_j/n$ e le quote relative di reddito $q(x_j) = n_j.x_j/n\ \bar{x}$, essendo:

$n = n_1 + n_2 + \ldots + n_s$ e $\bar{x} = (\sum_1^s x_j.n_j)/n$.

Per determinare i valori Z(p) basta ricavare la tabella di cograduazione fra la variabile X, che assume i valori x_j con pesi $f(x_j)$, e la variabile X*, che assume i valori x_j con pesi $q(x_j)$.

Esempio 2. Si abbia la seguente distribuzione di frequenza

valori x_j	1	4	8	11	18	Totali
frequenze n_j	9	4	3	3	1	20
prodotti $x_j.n_j$	9	16	24	33	18	100
pesi $f(x_j)$	0,45	0,20	0,15	0,15	0,05	1,00
pesi $q(x_j)$	0,09	0,16	0,24	0,33	0,18	1,00

Per ricavare la tabella di cograduazione (Tab. III) si procede come nell'esempio 1.

Tab. III: Tabella di cograduazione per la distribuzione dell'esempio 2.

	$f(x_i)$	0,45	0,20	0,15	0,15	0,05	1,00
	18	–	–	–	0,13	0,05	0,18
	11	–	0,16	0,15	0,02	–	0,33
x_j	8	0,20	0,04	–	–	–	0,24
	4	0,16	–	–	–	–	0,16
	1	0,09	–	–	–	–	0,09
X* / X		1	4	8	11	18	$q(x_j)$
			x_i				

Il prospetto per il calcolo di Z(p) è riportato nella tab. IV.

La tab IV indica che vi possono essere intervalli di p – a parte quelli estremi di cui si è già detto in precedenza – in cui Z(p) = 0. Ciò può accadere quando le frequenze relative si accentrano su pochi valori e può allora verificarsi che la tabella di cograduazione associ valori uguali.

Tab. IV: Prospetto per il calcolo di Z(p) e di x.

r	V_r	p_r	$x(p_r)$	$x^*(p_r)$	$Z(p_r)$	$Z(p_r).\ V_r$
1	0,09	0,09	1	1	0,000	0,000
2	0,16	0,25	1	4	0,750	0,120
3	0,20	0,45	1	8	0,875	0,175
4	0,04	0,49	4	8	0,500	0,020
5	0,16	0,65	4	11	0,636	0,102
6	0,15	0,80	8	11	0,273	0,041
7	0,02	0,82	11	11	0,000	0,000
8	0,13	0,95	11	18	0,389	0,051
9	0,05	1,00	18	18	0,000	0,000
						0,508

4. IL DIAGRAMMA Z(p) NEI CASI ESTREMI DI CONCENTRAZIONE

È ora utile considerare come si presenta il diagramma Z(p) in presenza di concentrazione nulla ed in presenza di massima concentrazione.

Concentrazione nulla.

Si ha concentrazione nulla se $x_1 = x_2 = \dots = x_n = \bar{x} > 0$.

In questo caso la tabella di cograduazione si presenta come indicato nella Tab. V e pertanto $Z(p) = 0$ per ogni $0 \le p \le 1$.

Si osservi che in questa situazione le funzioni F(x) e Q(x) coincidono ed assumono valore 0 per $x < \bar{x}$ e valore 1 per $x \ge \bar{x}$.

Massima concentrazione.

Si ha la concentrazione massima se

$x_1 = x_2 = \dots = x_{n-1} = 0$ e $x_n = n \,.\, \bar{x} > 0$.

La tabella di cograduazione in questo caso si presenta come indicato nella Tab. VI. Si ha allora $Z(p) = 1$ per $0 \le p \le (n-1)/n$ e, come di consueto, $Z(p) = 0$ per $(n-1)/n < p \le 1$. Consegue che in presenza della massima concentrazione $\xi = (n-1)/n$.

Tab.V: Tabella di cograduazione in presenza di concentrazione nulla.

f(X)	1	1
$\bar{x}$	1	1
X* / X	$\bar{x}$	q(x)

Tab. VI: Tabella di cograduazione in presenza di massima concentrazione.

f(x)	(n–1)/n	1/n	1
n $\bar{x}$	(n–1)/n	1/n	1
0	–	–	0
X* / X	0	n $\bar{x}$	q(X)

Ecco ora come si presenta il diagramma Z(p) nei due casi estremi considerati (Fig. 2 e Fig. 3).

5. IL DIAGRAMMA Z(p) IN PRESENZA DI VALORI ELEVATI DI n

La trattazione svolta ha permesso di definire in modo semplice e rigoroso gli indici Z(p) e ξ. In questo modo sarà possibile studiarne più agevolmente le proprietà. Nelle applicazioni reali il valore di n è solitamente molto elevato e per il calcolo dei due indici bisogna distinguere due casi.

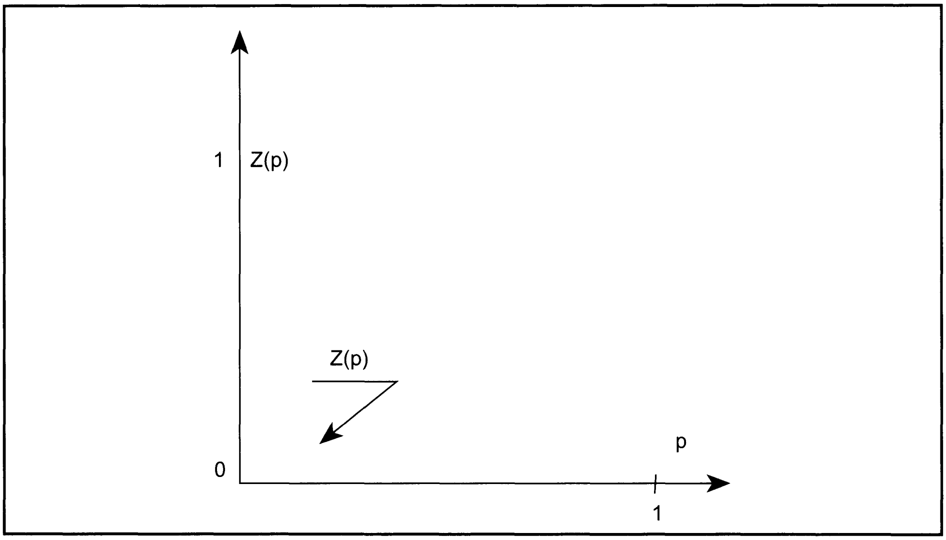

Fig. 2: Diagramma Z(p) in presenza di equiripartizione.

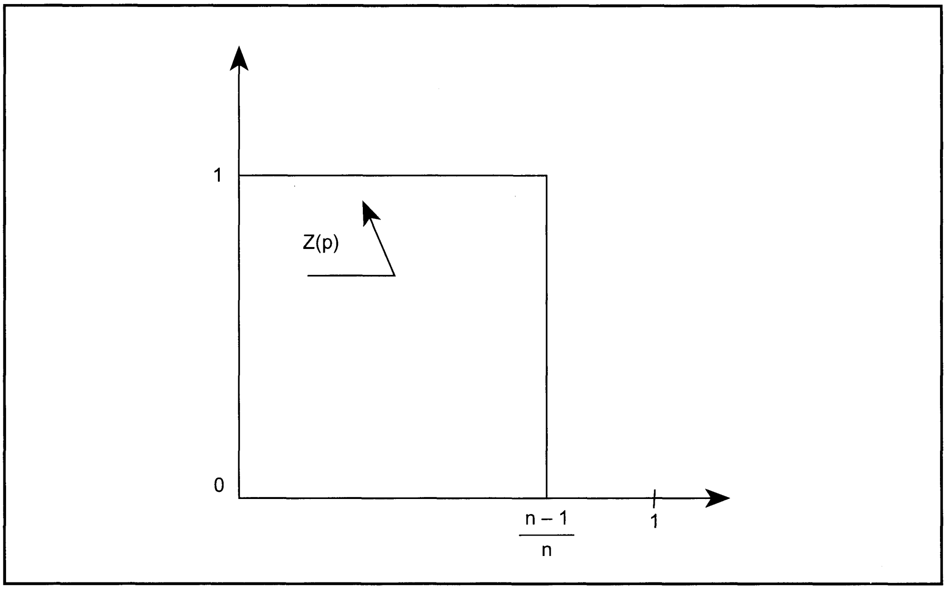

Fig. 3: Diagramma Z(p) in caso di massima concentrazione.

Si conoscono i dati elementari

In questo caso il ricorso alla tabella di cograduazione è molto oneroso. Una buona approssimazione del diagramma Z(p) si può ottenere scegliendo un opportuno insieme di valori p_i e, ricorrendo alle (1.1) e (1.2), ricavare $x(p_i)$ e $x^*(p_i)$. Da questi si ricaveranno le misure $Z(p_i)$ con le quali si potrà disegnare il grafico di

Z(p). Questa procedura è stata già impiegata molte volte (T. Salvaterra 1987, 1990).

Non si conoscono i dati elementari

I dati sono raggruppati in classi e le tabelle spesso riportano sia le frequenze di classe sia la somma dei redditi della classe. Per pervenire al grafico Z(p) si può procedere in vari modi, ciascuno dei quali però presenta sempre un certo grado di arbitrarietà. Se le classi sono numerose e ben "spaziate" si dovrebbero avere valori di Z(p) assai prossimi a quelli che si otterrebbero se si conoscessero i valori dei singoli redditi [Zenga 1984; Salvaterra 1987, 1990].

6. CONCLUSIONI

Con l'ausilio della tabella di cograduazione è possibile ricavare agevolmente i valori dell'indice di concentrazione puntuale Z(p), $0 \leq p \leq 1$, nonché dell'indice sintetico ξ.

Vengono così a cadere alcune arbitrarietà presenti nei procedimenti per il calcolo di Z(p) che fanno riferimento alla interpolazione della funzione di ripartizione e del primo momento incompleto. Il ricorso alla tabella di cograduazione permette inoltre di studiare più agevolmente le caratteristiche di Z(p) e di ξ.

La dottoressa A. Gnudi ha predisposto un programma di calcolo che fornisce: la tabella di cograduazione, il prospetto per il calcolo di Z(p) e di ξ ed il grafico di Z(p).

BIBLIOGRAFIA

Leti G.: *Statistica descrittiva,* Bologna, Il Mulino, 1983.

Salvaterra T.: *Analisi comparata dei procedimenti di calcolo dei rapporti di concentrazione Z(p) e dell'indice di concentrazione ξ di Zenga.* In La distribuzione personale del reddito: problemi di formazione, di ripartizione e di misurazione, ed. M. Zenga, Vita e Pensiero, Milano, 1987.

Salvaterra T.: *Comparisons Among Concentration Curves And Indexes In Some Empirical Distributions.* In Income and Wealth distribution, inequality and poverty, eds C. Dagum – M.Zenga, Springer–Verlag, Berlin–Heidelberg, 1990.

Zenga M.: *Proposta per un indice di concentrazione basato sui rapporti fra quantili di popolazione e quantili di reddito.* Giornale degli Economisti e Annali di Economia, pp. 301–326; 1984.

Zenga M.: *Un secondo indice di concentrazione basato sui rapporti fra quantili di reddito e quantili di popolazione.* Rivista di Statistica Applicata, n. 3, pp.143–154; 1985.

Zenga M.: *Introduzione alla statistica descrittiva.* Milano, Vita e Pensiero, 1988.

SUMMARY

The distribution function F(x) and the first incomplete moment Q(x) of a variable X which assumes the n ordered values $0 \leq x_1 \leq \ldots \leq x_i \leq \ldots \leq x_n > 0$ *are respectively given by* $F(x) = (\#\,(x_i \leq x))/n$ *and* $Q(x) = \sum_{(x_i \leq x)} x_i \,/\, \sum_1^n x_i$.

The inverse functions of F(x) and Q(x), for $0 < p \leq 1$, *are defined as* $F^{-1}(p) = x(p) = \inf \{x_i : F(x_i) \geq p\}$ *and* $Q^{-1}(p) = x^*(p) = \inf \{x_i : Q(x_i) \geq p\}$.

Zenga has recently (1984) proposed the point concentration measure

$Z(p) = 1 - x(p)/x^*(p)$.

The direct computation of Z(p) as a function of x(p) and x(p) is quite cumbersome.*

This paper shows that the values of the point concentration function Z(p) can be easily obtained using the cograduation table between the variables X and X, where X* takes the n ordered values* x_i *with relative weights* $q(x_i) = x_i / \sum_1^n x_i$.