Giancarlo Bettuzzi

Considerazioni su un indice di correlazione parziale tra graduatorie

Serie Ricerche 2001, n.1

Dipartimento di Scienze Statistiche "Paolo Fortunati"
Università degli studi di Bologna

Finito di stampare nel mese di Aprile 2001
presso le Officine Grafiche Tecnoprint
Via del Legatore 3, Bologna

1. Nel 1942 Kendall , in tema di correlazione parziale tra graduatorie, propose il *partial rank correlation*

$$\tau_{12.3} = \frac{\tau_{12} - \tau_{13} \cdot \tau_{23}}{\left[(1-\tau_{13}^2)(1-\tau_{23}^2)\right]^{\frac{1}{2}}} \tag{1}$$

espresso in funzione dell' indice di cograduazione τ ; com'è noto, il *partial rank correlation* è interpretabile alla stregua di un particolare coefficiente di associazione proposto da Yule (1912) relativamente ad una tabella 2×2 , e assume una sorprendente rassomiglianza formale col coefficiente di correlazione parziale. Nel commentare il risultato Kendall accenna ad alcune peculiari caratteristiche di questo indice quando annota che *"There was no reason to expect that partial τ , which is a pure function of disarrangements in rankings and is not expressible algebraically in terms of the ranks, should bear any analogy with the partial correlation of variates; but since it does so, we are evidently fortified in regarding partial τ as a convenient measure of rank correlation."*

Sempre nel corso del medesimo lavoro Kendall richiama anche, quale possibile espressione di un indice di cograduazione parziale, una struttura formale analoga alla precedente in cui, però, al posto del coefficiente τ figura il coefficiente di cograduazione ρ di Spearman

$$\rho_{12.3} = \frac{\rho_{12} - \rho_{13} \cdot \rho_{23}}{\left[(1-\rho_{13}^2)(1-\rho_{23}^2)\right]^{\frac{1}{2}}} \tag{2}$$

ma quasi subito manifesta la sua perplessità nei confronti di questa formula quando afferma che *" There is clearly very little justification for such a procedure, and it is far from easy to explain just what $\rho_{12.3}$ means."* In verità l'Autore non esclude del tutto la possibilità di utilizzare questo indice e ciò appare evidente nel momento in cui sostiene che può essere impiegato per approssimare il valore del coefficiente di correlazione parziale, rilevando nel contempo che *" we arrive , not at a partial rank coefficient, but at an estimate of a partial product moment coefficient in a normal population."* Ed ancora nel 1964, quando presenta l'espressione del suo τ parziale nel volume *An Introduction to the Theory of Statistics* pubblicato assieme a Yule, ribadisce che "*No similar results are known for Spearman's* ρ ."

Lo scopo di questo nostro lavoro è quello di soffermarci ulteriormente sull'approfondimento dell'analisi dell'indice $\rho_{12,3}$, non tanto per ribadire i limiti già chiaramente evidenziati dalle critiche sviluppate dal Kendall, quanto piuttosto con l'intento di coglierne ulteriori aspetti caratterizzanti che, se da un lato confermano la incapacità di questo indice di proporsi come misura della *partial rank correlation* , dall'altro possono forse rivelarsi utili nel nostro tentativo di definire nuove espressioni di indici di correlazione parziale. E dal momento che lo strumento concettuale che ci guiderà nell' indagine è costituito dal complesso delle nozioni riconducibili alla teoria della concordanza del Gini, convenientemente estesa al caso dell'associazione di tre caratteri graduabili, riteniamo opportuno introdurre qui di seguito alcune notazioni che ci consentiranno di sviluppare il discorso sul tema.

2. In una serie di memorie, pubblicate tra il 1914 e il 1916, Corrado Gini delineò i fondamenti metodologici di una teoria della concordanza basata sulla nozione di rassomiglianza inerente a due caratteri che si associano nelle unità di un collettivo. Nella concezione del Gini la teoria della concordanza si snoda attraverso la distinzione tra confronti riferiti a distribuzioni di quantità che possono essere *intensità*, oppure *scostamenti* dalla media, oppure ancora, *variazioni* (scostamenti standardizzati) ; si sviluppa tecnicamente nella distinzione tra indici semplici e indici quadratici, implica la nozione di carattere contrario, ed è sorretta dalla definizione dei criteri di concordanza. Passa altresì per la dicotomia tra massimo relativo e massimo assoluto di concordanza generando indici di omofilia accanto ad indici di correlazione, ricollegabili alle molteplici istanze conoscitive che si presentano nei contesti della ricerca sostanziale. Limitando il discorso agli indici quadratici, dei quali unicamente ci occuperemo nel presente articolo, è noto che Gini, avendo riguardo alla nozione di massimo relativo di concordanza, pervenne all'indice quadratico di omofilia

$$ {}^{2}O = \frac{\sigma(X,Y)}{\sigma^{(1)}(X,Y)} \tag{3} $$

e a quello quadratico di eterofilia

$$ {}^{2}E = -\frac{\sigma(X,Y)}{\sigma^{(2)}(X,Y)} \tag{4} $$

validi sia per le intensità che per gli scostamenti e per le variazioni. Al numeratore delle due espressioni figura la covarianza determinata rispetto ai valori osservati mentre al denominatore della (3) è posta la covarianza definita in conformità all'ipotesi di cograduazione delle quantità dei due caratteri e al denominatore della (4) la covarianza calcolata coerentemente all'ipotesi di contrograduazione delle medesime quantità. Il segno meno, che figura nel secondo membro della (4), consegue dalla circostanza che prevede il numeratore negativo per ipotesi e il denominatore negativo per costruzione. Il primo indice assume valori nell'intervallo [0 ; 1] ed è idoneo a misurare la concordanza in senso stretto, mentre il secondo indice realizza valori nell'intervallo [- 1 ; 0] e misura la discordanza. La caratteristica di tali indici è quindi quella di misurare la concordanza, o la discordanza, rispetto al massimo relativo che, per definizione, è condizionato dalla configurazione specifica delle distribuzioni dei due caratteri. Sotto questo profilo, le distribuzioni ora menzionate si propongono come un dato del problema e costituiscono un vincolo per l'operazione di cograduazione, o di contrograduazione, essenziale per la determinazione del denominatore degli indici. In questo contesto è del tutto evidente che la tendenza ad associarsi delle quantità corrispondenti, per quanto forte possa essere, non può oltrepassare il massimo vincolato dall'invarianza delle distribuzioni effettive dei due caratteri. In circostanze diverse, queste distribuzioni possono invece configurarsi come il risultato della tendenza associativa e, pertanto, non costituire un vincolo alle modalità concrete dell'associazione. In tale caso è concettualmente possibile, ed opportuno, modificare le distribuzioni statistiche originarie affinchè possa verificarsi tanto il massimo assoluto di concordanza quanto il massimo assoluto di discordanza. A questo proposito Gini indicò le condizioni che debbono sussistere perché ciò si realizzi, identificandole nell'uguaglianza di quattro distribuzioni : le distribuzioni statistiche dei due caratteri considerati e quelle dei loro caratteri contrari. Nel contempo suggerì la soluzione formale del problema ravvisabile nella costruzione di una distribuzione media delle quattro distribuzioni appena indicate, che risulterà simmetrica (Gini 1916 b). Il riferimento a distribuzioni statistiche semplici, componenti della distribuzione statistica doppia, coincidenti con la distribuzione statistica media sopra specificata consentirà di costruire un modello associativo che, sul piano logico-formale, prevede la realizzazione dei massimi teorici in questione se, in particolare, si abbia l'avvertenza di seguire il Fortunati quando, nella ricerca rivolta ad una impostazione unitaria della definizione di distribuzione contraria, suggerisce di assumere quale costante di complementarietà il doppio della media aritmetica della distribuzione originaria perché ciò consente di attribuire alla distribuzione contraria una media aritmetica coincidente con quella della distribuzione originaria (Fortunati 1967). Proprio al riguardo della nozione di massimo assoluto di concordanza Gini diede l'espressione

$$ {}^{2}r_{xy} = \frac{\sigma(X,Y)}{\frac{1}{2}\left[\sigma^2(X)+\sigma^2(Y)\right]+\frac{1}{4}\left[D^2(X)+D^2(Y)+D^2(\overline{X})+D^2(\overline{Y})\right]} \tag{5} $$

dell'indice quadratico di correlazione tra le intensità. In esso $\sigma^2(X)$ e $\sigma^2(Y)$ indicano le varianze delle distribuzioni del carattere X e del carattere Y. Per quanto concerne gli altri simboli che figurano nell'espressione (5) occorre innanzitutto precisare che $\overline{X}$ e $\overline{Y}$ rappresentano, rispettivamente, le distribuzioni dei caratteri contrari di X e di Y; le notazioni $D^2(X)$, $D^2(Y)$, $D^2(\overline{X})$ e $D^2(\overline{Y})$ indicano i quadrati delle differenze tra la media di ogni distribuzione e la media delle medie delle quattro distribuzioni, le due assegnate e le rispettive distribuzioni contrarie.

Infine, com'è noto, Gini pervenne all'espressione

$$ {}_{(s)}^{2}r_{xy} = \frac{\sigma(X,Y)}{\frac{1}{2}\left[\sigma^2(X)+\sigma^2(Y)\right]} \tag{6} $$

dell'indice quadratico di correlazione tra scostamenti dalla media, mentre nei confronti delle variazioni (scostamenti standardizzati) determinò l'indice quadratico di correlazione

$$ {}_{(v)}^{2}r_{xy} = \frac{\sigma(X,Y)}{\sigma(X)\sigma(Y)} \tag{7} $$

coincidente con la già conosciuta espressione del coefficiente di correlazione di Bravais-Pearson.

Venendo ora alla concordanza tra le graduatorie di due caratteri ordinabili è del tutto opportuno rammentare, come lo stesso Gini evidenziò, la coincidenza tra il massimo assoluto e il massimo relativo di concordanza (Gini 1916 b). Infatti le due distribuzioni dei posti sono tra loro uguali e simmetriche con costante di simmetria pari al doppio della loro media aritmetica, e così pure le distribuzioni contrarie coincidono con le rispettive distribuzioni originarie dei posti se, aderendo alle indicazioni del Fortunati (1967), si assume per costante di complementarietà il doppio della media aritmetica della distribuzione originaria. Tutto ciò evidentemente assicura la coincidenza del massimo relativo col massimo assoluto di concordanza e la conseguente identificazione dell'indice di omofilia con gli indici di correlazione giacché, in virtù della eguaglianza delle medie e delle varianze delle due graduatorie originarie e delle due rispettive graduatorie contrarie, gli indici (5), (6), (7) sono formalmente riconducibili ad una identica espressione che, come è noto, è quella del *rank correlation* ρ di Spearman. Sul piano formale, la coincidenza degli indici consentirebbe di attribuire a quella comune espressione, indifferentemente, il significato di indice di concordanza tra distribuzioni dei posti d'ordine, come pure quello di indici di concordanza tra distribuzioni di scostamenti dalla media dei posti oppure, ancora, tra distribuzioni di variazioni dei posti (scostamenti standardizzati) ma, nel contesto particolare e specifico della misura della cograduazione, si è spontaneamente indotti a porre la questione solo in termini di confronti tra le distribuzioni dei posti di graduatoria dal momento che non si può in alcun modo giustificare il confronto tra le distribuzioni degli scostamenti o delle variazioni dei posti.

3. Si considerino ora tre caratteri graduabili X_1, X_2, X_3 che si associano nelle N unità di un collettivo. La terna (p_{1i}, p_{2i}, p_{3i}) rappresenta i posti d'ordine che l' *i-esimo* individuo occupa, rispettivamente, nella graduatoria p_1 di X_1, nella graduatoria p_2 di X_2 e nella graduatoria p_3 di X_3. Come

è noto, le medie aritmetiche delle distribuzioni dei posti realizzano tutte il valore

$$M_1(p_1) = M_1(p_2) = M_1(p_3) = \frac{N+1}{2} \tag{8}$$

mentre le varianze valgono

$$\sigma^2(p_1) = \sigma^2(p_2) = \sigma^2(p_3) = \frac{N^2-1}{12}. \tag{9}$$

Coerentemente al procedimento adottato da Yule nella costruzione del coefficiente di correlazione parziale tra i caratteri quantitativi X_1 e X_2, dopo che sia stata eliminata l'influenza di X_3, si tratta di detrarre dalle graduatorie p_1 e p_2 quella parte dovuta a p_3 che corrisponde, rispettivamente, alla regressione lineare di p_1 su p_3 e di p_2 su p_3. Si ottengono, in tale modo, le due seguenti distribuzioni p_{13} e p_{23} :

$$p_{13,i} = p_{1i} - b_{13} \cdot p_{3i} \tag{10}$$

$$(i=1,2,\dots\dots,N)$$

$$p_{23,i} = p_{2i} - b_{23} \cdot p_{3i} \tag{11}$$

in cui figurano i coefficienti di regressione lineare

$$b_{13} = \frac{\sigma(p_1,p_3)}{\sigma^2(p_3)} \tag{12}$$

$$b_{23} = \frac{\sigma(p_2,p_3)}{\sigma^2(p_3)}. \tag{13}$$

Nei confronti delle distribuzioni (10) e (11) è agevole accertare i seguenti valori delle medie aritmetiche

$$M_1(p_{13}) = M_1(p_1) - b_{13} \cdot M_1(p_3) \tag{14}$$

$$M_1(p_{23}) = M_1(p_2) - b_{23} \cdot M_1(p_3) \tag{15}$$

quelli delle varianze

$$\sigma^2(p_{13}) = \sigma^2(p_1)(1-\rho_{13}^2) \tag{16}$$

$$\sigma^2(p_{23}) = \sigma^2(p_2)(1-\rho_{23}^2) \tag{17}$$

e della covarianza

$$\sigma(p_{13},p_{23}) = \sigma(p_1)\sigma(p_2)(\rho_{12} - \rho_{13} \cdot \rho_{23}) \tag{18}$$

espressi in funzione del coefficiente di cograduazione di Spearman.

Infine, soffermiamoci un attimo su una peculiare caratterizzazione delle distribuzioni (10) e (11) non priva di conseguenze sui possibili sviluppi metodologici che possono essere conferiti alla costruzione dell'indice di correlazione parziale tra graduatorie : chiaramente emerge che le distribuzioni in questione, non essendo necessariamente uguali e simmetriche, non consentono, in generale, che si realizzi la coincidenza tra il massimo assoluto e il massimo relativo di concordanza, come invece si è constatato in precedenza al riguardo degli indici quadratici di cograduazione definiti da Gini nei confronti di due soli caratteri. Ciò, ovviamente, consente di costruire alternativamente indici quadratici sia di correlazione parziale che di omofilia parziale tra graduatorie. Rispetto alla prima categoria di indici possono essere considerate le precedenti espressioni (5), (6) e (7) con l'avvertenza di sostituire le distribuzioni dei caratteri, che in esse figurano, con le distribuzioni (10) e (11) ottenute dalle graduatorie di X_1 e di X_2 attraverso l'eliminazione dell'effetto prodotto dalla graduatoria di X_3 ; per gli indici di omofilia parziale invece potranno essere considerate le precedenti formule (3) e (4).

4. Cominciamo col considerare l'espressione (7) del coefficiente di correlazione tra variazioni (o scostamenti standardizzati) riscrivendola in funzione delle distribuzioni (10) e (11) ; in virtù della sostituzione indicata potremo porre

$$ {}_{(v)}^{2}r_{12,3} = \frac{\sigma(p_{13}, p_{23})}{\sigma(p_{13}).\sigma(p_{23})} \quad ; \qquad (19) $$

se poi si considerano le espressioni (18), (16) e (17), si giunge rapidamente al risultato

$$ {}_{(v)}^{2}r_{12,3} = \frac{\rho_{12} - \rho_{13}.\rho_{23}}{\left[(1-\rho_{13}^2)(1-\rho_{23}^2)\right]^{\frac{1}{2}}} \qquad (20) $$

che coincide con l'espressione (2) criticata dal Kendall. L'indice (20) allora si qualificherebbe come indice quadratico di correlazione parziale tra distribuzioni di variazioni dei posti di graduatoria.
Se invece si considera l'espressione (6) del coefficiente di correlazione tra scostamenti, con l'avvertenza di sostituire le distribuzioni dei caratteri che in essa figurano con le sopra menzionate distribuzioni (10) e (11), risulterà

$$ {}_{(s)}^{2}r_{12,3} = \frac{\sigma(p_{13}, p_{23})}{\frac{1}{2}\left[\sigma^2(p_{13}) + \sigma^2(p_{23})\right]} \qquad (21) $$

quindi, ponendo mente ancora una volta, alle espressioni (18), (16) e (17) e tenuto conto della uguaglianza tra $\sigma^2(p_1)$ e $\sigma^2(p_2)$ di cui alla (9), si perviene alla espressione

$$ {}_{(s)}^{2}r_{12,3} = \frac{\rho_{12} - \rho_{13}.\rho_{23}}{\frac{1}{2}\left[(1-\rho_{13}^2) + (1-\rho_{23}^2)\right]} \qquad (22) $$

di un indice quadratico di correlazione parziale tra distribuzioni di scarti dalla media dei posti. Anche la (22), al pari della (20), risulta espressa in funzione dell'indice ρ di cograduazione di Spearman.

Ovviamente vale la relazione ${}_{(v)}^{2}r_{12,3} \geq {}_{(s)}^{2}r_{12,3}$, particolarmente evidente se si osservano la (19) e la (21), dal momento che i due indici hanno l'identico numeratore – la covarianza tra le distribuzioni (10) e (11) dei residui – mentre al denominatore figurano, rispettivamente, la media geometrica e la media aritmetica delle varianze delle stesse distribuzioni (10) e (11).
Ciò che ora necessita di fare è la ricerca dell' espressione di un indice di correlazione parziale che rifletta la spontaneità di un atteggiamento di ricerca che assuma le distribuzioni dei posti di graduatoria (e non quelle degli scostamenti dalla media dei posti o delle variazioni dei posti) quali elementi iniziali caratterizzanti della costruzione. La soluzione che intendiamo proporre ha come punto di riferimento generale la formula dell'indice riprodotto nella (5).

5. Alla luce di queste ultime precisazioni si tratta allora di considerare l'indice (5) espresso in funzione delle distribuzioni (10) e (11) ; ciò consente di scrivere l'espressione equivalente dal punto di vista formale

$$ {}^{2}r_{12,3} = \frac{\sigma(p_{13}, p_{23})}{\frac{1}{2}[\sigma^2(p_{13}) + \sigma^2(p_{23})] + \frac{1}{4}[D^2(p_{13}) + D^2(p_{23}) + D^2(\overline{p_{13}}) + D^2(\overline{p_{23}})]} \qquad (23) $$

nei cui riguardi è opportuno svolgere alcune specificazioni, in particolare per quanto concerne il secondo addendo del denominatore. Con i simboli $\overline{p_{13}}$ e $\overline{p_{23}}$ abbiamo rappresentato le distribuzioni contrarie delle distribuzioni p_{13} e p_{23} . Queste distribuzioni sono state costruite in base alle indicazioni del Fortunati che suggerisce di assumere le costanti di complementarietà $K_1 = 2M_1(p_{13})$ e $K_2 = 2M_1(p_{23})$; tale scelta comporta che le medie delle distribuzioni contrarie siano uguali alle medie delle rispettive distribuzioni originarie

$$ M_1(p_{13}) = M_1(\overline{p_{13}}) \qquad (24) $$

$$M_1(p_{23}) = M_1(\overline{p_{23}}) \qquad (25)$$

Il simbolo $D^2(p_{13})$ indica il quadrato della differenza tra la media della distribuzione p_{13} e la media M_1 delle quattro distribuzioni p_{13}, p_{23}, $\overline{p_{13}}$, $\overline{p_{23}}$. Analogo significato hanno i simboli $D^2(p_{23})$, $D^2(\overline{p_{13}})$ e $D^2(\overline{p_{23}})$. Le quantità che figurano al numeratore della (23) e quelle del primo addendo del denominatore sono già state determinate, pertanto ora si tratta di definire quelle del secondo addendo del denominatore. Iniziamo con la media M_1 delle quattro distribuzioni p_{13}, p_{23}, $\overline{p_{13}}$, $\overline{p_{23}}$:

$$M_1 = \frac{1}{4}\left[M_1(p_{13}) + M_1(p_{23}) + M_1(\overline{p_{13}}) + M_1(\overline{p_{23}})\right]$$

e, in virtù della (24) e della (25) risulta :

$$M_1 = \frac{1}{2}\left[M_1(p_{13}) + M_1(p_{23})\right] \quad . \qquad (26)$$

Ora determiniamo la seguente quantità

$$D^2(p_{13}) = \left[M_1(p_{13}) - M_1\right]^2$$

che, in conseguenza della (26), si specifica nel seguente modo

$$D^2(p_{13}) = \frac{1}{4}\left[M_1(p_{13}) - M_1(p_{23})\right]^2 \quad . \qquad (27)$$

Con procedimenti analoghi si perviene alle quantità

$$D^2(p_{23}) = \frac{1}{4}\left[M_1(p_{23}) - M_1(p_{13})\right]^2 \qquad (28)$$

$$D^2(\overline{p_{13}}) = \frac{1}{4}\left[M_1(p_{13}) - M_1(p_{23})\right]^2 \qquad (29)$$

$$D^2(\overline{p_{23}}) = \frac{1}{4}\left[M_1(p_{23}) - M_1(p_{13})\right]^2 \qquad (30)$$

Siamo adesso in grado di determinare il secondo addendo del denominatore della (23) :

$$\frac{1}{4}\left[D^2(p_{13}) + D^2(p_{23}) + D^2(\overline{p_{13}}) + D^2(\overline{p_{23}})\right] = \frac{1}{4}\left[M_1(p_{13}) - M_1(p_{23})\right]^2 \qquad (31)$$

e, in virtù di quest'ultimo risultato, possiamo riproporre la (23) stessa nella seguente forma :

$$^2r_{12,3} = \frac{\sigma(p_{13}, p_{23})}{\frac{1}{2}\left[\sigma^2(p_{13}) + \sigma^2(p_{23})\right] + \frac{1}{4}\left[M_1(p_{13}) - M_1(p_{23})\right]^2} \qquad (32)$$

Se ora si considera la seguente trasformazione del secondo addendo del denominatore della (32), operata sulla base delle espressioni (14) e (15) e delle eguaglianze tra le medie di cui alla (8),

$$\frac{1}{4}\left[M_1(p_{13}) - M_1(p_{23})\right]^2 = \frac{1}{4}\left[M_1(p_3)(b_{23} - b_{13})\right]^2 \qquad (33)$$

e, contemporaneamente, si rammentano le espressioni (18), (16) e (17), è possibile riscrivere la (32) nel seguente modo

$$^2r_{12,3} = \frac{\sigma(p_1)\sigma(p_2)(\rho_{12} - \rho_{13}.\rho_{23})}{\frac{1}{2}\left[\sigma^2(p_1)(1-\rho_{13}^2) + \sigma^2(p_2)(1-\rho_{23}^2)\right] + \frac{1}{4}\left[M_1(p_3)(b_{23} - b_{13})\right]^2} \qquad (34)$$

Infine, per effetto delle eguaglianze evidenziate mediante la (9) e della relazione $M_1(p_3) = (N+1)/2$ espressa nella (8), si può agevolmente pervenire all'espressione

$$^2r_{12,3} = \frac{\rho_{12} - \rho_{13} \cdot \rho_{23}}{\frac{1}{2}\left[(1-\rho_{13}^2) + (1-\rho_{23}^2)\right] + \frac{3(N+1)}{4(N-1)}(\rho_{23} - \rho_{13})^2} \tag{35}$$

dell'indice quadratico di correlazione parziale tra le distribuzioni dei posti di graduatoria espresso in funzione dell'indice di cograduazione di Spearman.
Si può immediatamente constatare che quest'ultimo indice, in genere, assume valori inferiori a quelli realizzati dai due precedenti coefficienti espressi mediante le formule (20) e (22) giacchè, a parità di numeratore, presenta un denominatore più elevato. Infine, è appena il caso di rilevare che nell'ipotesi di indipendenza lineare tra le graduatorie p_1 e p_3, da un lato, e tra p_2 e p_3, dall'altro, $^2r_{12,3}$ coincide con l'indice ρ_{12} di cograduazione di Spearman.

6. Anche per le graduatorie è possibile definire indici di omofilia parziale al pari di quanto è stato fatto per le distribuzioni di caratteri quantitativi (Bettuzzi 1997 e 1998). Per realizzare lo scopo possiamo riproporre le espressioni (3) e (4) del Gini nei confronti delle fondamentali distribuzioni p_{13} e p_{23}, di cui alla (10) e (11). In tale modo avremo la seguente coppia

$$^2O_{12,3} = \frac{\sigma(p_{13}, p_{23})}{\sigma^{(1)}(p_{13}, p_{23})} \tag{36}$$

$$^2E_{12,3} = -\frac{\sigma(p_{13}, p_{23})}{\sigma^{(2)}(p_{13}, p_{23})} \tag{37}$$

di indici quadratici di omofilia parziale tra le graduatorie p_1 del carattere X_1 e p_2 del carattere X_2 eliminata l'influenza della graduatoria p_3 del carattere X_3. A numeratore degli indici figura la covarianza tra le distribuzioni p_{13} e p_{23} che è ipotizzata positiva nella (36) e negativa nella (37) ; al denominatore della (36) compare la covarianza nell'ipotesi di cograduazione tra p_{13} e p_{23}, che è positiva per costruzione, mentre nella (37) c'è la covarianza nell'ipotesi di controgradazione tra p_{13} e p_{23}, che è negativa per costruzione e ciò rende conto del segno negativo che è posto al secondo membro della (37) stessa. Il primo indice, pertanto, è idoneo a misurare la concordanza in senso stretto, con riferimento alla nozione di massimo relativo di concordanza, e assume valori nell'intervallo [0 ; 1] ; il secondo quantifica la discordanza e presenta valori nell'intervallo [-1 ; 0]. Tramite la relazione (18) i numeratori degli indici di omofilia possono essere espressi in funzione dei coefficienti di cograduazione di Spearman, operazione non realizzabile, invece, per quanto concerne i denominatori dal momento che la cograduazione, o la controgradazione, condotta nei riguardi delle distribuzioni p_{13} e p_{23}, per la costruzione degli indici relativi al massimo concreto di concordanza in senso lato, non comporta necessariamente che risultino cograduate, o controgradute, le distribuzioni p_1 con p_2, p_1 con p_3 e p_2 con p_3. Pertanto, se l'indice (35) di correlazione parziale tra graduatorie viene calcolato senza che debbano essere realmente costruite le distribuzioni (10) e (11), per gli indici di omofilia tra graduatorie queste ultime distribuzioni diventano fondamentali non solo dal punto di vista concettuale ma anche sul piano computazionale.
Pure in questo contesto è agevole rendere esplicita la relazione tra indice quadratico di omofilia parziale e indice quadratico di correlazione parziale tra graduatorie; infatti, se dividiamo sia il numeratore che il denominatore della (36) per la stessa quantità corrispondente al denominatore della (23), si perviene al rapporto

$$^2O_{12,3} = \frac{^2r_{12,3}}{^2r^{(1)}_{12,3}} \tag{38}$$

al cui denominatore figura l'indice di correlazione tra graduatorie nell' ipotesi di cograduazione di p_{13} con p_{23}. Analogamente, per l'indice (37) risulta il rapporto

$$ {}^{2}E_{12,3} = -\frac{{}^{2}r_{12,3}}{{}^{2}r^{(2)}_{12,3}} \qquad (39) $$

il cui denominatore è caratterizzato dall'espressione dell'indice di correlazione parziale tra graduatorie nell'ipotesi di contrograduazione delle distribuzioni p_{13} e p_{23}.

7. Determiniamo ora i valori che assumono gli indici sopra specificati nei confronti delle tre seguenti distribuzioni di posti d'ordine

p_1:	3	1	4	2	6	5
p_2:	4	2	1	6	3	5
p_3:	1	2	3	4	5	6

quelle stesse considerate dal Kendall nel suo articolo del 1942. Dal calcolo risulta:

$$ {}^{2}_{(v)}r_{12,3} = \frac{\rho_{12} - \rho_{13}\cdot\rho_{23}}{\left[(1-\rho_{13}^{2})(1-\rho_{23}^{2})\right]^{\frac{1}{2}}} = -0{,}390 $$

$$ {}^{2}_{(s)}r_{12,3} = \frac{\sigma(p_{13}, p_{23})}{\frac{1}{2}\left[(1-\rho_{13}^{2})+(1-\rho_{23}^{2})\right]} = -0{,}381 $$

$$ {}^{2}r_{12,3} = \frac{\rho_{12} - \rho_{13}\cdot\rho_{23}}{\frac{1}{2}\left[(1-\rho_{13}{}^{2})+(1-\rho_{23}{}^{2})\right]+\frac{3(N+1)}{4(N-1)}(\rho_{23}-\rho_{13})^{2}} = -0{,}340 $$

mentre Kendall aveva determinato il valore di

$$ \tau_{12,3} = \frac{\tau_{12} - \tau_{13}\cdot\tau_{23}}{\left[(1-\tau_{13}^{2})(1-\tau_{23}^{2})\right]^{\frac{1}{2}}} = -0{,}185 $$

Infine per quanto concerne l'indice di omofilia parziale, dal momento che c'è discordanza tra le distribuzioni dei resdui dei posti, occorre riferirsi alla espressione (37) e dal calcolo risulta

$$ {}^{2}E_{12,3} = -\frac{\sigma(p_{13}, p_{23})}{\sigma^{(2)}(p_{13}, p_{23})} = -0{,}400 $$

Le relazioni di natura teorica richiamate nei precedenti paragrafi vengono, ovviamente, confermate dal calcolo poiché l'intensità della discordanza decresce quando dal coefficiente di correlazione tra distribuzioni di variazione di posti si passa, dapprima, al coefficiente di correlazione tra distribuzioni di scostamenti dalla media di posti e, successivamente, al coefficiente di correlazione tra distribuzioni di posti; inoltre risulta un valore di discordanza più elevato per l'indice di omoflia, che non per gli indici di correlazione, in considerazione del fatto che per il primo viene considerato il massimo relativo di discordanza mentre per i secondi interviene il massimo assoluto.

Il *partial rank correlation* $\tau_{12,3}$ di Kendall è l'indice che presenta il più basso valore della discordanza; tra gli altri indici, quello che maggiormente gli si avvicina, anche se la differenza tra i valori appare sensibile, è il coefficiente di correlazione tra distribuzioni di posti ${}^{2}r_{12,3}$.

8. Al termine del lavoro si impongono alcune annotazioni critiche che riguardano i risultati conseguiti. Innanzitutto si può rilevare che l'estensione della teoria giniana della concordanza al contesto dell'associazione dei posti di graduatoria di tre caratteri potrebbe portare alla individuazione di una pluralità di indici, definibili nei confronti delle nozioni di massimo assoluto e di massimo relativo di concordanza, a seconda che si consideri la concordanza tra le distribuzioni delle variazioni dei posti, oppure tra le distribuzioni degli scostamenti dalla media dei posti, oppure infine, tra le distribuzioni dei posti. Però, anche in questa circostanza, è opportuno trascendere i puri aspetti formali per fare emergere le questioni di sostanza. In questo particolare contesto di ricerca, a differenza di ciò che può prospettarsi per le distribuzioni dei caratteri quantitativi, riteniamo che possa assumere un rilievo predominante l'indice (35) che è il più coerente alla natura del problema di misura della cograduazione per il suo diretto riferimento iniziale ai posti di graduatoria, a differenza di quanto avviene per gli altri due indici (20) e (22) che invece si riferiscono a variazioni di posti e a scostamenti di posti. Come strumento alternativo semmai si potrebbe proporre l'indice di omofilia parziale, di cui alla (38) e (39), nel caso in cui si ravvisasse l'opportunità di interpretare la concordanza in termini di massimo relativo anziché di massimo assoluto. Certamente rimane aperta una questione interpretativa che non è di poco conto. Il problema su cui intendiamo brevemente soffermarci si ricollega al significato che può essere attribuito alle distribuzioni (10) e (11), che abbiamo qualificato come distribuzioni di " residui dei posti " in quanto si qualificano come il risultato del processo di eliminazione dell'influenza esercitata dal carattere X_3 nei confronti delle graduatorie dei caratteri X_1 e X_2. Generalmente queste distribuzioni, proprio per il procedimento di eliminazione adottato, non conservano nella forma l'originaria natura di distribuzioni di posti e la misura della concordanza che le coinvolge non potrebbe qualificarsi come indice di cograduazione nel significato appropriato del termine. L'interrogativo che si impone a questo punto consiste nel valutare se l'indice (35), che chiaramente risulta espresso in funzione dei coefficienti di cograduazione di Spearman e che senza ombra di dubbio è stato ottenuto considerando nella fase di impostazione le graduatorie dei tre caratteri, debba essere rifiutato per il fatto che un risultato intermedio, seppure di non scarsa rilevanza ai fini della costruzione dell'indice, non appare immediatamente interpretabile in termini di posti. Sovente accade nelle elaborazioni statistiche di pervenire a risultati , intermedi o finali, la cui espressione non è formalmente calzante con la natura delle grandezze coinvolte nel calcolo e ciò in genere si verifica per i caratteri discreti. Per tali caratteri, ad esempio, si potrebbe pensare, per non allontanarci dal problema specifico di cui stiamo trattando, al coefficiente di correlazione parziale di Yule; anche in questo caso le distribuzioni dei residui dei caratteri X_1 e X_2, eliminata l'influenza del carattere X_3 , non presentano in genere valori di modalità formalmente compatibili con la realtà fenomenica. Ancora, si pensi al procedimento di calcolo dell'indice di cogra-

duazione del Gini, o del *rank correlation* ρ di Spearman, quando nella serie vi siano termini ripetuti : ai termini uguali è consuetudine attribuire lo stesso "posto" di graduatoria, corrispondente alla media aritmetica dei posti che essi occupano, che, com'è risaputo, può anche non corrispondere a un numero intero e, quindi, dal punto di vista strettamente formale non rappresentare un posto effettivo della graduatoria. Ed ancora, a parziale sostegno di questa nostra posizione favorevole all'impiego dell'indice (35) potremmo avvalerci della stessa opinione manifestata dal Kendall, che, come già abbiamo avuto modo di esporre, non è di totale e assoluta preclusione all'uso dell'indice (2) , indice che chiaramente presenta alcune caratteristiche formali analoghe a quelle in discussione per l'indice (35) : l'Autore infatti ritiene che quell'indice potrebbe essere utilizzato per approssimare il valore del coefficiente di correlazione parziale in una popolazione normale in considerazione della nota relazione, dimostrata da K. Pearson, che lega il coefficiente di correlazione lineare al coefficiente di cograduazione di Spearman. D'altra parte, collocandosi in una nuova prospettiva di ricerca si potrebbe valutare l'opportunità di considerare un procedimento di costruzione di un indice di cograduazione parziale, inteso nel senso proprio del termine, che potrebbe svilupparsi attraverso la preliminare trasformazione in graduatorie delle distribuzioni (10) e (11) , sostituendo i valori dei residui che in esse figurano con i corrispondenti posti d'ordine, per poi procedere alla misura della concordanza tra questi. In tale evenienza sarebbe poi opportuno soffermarsi sulla ricerca di una ipotetica relazione intercorrente tra questo nuovo indice e l'indice (35), anche per accertare se l'indice (35), che indubbiamente è di più immediato calcolo, potesse qualificarsi come una misura approssimata, e in questo senso sostitutiva, di quella a cui si perverrebbe col nuovo indice a cui si è fatto un semplice accenno.

Giancarlo Bettuzzi

Dipartimento di scienze statistiche
"Paolo Fortunati"
Università degli Studi di Bologna

Riferimenti bibliografici

G. Bettuzzi (1996), *Correlazione parziale e teoria della concordanza di Gini*, Dipartimento di Scienze Statistiche "Paolo Fortunati" dell' Università degli Studi di Bologna, Serie Ricerche 1996 n.3.

G. Bettuzzi (1997), *Considerazioni sulla teoria della concordanza del Gini nel contesto delle distribuzioni statistiche multiple : gli indici di omofilia parziale*, Dipartimento di Scienze Statistiche "Paolo Fortunati" dell' Università degli Studi di Bologna, Serie Ricerche 1997, n .1

G. Bettuzzi (1998), *Correlazione e omofilia parziale secondo la teoria della concordanza di Gini : una rivisitazione,* Statistica, vol. LVIII, n. 4

P. Fortunati (1967), *Alcune considerazioni sulla impostazione giniana delle misure di concordanza,* Atti della XXV Riunione Scientifica della Società Italiana di Statistica ,Bologna 1967.

A. Gili (1978), *Una postilla sull'indice quadratico di correlazione tra intensità*, Statistica, vol. XXXVIII, n. 1

C. Gini (1914), *Di una misura della dissomiglianza tra due gruppi di quantità e delle sue applicazioni allo studio delle relazioni statistiche*, Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, Tomo LXXIV-parte seconda, Venezia 1914

C. Gini (1915 a), *Indici di omofilia e di rassomiglianza e loro relazioni col coefficiente di correlazione e con gli indici di attrazione,* Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, Tomo LXXIV-parte seconda, Venezia 1915

C. Gini (1915 b), *Nuovi contributi alla teoria delle relazioni statistiche*, Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, Tomo LXXIV-parte seconda, Venezia 1915

C. Gini (1916 a), *Sul criterio di concordanza tra due caratteri,* Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, Tomo LXXV-parte seconda, Venezia 1916

C. Gini (1916 b), *Indici di concordanza*, Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, Tomo LXXV-parte seconda, Venezia 1916

M.G. Kendall (1942), *Partial rank correlation*, Biometrica vol. XXXII

T. Salvemini (1939), *L'indice di cograduazione del Gini nel caso di serie statistiche con ripetizioni*, Metron vol. XIII- n. 4, 1939

C. Spearman (1906), *A footrule for measuring correlation*, British Journal of Psychology, July 1906

G.U. Yule (1897), *On the theory of correlation*, Journal of the Royal Statistical Society, vol. LX, 1897

G.U. Yule (1912), *On the methods of measuring association between two attributes* , Journal of the Royal Statistical Society, vol. LXXV , 1912

G.U. Yule (1932), *An introduction to the theory of statistics*, C. Griffin and Co., London 1932

G.U. Yule-M.G. Kendall (1965), *An introduction to the theory of statistics*, C. Griffin and Co., London 1965