UNA APPROSSIMAZIONE DELLA VARIABILE CASUALE G DI GINI

(An Approximation of Gini's Random Variable)

GIORGIO VITTADINI

Estratto da "Rivista Internazionale di Scienze Economiche e Commerciali"
anno XXXVIII Gennaio 1991 N. 1

Rivista Internazionale di Scienze Economiche e Commerciali
Volume 38 (1991), N. 1, 81-94

UNA APPROSSIMAZIONE DELLA VARIABILE CASUALE *G* DI GINI

di
GIORGIO VITTADINI *

1. Siano $(X_1, Y_1), ..., (X_n, Y_n)$ n v.c. (variabile casuale) bidimensionali indipendenti e identicamente distribuite generate dalle coppie $(x_1, y_1), ..., (x_n, y_n)$ ottenute con n prove nella v.c. continua (X, Y) con funzione di distribuzione $\phi(x, y)$. Per verificare se fra le marginali di (X, Y) esiste « associazione », Gini (1914) ha proposto un indice che appartiene alla famiglia delle misure di associazione di cui fanno parte, fra le altre, il coefficiente di correlazione di Kendall (1938) ed il coefficiente di correlazione fra ranghi di Spearman (1904). L'indice di Gini, chiamato « coefficiente di cograduazione », è così definito:

$$g = (1/D) \sum_{i=1}^{n} \{|r_{xi} - r_{yi}^*| - |r_{xi} - r_{yi}|\}, \tag{1}$$

dove, nell'ipotesi che fra gli n valori x_i e gli n valori y_i non esistano ripetizioni, r_{xi} e r_{yi} sono i ranghi attribuiti ai valori medesimi, $r_{yi}^* = [(n + 1) - r_{yi}]$, $(i = 1, ..., n)$, mentre:

$$\begin{aligned} D &= n^2/2, &\quad &\text{se } n \text{ è pari,} \\ D &= (n^2 - 1)/2, &\quad &\text{se } n \text{ è dispari} \end{aligned} \tag{2}$$

La v.c. G, generata da g al variare della serie di prove in (X, Y), assume valori dell'intervallo $[-1,1]$, e, in particolare i valori:

$$-1 \qquad \text{se e solo se } r_{yi} = (n + 1) - r_{xi},$$

* Unità predipartimentale Statistico-matematica, Facoltà di Economia e Commercio, Università di Brescia.

Si ringraziano il prof. G. Landenna per i suggerimenti e i consigli forniti e il dott. M. Sangrico per il supporto negli aspetti computazionali del problema.

$$+1 \quad \text{se e solo se } r_{yi} = r_{xi}$$
$$(i = 1, ..., n).$$

Ciò premesso sia $(X'_1, Y_{j1}), ..., (X'_n, Y_{jn})$ un ordinamento della n-upla $(X_1, Y_1), ..., (X_n, Y_n)$ tale che $X'_1 < ... < X'_n$, dove $(j_1, ..., j_n)$ è una delle $n!$ possibili permutazioni degli interi 1, ..., n.

Nei riguardi della famiglia di misure di cui si è detto vale la seguente assunzione: « in assenza di qualsiasi tipo di associazione fra X ed Y le $n!$ possibili permutazioni dei valori $Y_1, ..., Y_n$ sono fra loro alla pari; in altri termini, le $n!$ possibili n-uple formate da $y_1, ..., y_n$ che può assumere la v.c. (X', Y) hanno ciascuna associata probabilità $1/n!$ ».

Va comunque osservato che, stante un teorema dovuto a W. Hoeffding (1948), se $n \geq 5$ la suddetta assunzione equivale a ritenere le due componenti di (X', Y) fra loro indipendenti in un senso stocastico, ovvero tali che per la funzione di distribuzione della v.c. (X', Y) vale la relazione $\phi'(x', y) = \phi_1(x')\,\phi_2(y)$, essendo $\phi_1(x')$ e $\phi_2(y)$ le funzioni di distribuzione delle marginali di (X', Y).

Ne segue che se $n \geq 5$ i due concetti di « assenza di associazione » e di « indipendenza stocastica » si identificano e, cosí stando le cose, nel seguito si farà sempre riferimento a questo secondo concetto.

Alla v.c. di Gini hanno rivolto la loro attenzione Savorgnan (1915), Amato (1954), Salvemini (1951), Cucconi (1964), Rizzi (1971), Herzel (1972), Cifarelli e Regazzini (1977), Landenna e Scagni (1989).

In particolare, nel caso di indipendenza, Amato ha costruito per $n = 2, 3, ..., 7$ la distribuzione della v.c. G mostrando che la medesima è simmetrica rispetto allo zero e quindi che ha tutti i momenti di ordine dispari nulli: $E(G^{2k+1}) = 0, (k = 0, 1, ...)$.

Inoltre, Amato e Cucconi, indipendentemente uno dall'altro, hanno provato che:

$$E(G^2) = Var(G) = \left\{ \begin{array}{ll} \left[\frac{2}{3(n-1)}\right]\left[1+\frac{2}{n^2}\right] & \textit{per } n \textit{ pari} \quad n > 4; \\ \left[\frac{2}{3(n-1)}\right]\left[1+\frac{4}{n^2-1}\right] & \textit{per } n \textit{ dispari} \quad n > 5 \end{array} \right\} \quad (3)$$

mettendo in evidenza che detta varianza è funzione esclusivamente della numerosità campionaria. Herzel, invece, ha provato che:

$$E(G^4) = \begin{cases} \dfrac{4}{105n^7(n-1)(n-3)}(35n^7 - 111n^6 + 153n^5 - \\ \qquad - 366n^4 + 304n^3 - 456n^2 - 912n + 1248); \\ \\ \dfrac{4}{105(n+1)^3 n(n-1)^4(n-2)}(35n^7 - 76n^6 + 182n^5 \\ \qquad - 307n^4 + 315n^3 - 342n^2 - 420n - 315). \end{cases} \quad (4)$$

e, quindi, che anche il momento quarto è funzione solo di n.

Amato, Cucconi, Salvemini e Herzel, inoltre, hanno osservato come, al divergere di n, la v.c. standardizzata $\bar{G} = G/[Var(G)]^{1/2}$ converge in distribuzione alla v.c. Normale N (0, 1), mentre Cifarelli e Regazzini ne hanno fornito la dimostrazione rigorosa. Landenna e Scagni hanno identificato una v.c. continua $\tilde{G}$, simmetrica rispetto allo zero, la cui varianza coincide con quella della v.c. G, mentre il quarto momento $E(\tilde{G}^4)$ approssima rapidamente $E(G^4)$ al divergere di n. In particolare, con una semplice trasformazione, la v.c. G si riduce ad una v.c. di Student rendendo assai agevole la verifica della ipotesi di indipendenza fra le componenti di (X', Y).

Ciò premesso, nella presente nota viene innanzitutto mostrato che la v.c. G di Gini assume determinazioni e probabilità legate direttamente alla numerosità campionaria n. Si propone inoltre una v.c. G^* che, rispetto a quella di Landenna e Scagni, ha il pregio di essere di tipo discreto, in conformità con la natura della v.c. in questione; inoltre assume le stesse determinazioni della v.c. G, ha tutti i momenti dispari nulli e il secondo e quarto uguali a quelli della v.c. G, e il suo comportamento la approssima pressoché perfettamente.

2. Costruite le distribuzioni delle v.c. G per numerosità n da 5 a 12 (Tab. I, prime due colonne), è emerso che, dopo alcune semplici riduzioni, le determinazioni di ciascuna delle v.c. medesime risultano legate alla numerosità campionaria nel modo seguente: se n è pari, si identificano in frazioni il cui denominatore è $w = n^2/4$, mentre i numeratori sono gli interi i da $-w = -(n^2/4)$ a $w = (n^2/4)$ con in piú lo 0; conseguentemente il numero delle determinazioni è $[2(n^2/4) + 1]$.

Cosí, ad esempio, per $n = 6$ la successione delle determinazioni che può assumere la v.c. G è:

$$-\frac{9}{9}, -\frac{8}{9}, \ldots, -\frac{1}{9}, 0, \frac{1}{9}, \ldots, \frac{8}{9}, \frac{9}{9},$$

TABELLA I

n = 5

g_i	p_i	p_i^*	s_i
0.000000	0.160000	0.138984	0.011015
0.166666	0.133333	0.131471	0.001861
0.333333	0.108333	0.115391	- 0.007058
0.500000	0.086666	0.089337	- 0.022671
0.666666	0.075000	0.058005	0.016994
0.833333	0.033333	0.030027	0.003306
1.000000	0.008333	0.011782	- 0.003448

n = 6

g_i	p_i	p_i^*	s_i
0.000000	0.088888	0.099634	- 0.010746
0.111111	0.122222	0.100475	0.021746
0.222222	0.083333	0.095976	- 0.012643
0.333333	0.088888	0.084672	0.004215
0.444444	0.061111	0.067269	- 0.006158
0.555555	0.050000	0.046925	0.003074
0.666666	0.027777	0.028024	- 0.000246
0.777777	0.015277	0.013970	0.001307
0.888888	0.006555	0.006668	- 0.000113
1.000000	0.001388	0.001825	- 0.000436

n = 7

g_i	p_i	p_i^*	s_i
0.000000	0.086507	0.085703	0.000804
0.083333	0.088095	0.085117	0.002977
0.166666	0.080753	0.081760	- 0.001006
0.250000	0.075793	0.075057	0.000736
0.333333	0.063293	0.065072	- 0.001778
0.416666	0.050793	0.052646	- 0.001852
0.500000	0.035515	0.039277	- 0.003761
0.583333	0.027380	0.026701	0.000679
0.666666	0.018650	0.016344	0.002305
0.750000	0.010714	0.008902	0.001812
0.833333	0.004365	0.004262	0.000102
0.916666	0.001190	0.001773	- 0.000583
1.000000	0.000198	0.000633	- 0.000436

n = 8

g_i	p_i	p_i^*	s_i
0.000000	0.072172	0.072115	5.69E-005
0.062500	0.079216	0.071648	0.007567
0.125000	0.064087	0.069550	- 0.005462
0.187500	0.067906	0.065558	0.002347
0.250000	0.057961	0.059641	- 0.001680
0.312500	0.050793	0.052046	- 0.001252
0.375000	0.039384	0.043300	- 0.003915
0.437500	0.033730	0.034135	- 0.000404
0.500000	0.025000	0.025342	- 0.000342
0.562500	0.019047	0.017610	0.001437
0.625000	0.012599	0.011384	0.001214
0.687500	0.007738	0.006804	0.000933
0.750000	0.003943	0.003737	0.000205
0.812500	0.001736	0.001875	- 0.000138
0.875000	0.000595	0.000853	- 0.000258
0.937500	0.000148	0.000350	- 0.000202
1.000000	2.48E-005	0.000129	- 0.000104

TABELLA I

n = 9

g_i	p_i	p_i^*	s_i
0.000000	0.064710	0.062658	0.002051
0.050000	0.063425	0.061999	0.001425
0.100000	0.061337	0.060359	0.000978
0.150000	0.056525	0.057616	- 0.001090
0.200000	0.053409	0.053743	- 0.000334
0.250000	0.047630	0.048820	- 0.001190
0.300000	0.041512	0.043041	- 0.001529
0.350000	0.035471	0.036703	- 0.001231
0.400000	0.029582	0.030168	- 0.000586
0.450000	0.023357	0.023823	- 0.000466
0.500000	0.018160	0.018010	0.000149
0.550000	0.013348	0.012991	0.000357
0.600000	0.009551	0.008910	0.000640
0.650000	0.006481	0.005791	0.000689
0.700000	0.004103	0.003555	0.000548
0.750000	0.002215	0.002054	0.000161
0.800000	0.001025	0.001113	- 8.80E-005
0.850000	0.000374	0.000563	- 0.000189
0.900000	0.000107	0.000266	- 0.000158
0.950000	2.20E-005	0.000116	- 9.45E-005
1.000000	2.75E-006	4.72E-005	- 4.45E-005

n = 10

g_i	p_i	p_i^*	s_i
0.000000	0.054193	0.054317	- 0.000123
0.040000	0.056089	0.053781	0.002307
0.080000	0.052289	0.052606	- 0.000317
0.120000	0.052277	0.050730	0.001546
0.160000	0.047159	0.048136	- 0.000977
0.200000	0.045197	0.044851	0.000346
0.240000	0.039537	0.040954	- 0.001417
0.280000	0.036109	0.036576	- 0.000467
0.320000	0.030740	0.031885	- 0.001145
0.360000	0.026616	0.027078	- 0.000461
0.400000	0.021711	0.022356	- 0.000644
0.440000	0.017761	0.017909	- 0.000148
0.480000	0.013721	0.013893	- 0.000171
0.520000	0.010746	0.010415	0.000331
0.560000	0.007978	0.007530	0.000447
0.600000	0.005819	0.005241	0.000577
0.640000	0.003941	0.003504	0.000437
0.680000	0.002505	0.002246	0.000259
0.720000	0.001439	0.001377	6.21E-005
0.760000	0.000745	0.000806	- 6.07E-005
0.800000	0.000336	0.000450	- 0.000114
0.840000	0.000131	0.000238	- 0.000106
0.880000	4.29E-005	0.000120	- 7.74E-005
0.920000	1.12E-005	5.74E-005	- 4.61E-005
0.960000	2.20E-006	2.59E-005	- 2.37E-005
1.000000	2.75E-007	1.10E-005	- 1.07E-005

dove, per l'appunto, ciascuna frazione ha al denominatore il valore $9 = 6^2/4$, mentre i numeratori sono rappresentati dagli interi da $-9 = -(6^2/4)$ a $9 = (6^2/4)$ con in piú lo 0, essendo $19 = [2(6^2/4) + 1]$ il numero delle determinazioni medesime.

Nel caso n dispari vale ancora quanto sopra salvo la sostituzione di $w = n^2/4$ con $w = (n^2 - 1)/4$.

TABELLA I

n = 11

g_i	p_i	p_i^*	s_i
0.000000	0.048362	0.048080	0.000281
0.033333	0.048342	0.047557	0.000786
0.066666	0.047165	0.046607	0.000558
0.100000	0.045731	0.045201	0.000529
0.133333	0.043536	0.043328	0.000207
0.166666	0.041023	0.040998	2.47E-005
0.200000	0.037809	0.038248	-0.000439
0.233333	0.034654	0.035136	-0.000481
0.266666	0.031044	0.031744	-0.000700
0.300000	0.027512	0.028170	-0.000658
0.333333	0.023918	0.024526	-0.000607
0.366666	0.020477	0.020922	-0.000445
0.400000	0.017200	0.017466	-0.000266
0.433333	0.014169	0.014251	-8.24E-005
0.466666	0.011419	0.011351	6.75E-005
0.500000	0.009019	0.008815	0.000203
0.533333	0.006943	0.006666	0.000276
0.566666	0.005218	0.004903	0.000314
0.600000	0.003788	0.003503	0.000285
0.633333	0.002668	0.002428	0.000239
0.666666	0.001783	0.001630	0.000152
0.700000	0.001126	0.001059	6.69E-005
0.733333	0.000658	0.000665	-7.37E-006
0.766666	0.000347	0.000403	-5.59E-005
0.800000	0.000162	0.000236	-7.34E-005
0.833333	6.62E-005	0.000132	-6.67E-005
0.866666	2.29E-005	7.20E-005	-4.90E-005
0.900000	6.56E-006	3.74E-005	-3.09E-005
0.933333	1.50E-006	1.87E-005	-1.71E-005
0.966666	2.50E-007	8.94E-006	-8.69E-006
1.000000	2.50E-008	4.09E-006	-4.06E-006

n = 12

g_i	p_i	p_i^*	s_i
0.000000	0.042669	0.042782	-0.000112
0.027777	0.043170	0.042332	0.000837
0.055555	0.041729	0.041584	0.000145
0.083333	0.041299	0.040521	0.000777
0.111111	0.039171	0.039137	3.39E-005
0.138888	0.037846	0.037437	0.000408
0.166666	0.035105	0.035439	-0.000334
0.194444	0.033136	0.033173	-3.78E-005
0.222222	0.030118	0.030681	-0.000563
0.250000	0.027702	0.028014	-0.000312
0.277777	0.024621	0.025234	-0.000613
0.305555	0.022039	0.022404	-0.000364
0.333333	0.019119	0.019592	-0.000472
0.361111	0.016660	0.016860	-0.000200
0.388888	0.014075	0.014268	-0.000193
0.416666	0.011869	0.011865	4.36E-006
0.444444	0.009717	0.009686	3.06E-005
0.472222	0.007895	0.007757	0.000137
0.500000	0.006236	0.006090	0.000146
0.527777	0.004878	0.004683	0.000195
0.555555	0.003716	0.003524	0.000192
0.583333	0.002787	0.002593	0.000193
0.611111	0.002025	0.001865	0.000160
0.638888	0.001427	0.001309	0.000117
0.666666	0.000958	0.000897	6.15E-005
0.694444	0.000610	0.000599	1.18E-005
0.722222	0.000364	0.000389	-2.55E-005
0.750000	0.000201	0.000246	-4.50E-005
0.777777	0.000102	0.000151	-4.96E-005
0.805555	4.70E-005	9.08E-005	-4.37E-005
0.833333	1.94E-005	5.27E-005	-3.33E-005
0.861111	7.07E-006	2.97E-005	-2.26E-005
0.888888	2.23E-006	1.62E-005	-1.40E-005
0.916666	5.97E-007	8.59E-006	-7.99E-006
0.944444	1.29E-007	4.40E-006	-4.27E-006
0.972222	2.08E-008	2.18E-006	-2.16E-006
1.000000	2.08E-009	1.04E-006	-1.04E-006

Cosí, ad esempio, nel caso $n = 7$ la successione delle determinazioni che può assumere la v.c. G è:

$$-\frac{12}{12}, -\frac{11}{12}, \dots, -\frac{1}{12}, 0, \frac{1}{12}, \dots, \frac{11}{12}, \frac{12}{12},$$

dove ciascuna frazione ha al denominatore il valore $12 = [(7^2 - 1)/4]$ mentre i numeratori sono rappresentati dagli interi da -12 a $+12$ con in piú lo zero, essendo $25 = \{[2(7^2 - 1)/4] + 1\}$ il numero delle determinazioni medesime.

3. Interessa ora individuare la v.c. G^* con le caratteristiche di cui si è detto.

Tale ricerca, tenuto conto della natura simmetrica della v.c. G di Gini, equivale ad identificare per ogni n le probabilità $p_i^* = P(G^* = g^*)$ sotto le condizioni che:

$$i)\ p_0^* + 2\sum_{i=1}^{w} p_i^* = p_0 + 2\sum_{i=1}^{w} p_i = 1,$$

$$ii)\ 2\sum_{i=1}^{w} (i/w)^2 p_i^* = 2\sum_{i=1}^{w} (i/w)^2 p_i = Var(G),$$

$$iii)\ 2\sum_{i=1}^{w} (i/w)^4 p_i^* = 2\sum_{i=0}^{w} (i/w)^4 p_i = E(G^4),$$

dove i parametri a, b, c, d si ricavano col metodo dei momenti risolvendo il (i/w), $(i = 0, 1, \dots, w)$.

Ciò premesso si supponga che le probabilità p_i^* siano interpretate dalla seguente funzione esponenziale:

$$p_i^* = e^{a + b(i/w) + c(i/w)^2 + d(i/w)^3}, \tag{5}$$

dove i parametri a, b, c, d si ricavano col metodo dei momenti risolvendo il seguente sistema non lineare nei parametri:

$$\begin{cases} 2\sum_{i=1}^{w} (i/w)^2 p_i = 2\sum_{i=1}^{w} (i/w)^2 p_i^* \\ 2\sum_{i=1}^{w} (i/w)^4 p_i = 2\sum_{i=1}^{w} (i/w)^4 p_i^* \\ p_0 + 2\sum_{i=1}^{w} [(i/w)^2 + 1] p_i = e^a + 2\sum_{i=1}^{w} [(i/w)^2 + 1] p_i^* \\ p_0 + 2\sum_{i=1}^{w} [(i/w)^4 + 1] p_i = e^a + 2\sum_{i=1}^{w} [(i/w)^4 + 1] p_i^* \end{cases} \tag{6}$$

essendo:

$$2\sum_{i=1}^{w}(i/w)^2 p_i = Var(G),$$
$$2\sum_{i=1}^{w}(i/w)^4 p_i = E(G^4),$$

$$p_0 + 2\sum_{i=1}^{w}[(i/w)^2 + 1]p_i = \left(p_0 + 2\sum_{i=1}^{w} p_i\right) + \\ + 2\sum_{i=1}^{w}(i/w)^2 p_i = 1 + Var(G)$$

$$p_0 + 2\sum_{i=1}^{w}[(i/w)^4 + 1]p_i = \left(p_0 + 2\sum_{i=1}^{w} p_i\right) + \\ + 2\sum_{i=1}^{w}(i/w)^4 p_i = 1 + E(G^4)$$

cosicché nell'ordine risultano rispettati i vincoli *ii*), *iii*) ed *i*). In particolare, stante il secondo membro della (5), si ha $p_i{}^* > 0$, ($i = 0, 1, ..., w$); inoltre, le equazioni del sistema (6) assicurano che la somma delle probabilità relative a tutte le determinazioni è uguale all'unità, cioè:

$$p_0{}^* + 2\sum_{i=1}^{w} p_i{}^* = e^a + 2\sum_{i=1}^{w} e^{a + b(i/w) + c(i/w)^2 + d(i/w)^3} = 1.$$

Merita ancora di osservare che i primi membri del sistema (6) sono tutti esprimibili in funzione della numerosità campionaria come emerge, d'altra parte, dalla (3) e dalla (4).

Ovviamente il sistema (6) può essere risolto solo ricorrendo a qualche metodo di ottimizzazione numerica quale, ad esempio, quello di Gauss-Newton.

Ricorrendo, per l'appunto, a tale metodo si sono ottenuti i valori dei parametri *a*, *b*, *c*, *d* della funzione (5) per numerosità campionarie *n* da 5 a 12 riportati nelle prime otto righe della Tabella II. Tramite tali valori si sono poi costruite le distribuzioni delle v.c. G^* approssimanti le corrispondenti v.c. G di Gini.

Il grado di approssimazione di G^* a G è stato misurato tramite la somma dei quadrati degli scarti tra valori osservati e valori attesi (Efron, 1978):

$$I^2 = \sum_{i=0}^{w}(p_i - p_i{}^*)^2 \tag{7}$$

e tramite il seguente indice (Frosini, 1987):

$$I^2 = 1 - \frac{\sum_{i=0}^{w} (p_i - p_i^*)^2}{\sum_{i=0}^{w} (p_i - \bar{p})^2} \tag{8}$$

dove $\bar{p} = [1/(w + 1)] \sum_{i=0}^{w} p_i$.

In particolare, dalla stessa Tabella II (ultime due colonne) emerge che l^2 è sempre pressoché nullo e I^2 assume sempre valori elevatissimi e, comunque, mai inferiori a 0.98.

Nella Tabella I (terza e quarta colonna), sempre per n compreso fra 5 e 12, sono proposti anche i valori delle probabilità p_i^* calcolate in base alla (5) e degli scarti $s_i = (p_i - p_i^*)$, scarti che, come si osserva, sono pressoché trascurabili.

TABELLA II

	a	b	c	d	l^2	I^2
5	-1.97339	-0.209834	-0.438205	-1.819741	1.49E-015	0.984
6	-2.30624	0.243792	-1.171983	-3.071553	4.58E-016	0.981
7	-2.45686	0.070388	-1.545587	-3.432281	3.28E-015	0.998
8	-2.62948	0.049337	-2.190171	-4.183367	7.27E-015	0.999
9	-2.77006	-0.071341	-2.570616	-4.547327	5.20E-015	0.998
10	-2.91291	-0.111621	-3.188954	-5.200442	9.31E-016	0.997
11	-3.03487	-0.202385	-3.592225	-5.576558	3.26E-014	0.998
12	-3.15163	-0.258832	-4.197821	-6.164027	4.55E-015	0.998
20	-4.50586	-0.973122	-0.651973	-1.793312	1.75E-014	0.999
25	-4.81916	-1.124245	-1.04880	-2.239166	1.20E-015	0.999

4. I risultati conseguiti inducono a ritenere che, qualunque sia n, tramite la soluzione del sistema (6), la funzione (5) possa consentire di pervenire, alla v.c. G^* la cui distribuzione è un'ottima approssimazione della v.c. G di Gini.

Poiché la costruzione della v.c. di Gini per numerosità $n > 12$ si presenta laboriosissima anche col ricorso ad un elaboratore, si può ricordare che, sulla base dei risultati conseguiti da Cifarelli e Regazzini, per $n \geq 20$ la distribuzione della v.c. medesima, debitamente standardizzata, può ritenersi Normale. Approfittando di tale fatto, pertanto, si sono costruite le distribuzioni della v.c. G di Gini e delle v.c. G^* approssimanti la stessa nei casi

TABELLA III
n = 20

g_i	p_i	p_i^*	s_i	g_i	p_i	p_i^*	s_i
0.00000	0.010165	0.011044	0.000879	0.25000	0.008365	0.008083	-0.000282
0.01000	0.010165	0.010936	0.000771	0.26000	0.008231	0.007950	-0.000280
0.02000	0.010158	0.010828	0.000669	0.27000	0.008093	0.007817	-0.000276
0.03000	0.010145	0.010719	0.000574	0.28000	0.007952	0.007682	-0.000270
0.04000	0.010125	0.010610	0.000484	0.29000	0.007809	0.007546	-0.000262
0.05000	0.010099	0.010500	0.000400	0.30000	0.007663	0.007410	-0.000253
0.06000	0.010066	0.010389	0.000322	0.31000	0.007515	0.007272	-0.000242
0.07000	0.010027	0.010278	0.000250	0.32000	0.007366	0.007134	-0.000231
0.08000	0.009981	0.010165	0.000183	0.33000	0.007214	0.006986	-0.000218
0.09000	0.009930	0.010052	0.000121	0.34000	0.007061	0.006856	-0.000206
0.10000	0.009872	0.009937	6.48E-005	0.35000	0.006907	0.006716	-0.000191
0.11000	0.009808	0.009821	1.32E-005	0.36000	0.006752	0.006575	-0.000176
0.12000	0.009738	0.009704	-3.35E-005	0.37000	0.006596	0.006435	-0.000161
0.13000	0.009662	0.009587	-7.56E-005	0.38000	0.006439	0.006293	-0.000145
0.14000	0.009581	0.009468	-0.000113	0.39000	0.006282	0.006152	-0.000130
0.15000	0.009495	0.009348	-0.000146	0.40000	0.006125	0.006010	-0.000114
0.16000	0.009402	0.009227	-0.000175	0.41000	0.005968	0.005869	-9.93E-005
0.17000	0.009305	0.009104	-0.000200	0.42000	0.005811	0.005727	-8.40E-005
0.18000	0.009203	0.008981	-0.000222	0.43000	0.005655	0.005588	-6.90E-005
0.19000	0.009096	0.008856	-0.000240	0.44000	0.005499	0.005445	-5.44E-005
0.20000	0.008985	0.008730	-0.000254	0.45000	0.005344	0.005304	-4.03E-005
0.21000	0.008869	0.008603	-0.000265	0.46000	0.005190	0.005164	-2.66E-005
0.22000	0.008749	0.008475	-0.000273	0.47000	0.005038	0.005024	-1.36E-005
0.23000	0.008625	0.008345	-0.000279	0.48000	0.004886	0.004885	-1.26E-006
0.24000	0.008497	0.008215	-0.000281	0.49000	0.004736	0.004747	1.04E-005

$n = 20$ ed $n = 25$ (Tabella III, prime tre colonne) e si sono calcolati gli scarti $s_i = (p_i - p_i^*)$ (Tabella III, ultima colonna). In particolare, nelle ultime due righe della Tabella II figurano i valori dei parametri a, b, c, d e degli indici I^2 e I^2.

I risultati conseguiti assicurano che il grado di approssimazione risulta ancora piú elevato ($I^2 \approx 0$; $I^2 \approx 0.999$; scarti s_i pressoché trascurabili), e un tale fatto dà conforto sulla bontà della v.c. G^*.

TABELLA III

n = 20

g_i	p_i	p_i^*	s_i	g_i	p_i	p_i^*	s_i
0.50000	0.004588	0.004609	2.13E-006	0.76000	0.001597	0.001646	4.88E-006
0.51000	0.004441	0.004473	3.15E-006	0.77000	0.001520	0.001564	4.36E-006
0.52000	0.004297	0.004338	4.08E-006	0.78000	0.001446	0.001484	3.82E-006
0.53000	0.004154	0.004203	4.94E-006	0.79000	0.001374	0.001407	3.29E-006
0.54000	0.004013	0.004071	5.71E-006	0.80000	0.001306	0.001333	2.75E-006
0.55000	0.003875	[illegible]	6.40E-006	0.81000	0.001240	0.001262	2.21E-006
0.56000	0.003739	[illegible]	7.00E-006	0.82000	0.001176	0.001193	1.68E-006
0.57000	0.003606	0.003681	7.52E-006	0.83000	0.001115	0.001127	1.16E-006
0.58000	0.003475	0.003554	7.96E-006	0.84000	0.001057	0.001063	6.47E-006
0.59000	0.003346	0.003429	8.31E-006	0.85000	0.001000	0.001002	1.48E-006
0.60000	0.003220	0.003306	8.58E-006	0.86000	0.000947	0.000943	-3.33E-006
0.61000	0.003097	0.003185	8.78E-006	0.87000	0.000895	0.000887	-7.98E-006
0.62000	0.002977	0.003066	8.90E-006	0.88000	0.000846	0.000834	-1.24E-006
0.63000	0.002860	0.002949	8.95E-006	0.89000	0.000799	0.000782	-1.66E-006
0.64000	0.002746	0.002834	8.93E-006	0.90000	0.000754	0.000733	-2.06E-006
0.65000	0.002633	0.002722	8.86E-006	0.91000	0.000711	0.000687	-2.44E-006
0.66000	0.002524	0.002611	8.70E-006	0.92000	0.000670	0.000642	-2.79E-006
0.67000	0.002418	0.002503	8.50E-006	0.93000	0.000632	0.000600	-3.12E-006
0.68000	0.002315	0.002398	8.25E-006	0.94000	0.000594	0.000560	-3.42E-005
0.69000	0.002215	0.002295	7.95E-006	0.95000	0.000559	0.000522	-3.69E-006
0.70000	0.002118	0.002194	7.61E-006	0.96000	0.000526	0.000486	-3.94E-006
0.71000	0.002024	0.002096	7.22E-006	0.97000	0.000494	0.000452	-4.16E-006
0.72000	0.001933	0.002001	6.81E-006	0.98000	0.000464	0.000420	-4.35E-006
0.73000	0.001845	0.001908	6.38E-006	0.99000	0.000436	0.000390	-4.52E-006
0.74000	0.001759	0.001818	5.88E-006	1.00000	0.000408	0.000361	-4.66E-006
0.75000	0.001677	0.001731	5.39E-006				

TABELLA III
n ≃ 25

g_i	p_i	p_i^*	s_i
0.000000	0.007376	0.008073	0.000697
0.006410	0.007376	0.008015	0.000639
0.012821	0.007373	0.007956	0.000583
0.019231	0.007368	0.007897	0.000529
0.025641	0.007360	0.007838	0.000477
0.032051	0.007350	0.007778	0.000427
0.038462	0.007338	0.007718	0.000380
0.044872	0.007323	0.007658	0.000335
0.051282	0.007305	0.007597	0.000292
0.057692	0.007285	0.007536	0.000251
0.064103	0.007263	0.007475	0.000212
0.070513	0.007238	0.007413	0.000174
0.076923	0.007211	0.007351	0.000139
0.083333	0.007181	0.007288	0.000106
0.089744	0.007150	0.007225	7.55E-005
0.096154	0.007115	0.007162	4.61E-005
0.10256	0.007079	0.007098	1.86E-005
0.10897	0.007040	0.007033	-7.14E-006
0.11538	0.007000	0.006969	-3.11E-005
0.12179	0.006957	0.006903	-5.34E-005
0.12821	0.006912	0.006838	-7.41E-005
0.13462	0.006865	0.006771	-9.31E-005
0.14103	0.006815	0.006705	-0.000110
0.14744	0.006764	0.006638	-0.000126
0.15385	0.006711	0.006570	-0.000140
0.16026	0.006657	0.006503	-0.000153
0.16667	0.006600	0.006434	-0.000165
0.17308	0.006542	0.006366	-0.000175
0.17949	0.006481	0.006297	-0.000184
0.18590	0.006420	0.006227	-0.000192
0.19231	0.006358	0.006157	-0.000199
0.19872	0.006291	0.006087	-0.000204
0.20513	0.006225	0.006016	-0.000209
0.21154	0.006157	0.005945	-0.000212
0.21795	0.006088	0.005874	-0.000214
0.22436	0.006018	0.005802	-0.000215
0.23077	0.005946	0.005730	-0.000216
0.23718	0.005874	0.005658	0.000216
0.24359	0.005800	0.005585	-0.000214

g_i	p_i	p_i^*	s_i
0.25000	0.005725	0.005512	-0.000213
0.25641	0.005649	0.005439	-0.000210
0.26282	0.005572	0.005365	-0.000207
0.26923	0.005495	0.005292	-0.000203
0.27564	0.005416	0.005218	-0.000198
0.28205	0.005337	0.005143	-0.000193
0.28846	0.005258	0.005069	-0.000188
0.29487	0.005177	0.004995	-0.000182
0.30128	0.005097	0.004920	-0.000176
0.30769	0.005015	0.004846	-0.000169
0.31410	0.004934	0.004771	-0.000163
0.32051	0.004852	0.004696	-0.000155
0.32692	0.004770	0.004621	-0.000148
0.33333	0.004687	0.004546	-0.000140
0.33974	0.004605	0.004471	-0.000133
0.34615	0.004522	0.004396	-0.000125
0.35256	0.004439	0.004322	-0.000117
0.35897	0.004356	0.004247	-0.000109
0.36538	0.004274	0.004172	-0.000101
0.37179	0.004191	0.004098	-9.35E-005
0.37821	0.004109	0.004023	-8.56E-005
0.38462	0.004027	0.003949	-7.77E-005
0.39103	0.003945	0.003875	-6.98E-005
0.39744	0.003864	0.003802	-6.21E-005
0.40385	0.003783	0.003728	-5.45E-005
0.41026	0.003702	0.003655	-4.70E-005
0.41667	0.003622	0.003582	-3.98E-005
0.42308	0.003542	0.003510	-3.26E-005
0.42949	0.003463	0.003438	-2.57E-005
0.43590	0.003385	0.003366	-1.90E-005
0.44231	0.003307	0.003294	-1.26E-005
0.44872	0.003230	0.003224	-6.37E-006
0.45513	0.003154	0.003153	-3.83E-007
0.46154	0.003078	0.003083	5.34E-006
0.46795	0.003003	0.003014	1.08E-005
0.47436	0.002929	0.002945	1.60E-005
0.48077	0.002856	0.002877	2.09E-005
0.48718	0.002784	0.002809	2.55E-005
0.49359	0.002712	0.002742	2.99E-005

TABELLA III
n = 25

g_i	p_i	p_i^*	s_i	g_i	p_i	p_i^*	s_i
0.50000	0.002642	0.002676	3.39E-005	0.75641	0.000696	0.000718	2.17E-005
0.50641	0.002572	0.002610	3.77E-005	0.76282	0.000668	0.000688	1.95E-005
0.51282	0.002504	0.002545	4.12E-005	0.76923	0.000642	0.000659	1.73E-005
0.51923	0.002436	0.002480	4.44E-005	0.77564	0.000616	0.000631	1.52E-005
0.52564	0.002370	0.002417	4.73E-005	0.78205	0.000591	0.000604	1.31E-005
0.53205	0.002304	0.002354	4.99E-005	0.78846	0.000567	0.000578	1.11E-005
0.53846	0.002240	0.002292	5.23E-005	0.79487	0.000543	0.000553	9.11E-006
0.54487	0.002176	0.002231	5.43E-005	0.80128	0.000521	0.000528	7.16E-006
0.55128	0.002114	0.002170	5.61E-005	0.80769	0.000499	0.000504	5.27E-006
0.55769	0.002053	0.002110	5.77E-005	0.81410	0.000478	0.000481	3.43E-006
0.56410	0.001992	0.002051	5.89E-005	0.82051	0.000458	0.000459	1.65E-006
0.57051	0.001933	0.001993	6.00E-005	0.82692	0.000438	0.000438	-6.10E-008
0.57692	0.001875	0.001936	6.07E-005	0.83333	0.000419	0.000417	-1.70E-006
0.58333	0.001819	0.001880	6.13E-005	0.83974	0.000401	0.000398	-3.28E-006
0.58974	0.001763	0.001824	6.16E-005	0.84615	0.000383	0.000379	-4.79E-006
0.59615	0.001708	0.001770	6.17E-005	0.85256	0.000366	0.000360	-6.23E-006
0.60256	0.001655	0.001716	6.16E-005	0.85897	0.000350	0.000342	-7.60E-006
0.60897	0.001602	0.001664	6.13E-005	0.86538	0.000334	0.000325	-8.89E-006
0.61538	0.001551	0.001612	6.07E-005	0.87179	0.000319	0.000309	-1.01E-005
0.62179	0.001501	0.001561	6.01E-005	0.87821	0.000306	0.000293	-1.12E-005
0.62821	0.001452	0.001511	5.92E-005	0.88462	0.000291	0.000278	-1.23E-005
0.63462	0.001404	0.001462	5.82E-005	0.89103	0.000277	0.000264	-1.33E-005
0.64103	0.001357	0.001415	5.70E-005	0.89744	0.000264	0.000250	-1.42E-005
0.64744	0.001312	0.001368	5.57E-005	0.90385	0.000252	0.000237	-1.51E-005
0.65385	0.001267	0.001322	5.43E-005	0.91026	0.000240	0.000224	-1.58E-005
0.66026	0.001224	0.001277	5.27E-005	0.91667	0.000229	0.000212	-1.66E-005
0.66667	0.001182	0.001233	5.11E-005	0.92308	0.000218	0.000201	-1.72E-005
0.67308	0.001140	0.001190	4.93E-005	0.92949	0.000207	0.000190	-1.78E-005
0.67949	0.001100	0.001147	4.74E-005	0.93590	0.000197	0.000179	-1.83E-005
0.68590	0.001061	0.001106	4.55E-005	0.94231	0.000188	0.000169	-1.87E-005
0.69231	0.001023	0.001066	4.35E-005	0.94872	0.000178	0.000159	-1.91E-005
0.69872	0.000986	0.001027	4.14E-005	0.95513	0.000170	0.000150	-1.95E-005
0.70513	0.000950	0.000989	3.93E-005	0.96154	0.000161	0.000141	-1.97E-005
0.71154	0.000914	0.000952	3.72E-005	0.96795	0.000153	0.000133	-2.00E-005
0.71795	0.000880	0.000915	3.50E-005	0.97436	0.000145	0.000125	-2.01E-005
0.72436	0.000847	0.000880	3.28E-005	0.98077	0.000138	0.000118	-2.02E-005
0.73077	0.000815	0.000846	3.06E-005	0.98718	0.000131	0.000111	-2.03E-005
0.73718	0.000784	0.000812	2.84E-005	0.99359	0.000124	0.000104	-2.03E-005
0.74359	0.000754	0.000780	2.61E-005	1.00000	0.000118	9.79E-005	-2.03E-005

RIFERIMENTI BIBLIOGRAFICI

AMATO V., "Sulla distribuzione dell'indice di cograduazione del Gini", *Statistica*, 1954, 505-19.

CIFARELLI D.M.-REGAZZINI E., "On a Distribution Free Test of Independence Based on Gini's Rank Association Coefficient. Recent Developments in Statistics", *Proceedings of the European Meeting of Statistics, Grenoble 6-11, September 1976*, 1977.

CUCCONI O., "La distribuzione dell'indice di cograduazione del Gini", *Statistica*, 1964, 143-51.

EFRON B., "Regression and Anova with Zero-one Data: Measures of Residual Variation", *Journal of the American Statistical Association*, 1978, *73*, 113-21.

FROSINI B.V., *Connessione, regressione, correlazione*, Milano: Celuc, 1987.

GINI C., *Di una misura delle relazioni fra le graduatorie di due caratteri*, Roma: Tip. Cecchi, 1914.

HERZEL A., "Sulla distribuzione campionaria dell'indice di cograduazione del Gini", *Metron*, 1972, *30*, 137-53.

HOEFFDING W., "A Class of Statistics with Asymptotically Normal Distribution", *Annals of Mathematical Statistics*, 1948, *29*, 293-325.

KENDALL M.G., "A New Measure of Rank Correlation", *Biometrika*, 1938, *30*, 81-93.

LANDENNA G.-SCAGNI A., "An Approximated Distribution of the Gini's Rank Association Coefficient", *Communications in Statistics*, 1989, *18*, 2017-26.

RIZZI A., "Distribuzione dell'indice di cograduazione del Gini", *Metron*, 1971, *29*, 63-73.

SALVEMINI T., "Sui vari indici di cograduazione", *Atti della XI Riunione della Società Italiana di Statistica*, 1951.

SAVORGNAN F., *Sulla formazione dei valori dell'indice di cograduazione*, Studi economico-giuridici dell'Università di Cagliari, 1915.

SPEARMAN C., "The Proof and Measurement of Association Between Two Things", *American Journal of Psychology*, 1904, *15*, p. 88.

AN APPROXIMATION OF GINI'S RANDOM VARIABLE

An approximate distribution is proposed for the Gini rank association coefficient, which is a statistic to test independence between two random variables. With respect to other approximations of Gini's random variable, the purposed distribution has the advantage of being a discrete distribution and having all the moments of even order equal to zero where the second and fourth are equal to those of Gini's random variable. Furthermore it assumes the values and the probabilities closely related to the sample size.