

**INTERVALLI DI CONFIDENZA SIMULTANEI ASINTOTICI PER LE
 PROBABILITA' MARGINALI IN UNA TABELLA 2 X 2 (*)**

di ANGIOLA POLLASTRI (**)

1. Introduzione

Sia data una popolazione le cui unità siano portatrici dei caratteri A e B, ognuno dei quali presenta solo due modalità che sono rispettivamente \bar{a} e a, \bar{b} e b. Nella Tab. I sono indicate le quattro associazioni possibili fra i due caratteri e le rispettive probabilità che si verifichino.

Si supponga di estrarre dalla suddetta popolazione un campione con ripetizione di ampiezza n. Nella Tab. II, vengono riportate le numerosità con cui si verificano le possibili associazioni fra i due caratteri.

A B	\bar{a}	a	
\bar{b}	p_{00}	p_{10}	$p_{.0}$
b	p_{01}	p_{11}	$p_{.1}$
	$p_{0.}$	$p_{1.}$	1

Tab. I

A B	\bar{a}	a	
\bar{b}	n_{00}	n_{10}	$n_{.0}$
b	n_{01}	n_{11}	$n_{.1}$
	$n_{0.}$	$n_{1.}$	n

Tab. II

In questo lavoro ci si propone di costruire intervalli di confidenza simultanei asintotici per $p_{1.}$ e $p_{.1}$ sia nel caso in cui si conoscano tutti i dati della Tab. II, sia nel caso in cui se ne conoscano solo i marginali.

(*) Lavoro presentato al convegno di studi "Le statistiche dello Sviluppo" tenuto a Salerno dal 15 al 17 ottobre 1980.

(**) Libera Università degli Studi di Trento - Dipartimento di Metodi Quantitativi.

Per fare ciò si impiegherà la distribuzione asintotica della statistica

$$Z = \max \{ |X|, |Y| \},$$

dove X e Y sono rispettivamente gli scarti standardizzati delle frequenze con cui si presentano a e b , che si dimostrerà essere costituita dal miscuglio di due v.c. Arcotangenti. Si danno poi indicazioni riguardo alla numerosità campionaria necessaria per poter utilizzare gli intervalli di confidenza simultanei asintotici di $p_{1.}$ e $p_{.1}$. Si mostra infine che gli intervalli di confidenza qui proposti sono più brevi, a parità di livello di confidenza, di quelli che si possono ottenere utilizzando il metodo di Bonferroni.

2. Distribuzione asintotica della statistica Z

E' noto che la coppia di v.c.

$$X = \frac{n_{1.} - np_{1.}}{\sqrt{np_{1.}(1-p_{1.})}} \quad \text{e} \quad Y = \frac{n_{.1} - np_{.1}}{\sqrt{np_{.1}(1-p_{.1})}}$$

è asintoticamente distribuita come una v.c. Normale Bivariata Standardizzata con coefficiente di correlazione:

$$\rho = \frac{p_{1.1} - p_{1.}p_{.1}}{\sqrt{p_{1.}(1-p_{1.})} \sqrt{p_{.1}(1-p_{.1})}}$$

Il luogo dei punti in cui $Z = z$ corrisponde al luogo dei punti in cui $[(|X| = z \text{ e } |Y| \leq z) \cup (|X| \leq z \text{ e } |Y| = z)]$, ossia ai lati del quadrato rappresentato in Fig. 1. I vertici del quadrato ABCD sono di coordinate:

$$\begin{aligned} A &= (-z, -z); \\ B &= (+z, -z); \\ C &= (+z, +z); \\ D &= (-z, +z); \end{aligned}$$

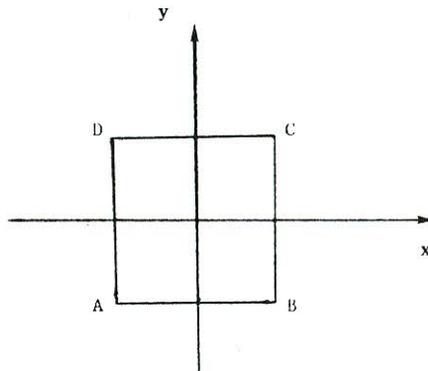


Fig. 1 : Luogo dei punti in cui $Z = z$ sul piano (x,y) .

Si deduce facilmente che la funzione di densità asintotica della v.c. Z si ottiene integrando la funzione di densità della v.c. (X, Y) sui lati del quadrato ABCD in modo analogo a [5] ed a [8]. Essa è data da:

$$g(z) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \left[\int_{-a_1 z}^{a_2 z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt + \int_{-a_2 z}^{a_1 z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt \right]$$

$$= \frac{2}{\pi} \theta_1 \frac{e^{-\frac{1}{2}z^2} \int_0^{a_1 z} e^{-\frac{1}{2}t^2} dt}{\theta_1} + \frac{2}{\pi} \theta_2 \frac{e^{-\frac{1}{2}z^2} \int_0^{a_2 z} e^{-\frac{1}{2}t^2} dt}{\theta_2}$$

dove: $a_1 = \frac{1-Q}{\sqrt{1-Q^2}}$, $a_2 = \frac{1+Q}{\sqrt{1-Q^2}}$,

$$\theta_i = \text{arctg } a_i, \quad i = 1, 2.$$

Se si indica con $f(x;a)$ la funzione di densità di una v.c. Arcotangente di parametro a (*) la funzione di densità asintotica della statistica Z diventa:

$$g(z) = \frac{2}{\pi} \theta_1 f(z;a_1) + \frac{2}{\pi} \theta_2 f(z;a_2) \quad (1)$$

Essendo possibile dimostrare che

$$\frac{2}{\pi} \theta_1 + \frac{2}{\pi} \theta_2 = 1, \quad (2)$$

si può concludere che la statistica Z è asintoticamente distribuita come un miscuglio di due v.c. Arcotangenti di parametro a_1 ed a_2 .

Si indichi con $h(\alpha; Q)$ l' $(1-\alpha)$ mo quantile della distribuzione asintotica della v.c. Z . Nella Tab. III, vengono presentati i valori di $h(\alpha; Q)$ in funzione di α e di $|Q|$. Essi sono stati ricavati servendosi dello sviluppo in serie della funzione $T(h; a)$ proposto da Owen [4], dove il legame tra $T(h; a)$ e la f.r. di una v.c. Arcotangente di parametro a è il seguente:

$$1 - F(h; a) = \int_h^\infty f(z;a) dz = \frac{2\pi}{\text{arctg } a} T(h;a). \quad (3)$$

Si osservi che le variazioni di $h(\alpha; Q)$, a parità di α , sono più rilevanti per valori elevati di $|Q|$.

(*) Per la definizione di v.c. Arcotangente si veda Zenga [8].

Tab. III : valori di $h(\alpha; \varrho)$ in funzione di α e di $|\varrho|$.

$\alpha \backslash \varrho $	0.01	0.02	0.05	0.10	0.15
0.000	2.8063	2.5741	2.2365	1.9488	1.7622
0.100	2.8059	2.5736	2.2357	1.9476	1.7607
0.200	2.8049	2.5721	2.2331	1.9439	1.7564
0.300	2.8030	2.5692	2.2285	1.9378	1.7491
0.400	2.7996	2.5647	2.2218	1.9287	1.7385
0.500	2.7943	2.5578	2.2121	1.9163	1.7241
0.525	2.7925	2.5556	2.2092	1.9126	1.7190
0.550	2.7906	2.5532	2.2060	1.9086	1.7153
0.575	2.7884	2.5506	2.2025	1.9043	1.7105
0.600	2.7860	2.5476	2.1987	1.8996	1.7053
0.625	2.7832	2.5444	2.1946	1.8947	1.6997
0.650	2.7802	2.5408	2.1901	1.8893	1.6937
0.675	2.7768	2.5368	2.1852	1.8836	1.6873
0.700	2.7730	2.5325	2.1799	1.8772	1.6804
0.725	2.7687	2.5276	2.1740	1.8705	1.6730
0.750	2.7639	2.5221	2.1676	1.8631	1.6649
0.775	2.7583	2.5160	2.1604	1.8550	1.6561
0.800	2.7522	2.5091	2.1525	1.8460	1.6465
0.825	2.7450	2.5012	2.1435	1.8361	1.6359
0.850	2.7368	2.4923	2.1334	1.8249	1.6241
0.875	2.7270	2.4818	2.1218	1.8123	1.6107
0.900	2.7154	2.4694	2.1081	1.7976	1.5953
0.925	2.7010	2.4541	2.0917	1.7801	1.5820
0.950	2.6822	2.4346	2.0709	1.7581	1.5543
0.975	2.6551	2.4065	2.0425	1.7276	1.5231
0.990	2.6282	2.3791	2.0133	1.6987	1.4937

Si vuole ora vedere qual'è il valore di ϱ per il quale la funzione di ripartizione della v.c. Z , ossia:

$$F(h; a_1, a_2) = F(h; \varrho) = \Pr(Z \leq h)$$

assume il valore minimo.

Ricordando la (1) e la funzione di densità della v.c. Arcotangente (Zenga [8]) si può scrivere:

$$F(h; a_1, a_2) = \frac{2}{\pi} \int_0^h \int_0^{a_1 x} e^{-\frac{1}{2}(x^2+y^2)} dy dx + \frac{2}{\pi} \int_0^h \int_0^{a_2 x} e^{-\frac{1}{2}(x^2+y^2)} dy dx \quad (4)$$

Ponendo nella (4) $y = r \sin \theta$ e $x = r \cos \theta$ si ottiene:

$$\begin{aligned} F(h; a_1, a_2) &= \frac{2}{\pi} \left[\int_0^{\theta_1} \int_0^{h/\cos\theta} r e^{-\frac{1}{2} r^2} dr d\theta + \int_0^{\theta_2} \int_0^{h/\cos\theta} r e^{-\frac{1}{2} r^2} dr d\theta \right] \\ &= 1 - \frac{2}{\pi} \left[\int_0^{\theta_1} e^{-\frac{1}{2} h^2 / \cos^2 \theta} d\theta + \int_0^{\frac{\pi}{2} - \theta_1} e^{-\frac{1}{2} h^2 / \cos^2 \theta} d\theta \right] \\ &= F(h; \theta_1). \end{aligned}$$

Derivando $F(h; \theta_1)$ rispetto a θ_1 si ottiene:

$$\frac{dF(h; \theta_1)}{d\theta_1} = -\frac{2}{\pi} e^{-\frac{1}{2} \frac{h^2}{\cos^2 \theta_1}} + \frac{2}{\pi} e^{-\frac{1}{2} \frac{h^2}{\cos^2 (\frac{\pi}{2} - \theta_1)}}$$

Tenendo presente che $a_1 > 0$ e $a_2 > 0$ poichè $-1 < Q < +1$, e tenendo presente la (2), deve essere

$$0 < \theta_1 < \frac{\pi}{2} \quad (5)$$

La derivata prima di $F(h; \theta_1)$ rispetto a θ_1 nell'intervallo fra 0 e $\frac{\pi}{2}$ si annulla in corrispondenza di $\frac{\pi}{4}$. Inoltre, per $\theta_1 = \frac{\pi}{4}$ la derivata seconda di $F(h; \theta_1)$ è maggiore di zero. L'equazione:

$$\theta_1 = \operatorname{arctg} a_1 = \operatorname{arctg} \frac{1 - Q}{\sqrt{1 - Q^2}} = \frac{\pi}{4}$$

è verificata per $Q = 0$. Pertanto la $F(h; Q)$ assume il valore minimo in corrispondenza di $Q = 0$ e cioè $a_1 = a_2 = 1$.

I valori di $h(\alpha; 0)$ si possono trovare nella Tab. III. Essi possono essere altresì ricavati dalle tavole di Owen [4].

Ricordando la (1) e la (3), si è interessati a quei valori di h per cui

$$T(h, 1) = \frac{\alpha}{8}$$

3. Intervalli di confidenza simultanei asintotici per $p_{1.}$ e $p_{.1}$

Si vogliono ora utilizzare i risultati del paragrafo precedente per costruire intervalli di confidenza simultanei asintotici per $p_{1.}$ e per $p_{.1}$.

Si indichi con I_1 l'intervallo casuale:

$$\hat{p}_{1.} \mp h(\alpha; Q) \frac{\sigma_x}{\sqrt{n}}, \quad (6)$$

dove $\hat{p}_{1.} = \frac{n_{1.}}{n}$ e $\sigma_x = \sqrt{p_{1.}(1-p_{1.})}$,

Si indichi con I_2 l'intervallo casuale:

$$\hat{p}_{.1} \mp h(\alpha; Q) \frac{\sigma_y}{\sqrt{n}} \quad (7)$$

dove $\hat{p}_{.1} = \frac{n_{.1}}{n}$ e $\sigma_y = \sqrt{p_{.1}(1-p_{.1})}$

L'evento $[(p_{1.} \in I_1) \cap (p_{.1} \in I_2)]$ corrisponde all'evento che la v.c. Z assuma valori minori od uguali ad $h(\alpha; Q)$. Infatti si può scrivere

$$\begin{aligned} P \{ Z \leq h(\alpha; Q) \} &= P \left\{ \max \left[|X|, |Y| \right] \leq h(\alpha; Q) \right\} \\ &= P \left\{ \left[|X| \leq h(\alpha; Q) \right] \cap \left[|Y| \leq h(\alpha; Q) \right] \right\} \\ &= P \left\{ (p_{1.} \in I_1) \cap (p_{.1} \in I_2) \right\} \cong 1 - \alpha \quad (8) \end{aligned}$$

Nella (6) e nella (7) non si conoscono gli scarti quadratici medi σ_x e σ_y . Ad essi si possono sostituire le rispettive stime consistenti date da:

$$s_x = \sqrt{\hat{p}_{1.} - (\hat{p}_{1.})^2} \quad \text{e} \quad s_y = \sqrt{\hat{p}_{.1} - (\hat{p}_{.1})^2}$$

Inoltre, dato che di solito anche il valore di Q è ignoto, non si è in grado di trovare $h(\alpha; Q)$. Si possono allora seguire tre vie. Se si conoscono tutti i dati della Tab. II, a Q si può sostituire la sua stima consistente data da:

$$\hat{Q} = \frac{\frac{n_{11}}{n} - \hat{p}_{1.} \hat{p}_{.1}}{s_x \cdot s_y}}$$

Si può, altrimenti, in $h(\alpha; Q)$ porre $Q = 0$ ossia il valore di Q che rende minima la funzione di ripartizione della v.c. Z . Ovviamente, se si conoscono solo i marginali della Tab. II, si è costretti a seguire questa seconda via.

Infine, dato che i valori di $h(\alpha; Q)$ incominciano a variare in modo sensibile per valori di $|Q| > 0.5$ si potrebbe adottare la seguente procedura. Si usa $h(\alpha; 0)$ quando $|\hat{Q}| \leq 0.5$ mentre si usa $h(\alpha; \hat{Q})$ quando $|\hat{Q}| > 0.5$.

Nel par. 5 si confrontano i livelli di confidenza effettivi per i metodi qui proposti.

4. Intervalli di confidenza simultanei asintotici per p_1 e $p_{.1}$ col metodo di Bonferroni

Si ricordi innanzitutto che X e Y sono asintoticamente distribuite come v.c. Normali Standardizzate.

Si indichi con E_1 l'intervallo casuale:

$$\hat{p}_1 \pm z\left(\frac{\alpha}{2}\right) \frac{\sigma_x}{\sqrt{n}}, \quad (9)$$

dove $z(\alpha/2)$ è l' $(1 - \frac{\alpha}{2})$.mo quantile della v.c. Normale standardizzata.

Si indichi con E_2 l'intervallo casuale:

$$\hat{p}_{.1} \pm z\left(\frac{\alpha}{2}\right) \frac{\sigma_y}{\sqrt{n}}. \quad (10)$$

Servendosi della diseuguaglianza di Bonferroni (si veda [3] e [6]) si può affermare che:

$$P \left[(p_1 \in E_1) \cap (p_{.1} \in E_2) \right] \geq 1 - 2\alpha. \quad (11)$$

Gli intervalli di confidenza simultanei ottenuti dalla (9) e dalla (10) aventi coefficienti $\geq 1 - 2\alpha$ sono più ampi degli intervalli di confidenza con coefficiente di confidenza pari a $1 - 2\alpha$ proposti nel par. 3 anche nel caso più sfavorevole in cui si conoscano solo i dati marginali della Tab. II, come verrà mostrato nel par. 6.

Poichè non è possibile calcolare gli intervalli (9) e (10) in quanto non si conoscono gli scarti quadratici medi σ_x e σ_y , si sostituiscono ad essi le rispettive stime consistenti s_x e s_y .

5. Confronto tra livelli di confidenza nominali ed effettivi

Si vuol ora vedere qual'è la numerosità campionaria necessaria per avere buone approssimazioni fra i coefficienti di confidenza effettivi e quelli nominali. Per fare ciò si è proceduto come segue.

Si sono innanzitutto fissati i livelli di α_j pari a:

j	1	2	3	4	5
α_j	0.01	0.02	0.05	0.10	0.15

Si sono scelte le seguenti terne di probabilità:

s	$p_{1.}$	$p_{.1}$	p_{11}
1	0.5	0.5	0.25;
2	0.5	0.5	0.375;
3	0.5	0.5	0.45;
4	0.9	0.9	0.81;
5	0.9	0.9	0.855;
6	0.9	0.9	0.872;
7	0.6	0.6	0.3;
8	0.2	0.2	0.1.

Si osservi che, una volta fissati $p_{1.}$ e $p_{.1}$ e p_{11} , la Tab. I è completamente determinata.

Per ogni terna di probabilità e per ogni numerosità campionaria si sono costruite, tramite elaboratore elettronico, le terne di possibili risultati $(n_{1.}, n_{.1}, n_{11})$.

Si è costruita poi la regione di confidenza effettiva per $p_{1.}$ e $p_{.1}$ data da:

$$\frac{n_{1.}}{n} \pm h(\alpha_j; Q_k) \frac{\sqrt{\frac{n_{1.}}{n} - \left(\frac{n_{1.}}{n}\right)^2}}{\sqrt{n}}, \quad (12)$$

$$\frac{n_{.1}}{n} \pm h(\alpha_j; Q_k) \frac{\sqrt{\frac{n_{.1}}{n} - \left(\frac{n_{.1}}{n}\right)^2}}{\sqrt{n}}, \quad k=1, 2, 3 \quad (13)$$

dove:

$$\begin{aligned} Q_1 &= \hat{Q} & ; \\ Q_2 &= 0 & ; \\ &= 0 & \text{ se } |\hat{Q}| \leq 0.5 \\ Q_3 &= \hat{Q} & \text{ se } |\hat{Q}| > 0.5. \end{aligned}$$

Si controlla poi se l'intervallo (12) contiene $p_{1.}$ e, contemporaneamente, se l'intervallo (13) contiene $p_{.1}$. Se ciò avviene si fa assumere valore 1 ad una variabile casuale indicatore della copertura della regione di confidenza, che verrà indicata con C_{kjs} . In caso contrario tale v.c. assume il valore 0. Si calcolano poi le probabilità che la v.c. Binomiale Bivariata assuma valori $n_{1.}$, $n_{.1}$ ed n_{11} (si veda Capobianco [1] e Zenga [7]). In base ad esse si calcola il valore atteso della v.c. C_{kjs} dato da:

$$E(C_{kjs}) = \sum_{n_{1.}=0}^n \sum_{n_{.1}=0}^n \sum_{n_{11}=\max\{0, n_{1.}+n_{.1}-n\}}^{\min\{n_{1.}, n_{.1}\}} c_{kjs} P(n_{1.}, n_{.1}, n_{11})$$

$$= \sum_{n_{1.}=0}^n \sum_{n_{.1}=0}^n \frac{\min\{n_{1.}, n_{.1}\}}{\sum_{n_{11}=\max\{0, n_{1.} + n_{.1} - n\}}^{\min\{n_{1.}, n_{.1}\}}} c_{kjs} \frac{n!}{n_{11}!(n_{1.}-n_{11})!(n_{.1}-n_{11})!(n-n_{1.}-n_{.1}+n_{11})!} \times$$

$$\times p_{11}^{n_{11}} p_{10}^{n_{1.}-n_{11}} p_{01}^{n_{.1}-n_{11}} p_{00}^{n-n_{1.}-n_{.1}+n_{11}}$$

Il valore atteso della v.c. C_{kjs} costituisce il livello effettivo di confidenza per le regioni di confidenza presentate in questo lavoro. Nella Tab. IV, V, VI vengono riportati tali valori seguendo le tre procedure proposte. I livelli effettivi di confidenza vengono considerati accettabili se non differiscono in più o in meno del 2% dal livello nominale di confidenza. I valori non considerati accettabili vengono contrassegnati con asterisco.

Come si può notare, l'approssimazione migliore fra i livelli effettivi e nominali si ha allorchè $p_{1.}$ e $p_{.1}$ non si discostano eccessivamente da 0.5 e per valori di α_j non troppo elevati. In tale situazione già $n=150$ porta ad approssimazioni soddisfacenti. Se ci si allontana da tali valori è preferibile aumentare la numerosità campionaria. I tre metodi non si discostano fra di loro sensibilmente quanto a bontà di approssimazione del livello di confidenza. Il I e il III metodo hanno livelli effettivi di confidenza praticamente uguali.

6. Conclusioni

Nel presente lavoro si sono trovate regioni di confidenza asintotiche per $p_{1.}$ e per $p_{.1}$ in una distribuzione Binomiale Bivariata che, a parità di livelli di confidenza, sono più brevi di quelli che si possono trovare col metodo di Bonferroni.

Il seguente prospetto mostra i valori di $z(\alpha/2)$ e di $h(2\alpha, \rho)$ per $\rho=0; 0.7; 0.99$.

2α	$z(\alpha/2)$	$h(2\alpha;0)$	$h(2\alpha;0.7)$	$h(2\alpha;0.99)$
0.01	2.8071	2.8063	2.7730	2.6282
0.02	2.5759	2.5741	2.5325	2.3791
0.05	2.2414	2.2365	2.1799	2.0133
0.10	1.9600	1.9488	1.8772	1.6987
0.15	1.7805	1.7622	1.6804	1.4937

Si può notare che anche nel caso più sfavorevole in cui $\rho=0$ gli intervalli di confidenza qui proposti sono migliori di quelli ottenuti col metodo Bonferroni. Il miglioramento diventa ancora più evidente all'aumentare di α e di $|\rho|$.

Gli intervalli di confidenza proposti sono applicabili per $n=150$ nel caso in cui $p_{1.}$ e $p_{.1}$ non si discostino eccessivamente da 0.5. Altrimenti conviene prendere un'ampiezza campionaria di $n=250$.

Tab. IV: Livelli di confidenza effettivi quando si usa $h(\alpha; \hat{\rho})$

$n \backslash 1-\alpha$	0.99	0.98	0.95	0.90	0.85
$(p_{1.}, p_{.1}, p_{11}) = (0.5; 0.5; 0.25)$					
150	0.98634	0.97788	0.94643	0.88355	0.83527
200	0.98866	0.97391	0.94454	0.89120	0.85223
250	0.98720	0.98134	0.94739	0.90306	0.83265 *
300	0.98692	0.98156	0.95222	0.89007	0.85875
$(p_{1.}, p_{.1}, p_{11}) = (0.5; 0.5; 0.375)$					
150	0.98689	0.97891	0.94966	0.89193	0.84774
200	0.98909	0.97519	0.94791	0.89893	0.85842
250	0.98770	0.97534	0.95055	0.90072	0.84532
300	0.98744	0.97559	0.94920	0.89788	0.83497
$(p_{1.}, p_{.1}, p_{11}) = (0.5; 0.5; 0.45)$					
150	0.98849	0.97894	0.94268	0.90122	0.83669
200	0.98913	0.97839	0.94646	0.89100	0.84457
250	0.98919	0.97783	0.94253	0.89782	0.84044
300	0.98896	0.97873	0.94832	0.88982	0.85288
$(p_{1.}, p_{.1}, p_{11}) = (0.9; 0.9; 0.81)$					
150	0.97135	0.97037	0.93189	0.85682 *	0.83597
200	0.98315	0.96368	0.92660 *	0.85954 *	0.83969
250	0.98026	0.96126	0.92716 *	0.86818 *	0.85002
300	0.97909	0.96330	0.93059 *	0.87901 *	0.81843 *
$(p_{1.}, p_{.1}, p_{11}) = (0.9; 0.9; 0.855)$					
150	0.97304	0.96921	0.93546	0.86882 *	0.84880
200	0.98010	0.96592	0.93201	0.87099 *	0.85205
250	0.98126	0.96368	0.93244	0.87862 *	0.85185
300	0.98017	0.96377	0.93549	0.88832	0.83149 *
$(p_{1.}, p_{.1}, p_{11}) = (0.9; 0.9; 0.872)$					
150	0.97489	0.95883	0.93592	0.87852 *	0.85292
200	0.97227	0.96829	0.93617	0.88094 *	0.85323
250	0.98247	0.96622	0.93723	0.88790	0.82662 *
300	0.98146	0.96623	0.94008	0.89479	0.82873 *
$(p_{1.}, p_{.1}, p_{11}) = (0.6; 0.6; 0.3)$					
150	0.98758	0.97573	0.94107	0.89509	0.84967
200	0.98809	0.97630	0.94088	0.88421	0.84236
250	0.99024	0.97988	0.95485	0.90623	0.86255
300	0.98859	0.97714	0.95098	0.89436	0.85366
$(p_{1.}, p_{.1}, p_{11}) = (0.2; 0.2; 0.1)$					
150	0.98412	0.97102	0.94433	0.89579	0.83883
200	0.98457	0.97404	0.94624	0.87983 *	0.82275 *
250	0.98645	0.97809	0.94339	0.90567	0.84767
300	0.98669	0.97339	0.93719	0.90091	0.84767

Tab. V: Livelli di confidenza effettivi quando si usa $h(\alpha; 0)$.

$1-\alpha$	0.99	0.98	0.95	0.90	0.85
n					
	$(P_{1..}, P_{.1}, P_{11}) = (0.5; 0.5; 0.25)$				
150	0.98634	0.97788	0.94643	0.88355	0.83527
200	0.98866	0.97391	0.94454	0.89120	0.85223
250	0.98720	0.98134	0.94739	0.90306	0.83265 *
300	0.98692	0.98156	0.95222	0.89007	0.85875
	$(P_{1..}, P_{.1}, P_{11}) = (0.5; 0.5; 0.375)$				
150	0.98689	0.97891	0.94966	0.89193	0.84774
200	0.98909	0.97519	0.94791	0.89893	0.86326
250	0.98770	0.98216	0.95055	0.90978	0.84532
300	0.98744	0.98237	0.95501	0.89788	0.86922 *
	$(P_{1..}, P_{.1}, P_{11}) = (0.5; 0.5; 0.45)$				
150	0.98849	0.98160	0.95653	0.90730	0.86961 *
200	0.99040	0.97839	0.95502	0.91323	0.88281 *
250	0.98919	0.98440	0.95726	0.92246 *	0.86746 *
300	0.98896	0.98458	0.96107	0.91230	0.88785 *
	$(P_{1..}, P_{.1}, P_{11}) = (0.9; 0.9; 0.81)$				
150	0.97135	0.97037	0.93189	0.85682 *	0.83597
200	0.98315	0.96368	0.92660 *	0.85954 *	0.83969
250	0.98026	0.96126	0.92716 *	0.86818 *	0.85002
300	0.97909	0.96330	0.93059 *	0.87901 *	0.81843 *
	$(P_{1..}, P_{.1}, P_{11}) = (0.9; 0.9; 0.855)$				
150	0.97304	0.97208	0.93689	0.86882 *	0.84901
200	0.98369	0.96592	0.93201	0.87099 *	0.85223
250	0.98126	0.96368	0.93244	0.87862 *	0.86154
300	0.98017	0.96562	0.93549	0.88833	0.83345 *
	$(P_{1..}, P_{.1}, P_{11}) = (0.9; 0.9; 0.872)$				
150	0.97490	0.97400	0.94162	0.87918 *	0.86099
200	0.98499	0.96831	0.93710	0.88110 *	0.86387
250	0.98248	0.96622	0.93746	0.88806	0.87237 *
300	0.98146	0.96802	0.94025	0.89693	0.84658
	$(P_{1..}, P_{.1}, P_{11}) = (0.6; 0.6; 0.3)$				
150	0.98758	0.97573	0.94107	0.89509	0.84974
200	0.98809	0.97630	0.94088	0.90219	0.84236
250	0.98875	0.98075	0.94513	0.89851	0.84722
300	0.98859	0.97714	0.95098	0.90111	0.85386
	$(P_{1..}, P_{.1}, P_{11}) = (0.2; 0.2; 0.1)$				
150	0.98412	0.97102	0.94433	0.89579	0.83883
200	0.98579	0.97404	0.95324	0.89097	0.85721
250	0.98645	0.97828	0.94339	0.90643	0.86751 *
300	0.98868	0.97339	0.93719	0.90091	0.86432

Tab. VI: Livelli di confidenza effettivi quando si usa $h(\alpha; 0)$ per $|\hat{\varrho}| < 0.5$ e $h(\alpha; \hat{\varrho})$ per $|\hat{\varrho}| > 0.5$.

$n \backslash 1-\alpha$	0.99	0.98	0.95	0.90	0.85
$(p_{1.}, p_{.1}, p_{11}) = (0.5; 0.5; 0.25)$					
150	0.98634	0.97788	0.94643	0.88355	0.83527
200	0.98866	0.97391	0.94454	0.89120	0.85223
250	0.98720	0.98134	0.94739	0.90306	0.83265 *
300	0.98692	0.98156	0.95222	0.89007	0.85875
$(p_{1.}, p_{.1}, p_{11}) = (0.5; 0.5; 0.375)$					
150	0.98689	0.97891	0.94966	0.89193	0.84774
200	0.98909	0.97519	0.94791	0.89893	0.85842
250	0.98770	0.97855	0.95055	0.90072	0.84532
300	0.98744	0.97914	0.94920	0.89788	0.85206
$(p_{1.}, p_{.1}, p_{11}) = (0.5; 0.5; 0.45)$					
150	0.98849	0.97894	0.94268	0.90122	0.83669
200	0.98913	0.97839	0.94646	0.89100	0.84457
250	0.98919	0.97783	0.94253	0.89782	0.84044
300	0.98896	0.97873	0.94832	0.88982	0.85288
$(p_{1.}, p_{.1}, p_{11}) = (0.9; 0.9; 0.81)$					
150	0.97135	0.97037	0.93189	0.85682 *	0.83597
200	0.98315	0.96368	0.92660 *	0.85954 *	0.83969
250	0.98026	0.96126	0.92716 *	0.86818 *	0.85002
300	0.97909	0.96330	0.93059 *	0.87901 *	0.81843 *
$(p_{1.}, p_{.1}, p_{11}) = (0.9; 0.9; 0.855)$					
150	0.97304	0.96921	0.93546	0.86882 *	0.84880
200	0.98010	0.96592	0.93201	0.87099 *	0.85205
250	0.98126	0.96368	0.93244	0.87862 *	0.85185
300	0.98017	0.96422	0.93549	0.88832	0.83149 *
$(p_{1.}, p_{.1}, p_{11}) = (0.9; 0.9; 0.872)$					
150	0.97489	0.95883	0.93592	0.87852 *	0.85292
200	0.97227	0.96829	0.93617	0.88094 *	0.85323
250	0.98247	0.96622	0.93723	0.88790	0.82662 *
300	0.98146	0.96623	0.94008	0.89479	0.82873 *
$(p_{1.}, p_{.1}, p_{11}) = (0.6; 0.6; 0.3)$					
150	0.98758	0.97573	0.94107	0.89509	0.84974
200	0.98809	0.97630	0.94088	0.90219	0.84236
250	0.99024	0.97988	0.95485	0.90623	0.86255
300	0.98859	0.97714	0.95098	0.90111	0.85386
$(p_{1.}, p_{.1}, p_{11}) = (0.2; 0.2; 0.1)$					
150	0.98412	0.97102	0.94433	0.89579	0.83883
200	0.98568	0.97404	0.95236	0.88986	0.85586
250	0.98645	0.97813	0.94339	0.90585	0.86575
300	0.98862	0.97339	0.93719	0.90091	0.86350

Bibliografia

- [1] M.F. CAPOBIANCO (1964). *On the Bivariate Binomial distribution and related problems*, Ph. D. Thesis Polytechnic Institute of Brooklyn, Department of Mathematics.
- [2] M. FEDRIZZI, M. ZENGA (1980). The Arctangent distribution and the simultaneous confidence intervals for trinomial proportions, *Redazione provvisoria*, Dipartimento di Metodi Quantitativi, Libera Università degli Studi di Trento, pp. 1-8.
- [3] R.G. MILLER (1966), *Simultaneous Statistical Inference*, McGraw Hill, New York.
- [4] D.B. OWEN (1956), Tables for computing bivariate normal probabilities, *Annals of Mathematical Statistics*, Vol. 27, pp. 1075-1090.
- [5] A. POLLASTRI (1980), Intervalli di confidenza simultanei asintotici fra probabilità in una trinomia, *Atti della XXX Riunione della Società Italiana di Statistica*, Vol. II, pp. 59-68.
- [6] S.S. WILKS (1962), *Mathematical Statistics*, J. Wiley, New York.
- [7] M. ZENGA (1968), La distribuzione binomiale bivariata, *Statistica*, Vol. 28, pp. 83-101.
- [8] M. ZENGA (1978), L'impiego della funzione Arcotangente incompleta nello studio della distribuzione asintotica dello scarto standardizzato assoluto massimo di una trinomia, *Statistica*, Vol. 39, pp. 269-286.

Ringraziamenti: Ringrazio il Prof. Michele Zenga per le utili indicazioni fornitemi.

SUMMARY**Asymptotique simultaneous confidence intervals for the marginal probabilities in a 2x2 table**

In the present article we present a method for obtaining simultaneous confidence intervals for the probabilities of success of two qualities of a population when we draw with replacement n elements from it.

We show that the maximum of the standardized frequencies of the two qualities is di-

distributed as a mixture of two Arcotangent random variables and we use this distribution for obtaining asymptotique simultaneous confidence intervals for the marginal probabilities in a 2×2 table. We compare the confidence intervals obtained through this method with those we can obtain using the Bonferroni inequality. Then we give indications about the sample size for obtaining the intervals proposed.

RESUME

Intervalles de confiance simultanés asymptotiques des probabilités marginales dans un tableau 2×2 .

L'article traite le problème de la construction des intervalles de confiance des probabilités de succès pour deux qualités dichotomiques d'une population quand on tire au hasard n éléments.

On démontre que la distribution asymptotique du maximum des fréquences standardisées est le mélange de deux variables aléatoires Arcotangents et on utilise cette distribution pour obtenir les intervalles de confiance.

On montre que les intervalles ici obtenus sont toujours meilleurs de ceux qu'on peut obtenir avec l'inégalité de Bonferroni et on donne des indications sur la dimension de l'échantillon.