SUGGERIMENTI PER L'IMPIEGO DEL TEST CHI-QUADRATO *

di *Michele Zenga* **

1. *Introduzione e sommario*

Una delle critiche più frequentemente fatte al test chi-quadrato è che per n (numerosità campionaria) grande, differenze dalla ipotesi nulla anche di lieve entità portano quasi sempre all'accettazione della ipotesi alternativa (il test sarebbe troppo sensibile).

In questa nota, dopo aver mostrato l'opportunità di misurare la distanza delle probabilità p_i $(i = 1, 2, \ldots, k)$ di una alternativa dalle probabilità π_i della ipotesi nulla, facendo riferimento alla funzione

$D = \max \left\{ \dfrac{|p_i - \pi_i|}{\pi_i} ; i = 1, 2, \ldots k \right\}$, si è proceduto alla determinazione delle numerosità campionarie necessarie per avere valori piccoli di α (probabilità errore di prima specie) e di max β (β è la probabilità di un errore di seconda specie; max β è riferito alle alternative con $D \geqslant \delta$, essendo $\delta > 0$).

Le numerosità campionarie così ottenute sono molto elevate ed indicano che il test chi-quadrato è meno sensibile di quanto si ritiene e che quindi le critiche di cui sopra non sembrano molto fondate.

Da ultimo, si è mostrato come utilizzare la statistica X^2 di Karl-Pearson e la funzione D per ottenere un test con errori proporzionali, cioè un test che soddisfa la condizione $\dfrac{\max \beta}{\alpha} = \nu$, essendo ν un valore positivo prefissato.

2. *Distanza dell'alternativa*

Si abbia una multinomiale a k classi con probabilità $p_1, p_2, \ldots, p_k$ e con numero di prove n.

Si supponga di voler verificare l'ipotesi $H_0 : (p_1 = \pi_1)$ e $(p_2 = \pi_2)$

---

* Lavoro presentato alla XXIX riunione scientifica della Società Italiana di Statistica, Bologna 20 - 22 marzo 1978.

** Libera Università degli Studi di Trento - Facoltà di Economia e Commercio.

e . . . . e $(p_k = \pi_k)$, contro l'alternativa $H_1 : (p_1 \neq \pi_1)$ o $(p_2 \neq \pi_2)$ o . . . . o $(p_k \neq \pi_k)$.

Come è noto, tale verifica si esegue ricorrendo alla statistica $X^2 = \Sigma \dfrac{(n_i - n\pi_i)^2}{n\,\pi_i}$, essendo n_i il numero delle volte in cui nelle n prove si è verificata la classe i^{ma}, ovviamente $\Sigma n_i = n$. Sotto Ho la statistica X^2 tende a distribuirsi, per n grande, secondo una v.c. chi-quadrato centrale con $k - 1$ gradi di libertà. Sotto una qualsiasi alternativa $(p_1, p_2, \ldots, p_k)$ la variabile X^2 tende a distribuirsi, per n grande, secondo una v.c. chi-quadrato non centrale con $k - 1$ gradi di libertà e parametro di non centralità $\lambda^2 = n\,\Sigma \dfrac{(p_i - \pi_i)^2}{\pi_i}$.

Alcuni studiosi sostengono [1] che per valori elevati di n, anche ai più bassi livelli di significatività, il test basato su X^2 porta a respingere l'ipotesi Ho anche per alternative assai prossime ad Ho. Queste affermazioni sono basate, più che sull'effettivo calcolo della funzione di potenza, sulla proprietà della consistenza del test X^2. La consistenza è però una proprietà asintotica per cui non è possibile sapere da quali valori di n si possa ritenere trascurabile la probabilità β.

Si mostrerà in seguito che, se fra l'alternativa ed Ho vi è una piccola distanza e si vogliono avere valori di α e di β piccoli, è necessario impiegare valori molto elevati di n.

Non potendo però disporre nella pratica di valori tanto elevati di n, si impone l'esigenza di considerare le alternative aventi almeno una certa distanza da Ho e di trascurare le alternative più vicine. Trascurare le alternative molto prossime ad Ho non sembra possa portare a conseguenze pratiche, perché al ricercatore interessa scoprire se l'alternativa si discosti o meno di una certa entità da Ho.

Inoltre, si tenga presente che una certa distanza fra Ho e l'alternativa è necessaria anche per motivi strettamente tecnici, in quanto anche per valori elevati di n, se fra una alternativa ed Ho non vi è una certa distanza, le distribuzioni di X^2 sotto le due ipotesi tendono a confondersi ed è allora difficile discriminare fra le stesse.

Vi sono diversi modi per valutare la distanza di una alternativa da Ho. Solitamente [2] si ricorre al parametro di non centralità $\lambda^2 = n\,\Sigma \dfrac{(p_i - \pi_i)^2}{\pi_i}$. Ciò è forse dovuto al fatto che la distribuzione di X^2 dipende solo da $(k - 1)$ e da λ^2.

A fronte di questi vantaggi, λ^2 presenta alcuni inconvenienti. Innanzitutto λ^2 fa dipendere la distanza di una alternativa da Ho dalla numerosità campionaria n. Questo inconveniente è facilmente eliminabile riferendosi ad $\eta^2 = \Sigma \dfrac{(p_i - \pi_i)^2}{\pi_i}$. Ma anche η^2 presenta ancora degli inconvenienti in quanto il ricercatore non riesce a valutare che cosa voglia significare un certo valore di η^2, essendo lo stesso, generalmente, più abituato a ragionare in termini di differenze percentuali fra p_i e π_i, piuttosto che sulle differenze al quadrato fra p_i e π_i rapportate a π_i.

Sembra, pertanto, più utile fornire un procedimento, per misurare la distanza, basato sulle differenze percentuali assolute $\dfrac{|p_i - \pi_i|}{\pi_i}$.

Ad esempio, si potrebbe considerare la seguente funzione degli scarti percentuali assoluti: $D = \max \left\{ \dfrac{|p_i - \pi_i|}{\pi_i} ; \ i = 1, 2, \ldots, k \right\}$.

Al ricercatore, potrebbero interessare solo quelle alternative per le quali $D \geqslant \delta$, essendo $\delta > 0$. D'ora in avanti si indicherà con Ω_δ il sottospazio dello spazio parametrico caratterizzato da $D \geqslant \delta$. Affermare che $D \geqslant \delta$ è equivalente ad affermare che almeno una differenza percentuale assoluta $\dfrac{|p_i - \pi_i|}{\pi_i}$ sia maggiore od uguale a δ.

Ovviamente, è possibile far riferimento anche ad altre funzioni delle differenze percentuali assolute per misurare la distanza fra una alternativa ed Ho. Ad esempio, si potrebbe considerare la media aritmetica delle percentuali assolute $\dfrac{1}{k} \Sigma \dfrac{|p_i - \pi_i|}{\pi_i}$ e considerare le alternative per le quali detta media sia maggiore o uguale a v, essendo $v > 0$. Come anche, si potrebbe far riferimento ad una funzione delle differenze assolute. In questo lavoro verrà considerata la funzione:

$$D = \max \left\{ \frac{|p_i - \pi_i|}{\pi_i} ; \ 1, 2, \ldots\ldots, k \right\}.$$

Sembra ora utile avere qualche idea sulla forma della porzione dello spazio parametrico caratterizzato da $D \leqslant \delta$.

Lo spazio parametrico Ω può considerarsi un sottospazio dello spazio euclideo K-dimensionale E^k i cui punti $(p_1, p_2, \ldots\ldots p_k)$ sono tali che: $\Sigma p_i = 1$, $p_i \geqslant 0$ $(i = 1, 2, \ldots\ldots k)$. Sia $(\pi_1, \ldots\ldots, \pi_k)$ un punto prefissato di Ω e si definisca la funzione

$$D = \max \left\{ \frac{|p_i - \pi_i|}{\pi_i} ; \ i = 1, 2, \ldots, k \right\}, \tag{1}$$

essendo $(p_1, p_2, \ldots\ldots, p_k)$ un punto arbitrario di Ω.

Fissato un numero reale $\delta > 0$ consideriamo il luogo dei punti che soddisfano la condizione: $\max \left\{ \frac{|p_i - \pi_i|}{\pi_1} ; i = 1, 2, \ldots, k \right\} = \delta$. (2)

Si controlla subito che trattasi di una ipersuperficie di dimensione $(k-1)$, e precisamente di un poliedro le cui facce sono di dimensione $(k-2)$ e sono caratterizzate dall'avere una coordinata p_i $(i = 1, 2, \ldots, k)$ costante. In particolare le facce sono a due a due parallele, poiché giacciono su iperpiani di equazioni: $p_i = \pi_i + \delta\pi_i$ e $p_i = \pi_i - \delta\pi_i$.

Non sempre però le facce sono 2k, come si riscontrerà negli esempi che seguono. Tuttavia, come verrà mostrato in seguito, ciò non ha nessuna conseguenza operativa nell'impiego del test. Per esemplificare consideriamo il caso $k = 3$ con $\pi_1 = 0{,}4$, $\pi_2 = 0{,}3$ e $\pi_3 = 0{,}3$ e con $\delta = 0{,}2$.

I punti dello spazio parametrico in cui $D \leqslant \delta$ sono caratterizzati dal contemporaneo realizzarsi delle seguenti relazioni:

$$0{,}32 \leqslant p_1 \leqslant 0{,}48$$
$$0{,}24 \leqslant p_2 \leqslant 0{,}36$$
$$0{,}64 \leqslant p_1 + p_2 \leqslant 0{,}76$$

Conseguentemente, il contemporaneo verificarsi delle relazioni precedenti dà luogo in Ω all'esagono indicato nel grafico 1.

Non sempre però, per $k = 3$, il luogo dei punti che si ottiene ponendo $D = \delta$ dà luogo ad un esagono.

Per mostrare ciò e senza perdere in generalità si supponga $\pi_1 \geqslant \pi_2 \geqslant \pi_3$, il che implica $\pi_1 \geqslant 1/3$. Si ricava facilmente che se $1/3 \leqslant \pi_1 < 1/2$ si ha un esagono e se $\pi_1 \geqslant 1/2$ si ha un quadrilatero, come è indicato nel grafico 2 in cui si è supposto $\pi_1 = 0{,}6$, $\pi_2 = 0{,}2$ e $\pi_3 = 0{,}2$ e $\delta = 0{,}20$.

3. *Riformulazione del problema ed alternativa più sfavorevole.*

È possibile riformulare il problema di verifica di ipotesi riguardante la multinomiale ipotizzando che il sottospazio parametrico dell'ipotesi alternativa H_1 sia Ω_δ.

Per il problema di verifica di ipotesi di cui sopra è possibile far ancora riferimento alla statistica X^2.

Si supponga che il valore critico della variabile test sia c. Per n sufficientemente elevato il valore di α è approssimato bene dalla probabilità che la variabile chi-quadrato con $k-1$ gradi di libertà, che si indica con $X^2_{(k-1)}$, sia maggiore o uguale a c. Analogamente per la alternativa $(p_1, p_2, \ldots, p_k) \in \Omega_\delta$ e sempre per n grande il valore di β

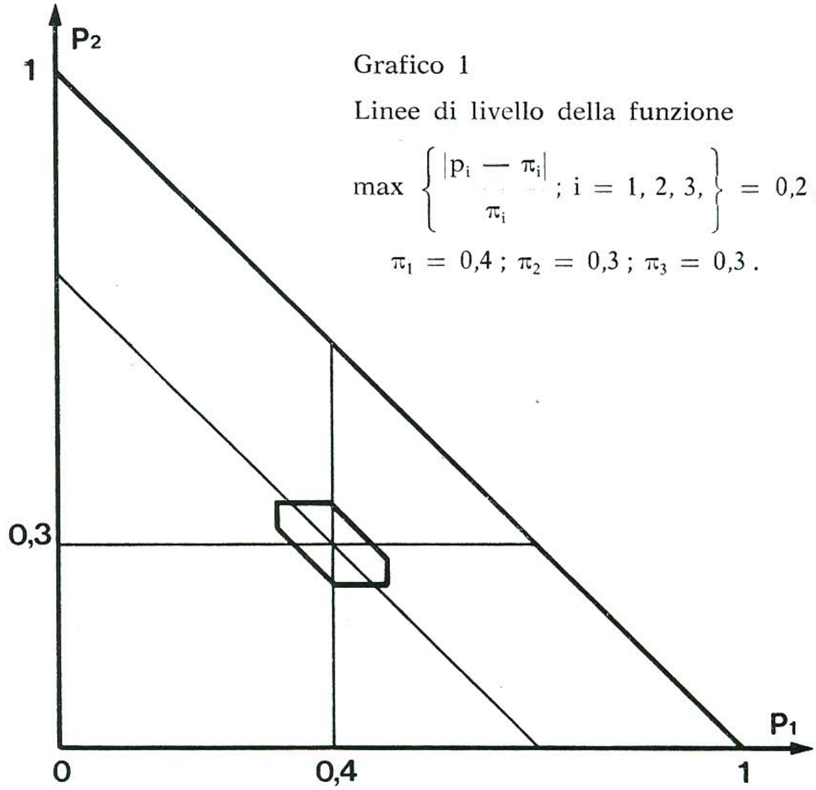

Grafico 1

Linee di livello della funzione

$$\max \left\{ \frac{|p_i - \pi_i|}{\pi_i} ; \ i = 1, 2, 3, \right\} = 0{,}2 \text{ , per:}$$

$$\pi_1 = 0{,}4 \ ; \ \pi_2 = 0{,}3 \ ; \ \pi_3 = 0{,}3 \ .$$

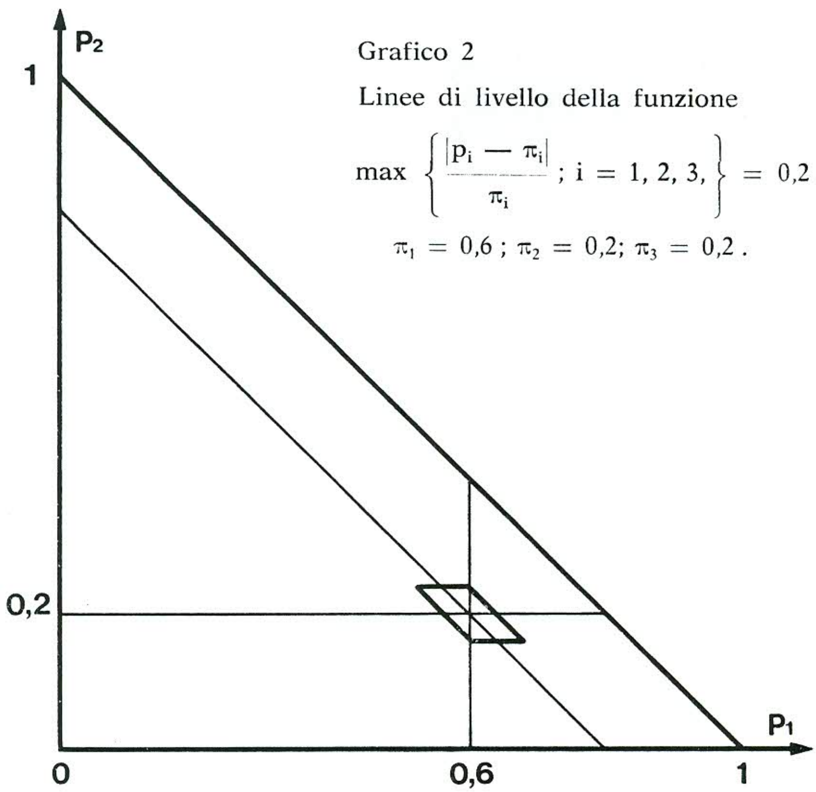

Grafico 2

Linee di livello della funzione

$$\max \left\{ \frac{|p_i - \pi_i|}{\pi_i} ; \ i = 1, 2, 3, \right\} = 0{,}2 \text{ , per:}$$

$$\pi_1 = 0{,}6 \ ; \ \pi_2 = 0{,}2; \ \pi_3 = 0{,}2 \ .$$

è approssimato dalla probabilità che la variabile casuale chi-quadrato non centrale con $k-1$ gradi di libertà e con parametro di non centralità $\lambda^2 = n \Sigma \dfrac{(p_i - \pi_i)^2}{\pi_i}$ assuma un valore inferiore a c.

È noto che la distribuzione della v.c. chi-quadrato non centrale dipende, a parità di gradi di libertà, solo dal valore assunto da λ^2. Inoltre è altresì noto che, per un prefissato c, il valore di β diminuisce al crescere di λ^2.

Conseguentemente l'alternativa $(p_1, p_2, \ldots, p_k) \in \Omega_\delta$ in corrispondenza della quale β assume il valore massimo è quella che minimizza λ^2, o, ciò che è lo stesso, che minimizza η^2.

La funzione η^2 è strettamente convessa ed ha un minimo assoluto in corrispondenza di $(\pi_1, \pi_2, \ldots, \pi_k)$; conseguentemente nello spazio Ω_δ la funzione η^2 assumerà il suo valore minimo su una o più facce del poliedro (2). Non resta che trovare il minimo di η^2 su ciascuna faccia del poliedro e poi il minimo di questi minimi.

Si è precisato in precedenza che le facce del poliedro giacciono su iperpiani di equazioni: $p_i = \pi_i + \delta\pi_i$ e $p_i = \pi_i - \delta\pi_i$.

Determiniamo il minimo di η^2 soggetto alla condizione $p_i = \bar{p}_i$ $(\bar{p}_i = \pi_i + \delta\pi_i$ o $\bar{p}_i = \pi_i - \delta\pi_i)$ nonché alla condizione $\Sigma p_i = 1$.

Ricorrendo ai moltiplicatori di Langrange si tratta di minimizzare la funzione ausiliaria:

$$F = \frac{(p_1 - \pi_1)^2}{\pi_1} + \frac{(p_2 - \pi_2)^2}{\pi_2} + \ldots + \frac{(\bar{p}_i - \pi_i)^2}{\pi_i} + \ldots + \frac{(p_k - \pi_k)^2}{\pi_k}$$

$+ \mu (p_1 + p_2 + \ldots + \bar{p}_i + \ldots + p_k - 1)$, essendo μ il moltiplicatore di Langrange. Si ricava immediatamente che il minimo condizionato di η^2 si ha in corrispondenza delle coordinate $(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_k)$ date da:

$\hat{p}_j = \pi_j \dfrac{1-\bar{p}_i}{1-\pi_i}$ per $j \neq i$ e $\hat{p}_i = \bar{p}_i$. Per $\bar{p}_i = \pi_i - \delta\pi_i$ le espressioni di $\hat{p}_j$ diventano: $\hat{p}_j = \pi_j \left(1 + \delta \dfrac{\pi_i}{1-\pi_i}\right)$ per $j \neq i$ e $\hat{p}_i = \pi_i - \delta\pi_i$.

Con facili passaggi si ricava che in corrispondenza di detto punto:

$\eta^2 = \delta^2 \dfrac{\pi_i}{1-\pi_i}$. Per $\bar{p}_i = \pi_i + \delta\pi_i$ le espressioni di $\hat{p}_j$ diventano:

$\hat{p}_j = \pi_j \left(1 - \delta \dfrac{\pi_i}{1-\pi_i}\right)$ per $j \neq i$ e $\hat{p}_i = \pi_i + \delta\pi_i$.

Anche in corrispondenza di detto punto risulta ancora $\eta^2 = \delta^2 \dfrac{\pi_i}{1-\pi_i}$.

È quindi indifferente, ai fini della determinazione del minimo di η^2, considerare l'iperpiano $p_i = \pi_i + \delta\,\pi_i$ o $p_i = \pi_i - \delta\,\pi_i$.

Il minimo fra i minimi relativi si ha in corrispondenza di $\pi^* = \min \{\pi_i\,;\ i = 1, 2, \ldots, k\}$.

Dalla convenzione $\pi_1 \geqslant \pi_2 \geqslant \ldots \geqslant \pi_k$ si deduce, infine, che il minimo di η^2 si ha sugli iperpiani: $p_k = \pi_k + \delta\,\pi_k$ o $p_k = \pi_k - \delta\,\pi_k$.

Resta ancora da dimostrare che il punto che minimizza η^2 sugli iperpiani sopra menzionati appartenga effettivamente a Ω_δ. Per provare ciò è sufficiente provare che le coordinate del punto $(\hat{p}_1, \hat{p}_2, \ldots .\hat{p}_k)$ giacente sull'iperpiano $p_k = \pi_k + \delta\,\pi_k$ o sull'iperpiano $p_k = \pi_k - \delta\,\pi_k$ siano comprese negli intervalli $\pi_i\,(1 - \delta) \leqslant \hat{p}_i \leqslant \pi_i\,(1 + \delta)$.
Per $p_k = \pi_k\,(1 + \delta)$ si ha $\hat{p}_i < \pi_i$ $(i = 1, 2, \ldots, k - 1)$, basterà allora mostrare che $\hat{p}_i \geqslant \pi_i\,(1 - \delta)$, per $i = 1, 2, \ldots., k - 1$.

Sappiamo che $\hat{p}_i = \pi_i\,(1 - \delta \dfrac{\pi_k}{1 - \pi_k})$. È allora sufficiente mostrare che $\dfrac{\pi_k}{1 - \pi_k} < 1$.

La funzione $\dfrac{\pi_k}{1 - \pi_k}$ cresce al crescere di π_k ed è pari ad 1 per $\pi_k = 1/2$. Essendo $\pi_k = \pi^*$, deve necessariamente essere $\pi_k < 1/2$. Conseguentemente: $\pi_i\,(1 - \delta) \leqslant \hat{p}_i \leqslant \pi_i$ $(i = 1, 2, \ldots, k - 1)$.

Analogamente si dimostra che se $p_k = \pi_k\,(1 - \delta)$ allora: $\pi_i \leqslant \hat{p}_i \leqslant \pi_i\,(1 + \delta)$ (per $i = 1, 2, \ldots, k - 1$).

Si può pertanto concludere che:

$$\min_{(p_1, p_2, \ldots., p_k) \in \Omega_\delta} \eta^2 = \delta^2 \frac{\pi^*}{1 - \pi^*}$$

4. *Numerosità campionarie ed errori di prima e seconda specie.*

Per constatare se il test chi-quadrato è più o meno sensibile si sono ricavate le numerosità campionarie in corrispondenza di alcuni valori di k, δ ed α che garantiscono l'uguaglianza fra α e max β. Per semplicità si è ipotizzato che sotto Ho sia $\pi_i = \dfrac{1}{k}$ $(i = 1, 2, \ldots. k)$. Le numerosità campionarie si sono ottenute ricorrendo alle tavole della v. c. chi-quadrato non centrale di Johnson e Pearson [3] e sono state riportate nella Tav. I.

Per comodità si riporta lo schema dei calcoli per ricavare n in corrispondenza di $k = 3$, $\delta = 0{,}05$ e $\alpha = 0{,}05$.

Tab. I

Numerosità campionarie n in corrispondenza di alcuni valori di k, δ ed α per avere $\alpha = \max \beta$

	$\delta = 0{,}01$	$\delta = 0{,}02$	$\delta = 0{,}05$	$\delta = 0{,}10$	$\delta = 0{,}20$
		$\alpha = 0{,}05$			
k = 2	129.960	32.490	5.198	1.300	325
3	309.148	77.287	12.365	3.091	772
4	520.385	129.118	20.608	5.204	1.288
6	988.775	247.194	39.551	9.888	2.472
10	2.124.468	531.117	85.132	21.245	5.311
		$\alpha = 0{,}025$			
k = 2	176.400	44.100	7.056	1.764	441
3	410.961	102.740	16.438	4.110	1.027
4	685.035	169.971	27.138	6.850	1.696
6	1.285.040	321.260	51.402	12.850	3.212
10	2.725.810	681.453	108.954	27.258	6.815
		$\alpha = 0{,}01$			
k = 2	240.296	60.074	9.612	2.403	601
3	547.880	136.970	21.915	5.479	1.369
4	903.712	224.229	35.801	9.037	2.237
6	1.674.989	418.737	66.998	16.750	4.187
10	3.495.535	873.873	139.719	34.923	8.738
		$\alpha = 0{,}005$			
k = 2	289.766	72.442	11.591	2.898	724
3	653.142	163.285	26.125	6.531	1.633
4	1.070.064	265.505	42.391	10.700	2.649
6	1.964.515	491.129	78.581	19.505	4.911
10	4.062.272	1.015.563	162.373	40.623	10.248

Per avere $\alpha = 0{,}05$ è necessario che la soglia critica c della variabile $\sqrt{X^2}$ sia $c = 2{,}4494$. In corrispondenza di detta soglia critica per avere $\beta = 0{,}05$ è necessario che il valore del parametro di non centralità λ^2 sia 15,457.

Sapendo che $\min \eta^2 = (0{,}05)^2 . \dfrac{1/3}{1 - 1/3} = 0{,}00125$ e che $\lambda^2 = n . \eta^2$ si ricava $n = 12.365$.

In maniera del tutto analoga si sono ricavati gli altri valori di n riportati nella Tav. I.

L'esame della Tav. I porta a fare le seguenti considerazioni. La numerosità campionaria:

a) diminuisce all'aumentare di δ, cioè della distanza fra Ho e H_1;

b) diminuisce all'aumentare di α e di $\max \beta$, cioè all'aumentare del rischio probabilistico di commettere un errore di prima o di seconda specie;

c) aumenta all'aumentare di k.

Per valori di $\delta \leqslant 0{,}02$ sono necessarie numerosità campionarie che solitamente non trovano riscontro nella pratica e viene così smentita l'affermazione che per valori elevati di n il test sia molto sensibile tanto da portare sempre all'accettazione della ipotesi H_1.

Anche per $\delta = 0{,}05$, specie per i valori più piccoli di α si hanno numerosità campionarie sufficientemente elevate. Sembra comunque potersi affermare che per $\delta = 0{,}05$ se si elevasse il valore di α e si considerassero rapporti probabilistici $\dfrac{\max \beta}{\alpha} > 1$ si dovrebbero avere numerosità campionarie accessibili.

Per $\delta = 0{,}10$ e $0{,}20$ si hanno numerosità campionarie non molto elevate per i valori più piccoli di k.

5. *Il test chi-quadrato con errori proporzionali.*

In un problema di verifica di ipotesi con numerosità campionaria fissata non è possibile controllare contemporaneamente il valore di α ed il valore di β. Spesso, ai consueti livelli di significatività α, si contrappongono valori molto elevati di β per valori di n non molto elevati e valori molto piccoli di β per valori molto elevati di n.

Per superare questo inconveniente è stato suggerito [4] di impiegare test che garantiscano un prefissato rapporto v fra β e α. Si mostrerà ora come impiegare la statistica X^2 per ricavare un test con errori proporzionali.

Si voglia verificare l'ipotesi Ho che le probabilità di una multinomiale siano: $(p_1 = \pi_1)$ e $(p_2 = \pi_2)$ e .... e $(p_k = \pi_k)$, contro l'alternativa H_1 che $\max \left\{ \frac{|p_i - \pi_i|}{\pi_i} ; i = 1, 2, \ldots, k \right\} \geqslant \delta$ essendo $\delta > 0$. Si vuol impiegare un test con errori proporzionali.

Nel caso in cui Ho sia una ipotesi semplice ed H_1 una ipotesi composita è stato suggerito [4] che il test proporzionale debba soddisfare la relazione $\frac{\max \beta}{\alpha} = v$. Per questo tipo di problema si può impiegare la consueta statistica X^2. Dalle pagine precedenti si sa che $\max \beta$ si ha per $\eta^2 = \delta^2 \frac{\pi^*}{1 - \pi^*}$ cui corrisponde un parametro di non centralità $\lambda^2 = n \cdot \eta^2$ della statistica X^2.

Si introduca ora, per un prefissato n, la funzione

$$F_n(c) = \frac{\max P\,[X^2 \leqslant c \,/\, H_1]}{P\,[X^2 > c \,/\, Ho]} = \frac{\max \beta\,(c)}{\alpha\,(c)}$$

che gode delle seguenti proprietà:

a) è sempre $\geqslant 0$,

b) è non decrescente.

Tenuto conto però che le probabilità $\max \beta(c)$ e $\alpha(c)$ si approssimano utilizzando le distribuzioni delle variabili casuali continue x^2 è possibile ritenere, per fini pratici, che $F_n(c)$ sia crescente e continua. Conseguentemente per un prefissato $v > 0$ si avrà una soglia critica $c_n(v)$ tale che $F_n(c_n(v)) = v$.

La zona di accettazione di H_0 è data da $X^2 < c_n(v)$. Non è necessario però costruire tavole ad hoc per poter impiegare il test chi-quadrato con errori proporzionali. Infatti, una volta conosciuto il valore sperimentale X^2 che indichiamo con x^2 sarà possibile ricavare:

a) $\alpha_s = P\,[X^2 \geqslant x^2 \,/\, Ho]$ tramite le tavole della v.c. chi-quadrato centrale;

b) $\max \beta_s = \max P\,[X^2 < x^2 \,/\, H_1]$ tramite le tavole della v.c. chi-quadrato non centrale con parametro di non centralità $\lambda^2 = n\,\delta^2 \frac{\pi^*}{1 - \pi^*}$.

Successivamente si ricaverà il rapporto probabilistico sperimentale $v_s = \frac{\max \beta_s}{\alpha_s}$. Se $v_s \geqslant v$ si respinge Ho, se è invece $v_s <$ si accetterà la predetta ipotesi.

Esempio : Numeri casuali.

Si supponga di voler verificare l'ipotesi che i 1000 numeri interi riportati in [5] a pag. 125 possano ritenersi casuali. Si può tentare di risolvere il quesito nel modo che segue. Sotto l'ipotesi Ho si ha un modello probabilistico che prevede 10 eventi rappresentati dalle cifre 0, 1, 2, ..., 9 cui sono associate probabilità pari a 1/10. Sotto l'ipotesi alternativa dette probabilità non sono tutte eguali ad 1/10. Le conseguenze di ritenere buone le tavole, da cui provengono i 1000 numeri, quando ciò non è vero, sembrano molto rilevanti e ovviamente tanto più rilevanti quanto più ci si allontana dalla ipotesi Ho. Dovendo, pur tuttavia, tollerare una sia pur minima discrepanza da Ho sembra ragionevole considerare solo quelle alternative per le quali almeno una probabilità differisca da 1/10 per una percentuale superiore al 5%. In altre parole almeno una probabilità deve trovarsi fuori dall'intervallo 0,095 $\div$ 0,105.

Indicando con $p_o, p_1, p_2, \ldots p_9$ le probabilità associate alle cifre 0, 1, 2, ... 9, lo spazio parametrico della ipotesi alternativa è caratterizzato da: $\max \left\{ \frac{|p_i - \pi_i|}{0,1}; \; i = 0, 1, 2, \ldots, 9 \right\} \geqslant 0,05$. Inoltre, sembra anche ragionevole far sì che α e max β assumano lo stesso valore; cioè sembra opportuno far riferimento ad un test con errori proporzionali in cui il rapporto probabilistico sia $\frac{\max \beta}{\alpha} = 1$.

Si noti che se la numerosità campionaria è data e si è fissata la minima distanza della alternativa non è possibile controllare numericamente sia α sia max β. In questi casi sembra quanto mai opportuno far riferimento ai test con errori proporzionali.

Conteggiando le frequenze delle cifre 0, 1, 2, ... 9 dal summenzionato volume si ottiene la seguente distribuzione:

numeri	0	1	2	3	4	5	6	7	8	9	tot.
frequenze	111	87	94	105	92	105	108	109	101	88	1000

Il valore sperimentale di X^2 è risultato 7,3. La probabilità α_s che detto valore sotto Ho venga superato è: $\alpha_s = P\,[X^2 \geqslant 7,3 / Ho] = 0,60$.

Dalla proprietà che la v.c. chi-quadrato non centrale è stocasticamente maggiore della corrispettiva v.c. chi-quadrato centrale, risulta:

$$\max \beta_s = \max P\,[X^2 < 7,3 / H_1] < 1 - \alpha_s .$$

Conseguentemente $\frac{\max \beta_s}{\alpha_s} \leqslant \frac{0,4}{0,6}$ e quindi si accetta l'ipotesi Ho. L'esempio in esame mostra che con i dati sperimentali a disposizione qualunque fosse stato il valore di δ bisogna accettare l'ipotesi Ho a meno di considerare un rapporto probabilistico $\frac{\max \beta}{\alpha} < \frac{2}{3}$.

BIBLIOGRAFIA

[1] J. BERKSON (1938). *Some difficulties of interpretation encountered in the application of Chi-square test.* J.A.S.A., 33; pag. 526-536.

[2] R. DE CRISTOFARO (1976). *Sulla valutazione della bontà di adattamento di una distribuzione.* Rivista di statistica applicata, 9; pag. 239-255.

[3] N. L. JOHNSON and E. S. PEARSON (1969). *Tables of percentage points of non-central* X. Biometrika, 56; pag. 255-272.

[4] M. ZENGA (1975). *I tests con errori proporzionali.* Statistica XXXV; pag. 731-750.

[5] J. L. HODGES e E. L. LEHMANN (1971). *I concetti fondamentali della probalità e della statistica.* Vol. I, Il Mulino, Bologna.