Quaderni di Statistica e Matematica
applicata alle Scienze Economico-Sociali
VOLUME I — N. 2 — OTTOBRE 1978

LA DISTANZA DELLE ALTERNATIVE NELLE VERIFICHE DI IPOTESI RIGUARDANTI LE DISTRIBUZIONI MULTINOMIALI *

di Michele ZENGA **

1. *Introduzione e sommario*

In un precedente lavoro [1] si è mostrata l'opportunità di misurare in una multinomiale a k classi la distanza di una alternativa, caratterizzata dalle probabilità $(p_1, p_2, \ldots, p_k)$, dalle probabilità $(\pi_1, \pi_2, \ldots, \pi_k)$ dell'ipotesi di nullità, utilizzando la funzione $\triangle = \max \left\{ \frac{|p_i - \pi_i|}{\pi_i} ;\ i = 1, 2, \ldots, k \right\}$.

In [1] si è anche proceduto alla determinazione delle numerosità campionarie necessarie, impiegando il consueto test basato sulla statistica X^2 di Karl-Pearson, per avere valori piccoli di α (probabilità di errori di prima specie) e di max β (β è la probabilità di errori di seconda specie; max β è riferito alle alternative con $\triangle \geqslant \delta$, essendo $\delta > 0$). Una caratteristica singolare riscontrata dall'esame di dette numerosità campionarie è che esse aumentano enormemente all'aumentare di k.

In questo lavoro si effettueranno dei confronti fra $\triangle$ e la consueta funzione di distanza $H^2 = \Sigma \frac{(p_i - \pi_i)^2}{\pi_i}$, e fra $\triangle$ e la funzione $\Gamma = \Sigma \mid p_i - \pi_i \mid$. In particolare si mostrerà come varia H^2 sul luogo dei punti ω_δ dello spazio parametrico Ω in cui $\triangle = \delta$, nonché come varia Γ su ω_δ. Si ricaveranno anche le variazioni di $\triangle$ sul luogo dei punti ω_η dello spazio parametrico in cui $H = \eta$, nonché sul luogo dei punti ω_γ in cui $\Gamma = \gamma$. Nel presente lavoro si verificherà inoltre se il test basato su X^2 è adeguato quando si vogliono caratterizzare le alternative con la funzione $\triangle$. Partendo dalla constatazione che su ω_δ la differenza fra max H^2 e min H^2 cresce al crescere di k, si riuscirà a spiegare anche il perché del considerevole aumento, all'aumentare di k, delle numerosità campionarie riscontrate in [1]. Ciò sarà sufficiente per ritenere inadeguato il test basato su X^2 nel caso in cui sia necessario caratterizzare le alternative, invece che con la consueta funzione H^2, tramite la funzione $\triangle$. Il lavoro porta così alla conclusione che in questo ultimo caso sarà forse meglio basare il test su una statistica analoga a $\triangle$ come è ad esempio la statistica

$$Z = \max \left\{ \frac{| n_i - n\,\pi_i |}{\sqrt{n\,\pi_i}} ;\ i = 1, 2, \ldots, k \right\};$$

* Lavoro sostenuto finanziariamente anche con il supporto del ministero della Pubblica Istruzione.

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

dove n è la numerosità campionaria ed n_i indica la frequenza osservata nella classe i^{ma} della multinomiale.

2. *Funzioni di distanza*

Si abbia una multinomiale a k classi con probabilità $(p_1, p_2, \ldots, p_k)$ e con numero di prove n. Lo spazio parametrico Ω può considerarsi un sottospazio dello spazio euclideo k-dimensionale E^k le cui coordinate $(p_1, \ldots, p_k)$ sono tali che $\sum_1^k p_i = 1, p_i \geqslant 0 \quad (i = 1, 2, \ldots, k)$.

Per misurare la distanza di un generico punto $(p_1, p_2, \ldots, p_k)$ da un prefissato punto di Ω, le cui coordinate si indicheranno con $(\pi_1, \pi_2, \ldots, \pi_k)$, esistono diversi criteri. Il più impiegato consiste nel calcolare la media quadratica ponderata delle differenze relative assolute $\zeta_i = \dfrac{|p_i - \pi_i|}{\pi_i}$ con pesi π_i, data dall'espressione

$$H = \sqrt{\Sigma \left(\frac{p_i - \pi_i}{\pi_i}\right)^2 \cdot \pi_i} = \sqrt{\Sigma \frac{(p_i - \pi_i)^2}{\pi_i}} \tag{1}$$

Un altro tipo di media delle differenze relative assolute potrebbe essere la media aritmetica ponderata data dall'espressione

$$\Gamma = \Sigma\, \zeta_i \cdot \pi_i = \Sigma\, |p_i - \pi_i| \,. \tag{2}$$

Un criterio del tutto diverso dalla (1) e dalla (2) è stato introdotto in [1] ed è dato dalla funzione

$$\triangle = \max \{ \zeta_i;\ i = 1, 2, \ldots, k \} \,. \tag{3}$$

Dal punto di vista formale tutti e tre i criteri sono ben definiti. La (1) e la (2), esendo delle medie, non tengono conto dei valori che possono essere assunti dalle singole ζ_i in corrispondenza ad un valore prefissato, rispettivamente, di H e di Γ. Vedremo in seguito che le singole ζ_i possono assumere valori di gran lunga più elevati dei valori prefissati di H e di Γ.
La (3) effettua un controllo delle singole deviazioni delle p_i da π_i. Il significato della (3) è immediato. Infatti, se $\triangle \geqslant \delta$ significa che almeno una differenza relativa assoluta è maggiore o uguale a δ. Se $\triangle = \delta$ significa che vi è almeno una ζ_i uguale a δ e che nessuna è maggiore di δ. Se $\triangle < \delta$ significa che tutte le differenze relative assolute sono inferiori a δ.

3. *Variazioni di* H *e di* Γ *su* ω_δ

Si indichi con $\zeta_{(1)}$ e con $\zeta_{(k)}$ rispettivamente la più piccola e la più grande fra le ζ_i. Dalle proprietà delle medie potenziate si deduce che

$$\zeta_{(1)} \leqslant \Gamma \leqslant H \leqslant \zeta_{(k)} = \triangle \tag{4}$$

Si considerino ora i punti dello spazio parametrico giacenti su ω_δ. Dalla (4) si ricava che su ω_δ si ha l'uguaglianza $\Delta = \delta$ e che vale la relazione $\Gamma \leqslant H \leqslant \delta$.

Per studiare le variazioni di Γ e H su ω_δ faremo, per semplicità, l'ipotesi che sotto H_0 si abbia

$$\pi_1 = \pi_2 = \ldots = \pi_k = \frac{1}{k}.$$

Cominciamo ad esaminare le variazioni di Γ su ω_δ. Dalla (2) si ricava la scomposizione che segue:

$$\Gamma = \Sigma\,|p_i - 1/k| = \sum_{(p_i \geqslant 1/k)} (p_i - 1/k) + \sum_{(p_i < 1/k)} (1/k - p_i).$$

Dovendo essere $\sum_1^k (p_i - 1/k) = 0$, si avrà

$$\sum_{(p_i \geqslant 1/k)} (p_i - 1/k) = - \sum_{(p_i < 1/k)} (p_i - 1/k).$$

Per comodità di scrittura si ponga ora :

$$\Gamma_1 = \sum_{(p_i \geqslant 1/k)} (p_i - 1/k) \quad \text{e} \quad \Gamma_2 = \sum_{(p_i < 1/k)} (1/k - p_i).$$

Risulta pertanto $\Gamma_1 = \Gamma_2$ ed inoltre $\Gamma = \Gamma_1 + \Gamma_2$. (5)
Pertanto massimizzare (minimizzare) Γ equivale a massimizzare (minimizzare) Γ_1 oppure Γ_2.

Minimo di Γ su ω_δ

Per studiare l'andamento di Γ su ω_δ bisogna tener presente che almeno uno scarto relativo $|p_i - 1/k| \,/\, \frac{1}{k}$ è uguale a δ. Dal momento che siamo alla ricerca del minimo di Γ, si ipotizzerà di avere, senza perdere in generalità, un solo scarto relativo pari a δ e inoltre si ipotizzerà che esso sia positivo.

Ne consegue che $\min\limits_{\omega_\delta} \Gamma_1 = \frac{1}{k}\delta$ e che si raggiunge se non vi sono altri scarti positivi.

Dall'uguaglianza $\Gamma_1 = \Gamma_2$ consegue che anche $\min\limits_{\omega_\delta} \Gamma_2 = \frac{1}{k}\delta$ e quindi

$$\min_{\omega_\delta} \Gamma = \frac{2}{k}\delta. \tag{6}$$

Si noti che l'uguaglianza $\Gamma_2 = \sum_{(p_i < 1/k)} (1/k - p_i) = \frac{1}{k} \delta$ si può conseguire in diversi modi. Ad esempio, si può avere una sola differenza $(1/k - p_i)$ uguale a $\frac{1}{k} \delta$ oppure avere $(k-1)$ scarti $(1/k - p_i) \geqslant 0$ la cui somma è $\frac{1}{k} \delta$. Si noti inoltre che $\min_{\omega_\delta} \Gamma$ è inversamente proporzionale al numero k delle classi della multinomiale.

Massimo di Γ *su* ω_δ

Bisogna distinguere fra k pari e k dispari. Se k è pari su ω_δ è possibile ottenere che tutti gli scarti relativi siano uguali al massimo consentito δ. Ciò si realizza con k/2 scarti relativi positivi e k/2 scarti relativi negativi. Consegue che per k pari $\max_{\omega_\delta} \Gamma = \delta$. Per k dispari non si possono avere k scarti relativi assoluti uguali a δ. Dal vincolo $\Gamma_1 = \Gamma_2$ e dal vincolo $\frac{|p_i - 1/k|}{\frac{1}{k}} \leqslant \delta$ deriva che $\max_{\omega_\delta} \Gamma_1 = \frac{k-1}{2} \frac{\delta}{k}$ e che si può raggiungere se vi sono almeno $\frac{k-1}{2}$ scarti positivi. Supponiamo per assurdo che $\Gamma_1 > \frac{k-1}{2} \cdot \frac{\delta}{k}$. Ciò si può realizzare se vi sono almeno $\frac{k-1}{2} + 1$ scarti positivi. Ne consegue che devono esservi allora al più $\frac{k-1}{2}$ scarti negativi. Quand'anche ciascuno di questi raggiungesse in valore assoluto il massimo consentito δ, la loro somma sarebbe inferiore a Γ_1, ma ciò non è possibile in quanto $\Gamma_1 = \Gamma_2$.

In definitiva $\max_{\omega_\delta} \Gamma_1 = \frac{k-1}{2} \cdot \frac{\delta}{k}$ e quindi $\max_{\omega_\delta} \Gamma = \frac{k-1}{k} \delta$. (7)

Deriva da tutto quanto precede che su ω_δ la funzione Γ oscilla da un valore minimo di $\delta \cdot \frac{2}{k}$ ad un valore massimo pari a δ o $\delta \frac{k-1}{k}$. Si noti come all'aumentare di k la differenza fra $\max_{\omega_\delta} \Gamma$ e $\min_{\omega_\delta} \Gamma$ aumenti.

Si analizzerà ora come varia H^2 su ω_δ. Nel caso in cui $\pi_i = 1/k$ $(i = 1, 2, \ldots, k)$

è $H^2 = k \sum_1^k (p_i - 1/k)^2$; pertanto minimizzare o massimizzare H^2 è equivalente a minimizzare o massimizzare $\sum_1^k (p_i - 1/k)^2$. È possibile scomporre questa ultima espresisone come segue

$$\sum (p_i - 1/k)^2 = \sum_{(p_i \geqslant 1/k)} (p_i - 1/k)^2 + \sum_{(p_i < 1/k)} (p_i - 1/k)^2 \qquad (8)$$

Minimo di H^2 *su* ω_δ

Per comodità poniamo

$$\sum_{(p_i \geqslant 1/k)} (p_i - 1/k)^2 = H_1^2 \text{ e } \sum_{(p_i < 1/k)} (p_i - 1/k)^2 = H_2^2 .$$

Sotto l'ovvia condizione $\Gamma_1 = \Gamma_2$, se si minimizza H_1^2 e H_2^2 su ω_δ, si minimizza anche H^2. Siccome $(p_i, \ldots, p_k) \in \omega_\delta$, allora almeno uno scarto relativo assoluto deve essere uguale a δ.

Dal momento che siamo interessati alla ricerca del minimo di H^2 si ipotizzerà di avere, senza perdere in generalità, uno scarto negativo relativo uguale a $-\delta$. Il min H_2^2 si avrà ovviamente quando non vi sono altri scarti negativi, per cui $\min_{\omega_\delta} H_2^2 = \frac{\delta^2}{k^2}$. Per la simmetria di H^2 rispetto alle p_i, si può immaginare che sia p_1 la probabilità il cui scarto negativo da $\frac{1}{k}$ è $-\frac{\delta}{k}$.

Si tratta di minimizzare $H_1^2 = \sum_k^2 (p_i - 1/k)^2$ sotto la condizione

$$\sum_{\substack{2 \\ (p_i \geqslant 1/k)}}^{k} (p_i - 1/k) = + \frac{\delta}{k} .$$

Dividendo per $k-1$ si tratta di minimizzare

$$\frac{\sum_2^k (p_i - 1/k)^2}{k-1} \text{ sotto la condizione } \frac{\sum_2^k (p_i - 1/k)}{k-1} = \frac{\delta}{k} \cdot \frac{1}{(k-1)} .$$

Si noti che la prima e la seconda espressione precedente indicano rispettivamente il quadrato della media quadratica e la media aritmetica dei $(k-1)$ scarti non negativi previsti nelle sommatorie. È noto che il quadrato della media quadratica è maggiore o uguale al quadrato della media aritmetica. L'uguaglianza si ha quando tutti gli scarti sono fra loro uguali e quindi nel nostro caso quando

$(p_i - 1/k) = \frac{\delta}{k} \cdot \frac{1}{k-1}$ $(i = 2, \ldots, k)$. Consegue che $\min\limits_{\omega_\delta} \frac{H_1^2}{k-1} =$

$= \left(\frac{\delta}{k} \cdot \frac{1}{k-1}\right)^2$ e quindi $\min\limits_{\omega_\delta} H_1^2 = \frac{\delta^2}{k^2} \cdot \frac{1}{k-1}$. In definitiva, $\min\limits_{\omega_\delta} H^2 =$

$$= \left\{\min_{\omega_\delta} H_1^2 + \min_{\omega_\delta} H_2^2\right\} \cdot k = \delta^2 \frac{1}{k-1} \text{ e quindi } \min_{\omega_\delta} H = \delta \sqrt{\frac{1}{k-1}} \tag{9}$$

Si perviene quindi allo stesso risultato conseguito altrove [1] per altra via. Si noti che, a parità di δ, il minimo di H è inversamente proporzionale alla radice quadrata di $(k-1)$.

Massimo di H *su* ω_δ.

Dovendo essere $(p_1, p_2, \ldots, p_k) \in \omega_\delta$, è chiaro che al più ogni scarto relativo assoluto può essere uguale a δ e quindi il corrispettivo scarto darà un apporto alla somma $\Sigma\,(p_i - 1/k)^2$ uguale a $\left(\frac{\delta}{k}\right)^2$. Se k è pari, è possibile che ogni scarto assoluto raggiunga il suo valore massimo ($\frac{k}{2}$ scarti positivi e $\frac{k}{2}$ scarti negativi) e quindi $\Sigma\,(p_i - 1/k)^2 = k\left(\frac{\delta}{k}\right)^2 = \delta^2 \frac{1}{k}$. Conseguentemente, per k pari, $\max\limits_{\omega_\delta} H^2 = \delta^2$. Se k è dispari, esiste qualche complicazione.

Ora non è più possibile che tutti gli scarti relativi assoluti siano uguali a δ in quanto, se ciò accadesse, non verrebbe più rispettata la condizione $\Sigma\,(p_i - 1/k) = 0$. Come si è già fatto osservare in precedenza, la somma degli scarti di un segno, ad esempio quelli positivi, può essere al più uguale a $\frac{k-1}{2} \cdot \frac{\delta}{k}$ e lo stesso deve accadere per quelli di segno opposto. Supponiamo che il numero degli scarti negativi la cui somma è $-\frac{(k-1)}{2} \cdot \frac{\delta}{k}$ sia $\frac{k-1}{2}$.

Il corrispettivo apporto alla $\Sigma\,(p_i - 1/k)^2$ è $\frac{k-1}{2}\left(\frac{\delta}{k}\right)^2$. Resta ora da massimizzare l'apporto degli altri $\frac{k-1}{2} + 1$ scarti non negativi, sempre sotto la condizione che la loro somma sia uguale a $\frac{k-1}{2} \cdot \frac{\delta}{k}$. Per la simmetria di H^2 rispetto a $(p_1, \ldots, p_k)$ possiamo supporre che gli scarti negativi si abbiano per $i = 1, 2, \ldots, \frac{k-1}{2}$.

In termini formali si tratta di massimizzare la funzione

$$\sum_{\frac{k-1}{2}+1}^{k} (p_i - 1/k)^2 \text{ sotto le condizioni:}$$

a) $0 \leqslant (p_i - 1/k) \leqslant \frac{\delta}{k}$ per $i = \frac{k-1}{2} + 1, \frac{k-1}{2} + 2, \ldots, k$;

b) $\displaystyle\sum_{\frac{k-1}{2}+1}^{k} (p_i - 1/k) = \frac{k-1}{2} \cdot \frac{\delta}{k}$.

Il massimo della funzione $\displaystyle\sum_{\frac{k-1}{2}+1}^{k} (p_i - 1/k)^2$ sotto le condizioni a) e b) sopra esplicitate si ottiene quando uno scarto è uguale a zero ed i restanti $\frac{k-1}{2}$ scarti raggiungono il masimo valore consentito δ/k.

Infatti, per la a) e per la b) è possibile porre:

1) $(p_i - 1/k) = (\delta/k - \varepsilon_i)$ per $i = \frac{k-1}{2} + 1, \ldots, k-1$, essendo

$$\varepsilon_i \geqslant 0 \text{ e } \sum_{\frac{k-1}{2}+1}^{k-1} \varepsilon_i \leqslant \frac{\delta}{k};$$

2) $(p_k - 1/k) = \displaystyle\sum_{\frac{k-1}{2}+1}^{k-1} \varepsilon_i$.

Facciamo il quadrato di questi scarti e sommiamo

$$\sum_{\frac{k-1}{2}+1}^{k} (p_i - 1/k)^2 = \sum_{\frac{k-1}{2}+1}^{k-1} (\delta/k - \varepsilon_i)^2 + \left(\sum_{\frac{k-1}{2}+1}^{k-1} \varepsilon_i \right)^2$$

$$= \Sigma \left[(\delta/k)^2 + \varepsilon_i^2 - 2\frac{\delta}{k} \varepsilon_i) \right] + (\Sigma\, \varepsilon_i)^2$$

$$= \frac{k-1}{2} (\delta/k)^2 + \Sigma\, \varepsilon_i^2 - 2\frac{\delta}{k} \Sigma\, \varepsilon_i + (\Sigma\, \varepsilon_i)^2 .$$

Esaminiamo ora $\Sigma\, \varepsilon_i^2 - 2\frac{\delta}{k} \Sigma\, \varepsilon_i + (\Sigma\, \varepsilon_i)^2$ che può scriversi

$$\Sigma\, \varepsilon_i^2 - \frac{\delta}{k} \Sigma\, \varepsilon_i - \frac{\delta}{k} \Sigma\, \varepsilon_i + (\Sigma\, \varepsilon_i)^2 .$$

Poniamo $\frac{\delta}{k} = \Sigma\, \varepsilon_i + \nu$, essendo $\nu > 0$. Conseguentemente l'espressione precedente diventa:

$$\Sigma\, \varepsilon_i^2 - (\Sigma\, \varepsilon_i + \nu)\, \Sigma\, \varepsilon_i - (\Sigma\, \varepsilon_i + \nu) \,.\, \Sigma\, \varepsilon_i + (\Sigma\, \varepsilon_i)^2 =$$

$$= \Sigma\, \varepsilon_i^2 - (\Sigma\, \varepsilon_i)^2 - \nu\, \Sigma\, \varepsilon_i - (\Sigma\, \varepsilon_i)^2 - \nu\, \Sigma\, \varepsilon_i + (\Sigma\, \varepsilon_i)^2 .$$

Essendo $\Sigma\, \varepsilon_i^2 - (\Sigma\, \varepsilon_i)^2 \leqslant 0$, tutta l'espressione precedente risulta inferiore o uguale a zero. Si ha l'uguaglianza se tutte le ε_i sono uguali a zero. In definitiva

$$\sum_{\frac{k-1}{2}+1}^{k} (p_i - 1/k)^2 \leqslant \frac{k-1}{2} \cdot (\delta/k)^2,$$ avendosi l'uguaglianza se tutte le ε_i sono uguali a zero.

Si può pertanto concludere che $\max\limits_{\omega_\delta} H^2$ si ha quando si hanno $\frac{k-1}{2}$ scarti relativi uguali a $-\delta$, $\frac{k-1}{2}$ scarti relativi uguali a δ ed uno scarto uguale a zero.

In conclusione $\max\limits_{\omega_\delta} H^2 = \delta^2 \dfrac{k-1}{k}$. (10)

Tornando alla funzione H si può ora affermare che su ω_δ essa oscilla da un minimo di $\delta \sqrt{\dfrac{1}{k-1}}$ ad un massimo di δ o $\delta \sqrt{\dfrac{k-1}{k}}$ a seconda che k sia pari o dispari.

Osservazione

Nel caso in cui la multinomiale di riferimento sotto H_0 non sia simmetrica, le divergenze fra $\max\limits_{\omega_\delta} H$ e $\min\limits_{\omega_\delta} H$ diventano ancora più rilevanti.

Infatti, mentre il $\max\limits_{\omega_\delta} H \leqslant \delta$ (come nel caso della multinomiale simmetrica), il $\min\limits_{\omega_\delta} H = \delta \sqrt{\dfrac{\pi^*}{1-\pi^*}}$ risulta inferiore al $\min\limits_{\omega_\delta} H = \delta \sqrt{\dfrac{1}{k-1}}$ che si ha nel caso della multinomiale simmetrica.

4. *Variazioni di ζ_i su ω_η e su ω_γ*

Può essere utile conoscere le variazioni delle singole p_i sul luogo dei punti dello spazio dell'alternativa in cui si abbia rispettivamente $H = \eta$ e $\Gamma = \alpha$. Cioè, anche quando si ritiene di dover misurare la distanza utilizzando la media quadratica o la media aritmetica di tutte le variazioni relative $\dfrac{|p_i - \pi_i|}{\pi_i} = \zeta_i$, può interessare conoscere le oscillazioni cui sono soggette le singole p_i per $(p_1, \ldots, p_k) \in \omega_\eta$, ovvero per $(p_1, \ldots, p_k) \in \omega_\gamma$.

Cominciamo dal luogo dei punti ω_η in cui $H = \eta$. In termini formali si tratta di ricavare il $\max \dfrac{|p_i - \pi_i|}{\pi_i}$ sotto le condizioni:

$$\text{a) } \sum_1^k (p_i - \pi_i) = 0 \quad \text{e} \quad \text{b) } \sum_1^k \frac{(p_i - \pi_i)^2}{\pi_i} = \eta^2 .$$

Conviene riscrivere le condizioni a) e b) nel modo che segue:

$$\text{a') } (p_i - \pi_i) = -\sum_{j \neq i} (p_j - \pi_j) \quad \text{e} \quad \text{b') } \frac{(p_i - \pi_i)^2}{\pi_i} + \sum_{j \neq i} \frac{(p_j - \pi_j)^2}{\pi_j} = \eta^2.$$

Dalla b') si deduce che massimizzare $\dfrac{|p_i - \pi_i|}{\pi_i}$ è equivalente a minimizzare $\sum\limits_{j \neq i} \dfrac{(p_j - \pi_j)^2}{\pi_j}$ sotto le condizioni a') e b'). Cominciamo col minimizzare $\sum\limits_{j \neq i} \dfrac{(p_j - \pi_j)^2}{\pi_j}$ rispetto alla condizione a').

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

$$L = \sum_{j \neq i} \frac{(p_j - \pi_j)^2}{\pi_j} + \mu \left[(p_i - \pi_i) + \sum_{j \neq i} (p_j - \pi_j) \right],$$

in cui μ è il moltiplicatore di Lagrange. Le derivate parziali sono

$$\frac{\partial L}{\partial p_j} = \frac{2\,(p_j - \pi_j)}{\pi_j} + \mu, \qquad \text{per } j \neq i;$$

$$\frac{\partial L}{\partial \mu} = (p_i - \pi_i) + \sum_{j \neq i} (p_j - \pi_j).$$

Uguagliando a zero le $\dfrac{\partial L}{\partial p_j}$, si ottiene: $(p_j - \pi_j) = -\pi_j \dfrac{\mu}{2}$ per $j \neq i$. **(11)**

Sommando rispetto a $j \neq i$, si ha: $\sum (p_j - \pi_j) = -\dfrac{\mu}{2} \sum\limits_{j \neq i} \pi_j$.

Tenuto conto della a'), la relazione precedente diventa:

$$-(p_i - \pi_i) = -\frac{\mu}{2}(1 - \pi_i), \qquad \text{da cui } \frac{\mu}{2} = \frac{(p_i - \pi_i)}{1 - \pi_i}.$$

Sostituendo quest'ultimo valore in (11), si ha:

$$(p_j - \pi_j) = -(p_i - \pi_i) \frac{\pi_j}{1 - \pi_i}.$$

Sostituendo questi valori in b', si ricava:

$$\frac{(p_i - \pi_i)^2}{\pi_i} + \sum_{j \neq i} \frac{\left\{- (p_i - \pi_i) \dfrac{\pi_j}{1 - \pi_i}\right\}^2}{\pi_j} = \eta^2 .$$

Dopo alcuni passaggi si ottiene

$$\max_{\omega_\eta} \left\{\frac{|p_i - \pi_i|}{\pi_i}\right\} = \eta \sqrt{\frac{1 - \pi_i}{\pi_i}} . \tag{12}$$

La (12) permette di valutare, per ogni classe della multinomiale, le oscillazioni massime relative delle p_i dai valori π_i. Fra tutte le classi quella che presenta la variazione relativa più elevata è quella in corrispondenza della quale la probabilità sotto H_0 è la più piccola. Indicando tale probabilità con π^*, si rileva che:

$$\max_{\omega_\eta} \Delta = \eta \sqrt{\frac{1 - \pi^*}{\pi^*}} .$$

Si noti come questa relazione sia « l'inversa » di quella ricavata in [1] per $\max\limits_{\omega_\delta} H$.

Nel caso in cui sotto H_0 la multinomiale sia simmetrica, la (12) diventa

$$\max_{\omega_\eta} \left\{\frac{|p_i - 1/k|}{1/k}\right\} = \eta \sqrt{k - 1} . \tag{13}$$

La (12) e la (13) permettono di calcolare le variazioni massime delle singole p_i, quando si assegna un valore η alla distanza media delle alternative rispetto ai valori nominali. Una caratteristica da tener presente è che all'aumentare di k dette variazioni massime aumentano. Ovviamente, per valori elevati di k ciò può essere preoccupante come si evince dalla Tav. I in cui, nell'ipotesi che sotto H_0 si abbia una multinomiale simmetrica, si sono riportati per $3 \leqslant k \leqslant 15$ e per $\eta = 0{,}10$ e $0{,}20$ i valori di $\max \dfrac{|p_i - 1/k|}{1/k}$ e $\max |p_i - 1/k|$.

Tav. I

Valori di $\max\limits_{\omega_\eta} \frac{|p_i - 1/k|}{1/k}$ e di $\max\limits_{\omega_\eta} |p_i - 1/k|$ per $3 \leqslant k \leqslant 15$ e per $\eta = 0,10$ e 0,20

k	$\sqrt{k-1}$	$\pi_i = 1/k$	$\eta = 0,10$		$\eta = 0,20$	
			$\max \frac{\lvert p_i - 1/k\rvert}{1/k}$	$\max \lvert p_i - 1/k\rvert$	$\max \frac{\lvert p_i - 1/k\rvert}{1/k}$	$\max \lvert p_i - 1/k\rvert$
3	1,414	0,333	0,141	0,047	0,282	0,094
4	1,732	0,250	0,173	0,043	0,346	0,087
5	2,000	0,200	0,200	0,040	0,400	0,080
6	2,236	0,166	0,224	0,037	0,447	0,075
7	2,449	0,143	0,245	0,035	0,490	0,070
8	2,646	0,125	0,265	0,033	0,529	0,066
9	2,828	0,111	0,283	0,031	0,566	0,063
10	3,000	0,100	0,300	0,030	0,600	0,060
11	3,162	0,091	0,316	0,029	0,632	0,057
12	3,317	0,083	0,332	0,028	0,663	0,055
13	3,464	0,077	0,346	0,027	0,693	0,053
14	3,606	0,071	0,361	0,026	0,721	0,052
15	3,742	0,067	0,374	0,025	0,748	0,050

Esaminiamo ora le variazioni delle singole p_i sul luogo dei punti ω_γ dello spazio dell'alternativa in cui $\Gamma = \gamma$. In termini formali si tratta di ricavare il $\max \frac{|p_i - \pi_i|}{\pi_i}$ sotto le condizioni:

$$\text{a'})\ (p_i - \pi_i) = -\sum_{j \neq i} (p_j - \pi_j) \quad \text{e} \quad \text{c'})\ |p_i - \pi_i| + \sum_{j \neq i} |p_j - \pi_j| = \gamma$$

Massimizzare $|p_i - \pi_i|$ equivale a minimizzare $\sum\limits_{j \neq i} |p_j - \pi_j|$. Cominciamo col minimizzare rispetto alla condizione a'.

Si sa che :

$$\sum_{j \neq i} |p_j - \pi_j| \geqslant |\sum_{j \neq i} (p_j - \pi_j)| = |-(p_i - \pi_i)| = |p_i - \pi_i| .$$

In definitiva: $\min \sum_{j \neq i} |p_j - \pi_j| = |p_i - \pi_i|$. Detto minimo si raggiunge se tutti i $(k-1)$ scarti $(p_j - \pi_j)$, sempre per $j \neq i$, hanno lo stesso segno che non può essere che il contrario del segno di $(p_i - \pi_i)$.

Tenuto conto ora della condizione c' si ha :

$$|p_i - \pi_i| + |p_i - \pi_i| = \gamma \text{, da cui } |p_i - \pi_i| = \frac{\gamma}{2} .$$

In definitiva

$$\max_{\omega_\gamma} \frac{|p_i - \pi_i|}{\pi_i} = \frac{\gamma}{2} \cdot \frac{1}{\pi_i} . \tag{14}$$

Anche in questo caso fra tutte le classi quella che presenta la variazione relativa più elevata è quella in corrispondenza della quale la probabilità sotto H_0 è la più piccola. Risulta pertanto :

$$\max_{\omega_\gamma} \Delta = \frac{\gamma}{2} \cdot \frac{1}{\pi^*} . \tag{15}$$

Per la multinomiale simmetrica la (15) diventa

$$\max_{\omega_\gamma} \left\{ \frac{|p_i - \pi_i|}{\pi_i} \right\} = \frac{\gamma}{2} \cdot k . \tag{16}$$

Si noti come dette variazioni relative massime aumentino proporzionalmente a k .

5. *Impiego della statistica* X^2 *nel caso in cui l'ipotesi alternativa sia caratterizzata da* Ω_δ

L'approccio classico al problema della verifica di ipotesi concernente le probabilità di una multinomiale assume che sotto H_0 sia $(p_1 = \pi_1)$ e ... e $(p_k = \pi_k)$ e che sotto H_1 sia $(p_1 \neq \pi_1)$ o ... o $(p_k \neq \pi_k)$. Sembra invece più realistico non preoccuparsi troppo delle alternative troppo vicine ad H_0 ; ad esempio, si potrebbero trascurare le alternative per le quali tutte le ζ_i siano inferiori a δ (cioè trascurare le alternative per le quali $\Delta < \delta$ e considerare invece come ipotesi alternativa H_1 quella caratterizzata da $\Delta \geqslant \delta$. In altre parole, H_1 dovrebbe essere caratterizzata da tutti i punti $(p_1, p_2, \ldots, p_k) \in \Omega_\delta$, essendo Ω_δ il sottospazio dello spazio parametrico in cui $\Delta \geqslant \delta$.

Esaminiamo ora il comportamento della statistica X^2 di Karl Pearson nel caso in cui lo spazio dell'alternativa sia caratterizzato da Ω_δ.

È noto che sotto H_0 la statistica $X^2 = \Sigma \dfrac{(n_i - n\,\pi_i)^2}{n\,\pi_i}$ tende a distribuirsi come un chi-quadrato con $(k-1)$ gradi di libertà. Sotto una alternativa qualsiasi $(p_1, p_2, \ldots, p_k)$ la variabile X^2 è approssimativamente distribuita, per n grande, secondo una v. c. chi-quadrato non centrale con $(k-1)$ gradi di libertà e parametro di non centralità $\lambda^2 = n\,\eta^2$, essendo η^2 il valore di H^2 in corrispondenza dell'alternativa $(p_1, \ldots, p_k)$. Ovviamente per la verifica di ipotesi in oggetto si respinge H_0 per $H^2 \geqslant c$, ove c è tale che

$P\,[X^2 \geqslant c / H_0] = \alpha$.

L'alternativa più sfavorevole, cioè quella in corrispondenza della quale β assume il valore più elevato, si ha in corrispondenza del punto (o dei punti) $(p_1, \ldots, p_k) \in \Omega_\delta$ per il quale λ^2 e quindi H^2 assume il valore più piccolo. È stato mostrato in [1] che detto punto si trova su ω_δ, inoltre è stato provato che

$$\min_{\omega_\delta} H^2 = \delta^2 \cdot \frac{\pi^*}{1 - \pi^*}, \quad \text{dove} \quad \pi^* = \min\,(\pi_1, \pi_2, \ldots, \pi_k)\,.$$

Dal valore del $\min\limits_{\omega_\delta} H^2$ è stato possibile determinare la numerosità campionaria necessaria per avere $\alpha = \max \beta$. Dall'esame di dette numerosità si è dedotto che il test basato su X^2 non è molto sensibile per le alternative caratterizzate da Ω_δ per valori piccoli di δ. Inoltre si è riscontrato un anomalo aumento di n all'aumentare di k.

Il fatto che si possa impiegare il consueto test basato su X^2 non autorizza però a ritenere che esso risolva il problema in modo soddisfacente. Sono state proprio le numerosità molto elevate che si hanno all'aumentare di k che hanno portato ad avere dei dubbi sulla bontà del test basato su X^2 per risolvere il problema in esame.

Una considerazione da fare è che, se in un problema interessa valutare la distanza delle alternative con la funzione $\triangle$, sembra ragionevole poi pretendere che il test impiegato ponga, dal punto di vista del valore di β, sullo stesso piano tutti i punti aventi la stessa distanza δ. Ciò non può essere assicurato dal test basato su X^2. Infatti siccome β dipende da $\lambda^2 = n \cdot H^2$ e siccome, come si mostrerà nel paragrafo che segue, su ω_δ il valore di H^2 oscilla molto (e dette oscillazioni aumentano all'aumentare di k) se ne deduce che β varia molto su ω_δ (e che dette variazioni aumentano all'aumentare di k).

Il test basato su X^2 è invece sicuramente adeguato se interessa caratterizzare le alternative con la funzione di distanza H^2.

5.1. *Variazioni di* β *su* ω_δ

Nelle pagine precedenti si è posto in evidenza che sarebbe auspicabile che un test per la verifica d'ipotesi in oggetto protegga allo stesso modo le alternative aventi la stessa distanza δ; in altre parole il valore di β dovrebbe essere costante su ω_δ. In questo paragrafo vedremo come si comporta a questo proposito il test basato su X^2. Per valori prefissati di k ed α il valore di β dipende dal parametro di non centralità $\lambda^2 = n \cdot \eta^2$. Quindi, per un n prefissato, per esaminare come varia β su ω_δ è sufficiente conoscere come varia η^2 su ω_δ. Per la multinomiale simmetrica η^2 varia dal valore minimo $\delta^2 \dfrac{1}{k-1}$ al valore massimo δ^2 o $\delta^2 \dfrac{k-1}{k}$. Il rapporto ϱ fra $\max\limits_{\omega_\delta} \eta^2$ e $\min\limits_{\omega_\delta} \eta^2$ è uguale a $(k-1)$ per k pari ed a $k-2+1/k$ per k dispari.
Nel prospetto che segue si sono calcolati i valori di ϱ per k da 2 a 10.

k	2	3	4	5	6	7	8	9	10
ϱ	1	1,33	3	3,2	5	5,14	7	7,11	9

Ovviamente i rapporti che si hanno fra i valori di η^2 sussistono anche quando si passa al parametro di non centralità λ^2.
Al crescere di k cresce anche ϱ e quindi dovrebbe crescere anche la divergenza fra $\max\limits_{\omega_\delta} \beta$ e $\min\limits_{\omega_\delta} \beta$. Per valutare queste divergenze si può procedere come segue. Si fissa il valore di k, α e $\max \beta$. Dalle tavole di chi-quadrato non centrale riportate in [2] si ricava il parametro di non centralità λ^2; il valore così ottenuto è ovviamente $\min\limits_{\omega_\delta} \lambda^2$. Moltiplicando questo valore per ϱ, si ottiene il valore di $\max\limits_{\omega_\delta} \lambda^2$. Utilizzando le stesse tavole e sempre per gli stessi valori di α e di k, si trova il valore di $\min\limits_{\omega_\delta} \beta$.

Esempio

Si abbia: $k = 3$, $\alpha = 0{,}10$ e $\max \beta = 0{,}10$. Si vuol conoscere $\min\limits_{\omega_\delta} \beta$. Dalle tavole riportate in [2] si trova che $\min\limits_{\omega_\delta} \lambda^2 = 10{,}458$. Ne consegue che $\max\limits_{\omega_\delta} \lambda^2 = \varrho \min\limits_{\omega_\delta} \lambda^2$ è uguale a 13,944. Sempre utilizzando le stesse tavole si trova che $\min\limits_{\omega_\delta} \beta = 0{,}039$. Utilizzando il procedimento indicato nell'esempio ora mostrato si è ricavata la Tav. II in cui in corrispondenza a diverse combinazioni di α, $\max\limits_{\omega_\delta} \beta$ e k si è indicato il valore di $\min\limits_{\omega_\delta} \beta$.

Tav. II

Valori di min λ^2, max λ^2 e min β in corrispondenza a diverse combinazioni di α, max β e k.

k	$\alpha = 0{,}05$; max $\beta = 0{,}05$			$\alpha = 0{,}10$; max $\beta = 0{,}10$		
	min λ^2	max λ^2	min β	min λ^2	max λ^2	min β
2	13,00	13,00	0,05	8,60	8,60	0,100
3	15,46	20,61	0,016	10,46	13,94	0,038
4	17,35	52,04	0,000	11,77	35,25	0,000
5	18,59	59,50	0,000	12,89	41,26	0,000
6	19,78	98,88	0,000	13,80	69,00	0,000
7	20,87	107,25	0,000	14,67	75,42	0,000
8	21,85	152,96	0,000	15,43	79,33	0,000
9	22,50	159,98	0,000	16,12	112,85	0,000
10	23,61	212,45	0,000	16,79	151,07	0,000

k	$\alpha = 0{,}05$; max $\beta = 0{,}10$			$\alpha = 0{,}10$; max $\beta = 0{,}20$		
	min λ^2	max λ^2	min β	min λ^2	max λ^2	min β
2	10,54	10,54	0,100	6,20	6,20	0,200
3	12,68	16,90	0,034	7,73	10,31	0,105
4	14,19	42,56	~ 0	8,81	26,44	0,002
5	15,43	49,36	~ 0	9,70	31,04	0,001
6	16,50	82,50	~ 0	10,47	52,33	~ 0
7	17,44	89,64	~ 0	11,14	57,27	~ 0
8	18,30	128,10	~ 0	11,76	94,11	~ 0
9	19,09	135,72	~ 0	12,34	87,73	~ 0
10	19,84	178,54	~ 0	12,87	115,81	~ 0

L'esame della Tav. II conferma quanto ipotizzato in precedenza. La differenza su ω_δ fra max β e min β è molto rilevante. Già per $k = 3$, min β è inferiore alla metà del max β (tranne il caso $\alpha = 0{,}10$ e max $\beta = 0{,}20$).

Per $k = 4$ il valore del min β è praticamente nullo anche per valori elevati di α e max β. È questa una conferma che il test basato su X^2 non è adeguato per il problema di verifica di ipotesi in cui interessa valutare le alternative con la metrica $\triangle$.

5.2. *Numerosità campionarie*

Per le elaborazioni relative alla Tav. II il valore di δ non entra in gioco. Esso però diviene essenziale per determinare la numerosità campionaria. Infatti, dalla relazione $\min_{\omega_\delta} \lambda^2 = n \, . \, \min_{\omega_\delta} \eta^2$ si ricava

$$n = \frac{\min\limits_{\omega_\delta} \lambda^2}{\min\limits_{\omega_\delta} \eta^2} \, ;$$

tenendo presente ora che per la multinomiale simmetrica $\min\limits_{\omega_\delta} \eta^2 = \delta^2 \cdot \dfrac{1}{k-1}$,

si ricava

$$n = \frac{\min\limits_{\omega_\delta} \lambda^2}{\delta^2} \cdot (k-1) \tag{17}$$

Dalla (17) si ricava che la numerosità campionaria è inversamente proporzionale al quadrato di δ ed è direttamente proporzionale a $(k-1)$. La (17) spiega allora il perché dell'anomalo aumento di n all'aumentare di k riscontrato in [1]. Perché ci si possa rendere conto ancora meglio dell'inadeguatezza del test basato su X^2 per le verifiche di ipotesi in cui l'alternativa è caratterizzata dalla funzione $\triangle$, confronteremo i valori dati dalla (17) con quelli che si avrebbero qualora il test X^2 dovesse discriminare su alternative caratterizzate da H.

Si sa che in questo caso β è costante su tutti i punti di Ω aventi la stessa distanza η da $(\pi_1, \ldots, \pi_k)$. Si ricordi che η è interpretabile come media quadratica ponderata degli scarti relativi $\dfrac{(p_i - \pi_i)}{\pi_i}$; può aver quindi senso assegnare lo stesso valore ad η per diversi valori di k. Per determinare la numerosità campionaria in questo caso bisognerà innanzitutto fissare come sempre α, max β e k. Le solite tavole daranno il valore di λ^2. Dalla relazione $\lambda^2 = n \cdot \eta^2$, una volta prefissato η, si ha che

$$n = \frac{\lambda^2}{\eta^2} \, . \tag{18}$$

Si osserva che la numerosità campionaria non è più direttamente proporzionata a k come accadeva con la (17). Ciò non significa che ora il k non abbia nessuna influenza nella determinazione di n, ma si vuol sottolineare che la sua influenza si esercita solo sul valore che viene ad assumere λ^2 che è funzione di α, di β e di k. Nel caso precedente k interveniva implicitamente nel valore che le tavole assegnavano a $\min_{\omega_\delta} \lambda^2$ ed esplicitamente come moltiplicatore (in verità il moltiplicatore è $k - 1$) del rapporto $\dfrac{\min \lambda^2}{\delta^2}$. Se ora, assegnato un valore a δ (ad esempio $\delta = 0{,}10$), si pone $\Delta = \delta$ e $H = \delta$, si può vedere che, usando sempre la stessa variabile test X^2, nel caso della caratterizzazione delle alternative con Δ la numerosità campionaria è ben $(k - 1)$ volte quella necessaria nel caso della caratterizzazione delle alternative con H.

Con la (17) e con la (18), ed utilizzando i valori di $\min_{\omega_\delta} \lambda^2$ già riportati nella Tav. II, si sono ricavate le numerosità campionarie, riportate nella Tav. III, necessarie per le due caratterizzazioni delle alternative per diversi valori di δ ed ovviamente con la condizione $\Delta = \delta$ e $H = \delta$.

Tav. III

Numerosità campionarie per le caratterizzazioni delle alternative con le funzioni Δ e H per diverse combinazioni di α, max β, k e δ.

k	$\alpha = 0{,}05$; max $\beta = 0{,}05$					
	$\delta = 0{,}05$		$\delta = 0{,}10$		$\delta = 0{,}20$	
	$H = 0{,}05$	$\Delta = 0{,}05$	$H = 0{,}10$	$\Delta = 0{,}10$	$H = 0{,}20$	$\Delta = 0{,}20$
2	5200	5200	1300	1300	325	325
3	6182	12364	1545	3090	386	772
4	6938	20814	1735	5205	434	1302
5	7437	29748	1859	7436	465	1860
6	7910	39550	1978	9890	494	2470
7	8346	58422	2087	12522	522	3654
8	8740	61180	2185	15295	546	3822
9	9103	72824	2275	18200	569	4552
10	9444	84996	2361	21249	590	5310

k	$\alpha = 0{,}10$; max $\beta = 0{,}10$					
	$\delta = 0{,}05$		$\delta = 0{,}10$		$\delta = 0{,}20$	
	$H = 0{,}05$	$\Delta = 0{,}05$	$H = 0{,}10$	$\Delta = 0{,}10$	$H = 0{,}20$	$\Delta = 0{,}20$
2	3439	3439	859	859	215	215
3	4183	8366	1046	2092	261	522
4	4708	14124	1177	3531	294	882
5	5156	20624	1289	5156	322	1288
6	5520	27600	1380	6900	345	1725
7	5868	35208	1467	8802	367	2202
8	6172	43204	1543	10801	386	2702
9	6448	51584	1612	12896	403	3224
10	6716	60444	1679	15111	420	3780

k	$\alpha = 0{,}05$; max $\beta = 0{,}10$					
	$\delta = 0{,}05$		$\delta = 0{,}10$		$\delta = 0{,}20$	
	$H = 0{,}05$	$\Delta = 0{,}05$	$H = 0{,}10$	$\Delta = 0{,}10$	$H = 0{,}20$	$\Delta = 0{,}20$
2	4214	4214	1054	1054	263	263
3	5071	10142	1268	2536	317	634
4	5675	17025	1419	4257	355	1065
5	6170	24680	1543	6172	386	1544
6	6600	33000	1650	8250	413	2065
7	6976	41856	1744	10464	436	2616
8	7320	51240	1830	12810	458	3206
9	7636	61088	1909	15272	477	3816
10	7935	71415	1984	17856	496	4464

k	$\alpha = 0{,}10$; max $\beta = 0{,}20$					
	$\delta = 0{,}05$		$\delta = 0{,}10$		$\delta = 0{,}20$	
	$H = 0{,}05$	$\Delta = 0{,}05$	$H = 0{,}10$	$\Delta = 0{,}10$	$H = 0{,}20$	$\Delta = 0{,}20$
2	2481	2481	620	620	155	155
3	3092	6184	773	1546	193	386
4	3525	10575	881	2643	220	660
5	3880	15520	970	3880	242	968
6	4186	20930	1047	5235	261	1305
7	4456	26736	1114	6684	278	1671
8	4704	32928	1176	8232	294	2058
9	4936	39488	1234	9872	308	2468
10	5148	46332	1287	11583	322	2896

L'esame della Tav. III mostra che l'aumento della numerosità campionaria dovuto all'aumento di k è contenuto nel caso in cui le alternative sono caratterizzate da H. Quando invece le alternative sono caratterizzate da $\triangle$, gli aumenti sono considerevoli a causa del fattore $(k-1)$ che interviene rispetto alle numerosità che si hanno con le caratterizzazioni della funzione $\triangle$.

7. *Conclusioni*

Da quanto esposto nei paragrafi precedenti risulta evidente che il test basato su X^2 non è adeguato quando si caratterizzano le alternative con la funzione $\triangle$. Questa constatazione mette in luce una rilevante « *limitazione* » del test basato su X^2 non abitualmente considerata in passato. Infatti si ritiene che il test chi-quadrato vada impiegato quando bisogna proteggersi equamente in tutte le direzioni rispetto all'ipotesi H_0.

Ora i punti giacenti su ω_δ si trovano in tutte le direzioni possibili rispetto al punto $(\pi_1, \ldots, \pi_k)$, eppure il test basato su X^2 non solo non protegge questi punti allo stesso modo, ma all'aumentare di k le cose peggiorano tanto da dover ricorrere a numerosità campionarie fuori dalla portata delle normali ricerche empiriche quando si vogliano impiegare gli usuali valori per α, max β e δ.

Sembra di poter concludere che, qualora interessi caratterizzare le alternative con funzioni diverse da H, bisogna basare il test su statistiche diverse da X^2. Nel caso in esame bisognerà presumibilmente basare il test su una statistica analoga alla funzione $\triangle$ come è ad esempio la statistica

$$Z = \max \left\{ \frac{|n_i - n\,\pi_i|}{\sqrt{n\,\pi_i}} \; ; \; i = 1, 2, \ldots, k \right\} .$$

Non è ovviamente possibile affermare però, solo in base alle considerazioni sopra svolte, che bisogna preferire il test basato sulla statistica Z a quello basato sulla statistica X^2. È necessario innanzitutto studiare la distribuzione campionaria di Z e fare poi il confronto fra i due test rispetto alla classe di alternative caratterizzata da $\triangle$.

BIBLIOGRAFIA

[1] MICHELE ZENGA: « Suggerimenti per l'impiego del test chi-quadrato » in: *Atti della XXIX riunione della SIS*, Bologna 1978, pubblicato anche in *Quaderni di Statistica e Matematica applicata alle Scienze economico-sociali* del Gruppo Statistico-Matematico della Libera Università degli Studi di Trento, Vol. I, 1978.

[2] Institute of Mathematical Statistics: *Selected Tables in Mathematical Statistics*, Vol. I, Markham Publishing Company, Chicago, 1970.

SUMMARY

Some meausures of distance between hypothesis and alternative in the case of multinomial distribution.

For a multinomial distribution with k classes one can meausure in several ways the distance between the probabilities $(\pi_1, \pi_2, \ldots, \pi_k)$ specified by the null hypothesis and those specified by the alternative hypothesis $(p_1, p_2, \ldots, p_k)$. In this paper the author makes some comparisons between the distance functions

$$H = \sqrt{\Sigma \frac{(p_i - \pi_i)^2}{\pi_i}} \quad \text{and} \quad \Delta = \max \left\{ \frac{|p_i - \pi_i|}{\pi_i}; \ i = 1, 2, \ldots, k \right\},$$

and shows that if one meausures the distance with Δ the test based on the K. Pearson X^2 statistics is not adeguate and that in this case it seems more reasonable to use a test statistics with a structure like Δ.

RÉSUMÉ

Fonctions de distance concernant les distributions multinomiales

Dans cet article on compare les fonctions de distance

$$H = \sqrt{\Sigma \frac{(p_i - \pi_i)^2}{\pi_i}} \quad \text{et} \quad \Delta = \max \left\{ \frac{|p_i - \pi_i|}{\pi_i}; \ i = 1, 2, \ldots, k \right\},$$

où $(\pi_1, \pi_2, \ldots, \pi_k)$ sont les probabilités de l'hypothèse nulle et $(p_1, p_2 \ldots, p_k)$ sont les probabilités de l'hypothèse alternative concernant une distribution multinomiale.

On montre aussi que le test construit sur la statistique de K. Pearson n'est pas adequat si l'on doit mesurer la distance des alternatives avec la fonction Δ.

Dans ce cas le test doit être construit sur une statistique qui a une structure similaire à la fonction Δ.