Rivista di Statistica Applicata Vol. 21 n. 2 Giugno 1988

Un test di qualità mediante campionamento sequenziale a gruppi in popolazione finita

Franco Caroti Ghelli, Dipartimento di Matematica, Università di Pisa

Viene proposto un procedimento Bayesiano di tipo sequenziale a gruppi per la scelta fra le ipotesi $H_0 : \vartheta \leq \vartheta^*$ e $H_1 : \vartheta > \vartheta^*$ con assegnate probabilità di errore, dove ϑ è la funzione di individui difettosi in una popolazione finita. Poiché si ipotizza che durante il campionamento (senza ripetizione) gli eventuali individui difettosi osservati siano allontanati dalla popolazione, la soglia ϑ^* viene confrontata non con la frazione iniziale, ma con quella risultante dopo l'allontanamento di quelli via via osservati. Il metodo presuppone che la probabilità a priori di H_0 sia molto maggiore di quella di H_1, perciò la numerosità gobale di campione è scelta in modo da risultare minima quando il procedimento termina con la scelta di H_0, trascurando così in media l'incidenza della non ottimatilità che si ha quando la scelta è favorevole ad H_1.

Viene proposto infine un algoritmo efficiente per il calcolo della probabilità a posteriori di H_0.

1. Introduzione

Nel Controllo di Qualità si incontra sovente il problema di decidere se una data popolazione P, miscuglio di individui appartenenti a due classi A (sani) e B (difettosi), la frazione ϑ di individui di classe B sia superiore a un valore fissato ϑ^* (di solito piuttosto piccolo).

Se non interessa un approccio decisionale, è noto che una soluzione ottimale (nel senso della minimizzazione del campione) è fornita dalla procedure sequenziali del rapporto di verosimiglianza, note anche come procedure di Wald, che minimizzano la numerosità di campione necessaria per arrivare o all'accettazione dell'ipotesi $H_0(\vartheta \leq \vartheta^*)$ con probabilità di errore α_0 o all'accettazione dell'alternativa $H_1(\vartheta > \vartheta^*)$ con probabilità di errore α_1, con α_0 e α_1 assegnate.

I metodi sequenziali di Wald però prevedono di procedere alla fase decisionale dopo ogni osservazione, e questo in molti casi si rivela poco pratico, fino a diventare a volte improponibile (De Mets D.L., Ware J.H., 1980; Pocock S.J., 1977). Perciò sono stati proposti alcuni metodi che prevedono un campionamento sequenziale a gruppi equinumerosi, e fanno seguire la fase decisionale a ciascun gruppo anziché a ciascuna osservazione (De Mets D.L., Ware J.H., 1980; De Mets D.L., Ware J.H., 1982; Lan K.K.G., De Mets D.L., 1983; Pocock S.J., 1977; Pocock S.J., 1982). Questo approccio tuttavia non si presenta ancora del tutto soddisfacente, perché il vincolo dell'equinumerosità dei gruppi ne limita non poco le prestazioni.

È stato anche proposto (Ehrenfeld S., 1972; Hayre L.S., 1985) di adottare un

campionamento sequenziale a gruppi di numerosità variabile, ma la scelta della numerosità di ciascun gruppo risulta o da complesse equazioni integrali di difficile soluzione (Ehrenfeld S., 1972; Wetherill B.G., 1975), o dall'introduzione di condizioni del tutto arbitrarie (Spahn M., Ehrenfeld S., 1974).

Nel tentativo di superare questi inconvenienti, nel presente lavoro verrà presentato un nuovo metodo di campionamento sequenziale a gruppi. Tale metodo, facendo riferimento a un campionamento senza rimpiazzo e scegliendo opportunamente la numerosità di ciascun gruppo sulla base dell'informazione ottenuta dai gruppi precedenti, raggiunge la minimizzazione del campione complessivo tutte le volte che il procedimento termina con l'accettazione di H_0.

Il metodo prevede l'uso di un calcolatore elettronico, anche se di piccole o piccolissime dimensioni, ma non si presenta più complicato di quanto non lo siano gli attuali metodi sequenziali a gruppi di uguale numerosità. Infatti esso richiede, al termine di ogni gruppo di osservazioni:

a) di espletare una fase decisionale consistente nel banale confronto del numero di individui B osservati in quel gruppo con due soglie, ricavate al passo precedente.
b) di ricavare, attraverso l'uso del calcolatore e qualora la decisione presa nella fase a) sia favorevole alla rilevazione di un successivo gruppo di osservazioni, la numerosità del gruppo da rilevare e le due soglie con le quali confrontare i dati da esso forniti (in realtà si tratta di ricavare una sola soglia, perché l'altra, per come è strutturato il metodo, è sempre uguale a zero).

Pertanto questo metodo si presenta vantaggioso, rispetto ai metodi sequenziali a gruppi di uguale numerosità, in tutti quei casi in cui la probabilità a priori di H_0 è nettamente superiore a quella di H_1, perché in tal caso la non ottimalità che si ha quando è vera H_1 viene ad esercitare un peso mediamente trascurabile. In considerazione di questo fatto, si è supposto che la densità a priori di ϑ sia decrescente (ad eccezione, al più, dei valori di ϑ molto piccoli) abbastanza rapidamente perché, pur essendo $\vartheta^* \ll 1$, si possa supporre $p(H_0) \gg p(H_1)$; questa ipotesi si concretizzerà nella scelta di alcuni modelli, precisati nel corso del par. 3).

Un altro aspetto del problema consiste nel fatto che, se la popolazione è finita, la eliminazione degli individui B man mano che durante il campionamento essi vengono osservati e riconosciuti, è un'operazione spesso attuata che porta un ovvio miglioramento alla popolazione. Di questa modifica graduale nella composizione della popolazione il metodo tiene conto sia per la determinazione della numerosità di gruppo che per quella delle soglie, in quanto ad ogni passo le ipotesi H_0 e H_1 vengono a riguardare la popolazione "depurata" e non quella originaria.

Nel prossimo paragrafo verrà mostrato come si possa tener conto di que-

st'ultimo problema, mediante una opportuna ridefinizione delle ipotesi H_0 e H_1; verrà inoltre presentata una procedura che consente un calcolo assai rapido ed efficiente della pobabilità a posteriori di H_0 facendo riferimento ad un campionamento senza ripetizione.

Nel par. 3) verranno poi dimostrate due proprietà della distribuzione a posteriori cui verrà fatto riferimento nel paragrafo successivo; nel par. 4) infine verrà presentato e discusso il metodo, sia per quanto riguarda la scelta della numerosità di gruppo sia per quanto riguarda la regola di arresto.

2. Calcolo della probabilità a posteriori di H_0

In accordo con quanto detto al punto 1) del par. precedente, ogni volta che durante il campionamento si riscontra un individuo di classe B esso viene allontanato dalla popolazione, alterandone così la composizione. In realtà, poiché il campionamento è supposto senza ripetizione, nessuno degli individui osservati verrà rimesso nella popolazione finché il procedimento non risulterà terminato: perciò quando si parla di allontanare un individuo B dalla popolazione, si intende di non reimmetterlo neppure a procedimento ultimato, ottenendo così che la popolazione che risulta dopo aver eseguito il test è uguale a quella originaria depurata di tutti gli individui B osservati durante il campionamento.

Perciò alla n–ma osservazione la richiesta che la percentuale di individui B sia $\leq \vartheta^*$ va riferita a quella che sarebbe la popolazione finale se il procedimento terminasse proprio con quella osservazione, e se m sono gli individui B fin allora osservati (ed eliminati), è chiaramente esprimibile in funzione di m:

$$H_0(m) \ : \ \frac{M-m}{N-m} \leq \vartheta^*$$

dove N è la numerosità iniziale di P ed M il numero iniziale di individui B in essa. Posto:

$$\vartheta = M/N \tag{1}$$

l'ipotesi $H_0(m)$ si può esprimere nella forma:

$$H_0(m) \ : \ \vartheta \leq h(m)/N \tag{2}$$

con:

$$h(m) = m + \vartheta^* \cdot (N-m)$$

Analogamente, anche H_1 essendo l'ipotesi complementare di H_0, risulterà funzione di m:

$$H_1(m) \ : \ \vartheta > h(m)/N$$

Pertanto, esprimendo le due ipotesi in termini di ϑ, esse dipendono dalle osservazioni precedenti, e così la procedura proposta si differenzia dagli schemi sequenziali abituali, nei quali le ipotesi contrapposte non variano al variare di n.

Poiché la popolazione è finita, i valori ammessi per ϑ sono gli $(N+1)$ valori dati da: $\vartheta = i/N$, con $i = 0, 1, \ldots, N$; perciò la distribuzione a priori assegnarà a questi valori una probabilità $\phi(i/N)$ diversa da zero, e zero ad ogni altro valore reale. Ne consegue che la probabilità a posteriori di ϑ, dopo n osservazioni e avendo riscontrato in esse m individui B, sarà anch'essa zero per i valori di ϑ non ammessi, mentre per quelli ammessi sarà proporzionale alla funzione di verosimiglianza moltiplicata per la probabilità a priori $\phi(\vartheta)$. Poiché il campionamento è effettuato senza ripetizione, la funzione di verosimiglianza è data dalla ipergeometrica:

$$p(m|n, M) = \binom{M}{m} \cdot \binom{N-M}{n-m} \Big/ \binom{N}{n}$$

e la probabilità a posteriori di $\vartheta = t/N$ sarà:

$$p(\vartheta = \frac{t}{N} |m, n) = K \cdot p(m|n, t)) \cdot \phi\left(\frac{t}{N}\right) \tag{3}$$

La costante di proporzionalità K si ottiene imponendo la condizione di normalizzazione:

$$K = 1/ \sum_{t=m}^{N-n+m} \phi(t/N) \cdot p(m|n, t)$$

dove la somma deve esser fatta fra m ed $N - n + m$ perché la distribuzione ipergeometrica è zero per $m > \vartheta \cdot N$ e per $(n - m) > N(1 - \vartheta)$.

Perciò, se h^* è il massimo intero $\leq h(m)$, la probabilità a posteriori di $H_0(m)$ avendo osservato (ed eliminato) m individui B nel campione di n individui, risulta:

$$\begin{aligned} p(H_0|m) &= \sum_{j=m}^{h^*} k\phi(j/N) \cdot p(m|n, j) = \\ &= \left[\sum_{j-m}^{h^*} \frac{\phi(j/N) \cdot j!(N-j)!}{(j-m)!(N-j-n+m)!} \right] \Big/ \\ &\Big/ \left[\sum_{i-m}^{N-n+m} \frac{\phi(i/N) \cdot i!(N-i)!}{(i-m)!(N-i-n+m)!} \right] \end{aligned} \tag{4}$$

Chiaramente, per il calcolo della (4) è indispensabile l'uso del calcolatore. Tuttavia a causa dei numerosi fattoriali presenti, la (4) impone gravi vincoli computazionali sia per l'eccessivo volume di calcolo richiesto sia per il pericolo di traboccamenti. Mostriamo adesso un algoritmo che consente invece un calcolo rapido ed efficiente di $p(H_0|m)$.

Posto:

$$z(t,m,n) = \frac{t!(N-t)!}{(t-m)!(N-t-n+m)!} = \frac{m!(n-m)!N!}{n!(N-n)!} \cdot p(m|n,t) \quad (5)$$

la (4) diviene:

$$p(H_0|m) = \sum_{j=m}^{h^*} \phi(j/N) \cdot z(j,m,n) / \sum_{i=m}^{N-n+m} \phi(i/N) \cdot z(i,m,n) \quad (6)$$

Il calcolo delle z, e quindi di $p(H_0|m)$, può essere notevolmente abbreviato facendo riferimento alle seguenti relazioni ricorsive, facilmente verificabili (Kotz S., Johnson N.L., 1969, pag. 145):

$$z(i,m,n) = \frac{(N-i)\cdot(i-m+1)}{(i+1)\cdot(N-i-n+m)} \cdot z(i+1,m,n) \quad (7)$$

$$z(i,m,n) = \frac{i\cdot(N-i-n+m+1)}{(N-i+1)\cdot(i-m)} \cdot z(i-1,m,n) \quad (8)$$

per cui basterà calcolare una delle z mediante la (5) per ottenere poi le altre molto rapidamente mediante la (7) o la (8).

Questo procedimento però non è sufficiente a garantire l'assenza di traboccamenti, a causa dei fattoriali presenti nella (5) che almeno una volta dev'essere calcolata, e anche a causa dei prodotti iterati dalla (7) o dalla (8). Per evitare questo inconveniente si può procedere nel modo che passiamo a descrivere.

Poiché $z(t,m,n)$, vista come funzione di t, è proporzionale a $p(m|n,t)$, il punto di massimo t^* di $z(t,m,n)$ coincide con il punto di massimo di $p(m|n,t)$ rispetto a t (ossia la stima di Massima Verosimiglianza di M), che è noto (Kotz S., Johnson N.L., 1969, pag. 146) essere il minimo intero minore o uguale a $t_0 = m\cdot(N+1)/n$. È noto anche che se t_0 è intero si ha un massimo anche in $t_1 = t_0 - 1$. La scelta fra t_0 e t_1 è inessenziale per l'uso che ci interessa, basta però tener presente che per $m = 0$ la scelta $t^* = t_1$ è inammissibile perché è $t_1 < 0$, mentre per $m = n$ è inammissibile $t^* = t_0$ perché risulta $t_0 > N$.

Dividendo allora numeratore e denominatore per $z(t^*,m,n)$, la (6) può esser posta nella forma:

$$p(H_0|m) = \sum_{j=m}^{h^*} \phi(j/N) \cdot w(j, m, n,) / \sum_{i=m}^{N-n+m} \phi(i/N) \cdot w(i, m, n) \qquad (9)$$

con:

$$w(t, m, n) = z(t, m, n)/z(t^*, m, n) \qquad (10)$$

Le quantità w sono ovviamente tutte comprese fra 0 e 1, ed è inoltre $w(t^*, m, n) = 1$. Le (7) ed (8) divengono:

$$w(i, m, n) = \frac{(N-i) \cdot (i-m+1)}{(i+1) \cdot (N-i-n+m)} \cdot w(i+1, m, n) \qquad (11)$$

$$w(i, m, n) = \frac{i \cdot (N-i-n+m+1)}{(N-i+1) \cdot (i-m)} \cdot w(i-1, m, n) \qquad (12)$$

Perciò, partendo da $w(t^*, m, n)$, si possono ottenere facilmente tutte le w senza pericolo di traboccamenti, e con un volume di calcolo alla portata anche dei più piccoli calcolatori.

3. Due proprietà della distribuzione a posteriori di ϑ

Per la (5), ponendo $m_2 = m_1 + 1$, si può porre:

$$\phi\left(\frac{i}{N}\right) \cdot z(i, m_2, n) = \ell(i-1) \cdot \phi\left(\frac{i-1}{N}\right) \cdot z(i-1, m_1, n) \qquad (13)$$

con:

$$\ell(j) = \frac{j+1}{N-j} \cdot R(j) \qquad (14)$$

$$R(j) = \phi\left(\frac{j+1}{N}\right) / \phi\left(\frac{j}{N}\right) \qquad (15)$$

È importante ora notare che, essendo crescente il rapporto $(j+1)/(N-j)$ *se* $R(j)$ *è non decrescente allora* $\ell(j)$ *cresce al crescere di* j.

La classe delle densità $\phi(\vartheta)$ che soddisfano a questa condizione è piuttosto ampia: si verifica facilmente che ad essa appartengono ad esempio la densità esponenziale, la densità Gamma, le densità proporzionali a $(\vartheta+a)^{-b}$ con $a, b > 0$. Densità di questi tipi sembrano adattarsi bene a rappresentare situazioni in cui si ha $p(H_0) \gg p(H_1)$, ossia in cui si ritengono molto più probabili i valori

piccoli di ϑ di quelli grandi. Perciò non sembra troppo limitativo supporre che $\phi(\vartheta)$ sia scelta in modo tale che $R(j)$ sia non decrescente.

Dimostriamo adesso la seguente proprietà:

A) – *Se $R(j)$ è non decrescente, allora al crescere di m, dati ϑ^*, N ed n, decresce la probabilità a posteriori dell'ipotesi H_0 definita dalla (2).*

DIM:

Notiamo anzitutto che la proprietà è dimostrata se è dimostrato che è $p(H_0|m_1) > p(H_0|m_2)$ con $m_2 = m_1 + 1$ ed m_1 qualsiasi. Perciò per la (2) la disuguaglianza da dimostrare è:

$$prob(\vartheta \leq h(m_1)/N|m_1) > prob(\vartheta \leq h(m_2)/N|m_2) \qquad (16)$$

$$m_2 = m_1 + 1$$

Se r è il massimo intero $\leq h(m_1)$, è:

$$prob(\vartheta \leq h(m_1)/N|m_1) = prob(\vartheta \leq r/N|m_1)$$

e inoltre, per la (2):

$$prob(\vartheta \leq h(m_2)/N|m_2) = prob(\vartheta \leq (h(m_1) + 1 - \vartheta^*)/N|m_2) <$$

$$< prob(\vartheta \leq (h(m_1) + 1)/N|m_2) = prob(\vartheta \leq (r+1)/N|m_2)$$

Combinando queste relazioni con la (16) è immediato vedere che condizione sufficiente perché la (16) sia verificata è che risulti:

$$prob(\vartheta \leq r/N|m_1) > prob(\vartheta \leq (r+1)/N|m_2) \qquad (17)$$

Posto infine:

$$C_{1j} = \sum_{i=m_j}^{r} \phi(i/N) \cdot z(i, m_j, n) \qquad j = 1,2$$

$$C_{2j} = \sum_{i=r+1}^{N-n+m_j} \phi(i/N) \cdot z(i, m_j, n) \qquad j = 1,2$$

la (17), per la (6), diviene:

$$C_{11}C_{22} > C_{21}C_{12} + (C_{11} + C_{21}) \cdot z(r+1, m_2, n) \cdot \phi\left(\frac{r+1}{N}\right) \qquad (18)$$

Poiché $R(j)$ è supposto non decrescente, $\ell(j)$ è crescente e perciò, applicando la (13), si ottiene:

$$C_{22} = \sum_{i=r+1}^{N-n+m_2} \ell(i-1) \cdot \phi\left(\frac{i-1}{N}\right) \cdot z(i-1, m_1, n) = \ell(r) \cdot \phi\left(\frac{r}{N}\right) \cdot$$

$$\cdot z(r, m_1, n) + \sum_{j=r+1}^{N-n+m_1} \ell(j) \cdot \phi\left(\frac{j}{N}\right) \cdot z(j, m_1, n) >$$

$$> \ell(r) \cdot \phi\left(\frac{r}{N}\right) \cdot z(r, m_1, n) + \ell(r) \cdot \sum_{j=r+1}^{N-n+m_1} \phi\left(\frac{j}{N}\right) \cdot$$

$$\cdot z(j, m_1, n) = \phi\left(\frac{r+1}{N}\right) \cdot z(r+1, m_2, n) + \ell(r) \cdot C_{21} \tag{19}$$

$$C_{12} = \sum_{i=m_2}^{r} \phi\left(\frac{i}{N}\right) \cdot z(i, m_2, n) = \sum_{i=m_2}^{r} \ell(i-1) \cdot \phi\left(\frac{i-1}{N}\right)$$

$$\cdot z(i-1, m_1, n) = \sum_{i=m_1}^{r} \ell(i)\phi\left(\frac{i}{N}\right) \cdot z(i, m_1, n) - \ell(r) \cdot \phi\left(\frac{r}{N}\right) \cdot$$

$$\cdot z(r, m_1, n) = \sum_{i=m_1}^{r} \ell(i)\phi\left(\frac{i}{N}\right) \cdot z(i, m_1, n) - \phi\left(\frac{r+1}{N}\right) \cdot$$

$$\cdot z(r+1, m_2, n) < \ell(r) \cdot C_{11} - \phi\left(\frac{r+1}{N}\right) \cdot z(r+1, m_2, n) \tag{20}$$

Moltiplicando per C_{11} il primo e l'ultimo membro della (19) e inserendo la (20), si ottiene facilmente la (18) e perciò la (17).

Dimostriamo adesso l'altra proprietà:

B) – *Al crescere di n, dati* ϑ^*, *N ed m cresce* $p(H_0|m)$

DIM:

posto:

$$p(H_0|m) = A(m,n)/B(m,n)$$

con:

$$A(m,n) = \sum_{j=m}^{h^*} \phi\left(\frac{j}{N}\right) \cdot z(j,m,n,)$$

$$B(m,n) = \sum_{j=m}^{N-n+m} \phi\left(\frac{j}{N}\right) \cdot z(j,m,n,) = A(m,n) + C(m,n)$$

$$C(m,n) = \sum_{j=h^*+1}^{N-n+m} \phi\left(\frac{j}{N}\right) \cdot z(j,m,n,)$$

basta dimostrare che B/A decresce al crescere di n.

Infatti, poiché è per la (5):

$$z(j,m,n+1) = (N-j-n+m) \cdot z(j,m,n)$$

ne deriva che è:

$$\frac{B(m,n+1)}{A(m,n+1)} = 1 + \frac{C(m,n+1)}{A(m,n+1)} = 1 +$$

$$+ \frac{\sum_{j=h^*+1}^{N-n-1+m} (N-j-n+m) \cdot \phi\left(\frac{j}{N}\right) \cdot z(j,m,n)}{\sum_{j=m}^{h^*} (N-j-n+m) \cdot \phi\left(\frac{j}{N}\right) \cdot z(j,m,n)} <$$

$$< 1 + \frac{N-h^*-n+m) \cdot \sum_{j=h^*+1}^{N-n-1+m} \phi\left(\frac{j}{N}\right) \cdot z(j,m,n)}{\sum_{j=m}^{h^*} (N-j-n+m) \cdot \phi\left(\frac{j}{N}\right) \cdot z(j,m,n)} <$$

$$< 1 + \frac{(N-h^*-n+m) \cdot C(m,n)}{(N-h^*-n+m) \cdot A(m,n)} = \frac{B(m,n)}{A(m,n)}$$

4. Descrizione del procedimento

All'operatore si richiede l'assegnazione iniziale di una soglia superiore α_0 per la probabilità di errore nella scelta di H_0 e una α_1 per la probabilità di errore nella scelta di H_1. La regola di decisione consisterà nell'accettare come vera H_0 se la probabilità di errore corrispondente, pari a $p(H_1|m)$, risulterà minore

di α_0, mentre si accetterà H_1 se $p(H_0|m)$ risulterà minore di α_1. Essendo $p(H_1|m) = 1 - p(H_0|m)$, le due condizioni sono esprimibili da:

$$p(H_0|m) > 1 - \alpha_0 \qquad \text{scelta di } H_0 \qquad (21)$$
$$p(H_0|m) < \alpha_1 \qquad \text{scelta di } H_1 \qquad (22)$$

Il procedimento risulta esente da possibili situazioni di equivoco a condizione che sia $\alpha_0 + \alpha_1 \leq 1$, altrimenti le disuguaglianze (21) e (22) potrebbero esser verificate contemporaneamente. Questa condizione però non è affatto stringente perché in generale i valori che si scelgono per α_0 e α_1 sono molto piccoli.

Se infine nessuna delle due disuguaglianze è verificata, ossia se risulta:

$$\alpha_1 \leq p(H_0|m) \leq 1 - \alpha_0 \qquad (23)$$

si conclude che l'informazione di cui si dispone è insufficiente per una scelta con i limiti imposti alle probabilità di errore. In tal caso si dovrà aumentare la numerosità di campione rilevando un altro gruppo di osservazioni.

Per quanto riguarda la scelta della numerosità di gruppo, si è seguito il seguente criterio:

- Per il primo gruppo, ossia alla prima interazione del procedimento, si sceglie una numerosità n uguale al minimo valore per il quale, con $m = 0$, si accetta H_0 con probabilità di errore $\leq \alpha_0$ (per la proprietà B del precedente paragrafo, tale valore esiste).
- Per l'i-mo gruppo, sia n_{i-1} la numerosità del gruppo precedente e:

$$\nu_{i-1} \equiv \sum_{j=1}^{i-1} n_j$$

il numero totale delle osservazioni già rilevate; sia inoltre μ_{i-1} il numero totale di individui B in esse riscontrati . Allora per n_i si sceglierà una numerosità tale che $\nu_i (= \sum_{j=1}^{i} n_j)$ sia la minima numerosità di campione per la quale si accetta H_0 con $m = \mu_{i-1}$ (cosa che accade se negli ultimi n_i non si riscontrano individui B).

Con questa scelta della numerosità di gruppo, si arresterà il procedimento alla fine della i–ma iterazione con l'accettazione di H_0, se e solo se risulterà $m_i = 0$.

Se negli n_i elementi estratti durante la $i - ma$ iterazione si riscontra un numero m_i di individui B tale che il campione globale di ν_i elementi, con $\mu_i = m_i + \mu_{i-1}$ di classe B, porti all'accettazione di H_1, il procedimento ovviamente si arresta. Poiché $p(H_0|m)$ decresce al crescere di m (proprietà A

del par. precedente), esiste un intero $\beta(\nu_i)$ tale che per $\mu_i \geq \beta(\nu_i)$ la (21) è la verificata e pertanto si accetta H_1.

Questo equivale a dire che se alla $i - ma$ iterazione si ha:

$$m_i \geq \beta(\nu_i) - \mu_{i-1} = m_i^*$$

si arresta il procedimento accettando H_1. Chiaramente, il procedimento verrà arrestato con l'accettazione di H_1 non appena si riscontrano, nel corso dell'$i - ma$ iterazione, m_i^* individui di classe B.

Infine, se risulta $0 < m_i < m_i^*$, non sono verificate le condizioni né per l'accettazione di H_0 né per quella di H_1, e pertanto si dovrà passare ad un'altra iterazione.

In questo modo, *ogni volta che il procedimento terminerà con l'accettazione di H_0, la numerosità globale di campione sarà risultata minima.*

Infatti, se fino alla $(i-1) - ma$ iterazione si sono riscontrati in tutto μ_{i-1} individui B e il procedimento termina alla $i - ma$ iterazione, allora:

1) Per come è stata scelta la numerosità di ciascun gruppo, alla fine della $i - ma$ iterazione la numerosità globale di campione è quella minima per l'accettazione di H_0 con μ_{i-1} individui B osservati nel campione. Perciò, avendo già osservato i μ_{i-1} individui B all'inizio dell'$i - ma$ iterazione, non era possibile accettare H_0 con la probabilità di errore assegnata prima della fine dell'iterazione stessa.
2) H_0 non poteva essere accettata nelle iterazioni precedenti. Infatti essa non poteva essere accettata prima della fine di ciascuna iterazione, per i motivi già esposti in 1); né poteva essere accettata alla fine di qualcuna di esse, perché il fatto che si sia arrivati all'$i - ma$ iterazione comporta che per ogni $j < i$ debba essere risultato $m_j > 0$.

Pertanto alla fine della $j - ma$ iterazione non poteva essere accettata H_0 in quanto la numerosità globale del campione, essendo minima per l'accettazione di H_0 con μ_{j-1} individui B (ossia quelli osservati nelle iterazioni precedenti) e con la probabilità di errore α_0, non poteva essere sufficiente con $\mu_{j-1} + m_j$ individui B. Notiamo infatti che in caso contrario tale numerosità sarebbe stata sufficiente indipendentemente dalla posizione occupata nel $j - mo$ gruppo dagli m_j individui B in esso riscontrati, quindi anche nell'ipotesi che essi fossero stati gli ultimi m_j individui del gruppo. Ma in tal caso avremmo dovuto poter accettare H_0 già m_j osservazioni prima, perché comunque fossero risultate le ultime m_j avremmo dovuto accettare H_0, per la prorpietà A. Questo evidentemente non è ammissibile, perché la numerosità di ogni gruppo è minimo per l'accettazione di H_0 se in esso non sono osservati individui B.

È ovvio infine che deve essere $\nu_i > \nu_{i-1}$, altrimenti H_0 sarebbe stata accettata alla iterazione $i - 1$; pertanto n_j risulta sempre ≥ 1. Ne deriva, che

essendo la popolazione finita, la procedura dovrà concludersi in un numero finito di iterazioni.

5. Considerazioni conclusive

Per quanto riguarda la densità a priori di ϑ, la costante di normalizzazione viene calcolata imponendo che sia:

$$\sum_{i=0}^{N} \phi\left(\frac{i}{N}\right) = 1 \tag{24}$$

Il calcolo della costante di normalizzazione tramite la (24) è preferibile all'uso delle ben note costanti di normalizzazione delle densità continue, per evitare di introdurre approssimazioni non necessarie. Notiamo qui che la scelta $a = 0$ nell'esponenziale $k \cdot e^{-a\vartheta}$ ci consente di ottenere la densità a priori uniforme come caso particolare. Questa possibilità acquista particolare interesse quando l'operatore, nonostante ritenga molto più attendibile H_0 che H_1 (condizione senza la quale il procedimento esposto non sarebbe per lui conveniente), voglia far derivare il risultato da una posizione a priori di tipo agnostico. Scelte di questo genere ad es. possono avere interesse quando l'operatore si identifica in un ente che deve stabilire la qualità di un prodotto sottoposto al suo giudizio e che pertanto, dovendo porsi al di sopra delle parti, non può inserire informazioni od opinioni soggettive.

Il metodo è stato applicato con successo per la verifica dello stato sanitario del materiale commerciale (ossia vivai di piantine) destinato a nuovi impianti viticoli.

Tipicamente, l'Ente preposto alla verifica si ritrova di fronte a un vivaio con $500 \div 2000$ piantine, e deve stabilire, senza superare assegnate probabilità di errore, se esso sia "sano", ossia se in esso la percentuale di individui infetti non supera una assegnata soglia. Poiché le analisi da eseguire su ogni piantina osservata costano in tempo e denaro, è essenziale minimizzare il campione e perciò si impongono metodi sequenziali. Tuttavia la procedura di Wald non è proponibile, perché dover decidere dopo ogni osservazione significherebbe passare continuamente dal vivaio al laboratorio (spesso assai distante), con un inammissibile dispendio di tempo.

D'altro canto in questi casi ci si aspetta che per lo più i vivai sottoposti all'esame siano sani, perché è appunto interesse e cura del vivaista mantenerli tali per evitare grosse perdite economiche; perció il metodo qui proposto è risultato adeguato, in quanto minimizza il campione ogni volta che il vivaio sottoposto ad esame è effettivamente sano.

In questo caso la scelta della densità a priori è stata per quella uniforme, perché anche se l'operatore ritiene molto probabilmente vera H_0, egli deve

fornire un responso al di sopra delle parti e quindi non può inserire opinioni soggettive o comunque preconcette.

Ringraziamenti

L'autore desidera ringraziare vivamente i professori Angelo Zanella e Mario di Bacco, per le loro preziose osservazioni.

Riferimenti bibliografici

De Mets D.L., Ware J.H., 1980, Group Sequential Methods for Clinical Trials with a One-sided Hypothesis, *Biometrika*, **67**, 651–660.
De Mets D.L., Ware J.H., 1982, Asymmetric Group Sequential Boundaries for Monitoring Clinical Trials, *Biometrika*, **69**, 661–663.
Ehrenfeld S., 1972, On Group Sequential Sampling, *Technometrics*, **14**, 167-174.
Hayre L.S., 1985, Group Sequential Sampling with Variable Group Sizes, *J.R. Statist. Soc. B*, **47**, 90-97.
Kotz S., Johnson N.L., 1969, *Discrete Distributions*, J. Wiley & Sons.
Lan K.K.G., De Mets D.L., 1983, Discrete sequential Boundaries for Clinical Trials, *Biometrika*, **70**, 659–663.
Pocock S.J., 1977, Group Sequential Method in the Design and Analysis of Clinical Trials, Biometrika, **64**, 191–199.
the Group Sequential Approach, *Biometrics*, 38, 153–162.
Pocock S.J., 1982, Interim Analyses for Randomized Clinical Trials: the Group Sequential Approach, *Biometrics*, 38, 153–162.
Spahn M., Ehrenfeld S., 1974, Optimal and Suboptimal Procedures i Group Sequential Sampling, *Nav. Res. Log. Q.*, **21**, 53–68.
Wetherill B.G., 1975, *Sequential Methods in Statistics*, J. Wiley & Sons.

Summary

Group sequential sampling in finite population

A group sequential Bayesian sampling procedure is proposed for choosing between the two hypotheses $H_0 : \vartheta \leq \vartheta^*$ and $H_1 : \vartheta > \vartheta^*$, with given error probabilities, where ϑ is the defective items rate in a finite population. It is hypothesized that the defective items, which are observed during the sampling (without repetition), are removed from the population; therefore the threshold ϑ^* has to be compared not with the initial rate, but with the rate of defective items in the final population.
The method requires that the prior of H_0 is much greater than that of H_1; this hypothesis allows us to choose the global size of the sample in such a way as to minimize it when the procedure ends with the choice of H_0, since in that case the mean incidence of the non–optimality for the choice of H_1 is negligible.
Finally, an efficient algorithm for computing the a posteriori of H_0 is proposed, which is fast and avoids the risk of overflow.