

STATISTICA COMPUTA- ZIONALE

Rubrica a cura di Natale Lauro

Un programma per un test di qualità sequenziale a gruppi in popolazione finita

Franco Caroti Ghelli, Dipartimento di Matematica, Università di Pisa

In un altro articolo dello stesso autore (Caroti Ghelli Franco, 1988) è presentato un metodo per un test di qualità sequenziale a gruppi di numerosità variabile in popolazione finita, che prevede che durante il campionamento, gli individui riconosciuti difettosi vengano eliminati dalla popolazione.

In questo lavoro viene descritto un programma di calcolo della numerosità del gruppo e della regola di arresto per ciascuno stadio della procedura.

1. Introduzione

In altra parte di questo numero della Rivista di Statistica Applicata, viene presentato un metodo di test sequenziale a gruppi in popolazione finita (Caroti Ghelli F., 1988), che si basa su un campionamento sequenziale senza rimpiazzo, eseguito a gruppi di numerosità variabile. In esso si fa l'ipotesi che gli individui di tipo B (difettosi), man mano che vengono riconosciuti durante il campionamento, vengano eliminati definitivamente dalla popolazione, alterandone così la composizione. Perciò l'ipotesi nulla H_0 (percentuale degli individui $B \leq \vartheta^*$, con ϑ^* data) viene definita con riferimento alla popolazione finale, ossia quella che si ottiene quando, al termine dell'intero procedimento, gli individui costituenti i vari gruppi via via rilevati vengono reimmessi nella popolazione, purché non siano risultati di tipo B .

Siano allora n_i il numero di individui da cui è composto il gruppo i -mo ed m_i il numero di individui B riscontrati in esso. Il procedimento si svolge a stadi successivi, l' i -mo dei quali consta di: una determinazione iniziale di n_i e

di una soglia m_i^* , una successiva fase di rilevamento nella quale si estraggono a caso dalla popolazione n_i individui e si osserva il numero m_i di essi che risultano di tipo B ; una fase finale di decisione, che consiste nel confronto di m_i con m_i^* (se risulta $m_i = 0$ si termina il procedimento accettando H_0 , se risulta $m_i \geq m_i^*$ si termina il procedimento rifiutando H_0 , se infine risulta $0 < m_i < m_i^*$ si esegue un altro stadio del procedimento).

I valori di n_i ed m_i^* sono scelti sulla base della richiesta che le probabilità di errore di prima specie e di seconda specie siano non superiori ai valori α_0 e α_1 rispettivamente assegnati e che la numerosità globale di campione sia minima se la procedura si arresta con l'accettazione di H_0 .

All'inizio di ogni stadio i , la determinazione di n_i e di m_i^* necessita dell'uso di un calcolatore, che può essere anche un piccolo "personal", in quanto non sono necessarie né alta velocità né elevata capacità di memoria. Allo scopo pertanto si è ritenuto opportuno realizzare un programma in linguaggio *GW-BASIC*, utilizzabile su una vasta classe di elaboratori di basso costo e perciò di larga diffusione. Alla descrizione di tale programma è rivolto questo lavoro, nella convinzione di fare cosa utile per gli eventuali utenti del metodo.

2. Descrizione dell'algoritmo

Inizialmente il programma richiede che vengano eseguite le seguenti scelte e specificazioni:

- a) La specificazioni della numerosità N iniziale (ossia al primo stadio) della popolazione.
- b) Se con M si indica il numero di individui difettosi inizialmente presenti nella popolazione e con ϑ il rapporto M/N , si richiede la scelta del modello matematico per la densità a priori di ϑ , fra i tre seguenti modelli:

$$\text{Esponenziale} \quad : \quad p(\vartheta) = K \cdot \exp(-a \cdot \vartheta)$$

$$\text{Gamma} \quad : \quad p(\vartheta) = K \cdot \vartheta^{a-1} \cdot \exp(-b \cdot \vartheta)$$

$$\text{Iperbolica} \quad : \quad p(\vartheta) = K \cdot (a + \vartheta)^{-b}$$

Del modello scelto, occorre specificare anche il valore dei parametri a e b , mentre invece la costante di normalizzazione viene calcolata all'interno del programma, tenendo conto che i valori consentiti per ϑ sono: $0, 1/N, 2/N, \dots, N/N$. La costante perciò è definita dalla condizione:

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

Si noti anche che la scelta $a = 0$ nel modello esponenziale corrisponde

- a una densità a priori uniforme. Inoltre nel modello Gamma non è ammessa la scelta $a \leq 1$, perché, per tali valori di a , $p(0)$ divergerebbe. Per analoghi motivi nella iperbolica dovrà essere $a > 0$.
- c) Si deve precisare se lo stadio che interessa sia il primo oppure no; se non è il primo, si dovrà dare il numero globale τ di individui osservati in tutti gli stadi precedenti, e il numero globale m di individui B in essi riscontrati. Qualora si tratti del primo stadio, allora τ ed m sono posti uguali a zero per programma.
- d) Si deve specificare il valore di ϑ^* , che dovrà essere uno dei valori consentiti per ϑ ; qualora l'operatore scegliesse un valore diverso, esso verrebbe rifiutato dal programma, che richiederebbe una scelta fra i due valori consentiti più vicini.
- e) Si debbono dare i valori di α_0 e α_1 .

Passiamo adesso a descrivere l'algoritmo. Faremo riferimento, per quanto riguarda risultati e notazioni, all'articolo precedentemente citato.

Il calcolo di n_i viene effettuato, cercando per tentativi (con il metodo della bisezione) quella numerosità globale di campione $\tau + n_i$ tale che con un totale di m individui B riscontrati (che corrisponde a 0 individui B fra gli n_i individui dell'ultimo gruppo) consenta l'accettazione di H_0 con probabilità di errore $\leq \alpha_0$, e tale che qualunque numerosità globale inferiore, con lo stesso numero m di individui B , comporterebbe per l'accettazione di H_0 una probabilità di errore superiore ad α_0 .

Analogamente, anche il calcolo di m_i^* è effettuato per tentativi, cercando il minimo intero ξ tale che, con numerosità globale di campione $\tau + n_i$ e con ξ individui B osservati in esso, si accetti H_1 con probabilità di errore $\leq \alpha_1$, e ponendo poi $m_i^* = \xi - m$.

Questi calcoli sono realizzati in tempi molto brevi, grazie al metodo di calcolo rapido delle probabilità a posteriori di H_0 , presentato e discusso nell'articolo già citato. Considerata quindi l'importanza dell'argomento, riteniamo opportuno evidenziare le linee lungo le quali si svolge l'algoritmo di calcolo di $p(H_0|m)$. Tale algoritmo si sviluppa nei seguenti passi:

- i) Calcolo di $h^* = (\max. \text{intero} \leq m + \vartheta^* \cdot (N - n))$.
- ii) Calcolo della stima Massima Verosimiglianza per ϑ , ossia del massimo intero t^* con $t^* \leq m(N + 1)/n$; per il caso $m = n$, si pone $t^* = N$.
- iii) A partire da $t = t^*$, e ponendo $w(t^*, m, n) = 1$, per ogni intero positivo $t < t^*$ si esegue iterativamente:

$$w(t, m, n) = \frac{(n - t) \cdot (t - m + 1)}{(t + 1) \cdot (N - t - n + m)} \cdot w(t + 1, m, n) \quad (1)$$

e analogamente, per ogni $t > t^*$ fino a $t = N - n + m$:

$$w(t, m, n) = \frac{t \cdot (N - t - n + m + 1)}{(N - t + 1) \cdot (t - m)} \cdot w(t - 1, m, n) \quad (2)$$

iiii) Si calcola $p(H_0|m)$ mediante la formula:

$$p(H_0|m) = \sum_{j=m}^{h^*} \phi(j/M) \cdot w(j, m, n) / \sum_{i=m}^{N-n+m} \phi(j/N) \cdot w(i, m, n) \quad (3)$$

dove $\phi(\cdot)$ è il valore della densità a priori scelta per ϑ , valutata in (\cdot) .

Questi passi traducono in forma algoritmica i risultati espressi in formule nell'articolo già più volte citato.

Un accorgimento che per lo più consente di abbreviare drasticamente i tempi, deriva dall'osservazione che per N molto grande, il calcolo della sommatoria a denominatore della (3) implica un gran numero di addendi, dei quali la maggior parte possono essere di entità del tutto trascurabile e pertanto possono essere posti = 0 senza modificare apprezzabilmente i risultati. Per stabilire a quale indice si può arrestare la somma, si osservi che $w(j, m, n)$ decresce al crescere di j per $j > t^*$, e inoltre $\phi(\vartheta)$ decresce o rimane costante con esclusione al più (nel caso del modello Gamma) dei valori di ϑ più piccoli. Perciò gli addendi hanno un valore decrescente al crescere di j e perciò, per ogni $j > t^*$ (e maggiore del valore, se $\neq 0$, cui corrisponde il massimo di $\phi(j/N)$), la somma dei termini rimanenti (ossia dall'indice j all'indice $N - n + m$) sarà \leq del termine $\phi(j/N) \cdot w(j, m, n)$ moltiplicato per il numero $(N - n + m - j)$ dei termini rimanenti. Perciò quando j raggiunge un valore per cui $\phi(j/N) \cdot w(j, m, n) \cdot (N - n + m - j)$ risulta trascurabile rispetto alla somma dei termini precedenti ($\sum_{i=m}^{j-1} \phi(i/N) \cdot w(i, m, n)$), si può essere certi che i termini rimanenti non daranno contributo apprezzabile e pertanto la somma viene arrestata.

Seguono il testo del programma e alcuni esempi di conversazione.

Nel primo di questi, si è ipotizzato di iniziare il procedimento in un caso in cui è $N = 1000$, $\vartheta^* = 0.02$, $\alpha_0 = \alpha_1 = 0.05$ e la densità a priori di ϑ è scelta esponenziale con $a = 20$.

Il risultato è: $n_1 = 115$ e $m_1^* = 6$. Nel secondo esempio di conversazione, si è ipotizzato di avere eseguito le 115 osservazioni e di avere osservato in esse 3 individui B ; allora il procedimento prevede di passare a uno stadio successivo, perché 3 individui B sui 115 sono > 0 e quindi non consentono l'accettazione di H_0 , ma sono < 6 e perciò non consentono neppure l'accettazione di H_1 . In questo secondo esempio, alla domanda se sia il primo gruppo di osservazioni, si risponde negativamente, comunicando che sono già state effettuate 115 osservazioni, nelle quali sono stati osservati 3 individui B . Si ottiene allora $n_2 = 174$ e $m_2^* = 12$. Il terzo esempio di conversazione è analogo

al secondo, ma ipotizza di aver ottenuto un individuo B anche nel secondo gruppo e perciò di dover passare ad un terzo stadio.

Riferimenti bibliografici

Caroti Ghelli F., 1988, Un test di qualità mediante campionamento sequenziale a gruppi in popolazione finita, *Rivista di statistica Applicata*, 21, 191–203.

Summary

A Program for a Group Sequential Quality Test in a Finite Population.

In another work of the same author (Caroti Ghelli F., 1988), a method is presented for a group sequential quality test with variable group sizes, which is operating when, during the sampling, the defective items are eliminated. In this work, a program is described for calculating the group sizes and the stopping rule for this method.

Appendice

```

1  PRINT "Controllo di qualità con campionamento senza ripetizione."
2  PRINT "Gli individui B(qualità inferiore) quando osservati"
3  PRINT "vengono eliminati":PRINT
4  PRINT "Dare 1 se dens. a priori = esponenziale, dare 2 se Gamma,"
5  INPUT "dare 3 se Iperbolica";IPR
6  PRINT
7  IF IPR=3 GOTO 14
8  IF IPR=2 GOTO 11
9  PRINT
10 INPUT "p=k*exp(-a*teta); dare a";AP:MAX=0:GOTO 16
11 INPUT "p=k*(teta^(a-1))*exp(-b*teta) ; dare a e b";AP,BP
12 IF AP<=1 THEN PRINT "ap?":IF AP<=1 THEN 11
13 MAX=INT(N*(AP-1)/BP+1):GOTO 16
14 INPUT "p=k*(a+teta)^(-b)) ; dare a e b";AP,BP:MAX=0
15 PRINT
16 INPUT "Qual'è la numerosità iniziale della popolazione?";N:PRINT
17 S=0
18 IF IPR=3 GOTO 34
19 IF IPR=2 GOTO 28
20 FOR I=0 TO N
21 S=S+EXP(-AP*(I/N))
22 NEXT I
23 KP=1/S: GOTO 41
24 FOR I=0 TO N
25 S=S+((I/N)^(AP-1))*EXP(-BP*I/N)
26 NEXT I:GOTO 26
27 FOR I=0 TO N
28 S=S+(AP+I/N)^(-BP)
29 NEXT I
30 GOTO 26
31 PRINT "è il primo gruppo di osservazioni per questa popolazione?"
32 INPUT "se si dare 1, altrimenti un altro intero."; IC:PRINT
33 M=0:NZX=0
34 IF IC=1 THEN 54
35 PRINT "quanti individui in tutto sono stati osservati finora"
36 INPUT "nella popolazione?";NZX:PRINT
37 PRINT "quanti individui b in tutto sono stati osservati finora"
38 INPUT "nella popolazione?";M:PRINT
39 MM=M
40 INPUT "dare la frazione massima di individui b ammessa";FI:PRINT
41 HH=INT(FI*N)
42 FP=HH/N
43 IF FP=FI GOTO 75
44 H1=HH+1
45 X1=H1/N
46 PRINT "frazioni possibili nella popolazione."
47 PRINT "prima frazione =";FP
48 PRINT "seconda frazione =";X1

```

```

69 PRINT "dare 1 se si preferisce la prima, altrimenti un altro"
70 INPUT "intero";ZZ
73 IF ZZ<>1 THEN HH=H1
75 HT=HH/N
77 NN=INT(N/2)+1:K=NN:DH=HH/N:HH=M+DH*(N-M)
78 PRINT
80 PRINT "dare la probabilità di errore ammessa per la scelta:"
82 PRINT "frazione dei b non superiore alla soglia (ipotesi h0)"
83 INPUT A2:PRINT
85 PRINT "dare la probabilità di errore ammessa per la scelta:"
86 PRINT "frazione dei b superiore alla soglia (ipotesi h1)"
87 INPUT A1:PRINT
96 GOSUB 210
100 IF K<=2 AND PO=>1-A2 GOTO 128
105 K=INT(K/2)+1
110 IF PO=1-A2 GOTO 128
115 IF PO<1-A2 GOTO 125
120 NN=NN-K:GOTO 96
125 NN=NN+K:GOTO 96
128 IF NN<=NZX THEN NN=NZX+1
130 PRINT
131 PRINT
132 PRINT "*****risultati*****"
133 PRINT:PRINT "popolazione di";N;"individui"
134 PRINT "individui esaminati nei gruppi precedenti =" ;NZX
135 PRINT "numero di b osservati nei gruppi precedenti =" ;M
136 PRINT "frazione massima di b ammessa =" ;HT
137 PRINT "probabilità di errore per la scelta di h0 =" ;A2
138 PRINT "probabilità di errore per la scelta di h1 =" ;A1
139 PRINT
140 NZZ=NN-NZX
141 PRINT "in questo gruppo vanno osservati " ;NZZ;"individui"
156 MA1=0
160 DM=2*M+2:MM=M+DM:IF MM-M>NZZ THEN MM=M+NZZ
164 HH=MM+DH*(N-MM):K=1
166 IF MM-M>NZZ AND DM>1 THEN MM=M+NZZ
167 IF MM-M=>NZZ THEN K=0
169 GOSUB 210
172 IF PO=A1 GOTO 184
176 IF PO>A1 GOTO 204
180 IF PO<A1 AND DM>1 GOTO 207
184 MA2=MM-M
185 PRINT "se nei";NZZ;"individui osservati in questo gruppo non si"
186 PRINT "riscontreranno individui b verrà accettata per vera h0"
187 IF MA2<NZZ GOTO 190
188 PRINT "altrimenti occorre " :GOTO 194
190 PRINT "se invece si osserveranno almeno";MA2;"individui b,"
192 PRINT "verrà accettata h1; se infine gli individui b saranno"

```

```

193 PRINT "più di";MA1;" ma meno di";MA2;" allora occorre"
194 PRINT "rilevare un altro gruppo di osservazioni."
200 END
204 IF K=0 THEN DM=INT((DM+1)/2)
205 MM=MM+DM:HH=MM+DH*(N-MM):IF K=1 THEN DM=2*DM
206 GOTO 166
207 DM=INT((DM+1)/2):MM=MM-DM:HH=MM+DH*(N-MM):K=0:GOTO 166
210 REM ***** calcolo della probabilità di h0 *****
215 MA=INT(MM*(N+1)/NN)
225 IF MA>N-NN+MM THEN MA=N-NN+MM
226 HH1=HH
227 IF HH<MAX THEN HH1=MAX
230 S=1:A=0:B=1
235 R=S:J=MA
236 GOSUB 350
237 B=B*PAPR
240 IF HH=>MA THEN A=PAPR
245 IF MA=MM GOTO 280
250 FOR I=1 TO MA-MM
255 J=J-1
257 GOSUB 350
260 S=S*(N-J)*(J-MM+1)/((J+1)*(N-J-NN+MM))
265 B=B+S*PAPR:IF J<=HH THEN A=A+S*PAPR
275 NEXT I
280 J=MA:S=R
285 KK=N-NN+MM-MA
290 IF KK<=0 GOTO 335
295 FOR I=1 TO KK
300 J=J+1
302 GOSUB 350
305 S=S*J*(N-J-NN+MM+1)/((N-J+1)*(J-MM))
310 B=B+S*PAPR:IF J<=HH THEN A=A+S*PAPR
325 IF S*(N-NN+MM-J)<B/100 AND J>HH1 THEN 335
330 NEXT I
335 PO=A/B
340 RETURN
350 REM***** calcolo probabilità a priori *****
355 IF IPR=3 GOTO 380:IF IPR=2 GOTO 370
360 PAPR=KP*EXP(-AP*J/N)
365 GOTO 385
370 PAPR=KP*((J/N)^(AP-1))*EXP(-BP*J/N)
375 GOTO 385
380 PAPR=KP*((AP+J/N)^(-BP))
385 RETURN

```



```
run
Controllo di qualità con campionamento senza ripetizione.
Gli individui B(qualità inferiore) quando osservati
vengono eliminati
```

```
Dare 1 se dens. a priori = esponenziale, dare 2 se Gamma,
dare 3 se Iperbolica? 1
```

```
p=k*exp(-a*teta); dare a? 20
```

```
Qual'è la numerosità iniziale della popolazione?? 1000
```

```
è il primo gruppo di osservazioni per questa popolazione?
se si dare 1, altrimenti un altro intero.? 1
```

```
dare la frazione massima di individui b ammessa? .02
```

```
dare la probabilità di errore ammessa per la scelta:
frazione dei b non superiore alla soglia (ipotesi h0)
? .05
```

```
dare la probabilità di errore ammessa per la scelta:
frazione dei b superiore alla soglia (ipotesi h1)
? .05
```

```
*****risultati*****
```

```
popolazione di 1000 individui
individui esaminati nei gruppi precedenti = 0
numero di b osservati nei gruppi precedenti = 0
frazione massima di b ammessa = .02
probabilità di errore per la scelta di h0 = .05
probabilità di errore per la scelta di h1 = .05
```

```
in questo gruppo vanno osservati 115 individui
se nei 115 individui osservati in questo gruppo non si
riscontreranno individui b verrà accettata per vera h0
se invece si osserveranno almeno 6 individui b,
verrà accettata h1; se infine gli individui b saranno
più di 0 ma meno di 6 allora occorre
rilevare un altro gruppo di osservazioni.
```

```
Ok
```

run

Controllo di qualità con campionamento senza ripetizione.
Gli individui B(qualità inferiore) quando osservati
vengono eliminati

Dare 1 se dens. a priori = esponenziale, dare 2 se Gamma,
dare 3 se Iperbolica? 1

$p = k \cdot \exp(-a \cdot \text{teta})$; dare a? 20

Qual'è la numerosità iniziale della popolazione?? 1000

è il primo gruppo di osservazioni per questa popolazione?
se si dare 1, altrimenti un altro intero.? 2

quanti individui in tutto sono stati osservati finora
nella popolazione?? 115

quanti individui b in tutto sono stati osservati finora
nella popolazione?? 3

dare la frazione massima di individui b ammessa? .02

dare la probabilità di errore ammessa per la scelta:
frazione dei b non superiore alla soglia (ipotesi h0)
? .05

dare la probabilità di errore ammessa per la scelta:
frazione dei b superiore alla soglia (ipotesi h1)
? .05

*****risultati*****

popolazione di 1000 individui
individui esaminati nei gruppi precedenti = 115
numero di b osservati nei gruppi precedenti = 3
frazione massima di b ammessa = .02
probabilità di errore per la scelta di h0 = .05
probabilità di errore per la scelta di h1 = .05

in questo gruppo vanno osservati 174 individui
se nei 174 individui osservati in questo gruppo non si
riscontreranno individui b verrà accettata per vera h0
se invece si osserveranno almeno 12 individui b,
verrà accettata h1; se infine gli individui b saranno
più di 0 ma meno di 12 allora occorre
rilevare un altro gruppo di osservazioni.

Ok

run

Controllo di qualità con campionamento senza ripetizione.
Gli individui B(qualità inferiore) quando osservati
vengono eliminati

Dare 1 se dens. a priori = esponenziale, dare 2 se Gamma,
dare 3 se Iperbolica? 1

$p=k*\exp(-a*teta)$; dare a? 20

Qual'è la numerosità iniziale della popolazione?? 1000

è il primo gruppo di osservazioni per questa popolazione?
se si dare 1, altrimenti un altro intero.? 2

quanti individui in tutto sono stati osservati finora
nella popolazione?? 289

quanti individui b in tutto sono stati osservati finora
nella popolazione?? 4

dare la frazione massima di individui b ammessa? .02

dare la probabilità di errore ammessa per la scelta:
frazione dei b non superiore alla soglia (ipotesi h0)
? .05

dare la probabilità di errore ammessa per la scelta:
frazione dei b superiore alla soglia (ipotesi h1)
? .05

*****risultati*****

popolazione di 1000 individui
individui esaminati nei gruppi precedenti = 289
numero di b osservati nei gruppi precedenti = 4
frazione massima di b ammessa = .02
probabilità di errore per la scelta di h0 = .05
probabilità di errore per la scelta di h1 = .05

in questo gruppo vanno osservati 38 individui
se nei 38 individui osservati in questo gruppo non si
riscontreranno individui b verrà accettata per vera h0
se invece si osserveranno almeno 13 individui b,
verrà accettata h1; se infine gli individui b saranno
più di 0 ma meno di 13 allora occorre
rilevare un altro gruppo di osservazioni.

Ok