CONSIDERAZIONI SULLE CONDIZIONI DI APPLICABILITÁ DEL TEOREMA CENTRALE LIMITE DA POPOLAZIONE FINITA

Antonio Gambini

Istituto di Statistica, Università Cattolica, Milano.

1. PREMESSA

Com'è noto, l'esigenza di applicare i teoremi centrali limite a casi concreti presenta, tra gli altri, dei problemi decisionali sulla scelta della numerosità campionaria adeguata per garantire che la Variabile Casuale (V.C.) media campionaria standardizzata sia approssimabile, in distribuzione, alla V.C. Normale standardizzata.

Tali problemi si presentano con caratteristiche simili ma distinte in dipendenza del tipo di campionamento che si impiega.

Nel caso del campione casuale semplice, ad esempio, essi si riducono a quello della individuazione di una numerosità campionaria sufficiente a garantire il soddisfacimento dell'obbiettivo in oggetto. La suddetta numerosità dipendenerà esclusivamente dalla "forma" della popolazione di estrazione del campione. Tuttavia, qualora il campione sia estratto in blocco da una popolazione finita, il problema della numerosità campionaria dipende non soltanto dalla forma della popolazione di estrazione del campione, ma anche dalla numerosità della popolazione stessa.

In questo contesto si inserisce la presente nota che si prefigge di puntualizzare alcuni aspetti sul problema delle relazioni che devono intercorrere tra numerosità campionaria e numerosità della Popolazione necessarie per il verificarsi delle condizioni di applicabilità del teorema centrale limite da popolazione finita.

Si mostrerà, in particolare, che la "bontà" della approssimazione ottenibile con un campione casuale in blocco di numerosità n estratto da una popolazione finita di N unità non è funzione crescente della numerosità campionaria n, $(1 \leq n \leq N)$.

Emergerà cioé che devono essere integrati i suggerimenti di alcuni autori (Cochran, 1977), (Zanella, 1974), (Diana e Salvan, 1987) ed altri, i quali, sulla base di differenti argomentazioni, forniscono esclusivamente elementi per individuare la numerosità campionaria minima n_0 sufficiente per l'ottenimento della approssimazione desiderata (pur con la condizione $n_0 << N$).

Infatti, come si dimostrerà, se n_0 è la numerosità minima suggerita, la numerosità adeguata n andrà scelta in realtà nel rispetto del vincolo:

$$n_0 \leq n \leq N - n_0.$$

2. CONSIDERAZIONI INIZIALI

Sia P una popolazione sulle cui N unità sono presenti le manifestazioni di un generico carattere quantitativo X del quale, mediante un campione casuale scelto in blocco in P, interessa stimare la media μ.

Indicata con x_i la manifestazione di X presente sulla i–esima unità di P, la media μ e la varianza σ^2 di X sono date da:

$$\begin{aligned} \mu &= \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{T}{N}, \\ \sigma^2 &= \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2. \end{aligned} \tag{1}$$

A priori, la cardinalità dell'insieme dei distinti campioni di ampiezza n del tipo suddetto ed associati a P è $\binom{N}{n}$ e questi risultano fra loro equiprobabili. Inoltre, le rispettive medie m_n costituiscono le determinazioni della v.c. M_n "media campionaria" caratterizzata per cose note da media e varianza:

$$\begin{aligned} E(M_n) &= \mu, \\ Var(M_n) &= \frac{N-n}{n}\frac{\sigma^2}{N-1} = \sigma^2_{M_n}. \end{aligned} \tag{2}$$

Ciò premesso, deve osservarsi che ad ogni campione di ampiezza n prelevabile in P corrisponde un campione di ampiezza $(N\!-\!n)$ formato dalle restanti $(N\!-\!n)$ unità della stessa P.

Ciascuno di tali campioni, che sono anch'essi in numerosità $\binom{N}{n} = \binom{N}{N-n}$, ha media m_{N-n} che si presenta come determinazione della v.c. M_{N-n} per la quale, sempre per cose note, risulta:

$$\begin{aligned} E(M_{N-n}) &= \mu, \\ Var(M_{N-n}) &= \frac{n}{N-n}\frac{\sigma^2}{N-1} = \sigma^2_{M_{N-n}} \end{aligned} \tag{3}$$

È immediato osservare dal confronto fra la seconda delle (2) e la seconda delle (3), che fra $\sigma^2_{M_n}$ e $\sigma^2_{M_{N-n}}$ intercorre il seguente legame:

$$\sigma^2_{M_{N-n}} = \left(\frac{n}{N-n}\right)^2 \sigma^2_{M_n}. \tag{4}$$

Verrà ora mostrato che anche per la media m_n relativa al generico campione di ampiezza *n* prelevabile in *P* e la media m_{N-n} relativa alle restanti *(N–n)* unità della stessa *P* intercorre il legame:

$$m_{N-n} = \mu - \frac{m_n - \mu}{(N-n)/n}. \tag{5}$$

Infatti, indicati con t_n, t_{N-n} e T l'ammmontare del carattere X rispettivamente nel campione di ampiezza *n,* nelle restanti *(N–n)* unità di *P* e nell'intera popolazione *P,* tenendo conto che:

$$t_n = nm_n,\; t_{N-n} = (N-n)m_{N-n},\; T = n\mu,$$

si ha:

$$m_{N-n} = \frac{t_{N-n}}{N-n} = \frac{T - t_n}{N-n}. \tag{6}$$

Ora, aggiungendo e togliendo nell'ultimo membro della (6) la quantità $n\mu$, con semplici riduzioni, si ottiene, per l'appunto:

$$m_{N-n} = \mu - \frac{m_n - \mu}{(N-n)/n},$$

in accordo con la (5).

La corrispondenza biunivoca che intercorre fra m_n ed m_{N-n} caratterizza ovviamente anche le v.c. M_n ed M_{N-n} descritte dalle medesime, essendo

$$M_{N-n} = \mu - \frac{M_n - \mu}{(N-n)/n}. \tag{7}$$

Si proverà ora che, per effetto di tale corrispondenza, fra le v.c. standardizzate:

$$Z_n = \frac{M_n - \mu}{\sigma_n} \qquad Z_{N-n} = \frac{M_{N-n} - \mu}{\sigma_{N-n}} \tag{8}$$

intercorre il legame:

$$Z_{N-n} = -Z_n \tag{9}$$

Infatti, impiegando la (7) e la (4), si ha:

$$Z_{N-n} = \frac{M_{N-n} - \mu}{\sigma_{N-n}} = \frac{\left[\mu - n(M_n - \mu)/(N-n)\right] - \mu}{\left[n/(N-n)\right]\sigma_{M_n}} = -\frac{M_n - \mu}{\sigma_{M_n}} = -Z_n,$$

dove Z_n è fornita dalla prima delle (8).

Ne segue che le due v.c. Z_n e Z_{N-n} sono fra loro speculari rispetto allo zero con la conseguenza che la probabilità $P(Z_n = z_n)$ con cui la v.c. Z_n assume la determinazione:

$$z_n = \frac{m_n - \mu}{\sigma_n}$$

è uguale alla probabilità $P(Z_{N-n} = z_{N-n})$ con cui la v.c. Z_{N-n} assume la determinazione:

$$z_{N-n} = \frac{m_{N-n} - \mu}{\sigma_{N-n}} = -z_n.$$

Si osservi da ultimo che poiché dalla (5) e dalla (4), mediante la posizione $n=N/2$, si ottiene:

$$\begin{aligned} m_{N-N/2} - \mu &= -\left(m_{N/2} - \mu\right), \\ \sigma^2_{N-N/2} &= \sigma^2_{N/2}, \end{aligned}$$

tenendo conto della (9) si ottiene anche l'uguaglianza in distribuzione delle V.C. $Z_{N/2}$ e $Z_{N-N/2}$. Sussiste pertanto la relazione:

$$Z_{N/2} \stackrel{d}{=} Z_{N-N/2}. \tag{10}$$

L'unione delle proprietà di relazione sopra evidenziate ed applicate ad $M_{N/2}$ e $M_{N-N/2}$ (specularità fornita dalla (9) ed uguaglianza fornita dalla (10)) porta infine a concludere che la distribuzione di probabilità $M_{N/2}$ è simmetrica.

4. CONCLUSIONI

Dai risultati sopra riportati possono essere dedotte alcune interessanti considerazioni. Infatti, siano:

$F_n(z)$: Funzione di ripartizione della V.C. Z_n,

$\phi(z)$: Funzione di ripartizione della V.C. Normale Standard.

Allora dalla (9) e dalla (10) è possibile dedurre che ogni norma *Cn* del tipo

$$C_n = \psi\left(\left| F_n(z) - \phi(z) \right|\right)$$

con ψ reale, non negativa, non decrescente ed uguale a zero se e solo se

$$\left| F_n(z) - \phi(z) \right| = 0 \qquad \forall z,$$

assume il minimo valore in corrispondenza di $n = N/2$.

Dalla (10) è immediato dedurre che sussiste inoltre la relazione

$$C_n = C_{N-n}.$$

La metrica C_n per specifiche scelte del funzionale ψ include la distanza di Kolmogorov

$$\psi = sup \left| F_n(z) - \phi(z) \right|,$$

quella di Von Mises Cramer Fisz

$$\psi = \int \left(F_n(z) - \phi(z)\right)^2 dz,$$

e quella di Anderson–Darling

$$\psi = \int \left(F_n(z) - \phi(z)\right)^2 \lambda(z) dz$$

con $\lambda(z)$ una qualsiasi funzione di ponderazione non negativa.*

Per quanto detto è possibile concludere, avendo presenti le proprietà di regolarità della metrica C_n per cui ogni allontanamento di n da $N/2$ comporta un conseguente possibile non avvicinamento alla simmetria, che la numerosità campionaria $n = N/2$ rappresenta, per una popolazione P di numerosità N, la condizione migliore per applicare, nei casi concreti, il teorema centrale del limite (da popolazione finita).

Numerosità campionarie differenti, sia maggiori che minori di $N/2$, renderanno meno buono, in egual misura, il grado C_n di approssimazione della V.C. $F_n(z)$ (Funzione di ripartizione della V.C. Media Campionaria standardizzata) alla V.C. $\varphi(z)$ in funzione esclusiva della distanza di n da $N/2$ ($|n-N/2|$).

Rimane in tal modo provato quanto anticipato nella introduzione a proposito delle relazioni che devono intercorrere tra numerosità campionaria n e numerosità

---

* *Solitamente viene posto* $\lambda(z) = 1/\phi(z)\ (1-\phi(z))$.

della popolazione P, ovvero che se n_0 è il limite minimo suggerito per un campione estratto da un popolazione P di numerosità N, in realtà, per un corretto impiego del teorema centrale del limite da popolazione finita, la numerosità adeguata n andrà scelta nel rispetto del vincolo seguente:

$$n_0 \leq n \leq N - n_0.$$

RIFERIMENTI BIBLIOGRAFICI

Cantelli F.P., (1933), *Sulla determinazione empirica delle leggi di probabilità,* Giorn. Ist. Ital. Attuari. 4, 421–424.

Cassel C.M., Sarndal C.E., Wretman J.M., (1977), *Foundations of Inference in Survey Sampling,* John Wiley & Sons, New York.

Cochran W.G., (1977), *Sampling Techniques,* John Wiley & Sons, New York.

Diana G., Salvan A., (1987), *Campionamento da popolazioni finite,* CLEUP, Padova.

Erdos P., Renyi A., (1959), *On the central limit theorem for samples from a finite population,* Matem. Kutato Intezet Kozlem, 4, 49.

Gambini A., (1984), *Problemi di numerosità campionaria relativi alla applicazione del teorema centrale del limite,* Atti della XXXII riunione scientifica della SIS, Sorrento, 1984.

Gambini A., (1988), *Sull'uso degli indici di forma nella applicazione del teorema centrale del limite,* Atti della XXXIV riunione scientifica della SIS, Siena 1988.

Glivenko V., (1933), *Sulla determinazione empirica delle leggi di probabilità,* Giorn. Ist. Ital. Attuari. 4, 92–99.

Hajek J., (1960), *Limiting distributions in simple random sampling from a finite population,* Publ. Math. Inst. Hungarian Acad. Sci., 5, 361–374.

Hansen M.H., Hurwitz W.N., Madow W.G., (1953), *Sample survey method and theory,* John Wiley & Sons, New York.

Landenna G., Marasini D., (1990), *Metodi Statistici non Parametrici,* Il Mulino, Bologna.

Zanella A., (1974), *Elementi di teoria del campionamento da popolazioni finite,* CLEUP, Padova.

SUMMARY

In this paper, the author outlines some key points about the necessary relations between sample size and population size to verify the applicability conditions of central limit theorem for samples from a finite population.

It is proved that goodness of approximation to Normal distribution is not an increasing function of sample size n.

In fact, under the conditions above mentioned, the distribution of the sample mean is symmetrical for $n = N/2$ and the goodness of approximation to Normal achieves its maximum level for this value of n, decreasing for different sample size both less and greater than $N/2$.

So the suggestion of different authors to choose a sample size greater than a minimum value n_0 is necessary but not sufficient.

Sample size must lie between n_0 and $N-n_0$.