Università degli Studi di Cassino, Anno accademico 2004-2005
Corso di Statistica 2, Prof. M. Furno

Esercitazione del 22/02/2005
dott. Claudio Conversano

Esercizio 1
Si supponga di estrarre con ripetizione 10 palline da un'urna contenente 1000 palline, delle quali il 10% sono di colore rosso. Si determini:
a) La distribuzione campionaria della proporzione di palline rosse nel campione;
b) La probabilità di ottenere un campione con almeno 7 palline rosse.

SVOLGIMENTO

a) La distribuzione campionaria della proporzione di palline rosse nel campione;

L'universo di campioni di ordine 10 è formato da 1000^{10} elementi. Se si contano le palline rosse estratte in ciascun campione si ottiene una v.c. X che può assumere valori 1,2,….,10 ciascuno con probabilità pari alla somma delle probabilità di tutti i campioni che danno luogo a quel risultato. La distribuzione della v.c. $X/10$ è la distribuzione campionaria cercata.
In questo caso non è possibile costruire l'universo di tutti i possibili campioni (da cui poi derivare la distribuzione di $X/10$). Poiché la distribuzione di frequenza del carattere dicotomico "estrazione di una pallina rossa" è rappresentata da una v.c. di Bernoulli $B(1,p=0.10)$, in cui p rappresenta la proporzione di palline rosse nella popolazione, la v.c. $X/10$ ha una distribuzione Binomiale $Bin(n=10, p=0.10)$, la cui funzione di probabilità è:

$$f\left(\frac{X}{10}\right)=\binom{10}{x}0.1^x\left(1-0.1\right)^{10-x}$$

Per $x=0,1,\ldots,10$ si ottiene la distribuzione campionaria della v.c. $X/10$:

$$f\left(\frac{0}{10}\right)=\binom{10}{0}0.1^0\left(1-0.1\right)^{10-0}=0.34868$$
$$f\left(\frac{1}{10}\right)=\binom{10}{1}0.1^1\left(1-0.1\right)^{10-1}=0.38742$$
$$f\left(\frac{2}{10}\right)=\binom{10}{2}0.1^2\left(1-0.1\right)^{10-2}=0.19371$$
$$f\left(\frac{3}{10}\right)=\binom{10}{3}0.1^3\left(1-0.1\right)^{10-3}=0.05739$$
$$f\left(\frac{4}{10}\right)=\binom{10}{4}0.1^4\left(1-0.1\right)^{10-4}=0.01116$$
$$f\left(\frac{5}{10}\right)=\binom{10}{5}0.1^5\left(1-0.1\right)^{10-5}=0.00149$$
$$f\left(\frac{6}{10}\right)=\binom{10}{6}0.1^6\left(1-0.1\right)^{10-6}=0.00014$$
$$f\left(\frac{7}{10}\right)=\binom{10}{7}0.1^7\left(1-0.1\right)^{10-7}=0.00001$$

$$f\left(\frac{8}{10}\right)=\binom{10}{8}0.1^{8}\left(1-0.1\right)^{10-8}=0.00000$$
$$f\left(\frac{9}{10}\right)=\binom{10}{9}0.1^{9}\left(1-0.1\right)^{10-9}=0.00000$$
$$f\left(\frac{10}{10}\right)=\binom{10}{10}0.1^{10}\left(1-0.1\right)^{10-10}=0.00000$$

b) La probabilità di ottenere un campione con almeno 7 palline rosse.

La probabilità di estrarre un campione con almeno 7 palline rosse si ottiene sommando le probabilità:

$$\sum_{i=7}^{10} f\left(\frac{i}{10}\right)=\binom{10}{i}0.1^{i}\left(1-0.1\right)^{10-i}=0.00001$$

Esercizio 2
La durata delle telefonate effettuate ad un call-center nella fascia oraria 9-11 ha una distribuzione normale con media 175 secondi e varianza 16. Si determini la probabilità che le prime 10 chiamate effettuate nella fascia oraria di riferimento nella giornata di domani abbiano una durata media compresa tra 173 e 176 secondi.

SVOLGIMENTO

Anche in questo caso non è possibile costruire l'universo di tutti i possibili campioni di ordine 10. Si può ritenere che la distribuzione delle chiamate effettuate nella fascia oraria….., segua una distribuzione normale con media 175 e varianza 16. Sotto questa ipotesi la media campionaria ha anch'essa una distribuzione normale con media μ e varianza σ^2/n.

La probabilità richiesta sarà quindi data da:

$$P\left(173 \le \bar{X} \le 176\right) = P\left(\frac{173-175}{4/\sqrt{10}} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le \frac{176-175}{4/\sqrt{10}}\right) =$$

$$= P\left(-1.58 \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le 0.79\right) = 0.78542 - \left(1 - 0.94295\right) = 0.72819$$

Esercizio 3
Nella giornata di ieri il programma in onda in prima serata sulla rete X ha registrato uno *share* del 40%. Si determini la probabilità che, estraendo con ripetizione un campione di 1000 telespettatori, almeno il 38% di essi abbia guardato il programma in oggetto.

SVOLGIMENTO

L'universo di campioni di ordine 1000 è formato da N^{1000} elementi. Se si contano i telespettatori che hanno assistito al programma della rete X in ciascun campione si ottiene una v.c. X che può assumere valori 1,2,….,1000 ciascuno con probabilità pari alla somma delle probabilità di tutti i campioni che danno luogo a quel risultato. La distribuzione della v.c. $X/1000$ è la distribuzione campionaria cercata.

In questo caso non è possibile costruire l'universo di tutti i possibili campioni (da cui poi derivare la distribuzione di $X/1000$). Poiché la distribuzione di frequenza del carattere dicotomico "estrazione di un telespettatore che ha assistito al programma della rete X" è rappresentata da una v.c. di Bernoulli $B(1,p=0.40)$, in cui p rappresenta la proporzione di telespettatori che ha assistito al programma in oggetto nella popolazione, la v.c. $X/1000$ ha una distribuzione Binomiale $Bin(n=1000, p=0.40)$, la cui funzione di probabilità è:

$$f\left(\frac{X}{1000}\right)=\binom{1000}{x}0.4^{x}\left(1-0.4\right)^{1000-x}$$

La probabilità richiesta può essere calcolata nel seguente modo:

$$P\left(\frac{X}{1000}\geq 0.38\right)=\sum_{x=380}^{1000}\binom{1000}{x}0.4^{x}\left(1-0.4\right)^{1000-x}$$

Poiché per n che tende all'infinito la v.c. $X/1000$ tende ad una distribuzione normale $N\left(\mu=p,\sigma^2=p(1-p)/n\right)$ quando si ha un campione sufficientemente grande si può approssimare la distribuzione binomiale con la distribuzione normale. La probabilità cercata risulta:

$$P\left(Z\geq\frac{0.38-0.40}{\sqrt{\frac{(0.4)(0.6)}{1000}}}\right)=P\left(Z\geq -1.29\right)=0.90147$$

Esercizio 4

Sia X il tempo di percorrenza di una intera corsa dell'autobus "128". Si suppone che X sia distribuito modo normale con media incognita e varianza $\sigma^2 = 16$. Si determini la probabilità che la devianza campionaria nS^2 in un campione di $n=13$ corse risulti superiore a 6.304 σ^2.

SVOLGIMENTO

La devianza campionaria nS^2 è il numeratore della varianza campionaria S^2. Poiché il rapporto $nS^2/\sigma^2 \approx \chi_{(n-1)}$ avremo:

$$P\left(nS^2 \geq 6.304\sigma^2\right) = P\left(nS^2/\sigma^2 \geq 6.304\right) = P\left(\chi_{(n-1)} \geq 6.304\right) = P\left(\chi_{(12)} \geq 6.304\right) = 0.90$$

Esercizio 5
Sia $X_n = (x_1, x_2, \dots, x_n)$ un campione casuale proveniente da una popolazione con media uguale a μ e varianza σ^2.
Determinare l'efficienza relativa del valore centrale campionario (semisomma dei due valori estremi campionari) rispetto alla media campionaria.

SVOLGIMENTO

Sia T_n lo stimatore valore centrale campionario: $T_n = \frac{X_{(1)} + X_{(n)}}{2}$.

Sia il valore centrale campionario che la media campionaria sono non distorti rispetto al parametro μ.
Infatti:

a) $E(T_n) = E\left(\frac{X_{(1)} + X_{(n)}}{2}\right) = \frac{\mu + \mu}{2} = \mu$

b) $E(\bar{X}) = \mu$

L'efficienza relativa di T_n rispetto alla media campionaria è:

$$eff\left(T_n \mid \bar{X}\right) = \frac{\frac{1}{Var(T_n)}}{\frac{1}{Var(\bar{X}_n)}} = \frac{Var(\bar{X}_n)}{Var(T_n)}$$

In cui:

$$Var(T_n) = Var\left(\frac{X_{(1)} + X_{(n)}}{2}\right) = \frac{Var\left(X_{(1)}\right) + Var\left(X_{(n)}\right)}{4} = \frac{2\sigma^2}{4} = \frac{\sigma^2}{2}$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

Sostituendo questi risultati all'interno dell'espressione $eff\left(T_n \mid \bar{X}\right)$ si ottiene:

$$eff\left(T_n \mid \bar{X}\right) = \frac{Var(\bar{X}_n)}{Var(T_n)} = \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{2}} = \frac{2}{n} \leq 1$$

L'uguaglianza sussiste solo nel caso *n=2*. In tal caso il valore centrale campionario coincide con la media campionaria. Invece per *n>2* lo stimatore T_n risulta meno efficiente della media campionaria.

Esercizio 6

Data una popolazione normale con media nulla e varianza σ^2 ed estratto un campione casuale di ordine *n=2*, si stabilisca quale, fra i seguenti stimatori di σ^2 risulta preferibile e perché.

$$T_1 = X_1^2; \qquad T_2 = X_1^2 + X_2^2; \qquad T_3 = \frac{T_2}{2} = \left(\frac{X_1^2 + X_2^2}{2}\right)$$

SVOLGIMENTO

Calcoliamo la distorsione dei 3 stimatori:

$E(T_1) = E(X_1^2) = \sigma^2$ perché la media è nulla;

$$E(T_2) = E(X_1^2 + X_2^2) = 2\sigma^2$$

$$E(T_3) = \frac{E(T_2)}{2} = \frac{2\sigma^2}{2} = \sigma^2$$

Il primo ed il terzo stimatore risultano corretti mentre il secondo è distorto. Per quest'ultimo la distorsione è:

$$b(T_2) = 2\sigma^2 - \sigma^2 = \sigma^2$$

Il confronto tra i tre stimatori va fatto sulla base dei loro rispettivi Errori Quadratici Medi (EQM):

$$EQM(T_1) = Var(T_1) + [b(T_1)]^2 =$$
$$= Var(T_1) - 0^2 = E(T_1^2) - [E(T_1)]^2 =$$
$$= E(X_1^4) - [E(X_1^2)]^2 = \mu_4 - \sigma^4 = 3\sigma^4 - \sigma^4 = 2\sigma^4$$

perché per una v.c. normale $\mu_4 = 3\sigma^4$

$$Var(T_2) = E(T_2^2) - [E(T_2)]^2 = E\left[(X_1^2 + X_2^2)^2\right] - (2\sigma^2)^2 = E(X_1^4) + E(X_1^4) + 2E(X_1^2 \cdot X_1^2) - 4\sigma^4 =$$
$$= 2\mu_4 - 2E(X_1^2)E(X_2^2) - 4\sigma^4 = 6\sigma^4 + 2\sigma^4 - 4\sigma^4 = 4\sigma^4$$

$$EQM(T_2) = Var(T_2) + [b(T_2)]^2 = 4\sigma^4 + (\sigma^2)^2 = 5\sigma^4$$

$$EQM(T_3) = Var(T_3) = \frac{Var(T_2)}{2^2} = \frac{4\sigma^4}{4} = \sigma^4$$

Lo stimatore che presenta la minore variabilità intorno al parametro σ^2 e T_3.