Probabilità e Statistica

Campionamento

Marco Pietro Longhi

C.d.L.: Ingegneria Elettronica e delle Telecomunicazioni, Ingegneria Informatica
a.s. 2020/2021

disuguaglianza di Chebyschev

disuguaglianza di Markov

se $g(x) \geq 0$ ed è continua e $r > 0$ allora

$$P[g(X) \geq r] \leq \frac{1}{r} E[g(X)]$$

disuguaglianza di Chebyschev

se $r > 0$ si ha

$$P[|X - \mu_X| \geq r\,\sigma_X] \leq \frac{1}{r^2}$$

disuguaglianza di Chebyschev applicata alla media campionaria

$$P\left[\left|\overline{X}_n - \mu\right| \geq \epsilon\right] \leq \frac{var\left[\overline{X}_n\right]}{\epsilon^2} = \frac{\sigma^2}{n\,\epsilon^2}$$

$$P\left[\left|\overline{X}_n - \mu\right| < \epsilon\right] \geq 1 - \frac{\sigma^2}{n\,\epsilon^2}$$

Esercizio

Durante un processo di produzione vengono estratti 300 campioni di un pezzo e ne viene misurato lo spessore in *mm*, fornendo una media campionaria di $2,6\,mm$. Si sa inoltre che la varianza della variabile casuale che dá lo spessore é $0,4\,mm^2$. Qual é la probabilitá che il valore atteso dello spessore sia inferiore a $2,5\,mm$?

Sia X la variabile casuale che denota lo spessore in *mm*, posto

$$Z = \frac{\overline{X}_n - \mu_X}{\sigma_X/\sqrt{n}} \quad \text{con} \quad \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

dal Teorema del limite centrale risulta che $Z \sim N(0;1)$ per $n \geq 30$.

Essendo $n = 300$ possiamo utilizzare l'approssimazione indicata.

$$\mu_X < 2,5 \Longrightarrow -\mu_X > -2,5 \Longrightarrow 2,6 - \mu_X > 2,6 - 2,5$$

$$\text{da cui} \quad \frac{2,6 - 2,5}{\sqrt{0,4/300}} \Longrightarrow \frac{2,6 - \mu_X}{\sqrt{0,4/300}} > 2,739 \approx 2,74$$

$$\text{pertanto} \qquad \mu_X < 2,5 \qquad \text{se} \qquad Z > 2,74$$

$$P[\mu_X < 2,5] = P[Z > 2,74] = 1 - \Phi\,(2,74) = 1 - 0,99693 = 0,00307$$

Esercizio

In un processo di produzione vengono prodotte sbarre con una lunghezza data da una distribuzione normale, di cui si conosce la deviazione standard, pari a $0,2\,m$. Si esegue un campionamento per calcolare il valore atteso, misurando un numero n di sbarre e calcolando la media. Qual é la numerositá del campione che garantisce una probabilitá superiore al 97% che la media campionaria disti dal valore atteso per meno di 1 cm?

Sia X la v.c. che denota la lunghezza delle sbarre in m e sia

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

bisogna determinare n tale che

$$P\left[\left|\overline{X}_n - \mu_X\right| < 0,01\right] > 0,97$$

Essendo

$$\overline{X}_n \sim N\left(\mu_X; \frac{\sigma_X^2}{n}\right) \Longrightarrow Z \sim N(0;1) \quad \text{con} \quad Z = \frac{\overline{X}_n - \mu_X}{\sigma_X/\sqrt{n}}$$

da cui

$$\left|\overline{X}_n - \mu_X\right| = \frac{\sigma_X}{\sqrt{n}}|Z| = \frac{0,2}{\sqrt{n}}|Z|$$

quindi

$$P\left[\left|\overline{X}_n - \mu_X\right| < 0,01\right] = P\left[\frac{0,2}{\sqrt{n}}|Z| < 0,01\right] =$$

$$= P\left[|Z| < \frac{0,01}{0,2/\sqrt{n}}\right] = P\left[|Z| < 0,05\sqrt{n}\right]$$

pertanto

$$P\left[\left|\overline{X}_n - \mu_X\right| < 0,01\right] > 0,97 \Longrightarrow P\left[|Z| < 0,05\sqrt{n}\right] > 0,97$$

$$P\left[|Z| < 0,05\sqrt{n}\right] = \Phi\left(0,05\sqrt{n}\right) - \Phi\left(-0,05\sqrt{n}\right) =$$
$$= \Phi\left(0,05\sqrt{n}\right) - \left[1 - \Phi\left(0,05\sqrt{n}\right)\right] = 2\Phi\left(0,05\sqrt{n}\right) - 1$$

quindi

$$2\Phi\left(0,05\sqrt{n}\right) - 1 > 0,97 \Longrightarrow \Phi\left(0,05\sqrt{n}\right) > 0,985$$

dalle tavole $\quad \Phi(2,17) = 0,985$

da cui

$$0,05\sqrt{n} > 2,17 \Longrightarrow n > \left(\frac{2,17}{0,05}\right)^2 \approx 1884$$

Esercizio

[Tema d'esame del 07/12/2004-E5]
Si supponga che una distribuzione con media incognita abbia deviazione standard uguale a 1. Quanto deve essere grande un campione affinché si abbia una probabilitá almeno del 90% che la media campionaria $\overline{X}_n$ disti meno di 0.5 dalla media della popolazione?

Bisogna determinare n tale che

$$P\left[\left|\overline{X}_n - \mu_X\right| < 0,5\right] \geq 0,90$$

La *legge dei grandi numeri* ci garantisce che pur di prendere un campione sufficientemente numeroso, l'evento secondo cui la media campionaria $\overline{X}_n$ si discosta di "pochissimo" (nel nostro caso meno di $0,5$) dal valore atteso puó avere probabilitá arbitrariamente vicino ad 1. Pertanto esisterá sicuramente un valore opportuno della dimensione n del campione tale che la probabilitá di questo evento sia maggiore di $0,90$

$$P\left[\left|\overline{X}_n - \mu_X\right| < 0,5\right] = 1 - P\left[\left|\overline{X}_n - \mu_X\right| \geq 0,5\right] =$$

$$= 1 - P\left[\left|\overline{X}_n - \mu_X\right|^2 \geq 0,25\right] \underbrace{\geq}_{(Markov)} 1 - \frac{E\left(\left|\overline{X}_n - \mu_X\right|^2\right)}{0,25} = 1 - \frac{\sigma^2/n}{0,25}$$

Pertanto, poiché $\sigma = 1$,

$$1 - \frac{\sigma^2/n}{0,25} = 0,90 \quad \Longrightarrow \quad \frac{\sigma^2/n}{0,25} = 0,10 \quad \Longrightarrow n = 40$$

In conclusione la probabilitá che la media campionaria differisca dalla media incognita della popolazione meno di $0,5$ é almeno del 90% se si prende un campione casuale di ampiezza superiore a 40.

Esercizio

[Tema d'esame del 08/07/2019-C7]
I pacchi da sei bottiglie di una marca di acqua minerale hanno un peso medio di $9,3\,Kg$, con varianza $\sigma^2 = 0,4\,Kg^2$. Calcolare la probabilità che un campione di 20 pacchi abbia un peso totale maggiore di $184\,Kg$.

[Teorema del limite centrale 0.75804]

Esercizio

[Tema d'esame del 17/06/2019-C6]
Da una popolazione con varianza $\sigma^2 = 121$ si estrae un campione di numerosità 64. Calcolare la probabilità che la media campionaria differisca dalla media della popolazione per più di 2 unità.

[Teorema del limite centrale 0.14706]