Teorema del limite centrale - Correzione di continuità

Leonardo Bizzoni

May 3, 2024

Dato un campione aleatorio $X_1, X_2, \dots, X_n$ con $E[X_i] = \mu$ e $Var[X_i] = \sigma^2$ allora $\lim_{n \to +\infty} P\left(\frac{X_1 + \dots + X_n - \mu}{\sigma} \leq t\right) \simeq \Phi(t)$. Dove $\Phi(t)$ è la funzione di ripartizione di una variabile aleatoria normale standard e $\frac{X_1 + \dots + X_n - \mu}{\sigma}$ è la standardizzazione del campione.

È possibile utilizzare questa approssimazione se abbiamo n grande ($n \geq 30$). Nel caso particolare di un campione composto da variabili aleatorie $Be(p)$ si richiede anche che $np \geq 5$ e $n(1-p) \geq 5$.

1 Esempio con correzione di continuità

> Si lanci 100 volte una moneta equilibrata. Qual è la probabilità che il numero di teste sia compreso tra 40 e 70 estremi inclusi?

$X_k = \begin{cases} 1 & \text{se il k-esimo lancio da testa} \\ 0 & \text{altrimenti} \end{cases}$, $X_k \sim Be(p = \frac{1}{2})$ su 100 lancia abbiamo che $X \sim Bin(n = 100, p = \frac{1}{2})$.

Dobbiamo quindi calcolare $\sum_{k=40}^{70} \binom{100}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{100-k} = 0.9823$.

```
res <- 0
for (k in 40:70) {
    res <- res + choose(100, k) * (1/2)^k * (1/2)^(100-k)
}
res
```

Sapendo che $E[X_i] = \frac{1}{2}$, $Var[X_i] = \frac{1}{4}$, utilizzando il teorema del limite centrale possiamo affermare che $\frac{\overline{X_i} - 100 * \frac{1}{2}}{\frac{1}{2}\sqrt{100}} = \frac{\overline{X_i} - 50}{5} \sim N(0, 1)$. A questo punto $P\left(40 \leq \overline{X_i} \leq 70\right) \simeq P\left(\frac{40-50}{5} \leq \frac{\overline{X_i}-50}{5} \leq \frac{70-50}{5}\right) = P\left(-2 \leq \frac{\overline{X_i}-50}{5} \leq 4\right) = \Phi(4) - \Phi(-2) = 0.9772$, gran parte dell'errore è dovuto al fatto che X_i ha distribuzione discreta mentre l'approssimazione è continua.

Per migliorare l'appossimazione si sostituiscono 40 e 70 con dei valori compresi tra $(39, 40)$ e $(70, 71)$ rispettivamente, questo processo è detto **correzione di continuità**. Così facendo abbiamo che $P\left(39.5 \leq \overline{X_i} \leq 70.5.5\right) \simeq P\left(\frac{39.5-50}{5} \leq \frac{\overline{X_i}-50}{5} \leq \frac{70.5-50}{5}\right) = P(-2.1 \leq \frac{\overline{X_i}-50}{5} \leq 4.1) = \Phi(4.1) - \Phi(-2.1) = 0.9821$.

```
pnorm(4.1)-pnorm(-2.1)
```

2 Eserizi

$X_1, X_2, \dots X_{40} \sim Exp(\lambda)$ dove X_i è il tempo di vita della i-esima lampada

$E[X_i] = \frac{1}{\lambda} = 10 \qquad \lambda = \frac{1}{10}$

$\bar{X} = X_1 + \dots + X_{40}$

$$P(X_1 + \dots + X_{40} > 365) = 1 - P(X_1 + \dots + X_{40} \leq 365) = 1 - P\left(\frac{\bar{X} - E[\bar{X}]}{\sqrt{\frac{Var[\bar{X}]}{n}}} \leq \frac{365 - E[\bar{X}]}{\sqrt{\frac{Var[\bar{X}]}{n}}}\right)$$

dove $E[\bar{X}] = E[X_1 + \dots + X_{40}] = E[X_1] + \dots + E[X_{40}] = 40 \cdot 10 = 400$

$Var[\bar{X}] = Var[X_1 + \dots + X_{40}] = Var[X_1] + \dots + Var[X_{40}] = 40 \cdot 10^2 = 4000$

quindi:

$$1 - P\left(\frac{\bar{X} - E[\bar{X}]}{\sqrt{\frac{Var[\bar{X}]}{n}}} \leq \frac{365 - E[\bar{X}]}{\sqrt{\frac{Var[\bar{X}]}{n}}}\right)$$

$$= 1 - P\left(\frac{\bar{X} - 400}{\sqrt{4000}} \leq \frac{-35}{\sqrt{4000}}\right)$$

$$= 1 - P\left(\frac{\bar{X} - 400}{\sqrt{4000}} \leq -0{,}5534\right)$$

$$\simeq 1 - P(Z \leq -0{,}5534) \quad \text{con } Z \sim N(0,1)$$

$$= 1 - \Phi(-0{,}5534)$$

$$= 0{,}71$$

$$P(X_1 + \dots + X_{40} > 365) \simeq 0{,}71$$

$$P(X_1 + \dots + X_n > 365) \geq 0.95$$

$$1 - P(X_1 + \dots + X_n \leq 365) \geq 0.95$$

$$1 - P\left(\frac{\bar{X} - E[\bar{X}]}{\sqrt{Var[\bar{X}]}} \leq \frac{365 - E[\bar{X}]}{\sqrt{Var[\bar{X}]}}\right) \geq 0.95$$

dove $\bar{X} = X_1 + \dots + X_n$

$E[\bar{X}] = E[X_1] + \dots E[X_n] = 10n$

$Var[\bar{X}] = Var[X_1] + \dots + Var[X_n] = 100n$

$$1 - P\left(\frac{\bar{X} - 10n}{10\sqrt{n}} \leq \frac{365 - 10n}{10\sqrt{n}}\right) \geq 0.95$$

$$\simeq 1 - P\left(Z \leq \frac{365 - 10n}{10\sqrt{n}}\right) \geq 0.95$$

$$1 - \Phi\left(\frac{365 - 10n}{10\sqrt{n}}\right) \geq 0.95$$

$$\Phi\left(\frac{10n - 365}{10\sqrt{n}}\right) \geq 0.95$$

perché $n > 40$ quindi $365 - 10n$ è negativo

$$\frac{10n - 365}{10\sqrt{n}} \geq 1{,}645$$

$$10n - 365 \geq 16{,}45\sqrt{n}$$

$10n - 16,45\sqrt{n} - 365 \geq 0$

$10x^2 - 16,45x - 365 \geq 0$

$\sqrt{n} = x = \dfrac{16,45 \pm \sqrt{(-16,45)^2 - 40(-365)}}{20}$ — $6.9 \to n \geq 6.9^2 (= 47,883) \to n \geq 48$ lampadine ; $\cancel{-5.2}$

$X_i \sim Be\left(p = \frac{1}{2}\right) \qquad X_i = \begin{cases} 1 & \text{se } p \\ 0 & \text{se } 1-p \end{cases}$

$P(X_1 + \ldots + X_{50} \geq 29)$?

$\bar{X} := X_1 + \ldots + X_{50}$

$E[\bar{x}] = E[x_1] + \ldots + E[x_{50}] = 50\, E[x_i] = 50 \cdot \frac{1}{2} = 25$

$Var[\bar{x}] = Var[x_1] + \ldots + Var[x_{50}] = 50\, Var[x_i] = 50 \cdot \frac{1}{4} = 12,5$

$$P(\bar{x} \geq 28,5) = P\left(\frac{\bar{x} - E[\bar{x}]}{\sqrt{Var[\bar{x}]}} \geq \frac{28,5 - E[\bar{x}]}{\sqrt{Var[\bar{x}]}}\right)$$

$$= 1 - P\left(\frac{\bar{x} - E[\bar{x}]}{\sqrt{Var[\bar{x}]}} < \frac{28,5 - E[\bar{x}]}{\sqrt{Var[\bar{x}]}}\right)$$

$$= 1 - P\left(\frac{\bar{x} - 25}{\sqrt{12,5}} < 0.99\right)$$

$$\simeq 1 - P(z < 0.99) \qquad z \sim N(0,1)$$

TLC $= 1 - \Phi(0,99)$

$= 0,16$

$n \geq 30$ ✓

$np \geq 5 \to 25 \geq 5$ ✓

$n(1-p) \geq 5 \to 25 \geq 5$ ✓

$E[X_1 + \ldots + X_n] = n\mu$

$Var[X_1 + \ldots + X_n] = n \cdot 1 = n \qquad X_1 + \ldots + X_n \sim N(n\mu, n)$

$\cancel{\sqrt{n}}\, \dfrac{X_1 + \ldots + X_n - n\mu}{\cancel{\sqrt{n}}} = X_1 + \ldots + X_n - n\mu \qquad \sqrt{n}\, x$

$X_i \sim Be\left(p = \frac{1}{2}\right) \qquad X_i = \begin{cases} 1 & \text{se } F \\ 0 & \text{se } M \end{cases}$

$P\left(X_1+\ldots+X_{100} \geq 55\right)$?

$E[X_1+\ldots+X_{100}] = E[X_1]+\ldots+E[X_{100}] = 100\cdot E[X_i] = 50$

$Var[X_1+\ldots+X_{100}] = Var[X_1]+\ldots+Var[X_{100}] = 100\cdot Var[X_i] = 25$

$[55,+\infty)$

$$P\left(X_1+\ldots+X_{100} \geq 54{,}5\right) = 1-P\left(X_1+\ldots+X_{100} < 54{,}5\right)$$
$$= 1-P\left(\frac{X_1+\ldots+X_{100}-50}{\sqrt{25}} < \frac{54{,}5-50}{\sqrt{25}}\right)$$
$$= 1-P\left(\frac{X_1+\ldots+X_{100}-50}{5} < 0{,}88\right)$$
$$\overset{TLC}{\simeq} 1-P(Z < 0{,}88) \quad \text{con } Z\sim N(0,1)$$
$$= 1-\Phi(0{,}88)$$
$$= 0.1894$$

$X_1,\ldots,X_{100}$ v.a. i.i.d. con $E[X_i]=\frac{1}{2}$, $Var[X_i]>0$

$P\left(X_1+\ldots+X_{100} > 50\right)$?

$E[X_1+\ldots+X_{100}] = 100\cdot E[X_i] = 50$

$Var[X_1+\ldots+X_{100}] = 100\, Var[X_i] = 100\cdot\sigma$

$$P\left(X_1+\ldots+X_{100} > 50\right) = 1-P\left(X_1+\ldots+X_{100} \leq 50\right)$$
$$= 1-P\left(\frac{X_1+\ldots+X_{100}-50}{10\cdot\sigma} \leq \frac{50-50}{10\cdot\sigma}\right)$$
$$= 1-P\left(\frac{X_1+\ldots+X_{100}-50}{10\cdot\sigma} \leq 0\right)$$
$$\overset{TLC}{\simeq} 1-P(Z\leq 0) \quad \text{con } Z\sim N(0,1)$$
$$= 1-\Phi(0)$$
$$= 0.5$$

$X_i \sim Pois(\lambda = 13)$ indica il numero di vittime giornaliere

$P\left(X_1+\ldots+X_{30} < 360\right)$?

$E[X_1+\ldots+X_{30}] = 30\, E[X_i] = 390$

$$Var[X_1+\dots+X_{30}] = 30 \cdot Var[X_i] = 390$$

$$P\Big(X_1+\dots+X_{30} < \overset{[0,\,360)}{\downarrow}\!\!\!\!\! 359,5\Big) = P\left(\frac{X_1+\dots+X_{30}-390}{\sqrt{390}} < \frac{359,5-390}{\sqrt{390}}\right)$$

$$= P\left(\frac{X_1+\dots+X_{30}-390}{\sqrt{390}} < -1,54\right)$$

$$\overset{TLC}{\simeq} P(Z < -1,54) \quad \text{con } Z \sim N(0,1)$$

$$= \Phi(-1,54)$$

$$= 0.0612$$

$X_1,\dots,X_{100} \sim Be(0,2)$ v.a. i.i.d. che rappresenta un componente funzionante del sistema

$$X_i = \begin{cases} 1 & \text{se funziona} \\ 0 & \text{se non funziona} \end{cases}$$

$$X_1+\dots+X_{100} = \begin{cases} 1 & \text{se } X_1+\dots+X_n \geq 30 \\ 0 & \text{altrimenti} \end{cases}$$

$$E[X_1+\dots+X_{100}] = 100\, E[X_i] = 20$$

$$Var[X_1+\dots+X_{100}] = 100\, Var[X_i] = 16$$

$$P(X_1+\dots+X_{100} \geq 29,5) = 1 - P(X_1+\dots+X_{100} < 29,5)$$

$$= 1 - P\left(\frac{X_1+\dots+X_{100}-20}{\sqrt{16}} < \frac{29,5-20}{\sqrt{16}}\right)$$

$$= 1 - P\left(\frac{X_1+\dots+X_{100}-20}{4} < 2,375\right)$$

$$\overset{TLC}{\simeq} 1 - P(Z < 2,375) \quad \text{con } Z \sim N(0,1)$$

$$= 1 - \Phi(2,375)$$

$$= 0,0087$$

$$P(100 - X_1+\dots+X_{100} \in [72,88])$$

$$= P(71,5 \leq 100 - X_1+\dots+X_{100} \leq 88,5)$$

$$= P(-28,5 \leq -(X_1+\dots+X_{100}) \leq -71,5)$$

$$= P(11,5 \leq X_1+\dots+X_{100} \leq 28,5)$$

$$= P\left(\frac{11,5-20}{4} \leq \frac{X_1+\ldots+X_{100}-20}{4} \leq \frac{28,5-20}{4}\right)$$

$$\overset{TLC}{\simeq} P\left(\frac{11,5-20}{4} \leq Z \leq \frac{28,5-20}{4}\right) \text{ con } Z \sim N(0,1)$$

$$= P(-2,125 \leq Z \leq 2,125)$$

$$= \Phi(2,125) - (1-\Phi(2,125))$$

$$= 2\,\Phi(2,125) - 1$$

$$= 0,9664$$