Esercitazione 10

Docente: Mauro Piccioni. Tutor: Hlafo Alfie Mimun

Università di Roma "Sapienza"

Esercizio 2

Lanciamo n volte una moneta truccata con

$$\mathbb{P}(\text{Testa}) = \frac{3}{4}\,.$$

$S_n :=$ numero di teste ottenute in n lanci.

Domanda: utilizzando la disuguaglianza di Markov, si trovi il minimo valore di n affinché

$$\mathbb{P}(S_n \geq 3n^2) \leq \frac{1}{16}\,.$$

Soluzione dell'esercizio 2

Definiamo la successione di variabili aleatorie

$$X_1, X_2, \ldots, X_n\,,$$

dove per $i = 1, 2, \ldots, n$

$$X_i = \begin{cases} 1, & \text{se l'}i\text{-esimo lancio dà testa;} \\ 0, & \text{se l'}i\text{-esimo lancio dà croce.} \end{cases}$$

Allora per $i = 1, 2, \ldots, n$

$$\mathbb{P}(X_i = 1) = \frac{3}{4}\,, \qquad \mathbb{P}(X_i = 0) = \frac{1}{4}\,.$$

Si noti che

$$S_n := \text{numero di teste in } n \text{ lanci} = X_1 + \ldots + X_n\,.$$

Disuguaglianza di Markov

Sia $X \geq 0$ una variabile aleatoria con $\mathbb{E}[X] < \infty$ e sia $k > 0$. Allora

$$\mathbb{P}(X \geq k) \leq \frac{\mathbb{E}[X]}{k}.$$

$$\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \frac{3}{4} = \frac{3n}{4}.$$

Dunque, applicando la disuguaglianza di Markov con $X = S_n$ e $k = 3n^2$ si ha

$$\mathbb{P}(S_n \geq 3n^2) \leq \frac{\frac{3n}{4}}{3n^2} = \frac{1}{4n}.$$

Notiamo che $\frac{1}{4n} \leq \frac{1}{16}$ se $n \geq 4$. Dunque se $n = 4$ si ha

$$\mathbb{P}(S_n \geq 3n^2) \leq \frac{1}{4n} \leq \frac{1}{16}.$$

Esercizio 2

Lanciamo n volte una moneta truccata con

$$\mathbb{P}(\text{Testa}) = \frac{3}{4}\,.$$

$S_n :=$ numero di teste ottenute in n lanci.

Domanda:

(2a) Utilizzando la disuguaglianza di Chebyshev, si dia una stima della probabilità

$$\mathbb{P}\left(\left|4S_n - 3n\right| \geq t\sqrt{3n}\right)\,.$$

(2b) Utilizzando il teorema del limite centrale, si dia una stima della probabilità

$$\mathbb{P}\left(\left|4S_n - 3n\right| \geq t\sqrt{3n}\right)\,,$$

per ogni $t > 0$ e per n grande.

(2c) Si confrontino poi le stime ottenute nei punti (2a) e (2b) quando $t = 2$.

Soluzione dell'esercizio 2 (parte 2a)

Definiamo la successione di variabili aleatorie

$$X_1, X_2, \ldots, X_n \,,$$

dove per $i = 1, 2, \ldots, n$

$$X_i = \begin{cases} 1, & \text{se l'}i\text{-esimo lancio dà testa;} \\ 0, & \text{se l'}i\text{-esimo lancio dà croce.} \end{cases}$$

Allora per $i = 1, 2, \ldots, n$

$$\mathbb{P}(X_i = 1) = \frac{3}{4} \,, \qquad \mathbb{P}(X_i = 0) = \frac{1}{4} \,.$$

Si noti che

$$S_n := \text{numero di teste in } n \text{ lanci} = X_1 + \ldots + X_n \,.$$

Proprietà del valore atteso

$$\mathbb{E}[a \cdot X] = a \cdot \mathbb{E}[X]\,, \qquad \forall\, a \in \mathbb{R}\,.$$

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]\,.$$

$$\mathbb{E}\,[4S_n] = 4 \cdot \mathbb{E}[S_n] = 4 \cdot \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = 4 \cdot \sum_{i=1}^{n} \mathbb{E}[X_i]\,.$$

Abbiamo bisogno di $\mathbb{E}[X_i]$ per $i = 1, 2, \dots, n$.

Calcoliamo $\mathbb{E}[X_i]$

$$\mathbb{E}[X_i] = 1 \cdot \mathbb{P}(X_i = 1) + 0 \cdot \mathbb{P}(X_i = 0) = 1 \cdot \frac{3}{4} + 0 \cdot \frac{1}{4} = \frac{3}{4}\,.$$

Allora

$$\mathbb{E}\left[4S_n\right] = 4 \cdot \sum_{i=1}^{n} \mathbb{E}[X_i] = 4 \cdot \sum_{i=1}^{n} \frac{3}{4} = 4 \cdot n \cdot \frac{3}{4} = 3n\,.$$

Proprietà della varianza

$$\mathrm{Var}(a \cdot X) = a^2 \mathrm{Var}(X)\,, \qquad \forall\, a \in \mathbb{R}\,.$$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)\,, \quad \text{se } X, Y \text{ sono indipendenti}\,.$$

$$\mathrm{Var}\,(4S_n) = 4^2 \cdot \mathrm{Var}(S_n) = 16 \cdot \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = 16 \cdot \sum_{i=1}^{n} \mathrm{Var}(X_i)\,.$$

Abbiamo bisogno di $\mathrm{Var}(X_i)$ per $i = 1, 2, \ldots, n$.

Calcoliamo $\mathrm{Var}(X_i) = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2$. $\mathbb{E}[X_i^2] =?$

Descriviamo X_i^2 per calcolare $\mathbb{E}[X_i^2]$

$$X_i{}^2 = \begin{cases} 1^2, & \text{se l'}i\text{-esimo lancio dà testa;} \\ 0^2, & \text{se l'}i\text{-esimo lancio dà croce.} \end{cases}$$

Allora

$$\mathbb{E}[X_i^2] = 1^2 \cdot \mathbb{P}(X_i = 1) + 0^2 \cdot \mathbb{P}(X_i = 0) = 1^2 \cdot \frac{3}{4} + 0^2 \cdot \frac{1}{4} = \frac{3}{4}\,.$$

Dunque (essendo $\mathbb{E}[X_i] = \frac{3}{4}$)

$$\mathrm{Var}(X_i) = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 = \frac{3}{4} - \left(\frac{3}{4}\right)^2 = \frac{3}{16}\,.$$

Allora

$$\mathrm{Var}\,(4S_N) = 16 \cdot \sum_{i=1}^{n} \mathrm{Var}(X_i) = 16 \cdot \sum_{i=1}^{n} \frac{3}{16} = 16 \cdot n \cdot \frac{3}{16} = 3n\,.$$

Disuguaglianza di Chebyshev

Sia X una variabile aleatoria con media μ e varianza σ^2. Sia $\lambda > 0$. Allora

$$\mathbb{P}(|X - \mu| > \lambda\sigma) \leq \frac{1}{\lambda^2}\,.$$

Dunque applicando Chebyshev con $X = 4S_n$, $\mu = 3n$, $\sigma = \sqrt{3n}$, si ha per ogni $t > 0$

$$\mathbb{P}\left(\left|4S_n - 3n\right| \geq t\sqrt{3n}\right) \leq \frac{1}{t^2}$$

Soluzione dell'esercizio 2 (parte 2b)

Utilizziamo ora il Teorema del limite centrale (**TLC**) per calcolare la stessa probabilità

Teorema del limite centrale (**TLC**)

Se $X_1, \ldots, X_n$ sono variabili aleatorie i.i.d. con $\mathbb{E}[X_i] = \mu$ e $\mathrm{Var}(X_i) = \sigma^2$ per $i = 1, \ldots, n$. Allora

$$\mathbb{P}\left(\frac{\sum_{i=1}^n X_i - n \cdot \mu}{\sigma\sqrt{n}} \leq t\right) \underset{n\to+\infty}{=\!\!\longrightarrow} \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}\, dt\,.$$

In questo caso $S_n = \sum_{i=1}^n X_i$, $\mu = 3/4$ e $\sigma = \frac{\sqrt{3}}{4}$. Dunque

$$\mathbb{P}\left(\frac{4S_n - 3n}{\sqrt{3n}} \leq t\right) = \mathbb{P}\left(\frac{S_n - n\cdot\frac{3}{4}}{\frac{\sqrt{3n}}{4}} \leq t\right) \underset{n\to+\infty}{\longrightarrow} \Phi(t)\,.$$

I valori di Φ si trovano consultando le tavole della distribuzione normale standard.

Poiché

$$\mathbb{P}\left(\left|4S_n - 3n\right| \geq t\sqrt{3n}\right) = \mathbb{P}\left(4S_n - 3n \leq -t\sqrt{3n}\right) + \mathbb{P}\left(4S_n - 3n \geq t\sqrt{3n}\right)$$
$$= \mathbb{P}\left(4S_n - 3n \leq -t\sqrt{3n}\right) + 1 - \mathbb{P}\left(4S_n - 3n \leq t\sqrt{3n}\right)$$

dal TLC si ha

$$\mathbb{P}\left(\left|4S_n - 3n\right| \geq t\sqrt{3n}\right) \underset{n\to+\infty}{\longrightarrow} \Phi(-t) + 1 - \Phi(t)\,.$$

Per proprietà di simmetria della "curva a campana" si ha $\Phi(t) + \Phi(-t) = 1$ per ogni $t \in \mathbb{R}$.

Dunque

$$\mathbb{P}\left(\left|4S_n - 3n\right| \geq t\sqrt{3n}\right) \underset{n\to+\infty}{\longrightarrow} 2(1 - \Phi(t))\,.$$

Soluzione dell'esercizio 2 (parte (2c))

Quando $t = 2$ abbiamo

- usando Chebyshev

$$\mathbb{P}\left(\left|4S_n - 3n\right| \geq 2\sqrt{3n}\right) \leq \frac{1}{4}$$

- usando il TLC

$$\mathbb{P}\left(\left|4S_n - 3n\right| \geq 2\sqrt{3n}\right) \underset{n\to+\infty}{\longrightarrow} 2(1-\Phi(2)) \approx 2(1-0.97) = 0.06\,.$$

Exercise 3

$Y :=$ punteggio di uno studente alla sessione di laurea.

$$Y \sim \mathcal{N}(100, 25)\,.$$

Supponiamo che n studenti si laureino nella stessa sessione.

$S_n :=$ somma dei punteggi degli n studenti.

Domanda: Usare il teorema del limite centrale per calcolare il valore di n per cui si ha

$$\mathbb{P}\left(95 < \frac{S_n}{n} \leq 105\right) \gtrsim 0.9\,.$$

Soluzione dell'esercizio 3

Teorema del limite centrale (**TLC**)

Se $X_1, \dots, X_n$ sono variabili aleatorie i.i.d. con $\mathbb{E}[X_i] = \mu$ e $\mathrm{Var}(X_i) = \sigma^2$ per $i = 1, \dots, n$. Allora

$$\mathbb{P}\left(\frac{\sum_{i=1}^n X_i - n \cdot \mu}{\sigma\sqrt{n}} \le t\right) \xrightarrow[n\to+\infty]{} \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}\, dt\,.$$

$S_n :=$ somma dei voti di laurea degli n studenti.

$X_1, \dots, X_n$ sono i voti di laurea degli n studenti $\Rightarrow S_n = \sum_{i=1}^n X_i$.

Dunque $X_1, \dots, X_n$ sono variabili aleatorie i.i.d. con

$$\mathbb{E}[X_i] = 100\,, \qquad \mathrm{Var}(X_i) = \sigma^2 = 25\,.$$

$$\text{TLC} \;\Rightarrow \mathbb{P}\left(\frac{S_n - n \cdot 100}{5\sqrt{n}} \le t\right) \xrightarrow[n\to+\infty]{} \Phi(t)$$

$$\mathbb{P}\left(95 < \frac{S_n}{n} \leq 105\right) = \mathbb{P}\left(\frac{S_n}{n} \leq 105\right) - \mathbb{P}\left(\frac{S_n}{n} \leq 95\right)$$

Calcoliamo $\mathbb{P}\left(\frac{S_n}{n} \leq 105\right)$ e $\mathbb{P}\left(\frac{S_n}{n} \leq 95\right)$. Sappiamo che

$$\mathbb{P}\left(\frac{S_n - n \cdot 100}{5\sqrt{n}} \leq t\right) \xrightarrow[n\to+\infty]{} \Phi(t)\,.$$

Dunque per $a \in \mathbb{R}$

$$\begin{aligned}
\mathbb{P}\left(\frac{S_n}{n} \leq a\right) &= \mathbb{P}(S_n \leq a \cdot n) = \mathbb{P}(S_n - 100 \cdot n \leq a \cdot n - 100 \cdot n) = \\
&= \mathbb{P}\left(\frac{S_n - 100 \cdot n}{5\sqrt{n}} \leq \frac{(a-100) \cdot n}{5\sqrt{n}}\right) = \\
&= \mathbb{P}\left(\frac{S_n - 100 \cdot n}{5\sqrt{n}} \leq \frac{(a-100) \cdot \sqrt{n}}{5}\right) \underset{\text{TLC}}{\approx} \Phi\left(\frac{(a-100) \cdot \sqrt{n}}{5}\right)
\end{aligned}$$

Dunque per $a \in \mathbb{R}$ abbiamo ottenuto

$$\mathbb{P}\left(\frac{S_n}{n} \le a\right) \underset{\text{TLC}}{\approx} \Phi\left(\frac{(a-100)\cdot\sqrt{n}}{5}\right).$$

Dobbiamo calcolare

$$\begin{aligned}\mathbb{P}\left(95 < \frac{S_n}{n} \le 105\right) &= \mathbb{P}\left(\frac{S_n}{n} \le 105\right) - \mathbb{P}\left(\frac{S_n}{n} \le 95\right) \approx \\ &\underset{\text{TLC}}{\approx} \Phi\left(\frac{(105-100)\cdot\sqrt{n}}{5}\right) - \Phi\left(\frac{(95-100)\cdot\sqrt{n}}{5}\right) = \\ &= \Phi(\sqrt{n}) - \Phi(-\sqrt{n})\end{aligned}$$

Poiché $\Phi(-t) = 1 - \Phi(t)$ per ogni $t \in \mathbb{R}$, abbiamo che

$$\begin{aligned}\mathbb{P}\left(95 < \frac{S_n}{n} \le 105\right) &\underset{\text{TLC}}{\approx} \Phi(\sqrt{n}) - \Phi(-\sqrt{n}) = \\ &= \Phi(\sqrt{n}) - (1 - \Phi(\sqrt{n})) = 2\Phi(\sqrt{n}) - 1.\end{aligned}$$

Dunque abbiamo ottenuto

$$\mathbb{P}\left(95 < \frac{S_n}{n} \leq 105\right) \underset{\text{TLC}}{\approx} 2\Phi(\sqrt{n}) - 1 \,.$$

Vogliamo n tale che

$$\mathbb{P}\left(95 < \frac{S_n}{n} \leq 105\right) \gtrsim 0.9 \,.$$

Dunque

$$2\Phi(\sqrt{n}) - 1 \gtrsim 0.9 \Rightarrow \Phi(\sqrt{n}) \gtrsim 0.95 \,.$$

Usando le tavole della distribuzione normale standard si ha

$$\Phi(\sqrt{n}) \gtrsim 0.95 \Rightarrow \sqrt{n} \gtrsim 1.65 \Rightarrow n \gtrsim 2.7 \,.$$

Poiché n è intero, prendiamo $n = 3$ per avere $\mathbb{P}\left(95 < \frac{S_n}{n} \leq 105\right) \gtrsim 0.9$.

Esercizio 4

Un dato componente è fondamentale per il funzionamento di una macchina

Quando la componente si rompe, viene immediatamente sostituita con una componente nuova.

Il tempo di vita (in ore) di questa componente è una variabile aleatoria Y con $\mathbb{E}[Y] = 100$ e $\text{Var}(Y) = 30$.

Domanda: Usare il teorema del limite centrale per calcolare quante volte dobbiamo sostituire tale componente per avere

$$\mathbb{P}(\text{la macchina funziona per più di 2000 ore}) \gtrsim 0.95\,.$$

Soluzione dell'esercizio 4

Denotiamo con X_1 il tempo di vita della componente che si trova all'interno della macchina all'inizio.

Quando tale componente si rompe, la sostituiamo con una seconda componente. Denotiamo con X_2 il suo tempo di vita.

In questo modo, se sostituiamo la componente $n-1$ volte, definiamo i tempi di vita $X_1, X_2, \dots, X_n$ delle n componenti sostituite.

$S_n := \sum_{i=1}^{n} X_i$ è il tempo di vita totale della macchina dopo n sostituzioni della componente.

$X_1, \ldots, X_n$ sono variabili aleatorie i.i.d. con

$$\mathbb{E}[X_i] = 100, \qquad \mathrm{Var}(X_i) = 30\,.$$

Si noti la seguente equivalenza tra i due eventi

$$\{\text{la macchina funziona per più di 2000 ore}\} = \{S_n > 2000\}\,.$$

Dunque dobbiamo calcolare n tale che $\mathbb{P}(S_n > 2000) \geq 0.95$.

Teorema del limite centrale (**TLC**)

Se $X_1, \ldots, X_n$ sono variabili aleatorie i.i.d. con $\mathbb{E}[X_i] = \mu$ e $\mathrm{Var}(X_i) = \sigma^2$ per $i = 1, \ldots, n$. Allora

$$\mathbb{P}\left(\frac{\sum_{i=1}^n X_i - n \cdot \mu}{\sigma\sqrt{n}} \leq t\right) \underset{n\to+\infty}{\longrightarrow} \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}\, dt\,.$$

Voglimo usare il TLC per calcolare $\mathbb{P}(S_n > 2000)$.

$$\begin{aligned}\mathbb{P}(S_n > 2000) &= \mathbb{P}(S_n - 100n > 2000 - 100n) = \\ &= \mathbb{P}\left(\frac{S_n - 100n}{\sqrt{30n}} > \frac{2000 - 100n}{\sqrt{30n}}\right) = \\ &= 1 - \mathbb{P}\left(\frac{S_n - 100n}{\sqrt{30n}} \leq \frac{2000 - 100n}{\sqrt{30n}}\right) \approx \\ &\underset{\text{TLC}}{\approx} 1 - \Phi\left(\frac{2000 - 100n}{\sqrt{30n}}\right).\end{aligned}$$

Dunque abbiamo ottenuto

$$\mathbb{P}(S_n > 2000) \underset{\text{TLC}}{\approx} 1 - \Phi\left(\frac{2000 - 100n}{\sqrt{30n}}\right).$$

Vogliamo

$$\mathbb{P}(S_n > 2000) \geq 0.95 .$$

Dunque imponiamo (si ricordi che $1 - \Phi(t) = \Phi(-t)$)

$$1 - \Phi\left(\frac{2000 - 100n}{\sqrt{30n}}\right) \geq 0.95 \Rightarrow \Phi\left(-\frac{2000 - 100n}{\sqrt{30n}}\right) \geq 0.95 .$$

Usando le tavole della distribuzione normale standard, si ha $\Phi(1.66) \approx 0.95$. Dunque, essendo Φ una funzione crescente, si ha

$$\Phi\left(-\frac{2000 - 100n}{\sqrt{30n}}\right) \geq 0.95 \Rightarrow \Phi\left(-\frac{2000 - 100n}{\sqrt{30n}}\right) \geq \Phi(1.66) \Rightarrow$$
$$\Rightarrow -\frac{2000 - 100n}{\sqrt{30n}} \geq 1.66 .$$

Dobbiamo risolvere

$$-\frac{2000-100n}{\sqrt{30n}} \geq 1.66 \Rightarrow \frac{2000-100n}{\sqrt{30n}} \leq -1.66\,.$$

Abbiamo che

$$\frac{2000-100n}{\sqrt{30n}} \leq -1.66 \Rightarrow 2000-100n \leq -9.1\sqrt{n}$$
$$\Rightarrow 100n - 9.1\sqrt{n} - 2000 \geq 0\,.$$

Definendo $y = \sqrt{n}$, si ottiene

$$100y^2 - 9.1y - 2000 \geq 0 \Rightarrow y \geq 4.52 \Rightarrow \sqrt{n} \geq 4.52 \Rightarrow n \geq 20.41\,.$$

Essendo n un numero intero, prendiamo $n = 21$ per avere

$$\mathbb{P}(S_n > 2000) \geq 0.95\,.$$

Esercizio 5

Una compagnia assicurativa ha 10000 clienti che possiedono una polizza assicurativa sulla macchina.

Ogni cliente ha un compenso annuo (in euro) rappresentato da una variabile aleatoria Y con

$$\mathbb{E}[Y] = 240\,, \qquad \mathrm{Var}(Y) = 80^2\,.$$

Domanda: Usare il teorema del limite centrale per calcolare la probabilità che il totale dei compensi non superi i $2.41 \cdot 10^6$ euro.

Soluzione dell'esercizio 5

Ci sono 10000 clienti che possiedono una polizza assicurativa.

Denotiamo con X_i il compenso annuo dell'i-esimo cliente.

Dunque abbiamo che $X_1, \ldots, X_{10000}$ sono variabili aleatorie i.i.d. con

$$\mathbb{E}[X_i] = 240\,, \qquad \mathrm{Var}(X_i) = 80^2 \quad \text{per } i = 1, \ldots, 10000\,.$$

Definiamo $S_{10000} := \sum_{i=1}^{10000} X_i$. I seguenti eventi sono equivalenti

$$\{\text{compenso totale} \leq 2.41 \cdot 10^6 \text{ euro}\} = \{S_{10000} \leq 2.41 \cdot 10^6\}\,.$$

Dunque dobbiamo calcolare

$$\mathbb{P}(S_{10000} \leq 2.41 \cdot 10^6) = ?$$

Teorema del limite centrale (**TLC**)

Se $X_1, \ldots, X_n$ sono variabili aleatorie i.i.d. con $\mathbb{E}[X_i] = \mu$ e $\mathrm{Var}(X_i) = \sigma^2$ per $i = 1, \ldots, n$. Allora

$$\mathbb{P}\left(\frac{\sum_{i=1}^n X_i - n \cdot \mu}{\sigma\sqrt{n}} \leq t\right) \underset{n\to+\infty}{\longrightarrow} \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}\, dt\,.$$

Vogliamo usare il TLC per calcolare $\mathbb{P}(S_{10000} \leq 2.41 \cdot 10^6)$.

$$\mathbb{P}(S_{10000} \leq 2.41 \cdot 10^6) = \mathbb{P}(S_{10000} - 10000 \cdot 240 \leq 2.41 \cdot 10^6 - 10000 \cdot 240) =$$

$$= \mathbb{P}\left(\frac{S_{10000} - 10000 \cdot 240}{800 \cdot \sqrt{10000}} \leq \frac{2.41 \cdot 10^6 - 10000 \cdot 240}{800 \cdot \sqrt{10000}}\right) =$$

$$= \mathbb{P}\left(\frac{S_{10000} - 10000 \cdot 240}{800 \cdot \sqrt{10000}} \leq 1.25\right) \underset{TLC}{\approx} \Phi(1.25)\,.$$

Usando le tavole per la distribuzione normale standard abbiamo che $\Phi(1.25) \approx 0.89$.

Dunque

$$\mathbb{P}(S_{10000} \leq 2.41 \cdot 10^6) \underset{TLC}{\approx} \Phi(1.25) \approx 0.89\,.$$