⑤ TEOREMA LIMITE CENTRALE

Illustra l'uso "universale" delle v.a. gaussiane

Motivazione. Una moneta garantita equa.
La provo. $N = 10^4$ lanci escono 5500 Teste.
Devo sporgere reclamo?

Assumo sia effettivamente equa $X \sim Bin(N, \frac{1}{2})$. $X = \#$ Teste
Quant'è probabile aver osservato una fluttuazione di 500 rispetto al valor medio $\mathbb{E}(X) = \frac{1}{2} N = 5000$?

$$\mathbb{P}(|X - 5000| \geq 500) = ?$$

Vediamo cosa concludiamo usando Chebyshev

$$\mathbb{P}(|X - \underbrace{5000}_{= \mathbb{E}(X)} | \geq 5 \cdot 10^2) \leq \frac{1}{25 \cdot 10^4} \mathbb{V}(X)^{= N \frac{1}{2} \cdot \frac{1}{2}}$$

$$= \frac{1}{25 \cdot 10^4} \cdot 10^4 \cdot \frac{1}{2} = \frac{1}{100}$$

difficile che capiti, ma rientra nelle specifiche garantite dal venditore.

Quanto abbiamo perso usando Chebyshev?

Possiamo fare un'asintotica precisa per $N \to \infty$?

• Ricordo legge dei grandi numeri

$X_1, \dots, X_n$ iid $\qquad S_n = \sum_{i=1}^{n} X_i$

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}(X_1) \qquad \forall \delta > 0 \quad \lim_{n\to\infty} \mathbb{P}\left(\left| \frac{S_n}{n} - \mathbb{E}(X_1) \right| < \delta \right) = 1$$

Vogliamo capire la correzione per n grande

<u>TEO</u> (Teorema limite centrale)

$X_1, \dots, X_n$ iid $\quad [\mathbb{E}(X_1) \ \exists \quad \mathbb{V}(X_1) < +\infty]$

$S_n = \sum_{i=1}^{n} X_i$

Allora

$$\frac{S_n - \overbrace{\mathbb{E}(S_n)}^{= n\mathbb{E}(X_1)}}{\sqrt{\underbrace{\mathbb{V}(S_n)}_{= n\mathbb{V}(X_1)}}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mathbb{E}(X_i)}{\sqrt{\mathbb{V}(X_i)}} \xrightarrow[n\to\infty]{} \overbrace{\mathcal{N}(0,1)}^{\text{Gaussiana standard}}$$

nel senso che se $a < b$

$$\mathbb{P}\left(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathbb{V}(S_n)}} \in (a,b) \right) \longrightarrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

dimostrazione omessa nel caso generale, nel seguito il caso $X_i \sim Bern(p)$ $\quad S_n \sim Bin(n,p)$

- Ritorno dal venditore di monete eque (presunte tali)

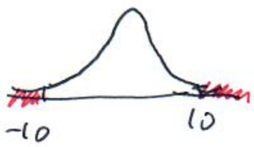
$$X \sim \mathrm{Bin}(N, \tfrac{1}{2}) \qquad N = 10^4$$

$$\mathbb{P}(\,|X - \mathbb{E}(X)| > 500\,)$$

$$= \mathbb{P}\left(\frac{|X - \mathbb{E}(X)|}{\sqrt{\underbrace{\mathbb{V}(X)}_{= N \frac{1}{4}}}} > \frac{500}{\frac{1}{2}\, 10^2} \right)$$

approx gaussiana

$$\approx \mathbb{P}(\,|Z| > 10\,) \qquad Z \sim \mathcal{N}(0,1)$$

= Area() = veramente risibile

[Nelle tavole con 4 cifre significative è zero]

- Altro contesto

$$X_i = \begin{cases} +1 & \text{prob } 1/2 \\ -1 & \text{" } 1/2 \end{cases} \qquad \sum_{i=1}^{n} X_i = ?$$

iid

$$\mathbb{E}(X_i) = 0 \qquad \mathbb{V}(X_i) = 1 \cdot \tfrac{1}{2} + 1 \cdot \tfrac{1}{2} = 1$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \longrightarrow \mathcal{N}(0,1)$$

$$\sum_{i=1}^{n} X_i \approx \sqrt{n}\, Z \qquad \text{con } Z \sim \mathcal{N}(0,1)$$

- In generale il teorema limite centrale motiva l'uso di v.a. gaussiane nella teoria degli errori di misura (somma di tanti piccoli effetti). Anche in altri contesti.

(5) Teoria cinetica dei gas. Distribuzione di ~~Maxwell~~ Maxwell - Boltzmann (= gaussiana)

Gas diluito all'equilibrio termico. Tante molecole.
Distribuzione (= statistica) delle velocità

$\underline{V} = (V_1, V_2, V_3)$ v.a. a valori $\mathbb{R}^3$ velocità

f = densità di $\underline{V}$ = ? $f : \mathbb{R}^3 \to \mathbb{R}$

Ipotesi (modellistiche)

- per isotropia dello spazio (tutte le direzioni sono uguali)

  $f = f(v_1, v_2, v_3)$

  dipende solo da $|\underline{v}|^2 = v_1^2 + v_2^2 + v_3^2$

  senz'altro giusta [trascurando la gravità]

- V_1, V_2, V_3 (componenti della velocità) sono iid

  Non è proprio giusto, ma andrà abbastanza bene se il gas è diluito

Ricaviamo la distribuzione di $\underline{V}$

$$f(v_1, v_2, v_3) = g(v_1^2 + v_2^2 + v_3^3)$$
$$\shortparallel$$
$$h(v_1^2)\, h(v_2^2)\, h(v_3^2)$$

Ora $v_2 = v_3 = 0 \Rightarrow h(0)\, h(0)\, h(v_1^2) = g(v_1^2)$

ovvero

$$h(v_1^2)\, h(v_2^2)\, h(v_3^2) = h(0)^2\, h(v_1^2+v_2^2+v_3^3)$$

equivalentemente, posto $\phi(v^2) = \dfrac{h(v^2)}{h(0)}$

$$\phi(v_1^2)\, \phi(v_2^2)\, \phi(v_3^2) = \phi(v_1^2+v_2^2+v_3^2)$$

per lo stesso argomento che perdita di memoria caratterizza v.a. esponenziale

$$\phi(v^2) = A\, e^{-\beta \frac{v^2}{2}}$$

quindi

$$f(v_1,v_2,v_3) \propto e^{-\frac{\beta}{2}(v_1^2+v_2^2+v_3^2)}$$ per un qualche $\beta > 0$

la costante la trovo imponendo la normalizzazione

$$\underset{\text{iid}}{V_1, V_2, V_3} \sim \mathcal{N}\left(0, \frac{1}{\beta}\right)$$

Quindi

$$E\left(\frac{1}{2}|\underline{v}|^2\right) = \frac{1}{2} E(v_1^2+v_2^2+v_3^2) = \frac{3}{2}\,\frac{1}{\beta}$$

~~$\beta = \frac{1}{KT}$ con K costante di Boltzmann, T temperatura~~

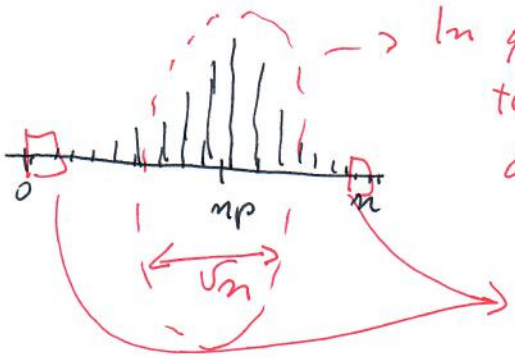
$\beta \propto \dfrac{1}{T}$ con T temperatura

⑥ Teorema de Moivre – Laplace

Conclusione analoga al teorema limite centrale nel caso $X_i \sim Ber(p)$ $\quad S_n = \sum_{i=1}^{n} X_i \sim Bin(n,p)$

ma in un senso più forte che nell'enunciato del teorema limite centrale.

$S_n \sim Bin(n,p)$ $\quad p$ fisso, n grande



In questa zona l'istogramma tende ad avere la forma della densità gaussiana

Qui l'approssimazione gaussiana non va bene

TEO (de Moivre – Laplace)

Sia

$$P_n(k) := \mathbb{P}(S_n = k) = \binom{n}{k} p^k (\overbrace{1-p}^{q})^{n-k}$$

Allora

$$P_n(k) \sim \frac{1}{\sqrt{2\pi \underbrace{npq}_{=\mathbb{V}(S_n)}}}\, e^{-\frac{1}{2}\frac{(k - \overbrace{np}^{\mathbb{E}(S_n)})^2}{\underbrace{npq}_{=\mathbb{V}(S_n)}}}$$

Nel senso che $\forall \delta > 0$

$$\lim_{n} \sup_{\substack{k: \\ |k-np| \le n^{\frac{2}{3}-\delta}}} \left| \frac{P_n(k)}{\frac{1}{\sqrt{2\pi npq}}\, e^{-\frac{(k-np)^2}{2npq}}} - 1 \right| = 0$$

Uniformemente sui k "non troppo lontani" da $\mathbb{E}(S_n)$

OSS Usando de Moivre - Laplace e usando che la somma di Riemann converge all'integrale, si ricava il teorema limite centrale per somme di Bernoulli iid

$$\mathbb{P}\left(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathbb{V}(S_n)}} \in (a,b) \right) = \sum_{k: \dots} \mathbb{P}(S_n = k)$$

$$\dots \longrightarrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \, dz$$

[dettagli a vostra cura]

dim: è un calcolo

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

si tratta di usare Stirling per trovare l'approssimazione dei ! e vedere quello che viene -

Ecco i dettagli

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \overset{\text{stirling}}{=} \frac{\sqrt{2\pi n}\, n^n e^{-n} (1 + R(n))}{\sqrt{2\pi k}\, k^k e^{-k} (1 + R(k)) \sqrt{2\pi(n-k)}\, (n-k)^{n-k} e^{-(n-k)} (1 + R(n-k))}$$

$$= \frac{1}{\sqrt{2\pi n \frac{k}{n}\left(1 - \frac{k}{n}\right)}} \cdot \frac{1}{\left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k}} \underbrace{\frac{1 + R(n)}{(1 + R(k))(1 + R(n-k))}}_{= 1 + \varepsilon(k, n, n-k)}$$

con $\varepsilon(k, n, n-k) \to 0$ se $k, n, n-k \to \infty$

introduco

$$\hat{p} := \frac{k}{n}$$

$$\binom{n}{k} = \frac{1}{\sqrt{2\pi n\, \hat p(1-\hat p)}} \quad \frac{1}{\hat p^{k}(1-\hat p)^{n-k}} \left(1+\varepsilon(n,k,n-k)\right)$$

quindi

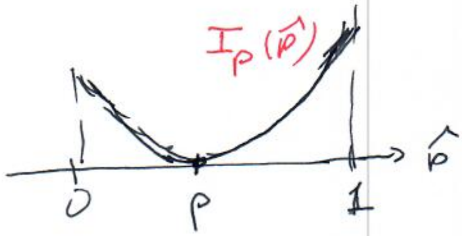
$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{\sqrt{2\pi n\, \hat p(1-\hat p)}} \left(\frac{p}{\hat p}\right)^k \left(\frac{1-p}{1-\hat p}\right)^{n-k} \left(1+\varepsilon(n,k,n-k)\right)$$

$$= \frac{1}{\sqrt{2\pi n\, \hat p(1-\hat p)}} \exp\left\{ k \lg \frac{p}{\hat p} + (n-k)\lg\frac{1-p}{1-\hat p} \right\} (1+\varepsilon)$$

$$= \frac{1}{\sqrt{2\pi n\, \hat p(1-\hat p)}}\, e^{-n \underbrace{\left\{ \hat p \lg \frac{\hat p}{p} + (1-\hat p)\lg \frac{1-\hat p}{1-p} \right\}}_{= I_p(\hat p)}} (1+\varepsilon)$$

se $|k - np| \leq n^{\frac{2}{3}-\delta}$ $\qquad |\hat p - p| \leq n^{-\frac{1}{3}-\delta} \underset{n\to\infty}{\longrightarrow} 0$

esercizio: $I_p(\cdot)$ è fatto così:



I_p convessa (strettamente)

caso

$I_p(\hat p) \geq 0$ e $=0$ sse $p = \hat p$

per lo sviluppo di Taylor

$$I_p(\hat p) = \underbrace{I_p(p)}_{=0} + \underbrace{I_p'(p)}_{=0}(\hat p - p) + \underbrace{\frac{1}{2} I_p''(p)}_{= \frac{1}{2}\frac{1}{p(1-p)}} (\hat p - p)^2 + O(|\hat p - p|^3)$$

ho quindi trovato

$$P_n(k) = \frac{1}{\sqrt{2\pi\, \hat{p}(1-\hat{p})\, n}}\, e^{-n \frac{1}{2} \frac{(p-\hat{p})^2}{p(1-p)} + n\,\mathcal{O}(|p-\hat{p}|^3)} (1+\varepsilon)$$

se $|k - np| \le n^{2/3 - \delta}$

$$|\hat{p} - p| = \left|\frac{k}{n} - p\right| \le n^{-1/3 - \delta} \Rightarrow n\,\mathcal{O}(|p-\hat{p}|^3) \xrightarrow[n\to\infty]{} 0$$

~~sviluppo~~

$$P_n(k) = \frac{1}{\sqrt{2\pi n\, p(1-p)}} \underbrace{\sqrt{\frac{p(1-p)}{\hat{p}(1-\hat{p})}}}_{\to 1} e^{-\frac{1}{2}\frac{(\overset{= k}{n\hat{p}} - np)^2}{n\,p(1-p)}} \underbrace{e^{+n\,\mathcal{O}(|p-\hat{p}|^3)}}_{\to 1} \underbrace{(1+\varepsilon)}_{\to 1}$$

(poiché $\hat{p} \to p$)

ovvero

$$\frac{P_n(k)}{\frac{1}{\sqrt{2\pi n\, p(1-p)}}\, e^{-\frac{1}{2}\frac{(k-np)^2}{n\,p(1-p)}}} \xrightarrow{n\to\infty} 1 \quad \text{unif per } k: \; |k - np| \le n^{\frac{2}{3} - \delta}$$