1020421 -Calcolo delle Probabilità, Laurea in INFORMATICA
a.a. 2025-26, canale 1 (A-L), Prof. Nappo
APPROSSIMAZIONE NORMALE E TEOREMA CENTRALE DEL LIMITE

1 Posizione del problema dell'approssimazione normale

Esempio 1.1. *Una fabbrica produce degli oggetti. Ogni oggetto può presentare dei difetti di fabbricazione con probabilità 0.05. Sappiamo che in un mese produce 10.000 oggetti.*
Si faccia l'ipotesi semplificativa che gli eventi

$$A_i = \{l\text{'}i\text{-simo oggetto prodotto è difettoso}\}, \qquad \text{per } i = 1, 2, \cdots, 10.000$$

sono (globalmente) indipendenti.
Vogliamo calcolare
a) la probabilità che il numero di oggetti difettosi sia 450.
b) la probabilità che il numero di oggetti difettosi sia al più 450.

La risposta esatta, come è noto, è data dall'osservare che, posto $S_n = \sum_{i=1}^n \mathbf{1}_{A_i}$ *si tratta di calcolare la probabilità* $\mathbb{P}(S_{10000} = 450)$ *e la probabilità* $\mathbb{P}(S_{10000} \leq 450)$. *La variabile aleatoria* S_{10000} *segue una legge binomiale di parametri* $n = 10000$ *e* $p = 0.05$.
Quindi le risposte esatte ai precedenti quesiti sono

a) $\mathbb{P}(S_{10000} = 450) = \binom{10000}{450}(0.05)^{450}(0.95)^{9550}$

b) $\mathbb{P}(S_{10000} \leq 450) = \sum_{k=0}^{450} \mathbb{P}(S_{10000} = k) = \sum_{k=0}^{450} \binom{10000}{k}(0.05)^k(0.95)^{(10000-k)}$

Tuttavia il calcolo esplicito di tali probabilità è decisamente complesso e nasce spontaneamente il problema di una sua approssimazione numerica.

Iniziamo con alcuni richiami. Come già sappiamo, la variabile aleatoria S_n numero di successi su n prove, in uno schema di prove ripetute (o schema di Bernoulli), segue una legge binomiale $B(n, p)$ dove $p \in (0, 1)$ è la probabilità di successo, per ciascuna prova. Ciò significa che

$$\text{per } k = 0,\, 1, \cdots, n \quad \mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}. \tag{1}$$

Calcolare questo valore quando n è grande, non è facile, neanche con l'ausilio di un computer. E a maggior ragione non è facile calcolare

$$\text{per } k = 0,\, 1, \cdots, n \quad \mathbb{P}(S_n \leq k) = \sum_{h=0}^{k} \binom{n}{h} p^h (1-p)^{n-h} = \sum_{h=0}^{k} \frac{n!}{h!(n-h)!} p^h (1-p)^{n-h}. \tag{2}$$

Un primo problema riguarda l'espressione del coefficiente binomiale. Questo si può in parte superare tenendo conto della seguente espressione ricorsiva

$$\mathbb{P}(S_n = 0) = (1-p)^n \tag{3}$$

$$\mathbb{P}(S_n = k+1) = \frac{n-k}{k+1} \frac{p}{1-p} \mathbb{P}(S_n = k) \tag{4}$$

Tuttavia rimane il problema che $(1-p)^n$ può essere molto piccolo, per n grande e così per $\mathbb{P}(S_n = k)$, che tende a zero per n che tende ad infinito.

L'espressione ricorsiva (4) per $\mathbb{P}(S_n = k)$ mostra anche il tipico andamento "*a campana*" della densità discreta di S_n $\big($cioè della funzione $k \mapsto \mathbb{P}\big(S_n = k\big)\big)$, ovvero prima crescente e poi decrescente, con il massimo che viene raggiunto per $k \in [np - (1-p), np + p]$, come si vede facilmente.

Per risolvere il problema posto nell'Esempio 1.1, si potrebbe pensare di utilizzare l'approssimazione di Poisson, che ricordiamo qui:
se consideriamo p dipendente da n ossia $p = p_n$ che tende a zero ed n che tende ad infinito, ma in modo che $np = \lambda$, si può usare l'approssimazione di Poisson. Tuttavia ciò **non funziona** in questo caso:

infatti tale approssimazione dal punto di vista numerico/empirico è soddisfacente/buona **solo se**
n è almeno 20 e p è minore di $0,05$,
e inoltre,
per $n \geq 100$, se vale $\lambda = np \leq 10$.

Nel nostro caso $p = 0,05$ è al limite dei valori ammissibili per p, ma n è decisamente troppo grande per poter utilizzare l'approssimazione di Poisson, infatti $\lambda = np = 10000\frac{5}{100} = 500$ e comunque sarebbe ugualmente poco agevole calcolare $\frac{\lambda^k}{k!}e^{-\lambda} = \frac{500^k}{k!}e^{-500}$.

Inoltre, ricordando un risultato[1] di Le Cam e che riportiamo qui sotto, possiamo solo dire che, per ogni insieme I,

$$\big|\mathbb{P}(S_n \in I) - \mathbb{P}(X \in I)\big| \leq np^2 = \lambda p = 500\,\frac{5}{100} = 25\,(!?!), \qquad \text{dove } X \sim Poisson(\lambda)$$

(la maggiorazione con 25 non ci dice nulla!! infatti in se $p_1, p_2 \in [0,1]$ allora $|p_1 - p_2| \leq 1$) e quindi sappiamo già che $\big|\mathbb{P}(S_n \in I) - \mathbb{P}(X \in I)\big| \leq 1$
Riprendiamo il problema di calcolare $\mathbb{P}(S_n = k)$. Siamo interessati almeno ad ottenere un valore approssimato. Prima di tutto notiamo che

$$\mathbb{P}(S_n = k) = \mathbb{P}\Big(k - \frac{1}{2} < S_n \leq k + \frac{1}{2}\Big)$$

e quindi, in entrambi i casi dell'Esempio 1.1 precedente, si tratta di calcolare delle probabilità del tipo

$$\mathbb{P}(a < S_n \leq b),$$

con $a = 450 - 0,5 = 449,5$ e $b = 450 + 0,5 = 450,5$ per il calcolo di $\mathbb{P}(S_{10000} = 450)$, e invece $a = -\infty$ e $b = 450$. Si osservi che in realtà sarebbe meglio prendere anche $b = 450 + 0,5 = 450,5$ e che si potrenne anche prendere $a = -0,5$.

Utilizzeremo ora il fatto che in generale dati due numeri reali $a < b$, e un numero reale w, le seguenti condizioni sono equivalenti

$$a < w \leq b \quad \Leftrightarrow \quad a - u < w - u \leq b - u,\ \forall\, u \in \mathbb{R}, \quad \Leftrightarrow \quad \frac{a-u}{v} < \frac{w-u}{v} \leq \frac{b-u}{v},\ \forall\, v > 0,$$

e quindi per ogni variabile aleatoria W, con $Var(W) \neq 0$, per ogni coppia di numeri reali $a < b$ si ha

$$\{a < W \leq b\} = \left\{\tfrac{a-\mathbb{E}(W)}{\sqrt{Var(W)}} < \tfrac{W-\mathbb{E}(W)}{\sqrt{Var(W)}} \leq \tfrac{b-\mathbb{E}(W)}{\sqrt{Var(W)}}\right\} = \left\{\tfrac{a-\mathbb{E}(W)}{\sqrt{Var(W)}} < W^* \leq \tfrac{b-\mathbb{E}(W)}{\sqrt{Var(W)}}\right\}$$

dove W^* è la standardizzata di W, e quindi W^* ha valore atteso nullo e varianza 1.

Possiamo quindi affermare che, per ogni intervallo $(a,b] \subseteq \mathbb{R}$ (sono inclusi in casi in cui $a = infty$ e/o $b = +\infty$) si ha

$$a < S_n \leq b \quad \Leftrightarrow \quad a - np < S_n - np \leq b - np \quad \Leftrightarrow \quad \frac{a-np}{\sqrt{np(1-p)}} < \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{b-np}{\sqrt{np(1-p)}}$$

e che, essendo $\mathbb{E}(S_n) = np$ e $Var(S_n) = np(1-p)$,

$$\tfrac{S_n - np}{\sqrt{np(1-p)}} = \tfrac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} = S_n^*$$

dove S_n^* è la standardizzata di S_n (ossia $\mathbb{E}(S_n^*) = 0$ e $Var(S_n^*) = 1$) otteniamo che

$$\mathbb{P}(a < S_n \leq b) = \mathbb{P}\Big(\frac{a-np}{\sqrt{np(1-p)}} < S_n^* \leq \frac{b-np}{\sqrt{np(1-p)}}\Big) = F_{S_n^*}\Big(\tfrac{b-np}{\sqrt{np(1-p)}}\Big) - F_{S_n^*}\Big(\tfrac{a-np}{\sqrt{np(1-p)}}\Big).$$

Fin qui sembra che abbiamo usato un problema semplice con uno più difficile, ma in realtà ci viene in aiuto un risultato fondamentale, ossia il Teorema Centrale del Limite, che come vedremo, garantisce che la funzione di distribuzione $F_{S_n^*}(x)$ della standardizzata S_n^* di una somma di variabili aleatorie indipendenti ed identicamente distribuite, con media e varianza finite (e varianza non nulla) è approssimata dalla funzione di distribuzione $\Phi(x)$ di una variabile aleatorie gaussiana standard, per cui

$$\mathbb{P}(a < S_n \leq b) = F_{S_n^*}\Big(\tfrac{b-np}{\sqrt{np(1-p)}}\Big) - F_{S_n^*}\Big(\tfrac{a-np}{\sqrt{np(1-p)}}\Big) \simeq \Phi\Big(\tfrac{b-np}{\sqrt{np(1-p)}}\Big) - \Phi\Big(\tfrac{a-np}{\sqrt{np(1-p)}}\Big).$$

Può essere interessante vedere la Figura 1 nella successiva pagina 3.

[1]Questo risultato è riportato anche nel file SCHEMA-VARIABILI-ALEATORIE-15maggio2023.pdf

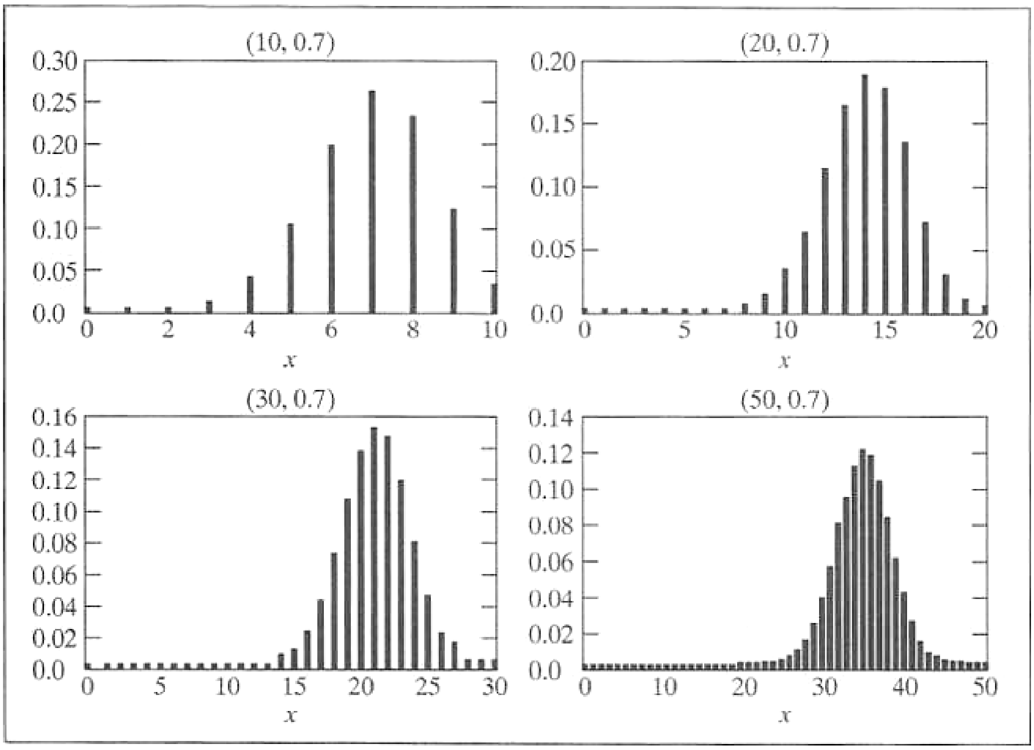


Figura 5.6 La densità discreta di una variabile aleatoria binomiale (n, p) diventa sempre più "normale" al crescere di n.

Figura 1: dal libro di Ross

Prima di enunciare il Teorema Centrale del Limite, vediamo come si utilizza nel caso dell'Esempio 1.1.

Esempio 1.2 (Esempio 1.1 - seconda parte)**.**
A questo punto possiamo calcolare in modo approssimato le probabilità dei punto a) e b) dell'Esempio 1.1. In particolare, essendo $np = 500$ *ed* $np(1-p) = 500 \cdot \frac{95}{100} = 5 \cdot 95 = 475$*, possiamo affermare che*

$$
\begin{aligned}
a)\quad \mathbb{P}(S_{10000} = 450) &= \mathbb{P}(449,5 < S_{10000} \leq 450,5) = \mathbb{P}\left(\tfrac{449,5-500}{\sqrt{475}} < \frac{S_{10000} - 500}{\sqrt{475}} \leq \tfrac{450,5-500}{\sqrt{475}}\right) \\
&\simeq \Phi\left(\tfrac{-49,5}{\sqrt{475}}\right) - \Phi\left(\tfrac{-50,5}{\sqrt{475}}\right) \simeq \Phi(-2,271) - \Phi(-2,317) = 1 - \Phi(2,271) - \big(1 - \Phi(2,317)\big) \\
&\simeq \Phi(2,32) - \Phi(2,27) \simeq 0,9898 - 0,9884 = 0,0014
\end{aligned}
$$

(Abbiamo approssimato $2,317$ *con* $2,32$ *e* $2,271$ *con* $2,27$ *e abbiamo usato il fatto che* $\Phi(-x) = 1 - \Phi(x)$*. Può essere interessare inoltre sapere che, il valore esatto di* $\mathbb{P}(S_{10000} = 450)$*, calcolato con Wolfram-Alpha è* 0.0012685984*)*

$$
b)\quad \mathbb{P}(S_{10000} \leq 450) = \mathbb{P}\left(\frac{S_{10000} - 500}{\sqrt{475}} \leq \tfrac{450-500}{\sqrt{475}}\right) \simeq \Phi\left(\tfrac{-50}{\sqrt{475}}\right) = \Phi(-2,2941) = 1 - \Phi(2,2941) \simeq 0,011.
$$

Nel caso b) si otterrebbe un'approssimazione leggermente migliore se si considerasse che

$$
\begin{aligned}
\mathbb{P}(S_{10000} \leq 450) = \mathbb{P}(S_{10000} \leq 450,5) &= \mathbb{P}\left(\frac{S_{10000} - 500}{\sqrt{475}} \leq \tfrac{450,5-500}{\sqrt{475}}\right) \\
&= \Phi(-2,2712) = 1 - \Phi(2,2712) \simeq 1 - \Phi(2,27) \simeq 1 - 0,9884 = 0,0116,
\end{aligned}
$$

Possiamo ora enunciare il Teorema Centrale del Limite, ma prima ricordiamo la definizione di successione di variabili aleatorie indipendenti.

Definizione 1.1. *Sia $\{X_n;\ n=1,2,\dots\}$ una successione di variabili aleatorie, tutte definite sullo stesso spazio di probabilità $(\Omega,\mathcal{F},\mathbb{P})$. Si dice che sono* **una successione di variabili aleatorie indipendenti** *se comunque scelto un numero finito di esse, queste risultano completamente indipendenti tra loro, ossia se per ogni $n\geq 2$, comunque scelti J_1, J_2, ..., J_n, intervalli (limitati o illimitati) di $\mathbb{R}$, si ha:*

$$\mathbb{P}(X_1\in J_1, X_2\in J_2,\dots,X_n\in J_n)=\mathbb{P}(X_1\in J_1)\cdot\mathbb{P}(X_2\in J_2)\cdot\ldots\cdot\mathbb{P}(X_n\in J_n).$$

Per il seguito è importante osservare che l'indipendenza (completa) di una successione di v.a. X_k, $k\geq 1$, implica l'indipendenza a due a due e quindi implica anche che $Cov(X_i,X_j)=0$. Infatti questo è il punto chiave che permette di calcolare la varianza della somma di $S_n=\sum_{k=1}^n X_k$ come somma delle varianze.
Proposizione 15.1 Se le n variabili aleatorie X_1, X_2, ..., X_n sono completamente (o globalmente) indipendenti fra loro, allora lo sono anche a due a due.
Dimostrazione Per semplicità di notazione mostriamo solamente che X_1 ed X_2 sono indipendenti, ma la dimostrazione è essenzialmente la stessa nel caso generale di X_i ed X_j. Il punto essenziale da osservare è che $\mathbb{R}$ è un intervallo, e che gli eventi del tipo $\{X_k\in\mathbb{R}\}$ coincidono con l'evento certo, di conseguenza

$$\{X_1\in J_1, X_2\in J_2\}=\{X_1\in J_1, X_2\in J_2, X_3\in\mathbb{R},\dots,X_n\in\mathbb{R}\}$$

e quindi

$$\begin{aligned}&\mathbb{P}(X_1\in J_1, X_2\in J_2)=\mathbb{P}(X_1\in J_1, X_2\in J_2, X_3\in\mathbb{R}\dots,X_n\in\mathbb{R})\\ =&\mathbb{P}(X_1\in J_1)\cdot\mathbb{P}(X_2\in J_2)\cdot\mathbb{P}(X_3\in\mathbb{R})\cdot\ldots\cdot\mathbb{P}(X_n\in\mathbb{R})=\mathbb{P}(X_1\in J_1)\cdot\mathbb{P}(X_2\in J_2).\end{aligned}$$

Teorema 1.1 (Teorema Centrale del Limite). *Sia $\{X_i,\ i\geq 1\}$ una successione di v.a. indipendenti ed identicamente distribuite, per le quali esistano finiti valore atteso e varianza. Posto $\mathbb{E}(X_i)=\mu_X$ e $Var(X_i)=\sigma_X^2$, si assuma che $\sigma_X^2>0$. Allora indicando con S_n^* variabile aleatoria standardizzata di $S_n:=\sum_{k=1}^n X_k$, si ha*

$$S_n^*=\frac{S_n-\mathbb{E}(S_n)}{\sqrt{Var(S_n)}}=\frac{S_n-n\mu_X}{\sqrt{n\sigma_X^2}}, \tag{5}$$

e, indicando con $F_{S_n^*}$ la funzione di distribuzione di S_n^*, ossia la funzione $\boxed{x\mapsto F_{S_n^*}(x):=\mathbb{P}(S_n^*\leq x),}$ *si ha, per ogni $x\in\mathbb{R}$,*

$$\lim_{n\to\infty}F_{S_n^*}(x)=\lim_{n\to\infty}\mathbb{P}(S_n^*\leq x)=\Phi(x),\qquad \textit{dove } \Phi(x)=\mathbb{P}(Y\leq x) \textit{ è la funzione di distribuzione di una v.a. } Y\sim N(0,1), \tag{6}$$

in altre parole, ricordando la defінzione della funzione di distribuzione Φ di una variabile aleatoria Gaussiana standard, si ha

$$\lim_{n\to\infty}\mathbb{P}\left(\frac{S_n-n\mu_X}{\sqrt{n\sigma_X^2}}\leq x\right)=\int_{-\infty}^x\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\,dy,\quad \textit{per ogni } x\in\mathbb{R}. \tag{7}$$

Inoltre il limite è uniforme per $x\in\mathbb{R}$, ovvero

$$\lim_{n\to\infty}\sup_{x\in\mathbb{R}}|\mathbb{P}(S_n^*\leq x)-\Phi(x)|=\lim_{n\to\infty}\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{S_n-n\mu_X}{\sqrt{n\sigma_X^2}}\leq x\right)-\int_{-\infty}^x\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\,dy\right|=0. \tag{8}$$

Va notato anche che
(i) Sia la standardizzata S_n^* che $Y\sim N(0,1)$ sono variabili aleatorie standard, ossia $\mathbb{E}(S_n^*)=\mathbb{E}(Y)=0$ e $Var(S_n^*)=Var(Y)=1$
(ii) **per ogni** $x_1<x_2$ si ha $\mathbb{P}(x_1<S_n^*\leq x_2)=\mathbb{P}(S_n^*\leq x_2)-\mathbb{P}(S_n^*\leq x_1)$ e quindi, per n abbastanza grande tale probabilità è approssimata da $\Phi(x_2)-\Phi(x_1)$. Inoltre l'approssimazione funziona anche se si vuole approssimare il valore di $\mathbb{P}(x_1\leq S_n^*\leq x_2)$. Inoltre, posto $E_n:=\sup_{x\in\mathbb{R}}|\mathbb{P}(S_n^*\leq x)-\Phi(x)|$, si ha

$$\Phi(x)-E_n\leq\mathbb{P}(S_n^*\leq x)\leq\Phi(x)+E_n,\qquad \text{e}\qquad |\mathbb{P}(S_n^*\leq x)-\Phi(x)|\leq E_n\xrightarrow[n\to\infty]{}0$$

1.1 Approssimazione normale per la somma di v.a. indipendenti

Supponiamo di essere interessati alla variabile aleatoria S_n data dalla somma di variabili aleatorie indipendenti ed identicamente distribuite, per le quali esistano finiti valore atteso e varianza: $\mathbb{E}(X_i) = \mu_X$ e $Var(X_i) = \sigma_X^2 > 0$ (ossia nelle stesse ipotesi del Teorema Centrale del Limite). Allora la funzione di distribuzione di S_n è la probabilità dell'evento

$$\{S_n \le x\} = \{S_n - \mathbb{E}(S_n) \le x - \mathbb{E}(S_n)\} = \left\{ \overbrace{\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}}}^{=S_n^*} \le \underbrace{\frac{x - \overbrace{\mathbb{E}(S_n)}^{=n\mu_X}}{\sqrt{Var(S_n)}}}_{=\sqrt{n\sigma_X^2}} \right\} = \left\{ S_n^* \le \frac{x - n\mu_X}{\sqrt{n\sigma_X^2}} \right\}, \tag{9}$$

dove l'ultima uguaglianza dipende dal fatto che, essendo S_n la somma di n variabili aleatorie X_i indipendenti, e tutte con la stessa distribuzione, e con valore atteso μ_X e varianza σ_X^2, finiti, con $\sigma_X^2 > 0$,

$$\mathbb{E}(S_n) = \sum_{i=1}^n \mathbb{E}(X_i) = n\mu_X \quad Var(S_n) = Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} \underbrace{Cov(X_i, X_j)}_{X_i \perp X_j \to Cov(X_i, X_j) = 0} = \sum_{i=1}^n Var(X_i) = n\sigma_X^2.$$

Dalla precedente uguaglianza (9) e dal Teorema Centrale del Limite, otteniamo quindi la seguente approssimazione per $F_{S_n}(x)$:

$$F_{S_n}(x) = \mathbb{P}(S_n \le x) = \mathbb{P}\left(S_n^* \le \tfrac{x - n\mu_X}{\sqrt{n\sigma_X^2}}\right) = F_{S_n^*}\left(\tfrac{x - n\mu_X}{\sqrt{n\sigma_X^2}}\right) \approx \Phi\left(\tfrac{x - n\mu_X}{\sqrt{n\sigma_X^2}}\right).$$

Di conseguenza vale anche che

$$\mathbb{P}(\alpha < S_n \le \beta) = F_{S_n}(\beta) - F_{S_n}(\alpha) \approx \Phi\left(\tfrac{\beta - n\mu_X}{\sqrt{n\sigma_X^2}}\right) - \Phi\left(\tfrac{\alpha - n\mu_X}{\sqrt{n\sigma_X^2}}\right), \qquad \text{per ogni } \alpha < \beta.$$

IMPORTANZA DELLA CONVERGENZA UNIFORME

Come è noto dall'Analisi Matematica, se $g_n(x)$ converge a $g(x)$ per ogni $x \in \mathbb{R}$ (ovvero $g_n(x) - g(x)$ tende a zero per ogni $x \in \mathbb{R}$), se $\{x_n\}$ è una successione **non è detto che** allora $g_n(x_n) - g(x_n)$ tenda a zero per n che tende ad infinito. Ad esempio se $g_n(x) = x(1 + 1/n)$ si ha $g(x) = x$, ma la convergenza non è uniforme: infatti $g_n(x) - g(x) = x/n$ e quindi, ad esempio, per $x_n = n$, si ha $g_n(x_n) - g(x_n) = n/n = 1$. Se però la convergenza è uniforme allora

$$|g_n(x_n) - g(x_n)| \le \sup_{x \in \mathbb{R}} |g_n(x) - g(x)| \underset{n \to \infty}{\longrightarrow} 0.$$

Basta ricordare che la convergenza uniforme di $g_n(x)$ a $g(x)$ equivale alla convergenza a zero di $\sup_{x \in \mathbb{R}} |g_n(x) - g(x)|$. In particolare, se la convergenza di $F_{S_n^*}(x)$ a $\Phi(x)$ non fosse uniforme, posto $g_n(x) = F_{S_n^*}(x)$, $g(x) = \Phi(x)$, $x_n = (x - n\mu_X)/\sqrt{n\sigma_X^2}$, **non potremmo quindi affermare** che $F_{S_n^*}(x_n)$ è approssimato da $\Phi(x_n)$.

Va detto inoltre che la precedente approssimazione si può usare anche per $\mathbb{P}(\alpha \le S_n \le \beta)$, per $\mathbb{P}(\alpha < S_n < \beta)$ e per $\mathbb{P}(\alpha \le S_n < \beta)$.

Il Teorema Centrale del Limite (TCL) si può usare anche per calcolare in modo approssimato $\mathbb{P}(|S_n - \mathbb{E}(S_n)| \leq \alpha) = \mathbb{P}(|S_n - n\mu_X| \leq \alpha)$, per ogni $\alpha > 0$. Nelle ipotesi del TCL si ha che

$$\boxed{\mathbb{P}(|S_n - n\mu_X| \leq \alpha) \simeq 2\Phi\Big(\tfrac{\alpha}{\sqrt{n\sigma_X^2}}\Big) - 1.} \tag{10}$$

Infatti

$$|S_n - \mathbb{E}(S_n)| \leq \alpha \quad\Leftrightarrow\quad -\alpha \leq S_n - \mathbb{E}(S_n) \leq \alpha \quad\Leftrightarrow\quad -\frac{\alpha}{\sqrt{Var(S_n)}} \leq \overbrace{\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}}}^{=S_n^*} \leq = \overbrace{\frac{\alpha}{\sqrt{Var(S_n)}}}^{\frac{\alpha}{\sqrt{n\sigma_X^2}}}$$

e quindi, per ogni $\alpha > 0$ si ha

$$\mathbb{P}(|S_n - \mathbb{E}(S_n)| \leq \alpha) = \mathbb{P}\Big(-\frac{\alpha}{\sqrt{Var(S_n)}} \leq S_n^* \leq \frac{\alpha}{\sqrt{Var(S_n)}}\Big) \simeq \Phi\Big(\tfrac{\alpha}{\sqrt{Var(S_n)}}\Big) - \Phi\Big(-\tfrac{\alpha}{\sqrt{Var(S_n)}}\Big) = 2\Phi\Big(\tfrac{\alpha}{\sqrt{Var(S_n)}}\Big) - 1,$$

dove l'ultima uguaglianza si basa sul fatto che

$$\Phi(x) - \Phi(-x) = \Phi(x) - \big(1 - \Phi(x)\big) = 2\Phi(x) - 1.$$

Esempio 1.3 (Esempio 1.1 - terza parte)**.**
Nelle condizioni dell'Esempio 1.1 vogliamo trovare l'espressione e calcolare un'approssimazione per

$$\mathbb{P}(475 \leq S_{10000} \leq 525).$$

Chiaramente

$$\mathbb{P}(475 \leq S_{10000} \leq 525) = \sum_{k=475}^{525} \mathbb{P}(S_{10000} = k) = \sum_{k=475}^{525} \binom{10000}{k} (0,05)^k (0,95)^{10000-k}$$

Inoltre, ricordando che $\mu_X = p = 0,05$ *e quindi* $\mathbb{E}(S_{10000}) = 500$*, e la varianza delle* $X_i = \mathbf{1}_{A_i}$ *vale* $\sigma_X^2 = Var(X_i) = p(1-p) = 0,05 \cdot 0,95 = \frac{475}{10000} = \frac{5^2 \cdot 19}{100^2}$*, e quindi* $\sqrt{n\sigma_X^2} = \sqrt{10000 \frac{5^2 \cdot 19}{10000}} = 5 \cdot \sqrt{19}$*, da cui*

$$\begin{aligned}\mathbb{P}(475 \leq S_{10000} \leq 525) &= \mathbb{P}(-25 \leq S_{10000} - 500 \leq 25) = \mathbb{P}\Big(-\frac{25}{5\cdot\sqrt{19}} \leq \frac{S_{10000} - 500}{5\cdot\sqrt{19}} \leq \frac{25}{5\cdot\sqrt{19}}\Big) \\ &= \mathbb{P}\big(-\tfrac{5}{\sqrt{19}} \leq S_{10000}^* \leq \tfrac{5}{\sqrt{19}}\big) \simeq 2\Phi\big(\tfrac{5}{\sqrt{19}}\big) - 1 = 2\Phi\big(\tfrac{5}{\sqrt{19}}\big) - 1 \\ &\simeq 2\Phi(1,147) - 1 \simeq 2 \cdot 0,874 - 1 = 1,748 - 1 = 0,748\end{aligned}$$

2 Relazione con la Legge dei Grandi Numeri

Iniziamo riportando la Legge dei Grandi Numeri nella versione finita e che avevamo dimostrato per variabili aleatorie in spazi di probabilità finiti. Grazie al fatto (non dimostrato) che la disuguaglianza di Chebyshev vale anche in spazi di probabilità generali, purché le variabili aleatorie abbiano valore atteso e varianza finiti.
Qui sotto ricordiamo la versione generale e la sua dimostrazione. Nell'enunciato si deve aggiungere la condizione che le variabili aleatorie abbiano valore atteso e varianza finita: questa condizione non serviva negli spazi di probabilità finiti, in quanto in tali spazi tutte le variabili aleatorie hanno valore atteso e varianza finiti.

Teorema 2.1 (Legge (debole) dei Grandi numeri, con n fissato)**.** *Se* X_1*,* X_2*, ...,* X_n *sono variabili aleatorie indipendenti a due a due, e con la stessa distribuzione, e se esistano finiti valore atteso* $\mathbb{E}(X_i) = \mu_X$ *e varianza* $Var(X_i) = \sigma_X^2$ *allora, posto* $S_n = \sum_{i=1}^n X_i$ *la somma delle* n *variabili aleatorie e* $Y_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n) = \frac{S_n}{n}$*, la loro media aritmetica, allora, per ongi* $\varepsilon > 0$*, si ha*

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu_X\right| > \varepsilon\right) = \mathbb{P}(|Y_n - \mu_X| > \varepsilon) \leq \frac{1}{n}\frac{\sigma_X^2}{\varepsilon^2} \quad\Leftrightarrow\quad \mathbb{P}\left(\left|\frac{S_n}{n} - \mu_X\right| \leq \varepsilon\right) = \mathbb{P}(|Y_n - \mu_X| \leq \varepsilon) \geq 1 - \frac{1}{n}\frac{\sigma_X^2}{\varepsilon^2}$$

Dimostrazione. Prima di tutto $\mathbb{E}(Y_n) = \mu_X$, come si vede subito:

$$\mathbb{E}(Y_n) = \mathbb{E}\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n}\mathbb{E}(X_1 + \ldots + X_n) = \frac{1}{n}\left[\mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n)\right] = \frac{1}{n} n\mu_X = \mu_X.$$

Inoltre $Var(Y_n) = \frac{\sigma_X^2}{n}$, come si vede subito: l'indipendenza di X_i e X_j per ogni $i \neq j$ implica che $Cov(X_i, X_j) = 0$ (e basterebbe questa condizione di non correlazione) e quindi la varianza della somma è la somma delle varianze:

$$Var(Y_n) = Var\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n^2} Var(X_1 + \ldots + X_n) = \frac{1}{n^2}\left[Var(X_1) + \ldots + Var(X_n)\right] = \frac{1}{n^2} n\sigma_X^2 = \frac{\sigma_X^2}{n}.$$

Basta poi applicare la disuguaglianza di Chebyshev alla v.a. Y_n:

$$\mathbb{P}(\{|Y_n - \mathbb{E}(Y_n)| > \varepsilon\}) \leq \frac{Var(Y_n)}{\varepsilon^2} \quad \Leftrightarrow \quad \mathbb{P}(\{|Y_n - \mu_X| > \varepsilon\}) \leq \frac{\sigma_X^2}{n\varepsilon^2}$$

$\square$

Teorema 2.2 (Legge (debole) dei Grandi Numeri). *Sia $\{X_i,\ i \geq 1\}$ una successione di v.a. indipendenti a due a due ed identicamente distribuite, per le quali esistano finiti valore atteso e varianza. Posto $\mathbb{E}(X_i) = \mu_X$ e $Var(X_i) = \sigma_X^2$, $S_n = \sum_{i=1}^n X_i$ e $Y_n = \frac{S_n}{n}$, si ha, qualunque sia $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu_X\right| > \varepsilon\right) = \lim_{n\to\infty} \mathbb{P}(|Y_n - \mu_X| > \varepsilon) = 0 \quad \Leftrightarrow \quad \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu_X\right| \leq \varepsilon\right) = \lim_{n\to\infty} \mathbb{P}(|Y_n - \mu_X| \leq \varepsilon) = 1$$

Dimostrazione. Basta osservare che

$$0 \leq \mathbb{P}(|Y_n - \mu_X| > \varepsilon) \leq \frac{1}{n}\frac{\sigma_X^2}{\varepsilon^2},$$

mandare n all'infinito ed usare il Teorema del confronto per le successioni numeriche:

$$0 \leq \lim_{n\to\infty} \mathbb{P}(|Y_n - \mu_X| > \varepsilon) \leq \lim_{n\to\infty} \frac{1}{n}\frac{\sigma_X^2}{\varepsilon^2} = 0.$$

$\square$

Legato alla Legge dei Grandi Numeri c'è il problema di calcolare, per $\varepsilon > 0$ fissato,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu_X\right| \leq \varepsilon\right) = \mathbb{P}\left(\frac{|S_n - \mathbb{E}(S_n)|}{n} \leq \varepsilon\right).$$

Tenendo conto che $\mathbb{E}(S_n) = n\mu_X$, si vede facilmente che $\left\{\left|\frac{S_n}{n} - \mu_X\right| \leq \varepsilon\right\} = \left\{\frac{|S_n - \mathbb{E}(S_n)|}{n} \leq \varepsilon\right\}$

Se siamo nelle condizioni del Teorema Centrale del Limite, ossia S_n è la somma di n variabili aleatorie X_i (completamente) indipendenti, e tutte con la stessa distribuzione, e con valore atteso μ_X e varianza σ_X^2, finiti, con $\sigma_X^2 > 0$, allora, le variabili aleatorie X_i sono anche indipendenti a due a due. Siamo quindi nella condizioni sia del Teorema 2.1 e possiamo affermare che

$$\boxed{\mathbb{P}\left(\left|\frac{S_n}{n} - \mu_X\right| \leq \varepsilon\right) \geq 1 - \frac{n\sigma_X^2}{n^2\varepsilon^2} = 1 - \frac{\sigma_X^2}{n\varepsilon^2}.} \tag{11}$$

Inoltre, grazie al TCL, e come mostreremo qui sotto, possiamo anche affermare che

$$\boxed{\mathbb{P}\left(\left|\frac{S_n}{n} - \mu_X\right| \leq \varepsilon\right) \simeq 2\Phi\left(\frac{\sqrt{n}\varepsilon}{\sigma_X}\right) - 1.} \tag{12}$$

Si osservi che nella disuguaglianza (11) compare il valore $\frac{\sigma_X^2}{n\varepsilon^2}$, e che invece nell'approssimazione (12) la funzione di distribuzione Φ viene calcolata in $\frac{\sqrt{n}\varepsilon}{\sigma_x}$, e tale numero è proprio il reciproco di $\sqrt{\frac{\sigma_X^2}{n\varepsilon^2}}$.

L'approssimazione (12) si potrebbe ottenere dal fatto che

$$\left|\frac{S_n}{n} - \mu_X\right| \leq \varepsilon \quad \Leftrightarrow \quad |S_n - n\mu_X| \leq \varepsilon n$$

e prendere $\alpha = n\varepsilon$ nella (10). Oppure possiamo rifare tutti i passaggi osservando che

$$\Big|\frac{S_n}{n} - \mu_X\Big| \le \varepsilon \quad \Leftrightarrow \quad \frac{|S_n - \mathbb{E}(S_n)|}{n} \le \varepsilon \quad \Leftrightarrow \quad |S_n - \mathbb{E}(S_n)| \le n\varepsilon \quad \Leftrightarrow \quad -n\varepsilon \le S_n - \mathbb{E}(S_n) \le n\varepsilon$$
$$\Leftrightarrow \quad -\frac{n\varepsilon}{\sqrt{Var(S_n)}} \le \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \le \frac{n\varepsilon}{\sqrt{Var(S_n)}} \quad \Leftrightarrow \quad -\frac{n\varepsilon}{\sqrt{n\sigma_X^2}} \le S_n^* \le \frac{n\varepsilon}{\sqrt{n\sigma_X^2}}$$
$$\Leftrightarrow \quad -\frac{\sqrt{n}\varepsilon}{\sigma_X} \le S_n^* \le \frac{\sqrt{n}\varepsilon}{\sigma_X},$$

e quindi, per ogni $\varepsilon > 0$ si ha

$$\mathbb{P}\Big(\Big|\frac{S_n}{n} - \mu_X\Big| \le \varepsilon\Big) = \mathbb{P}\Big(-\tfrac{\sqrt{n}\varepsilon}{\sigma_X} \le S_n^* \le \tfrac{\sqrt{n}\varepsilon}{\sigma_X}\Big) \simeq \Phi\Big(\tfrac{\sqrt{n}\varepsilon}{\sigma_X}\Big) - \Phi\Big(-\tfrac{\sqrt{n}\varepsilon}{\sigma_X}\Big) = 2\Phi\Big(\tfrac{\sqrt{n}\varepsilon}{\sigma_X}\Big) - 1,$$

dove, come in precedenza, l'ultima uguaglianza si basa sul fatto che

$$\Phi(x) - \Phi(-x) = \Phi(x) - \big(1 - \Phi(x)\big) = 2\Phi(x) - 1.$$

Come vedremo ora, con l'ausilio di questa approssimazione, si ottiene di nuovo la stessa tesi della Legge debole dei Grandi Numeri, (nella versione per le successioni) ma sotto l'ipotesi ***più restrittiva*** che le variabili aleatorie siano completamente indipendenti. Infatti mandando n all'infinito nella precedente relazione si ottiene

$$\lim_{n\to\infty} \mathbb{P}(\{|Y_n - \mu_X| \le \varepsilon\}) = \lim_{n\to\infty} \left[2\Phi\left(\sqrt{\tfrac{n}{\sigma_X^2}}\,\varepsilon\right) - 1\right] = 2\cdot 1 - 1 = 1,$$

in quanto $\lim_{y\to\infty} \Phi(y) = 1$.

Esempio* 2.1. *Sia X_1 una variabile aleatoria che può assumere i valori 0, $\frac{1}{2}$, 1, $\frac{3}{2}$ e con*

$$p_{X_1}(0) = \mathbb{P}(X_1 = 0) = \frac{1}{10}, \qquad p_{X_1}(\tfrac{1}{2}) = \mathbb{P}(X_1 = \tfrac{1}{2}) = \frac{1}{10},$$
$$p_{X_1}(1) = \mathbb{P}(X_1 = 1) = \frac{4}{10}, \qquad p_{X_1}(\tfrac{3}{2}) = \mathbb{P}(X_1 = \tfrac{3}{2}) = \frac{4}{10}.$$

Si ponga il valore atteso di X_1 uguale a μ_X e la sua varianza uguale a σ_X^2.
Siano X_1, X_2, $X_3, \ldots, X_{100}$ delle variabili aleatorie con la stessa distribuzione di X_1 e completamente (o globalmente) indipendenti tra loro e si ponga $Y_{100} \equiv \frac{\sum_{j=1}^{100} X_j}{100}$. Utilizzando il Teorema Centrale del Limite, approssimare la probabilità

$$\mathbb{P}\left(\left\{\mu_X - \frac{1}{10} \le Y_{100} \le \mu_X + \frac{1}{10}\right\}\right).$$

Soluzione: Innanzi tutto come si trova facilmente si ha $\mu_X = \frac{21}{20}$ e $\sigma_X^2 = \frac{89}{400}$ (vedere qui sotto per la verifica). Quindi la probabilità cercata è approssimata con

$$\begin{aligned}\mathbb{P}\left(\left\{\mu_X - \frac{1}{10} \le Y_{100} \le \mu_X + \frac{1}{10}\right\}\right) &\simeq 2\Phi\left(\sqrt{\tfrac{n}{\sigma_X^2}}\,\varepsilon\right) - 1 = 2\Phi\left(\sqrt{\tfrac{100}{\frac{89}{400}}}\,\tfrac{1}{10}\right) - 1\\ &= 2\Phi\left(\sqrt{\tfrac{400}{89}}\right) - 1 = 2\Phi(\tfrac{20}{\sqrt{89}}) - 1 \simeq 2\Phi(2,1199) - 1\\ &\simeq 2\Phi(2,12) - 1 \simeq 2\cdot 0,9830 - 1 = 1,9660 = 0,9660\end{aligned}$$

Calcolo del valore atteso e della varianza delle variabili aleatorie X_i:

$$\mu_X := \mathbb{E}(X_i) = 0\cdot p_X(0) + \frac{1}{2}\cdot p_X(\frac{1}{2}) + 1\cdot p_X(1) + \frac{3}{2}\cdot p_X(\frac{3}{2}) = 0\cdot\frac{1}{10} + \frac{1}{2}\cdot\frac{1}{10} + 1\cdot\frac{4}{10} + \frac{3}{2}\cdot\frac{4}{10} = \frac{1+8+12}{20} = \frac{21}{20},$$

e inoltre

$$\mathbb{E}(X_i^2) = 0^2\cdot p_X(0) + \left(\frac{1}{2}\right)^2\cdot p_X(\frac{1}{2}) + 1^2\cdot p_X(1) + \left(\frac{3}{2}\right)\cdot p_X(\frac{3}{2}) = 0\cdot\frac{1}{10} + \frac{1}{4}\cdot\frac{1}{10} + \frac{4}{10} + \frac{9}{4}\cdot\frac{4}{10} = \frac{1+16+36}{40} = \frac{53}{40}.$$

da cui

$$\sigma_X^2 = Var(X_i) = \mathbb{E}(X_i^2) - \big(\mathbb{E}(X_i)\big)^2 = \frac{53}{40} - \left(\frac{21}{20}\right)^2 = \frac{530-441}{400} = \frac{89}{400}$$

Determinazione del numero di prove in modo che la media aritmetica sia vicina al valore atteso con probabilità *"alta"*.

Finora ci siamo posti il problema del tipo: **fissati n (*grande*) ed $\varepsilon > 0$, quanto vale approssimativamente la probabilità che media aritmetica e valore atteso differiscano di meno di ε?**

Supponiamo ora di voler rispondere ***in modo approssimato*** alla domanda:
***siano ε e $\delta \in (0,1)$ fissati,* per quale valore di $n = n(\varepsilon,\delta)$ si può affermare che**

$$\mathbb{P}(\{|Y_n - \mu_X| \leq \varepsilon\}) \geq 1-\delta = \quad \text{o almeno} \quad \mathbb{P}(\{|Y_n - \mu_X| \leq \varepsilon\}) \simeq 1-\delta?$$

Dalla disuguaglianza di Chebyshev e dal TCL ossia

$$\mathbb{P}\Big(\Big|\frac{S_n}{n} - \mu_X\Big| \leq \varepsilon\Big) \geq 1 - \frac{\sigma_X^2}{n\varepsilon^2}$$

e

$$\mathbb{P}\Big(\big|\frac{S_n}{n} - \mu_X\big| \leq \varepsilon\Big) \simeq 2\Phi\Big(\tfrac{\sqrt{n}\varepsilon}{\sigma_X}\Big) - 1$$

Possiamo rispondere in due modi:

I. Con la disuguaglianza di Chebyshev possiamo affermare che se n è tale che $1 - \frac{\sigma_X^2}{n\varepsilon^2} \geq 1-\delta$ allora

$$\mathbb{P}\Big(\big|\frac{S_n}{n} - \mu_X\big| \leq \varepsilon\Big) \geq 1 - \frac{\sigma_X^2}{n\varepsilon^2} \geq 1-\delta.$$

Quindi, utilizzando la disuguaglianza di Chebyshev, prendendo

$$\boxed{n \geq n_{Ch}(\varepsilon;\delta) := \frac{\sigma_X^2}{\delta\varepsilon^2},}$$

sicuramente $\mathbb{P}\Big(\big|\frac{S_n}{n} - \mu_X\big| \leq \varepsilon\Big) \geq 1-\delta$, infatti

$$1 - \frac{\sigma_X^2}{n\varepsilon^2} \geq 1-\delta \quad \Leftrightarrow \quad \frac{\sigma_X^2}{n\varepsilon^2} \leq \delta \quad \Leftrightarrow \quad \sigma_X^2 \leq \delta n \varepsilon^2 \quad \Leftrightarrow \quad \frac{\sigma_X^2}{\delta\varepsilon^2} \leq n.$$

II. Con il TCL avremo solo una condizione approssimata, infatti possiamo affermare che se n è tale che

$$2\Phi\Big(\tfrac{\sqrt{n}\varepsilon}{\sigma_X}\Big) - 1 \geq 1-\delta$$

allora

$$\mathbb{P}\Big(\Big|\frac{S_n}{n} - \mu_X\Big| \leq \varepsilon\Big) \simeq 2\Phi\Big(\tfrac{\sqrt{n}\varepsilon}{\sigma_X}\Big) - 1 \geq 1-\delta.$$

A questo scopo, dopo aver definito, per ogni $\alpha \in (0,1)$, l'α-quantile x_α come quell'unico valore tale che

$$\Phi(x_\alpha) = \alpha,$$

possiamo affermare che basta prendere

$$\boxed{n \geq n_{TCL}(\varepsilon;\delta) := \frac{x_{1-\delta/2}^2 \sigma_X^2}{\varepsilon^2}.}$$

Infatti

$$\begin{aligned} 2\Phi\Big(\tfrac{\sqrt{n}\varepsilon}{\sigma_X}\Big) - 1 \geq 1-\delta \quad &\Leftrightarrow \quad 2\Phi\Big(\tfrac{\sqrt{n}\varepsilon}{\sigma_X}\Big) \geq 2-\delta \quad \Leftrightarrow \quad \Phi\Big(\tfrac{\sqrt{n}\varepsilon}{\sigma_X}\Big) \geq 1 - \frac{\delta}{2} = \Phi\left(x_{1-\delta/2}\right) \\ &\Leftrightarrow \quad \tfrac{\sqrt{n}\varepsilon}{\sigma_X} \geq x_{1-\delta/2} \quad \Leftrightarrow \quad \sqrt{n}\varepsilon \geq x_{1-\delta/2}\,\sigma_X \\ &\Leftrightarrow \quad n \geq \frac{x_{1-\delta/2}^2 \sigma_X^2}{\varepsilon^2} =: n_{TCL}(\varepsilon;\delta) \end{aligned}$$

Esempio 2.2 (INTERVALLI DI FIDUCIA o CONFIDENZA).
Si condisideri una moneta truccata con parametro p incognito, dove p rappresenta la probabilità che esca testa. Al fine di determinare p si lancia una la moneta n volte e si stima p con la frequenza di testa S_n/n, dove S_n è il numero delle volte in cui esce testa, tra gli n lanci.

Dato $\varepsilon > 0$ vogliamo determinare quanto si deve prendere n, cioè quante volte si deve lanciare la moneta, in modo che
(a) sicuramente la probabilità che $|S_n/n - p| \leq \varepsilon$ sia almeno il 95% (usando la disuguaglianza di Chebyshev).
(b) la probabilità che $|S_n/n - p| \leq \varepsilon$ sia circa il 95% (usando l'approssimazione normale, ovvero con il Teorema Centrale del Limite)
(c) per $\varepsilon = 0.1$ e per $\varepsilon = 0.01$, calcolare esplicitamente il valore n trovato in entrambi i casi (a) e (b).
Da quanto visto in precedenza consideriamo che $95\% = 0.95 = 1 - \delta$, per $\delta = 0.05$.
Per la risposta alla domanda (a) prendiamo, per tale valore di $\delta = 0.05$, e $\sigma_X^2 = p(1-p)$

$$n \geq n_{Ch,p}(\varepsilon;\delta) := \frac{\sigma_X^2}{\delta\varepsilon^2} = \frac{p(1-p)}{0.05\varepsilon^2}.$$

Abbiamo aggiunto la dipendenza da p in $n_{Ch,p}(\varepsilon;\delta)$ per mettere in evidenza che tale numero cambia al variare del parametro p da stimare.
Il problema è che non conosciamo p, ma sappiamo solo che per $p \in [0,1]$ il valore $\sigma_X^2 = p(1-p) \leq \frac{1}{2}(1-\frac{1}{2}) = 1/4$, e quindi, osservando che

$$n_{Ch,1/2}(\varepsilon;\delta) = \frac{\frac{1}{4}}{0.05\,\varepsilon^2} \geq \frac{p(1-p)}{0.05\,\varepsilon^2} = n_{Ch,p}(\varepsilon;\delta)$$

e che, per $\delta = 0.05$, $\frac{\frac{1}{4}}{0.05} = \frac{\frac{1}{4}}{\frac{5}{100}} = \frac{100}{20} = 5$. Basterà quindi prendere

$$n \geq n_{Ch,1/2}(\varepsilon;0.05) \equiv \frac{5}{\varepsilon^2} \geq n_{Ch,p}(\varepsilon,\delta) = \frac{p(1-p)}{0.05\,\varepsilon^2}.$$

Invece per la risposta alla domanda (b) prendiamo, sempre per $\delta = 0.05$

$$n \geq n_{TCL,p}(\varepsilon;\delta) := \frac{x^2_{1-\delta/2}\,p(1-p)}{\varepsilon^2} = \frac{x^2_{0.975}\,p(1-p)}{\varepsilon^2}$$

Di nuovo, non conosciamo p, ma tenendo presente che $p(1-p) \leq 1/4$ e che quindi

$$n_{TCL,(1/2)}(\varepsilon;\delta) = \frac{x^2_{1-\delta/2}(1/4)}{\varepsilon^2} \geq n_{TCL,p}(\varepsilon;\delta) = \frac{x^2_{1-\delta/2}\,p(1-p)}{\varepsilon^2},$$

si ottiene che basta prendere

$$n \geq \frac{x^2_{0.975}}{4\,\varepsilon^2} = n_{TCL,1/2}(\varepsilon,0.05) \geq n_{TCL,p}(\varepsilon,0.05) = \frac{x^2_{0.975}\,p(1-p)}{\varepsilon^2}$$

Per trovare $x_{0.975}$ basta consultare la tavola della gaussiana e osservare che per $\Phi(1.96) = 0.9750$ per ottenere che quindi basta prendere

$$n \geq \frac{x^2_{0.975}}{4\,\varepsilon^2} = \frac{1.96^2}{4\,\varepsilon^2} \equiv 0.9604\frac{1}{\varepsilon^2}.$$

Quindi, per rispondere al punto (c), essendo $\varepsilon = 0.1 = 1/10$ e $\varepsilon = 0.01 = 1/100$ si ha, rispettivamente che

$$n \geq n_{Ch,1/2}(0.1;0.05) \geq \frac{5}{(0.1)^2} = \frac{5}{(1/100)} = 500, \quad n \geq n_{Ch,1/2}(0.01,0.05) \geq \frac{5}{(0.01)^2} = \frac{5}{(1/10\cdot000)} = 50\cdot000,$$

$$n \geq 97, \text{ in quanto } n \geq n_{TCL,1/2}(0.1;0.05) \geq \frac{x^2_{0.975}}{1/100} = 96.04, \quad n \geq n_{TCL,1/2}(0.01;0.05) \geq \frac{x^2_{0.975}}{1/10000} = 9\cdot604.$$

ATTENZIONE:
La tavola che segue calcola $\Phi(z) = \mathbb{P}(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\,dy$, dove $Z \sim N(0,1)$ è una variabile aleatoria gaussiana standard, per centinaia valori di $z \geq 0$. In altre parole calcola l'area al di sotto del grafico della densità di probabilità gaussiana standard $y = \varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$ come descritta nella figura qui accanto.

NOTA BENE:
Ci sono delle tavole che calcolano invece
$\mathbb{P}(0 < Z \leq z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\,dy$.
Le due tavole si distinguono facilmente: basta osservare il valore che hanno nel punto 0.00, la nostra mette il valore $\mathbb{P}(Z \leq 0) = 0.5 = 1/2$, mentre l'altro tipo mette il valore 0, in quanto $\mathbb{P}(0 < Z \leq 0) = 0$

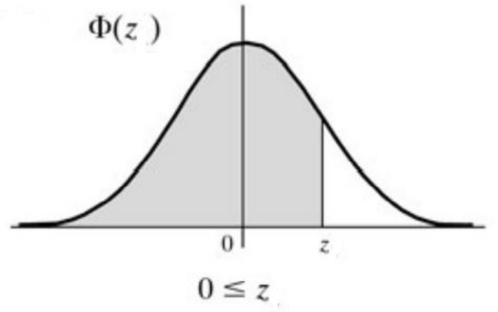


Grafico della funzione $\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$.
$\Phi(z)$ è l'area in grigio, delimitata dall'asse x, la retta $x = z$ e il grafico di φ

Tavola della funzione di distribuzione gaussiana standard

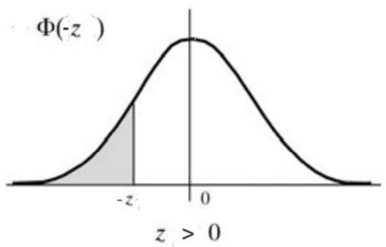
$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy$$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

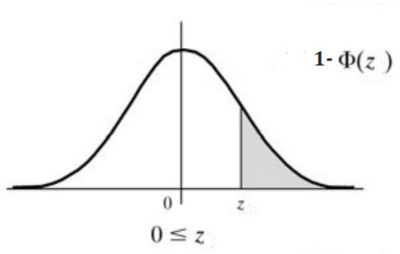
Spiegazione dell'uso della tavola della gaussiana standard:
Per iniziare si noti che gli indici di riga sono i 35 numeri $\{0.0,\ 0.1,\ \ldots,\ 3.3,\ 3.4\}$ che vanno da 0 a 3.4 e che differiscono tra loro di un decimo, mentre gli indici di colonna sono i 10 numeri $\{0.00,\ 0.01, \ldots, 0.09\}$, che vanno da 0 a 0.09 e differiscono tra loro di un centesimo. Sommando un numero di riga, con uno di colonna si può ottenere uno tra i 350 valori di x che vanno da 0 a 3.49, e che differiscono tra loro di un centesimo. Viceversa ognuno di tali valori z, ad esempio $z = 1.43$, si può considerare come la somma della parte fino ai decimi più la parte dei centesimi, nell'esempio $z = 1.43 = 1.4 + 0.03$, individuando così un indice di riga, nell'esempio 1.4, ed uno di colonna, nell'esempio 0.03. Nella tavola, al posto di riga 1.4 e di colonna 0.03 si trova il valore di $\Phi(1.43) = 0.9236$, approssimato alla quarta cifra decimale, ovvero

$$\Phi(1.43) = \int_{-\infty}^{1.43} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\, dy = 0.9236.$$

I valori di $\Phi(z)$ per $z \geq 3.50$ si possono[2] approssimare con 1. Per quanto riguarda i valori di $\Phi(z)$ per valori negativi si usa la seguente relazione, che dipende dalla simmetria rispetto a 0 della densità di una gaussiana standard, come illustrato nelle figure:



$$\boxed{\Phi(-z) = 1 - \Phi(z)}$$



In questo modo si possono ottenere[3] i valori della funzione $\Phi(x)$ in 699 valori tra $-3,49$ e $3,49$, equispaziati di un centesimo, ossia in

$$x = \frac{k}{100}, \qquad \text{per } -349 \leq k \leq 349.$$

16.3 Esercizi di Verifica della Lezione 16 di [SN]

ATTENZIONE: si tratta della Lezione dal titolo Variabili aleatorie in casi più generali: Legge dei Grandi Numeri e Teorema Centrale del Limite, che nella versione di settembre 2024 appare come Lezione 17

Esercizio 16.1. *(Stessa situazione dell'Esercizio 6.1)*
Un candidato ad un'elezione ha bisogno di almeno 50 *voti per essere eletto. Prepara allora una lettera per informare i potenziali elettori circa la sua candidatura, il suo programma elettorale, etc....*
Egli valuta che ogni persona che riceve la lettera si recherà effettivamente a votare per lui con una probabilità del 40%, *indipendentemente dal comportamento degli altri (e si sottointende che egli certamente non ottiene voti da coloro cui non ha inviato la lettera).*
(a) Calcolare approssimativamente la probabilità che egli riceva esattamente 71 *voti se invia la lettera a 200 persone.*
(b) Calcolare approssimativamente la probabilità di essere eletto se invia la lettera a 100 persone.
*(c) Calcolare approssimativamente il numero minimo di persone alle quali deve inviare copia della lettera affinché la probabilità di essere eletto sia superiore all'*80%

Esercizio 16.2. *Siano X_i variabili aleatorie indipendenti e uniformi in* $\{1,2,3\}$.
Posto $S_n = \sum_{i=1}^{n} X_i$, e prendendo $n = 54$
(a) Calcolare $\mathbb{E}[S_n]$ e $Var(S_n)$.
(b) Calcolare approssimativamente $\mathbb{P}(S_n \leq 116)$.
(c) Calcolare approssimativamente $\mathbb{P}(S_n > 100)$.
(d) Calcolare approssimativamente $\mathbb{P}(100 < S_n \leq 116)$.
(e) (domanda aggiuntiva)
Si può usare la disuguaglianza di Chebyshev per ottenere una minorazione di $\mathbb{P}(100 \leq S_n \leq 116)$?
in altre parole: potete trovare un valore $\alpha > 0$ tale che $\mathbb{P}(100 \leq S_n \leq 116) > \alpha$?

---

[2] Ovviamente approssimare con 1 numeri maggiori o uguali a 0.9998 ha senso solo in problemi in cui la precisione non è fondamentale.

[3] Ricordiamo che la funzione $\Phi(x)$ si interpreta come la funzione di distribuzione di una variabile aleatoria Gaussiana standard, ossia una v.a. Z continua per la quale

$$\mathbb{P}(a < Z \leq b) = \Phi(b) - \Phi(a) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy.$$

Si noti inoltre che $\Phi(0) = \Phi(-0) = 1/2$ e quindi non si tratta di 700 valori, ma solo di 699

ALTRI ESERCIZI

Esercizio 16.3. *Siano X_i variabili aleatorie indipendenti e tutte con distribuzione di Poisson di parametro λ, e con valore atteso $\mathbb{E}(X_i) = 3$.*
Dopo aver trovato il valore di λ e di $Var(X_i)$, posto $S_n = \sum_{i=1}^n X_i$, dire che tipo di distribuzione ha S_n e scrivere la formula per calcolare $\mathbb{P}(S_n = k)$, specificando, come usuale e necessario, per quali valori di k, $\mathbb{P}(S_n = k) > 0$.
Prendendo $n = 108$,
(a) Calcolare $\mathbb{E}[S_n]$ e $Var(S_n)$.
(b) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(S_n \leq 350)$.
(c) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(S_n > 288)$.
(d) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(288 < S_n \leq 350)$.

Esercizio 16.4. *Siano X_i variabili aleatorie indipendenti e tutte $Geom(p)$, e con valore atteso $\mathbb{E}(X_i) = 3$.*
Dopo aver trovato il valore di p e di $Var(X_i)$, posto $S_n = \sum_{i=1}^n X_i$, dire che tipo di distribuzione ha S_n e scrivere la formula per calcolare $\mathbb{P}(S_n = k)$, specificando, come usuale e necessario, per quali valori di k, $\mathbb{P}(S_n = k) > 0$. Prendendo $n = 96$
(a) Calcolare $\mathbb{E}[S_n]$ e $Var(S_n)$.
(b) Calcolare approssimativamente $\mathbb{P}(S_n \leq 324)$.
(c) Calcolare approssimativamente $\mathbb{P}(S_n > 252)$.
(d) Calcolare approssimativamente $\mathbb{P}(252 < S_n \leq 324)$.

Esercizio 16.5. *Sia S una variabile aleatoria di Poisson di parametro $\lambda = 324$.*
(a) Calcolare $\mathbb{E}[S]$ e $Var(S)$.
(b) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(S \leq 350)$.
(c) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(S > 288)$.
(d) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(288 < S \leq 350)$.
suggerimento: usare il fatto che la somma di variabili aleatorie indipendenti e tutte con distribuzione di Poisson ha distribuzione di Poisson e tenere presente l'Esercizio 16.3

Esercizio 16.6. *Sia T_n una variabili aleatoria di Pascal di parametri n e p. Prendendo $n = 96$ e $p = 1/3$*
(a) Calcolare $\mathbb{E}[T_n]$ e $Var(T_n)$.
(b) Scrivere l'espressione esatta di e calcolare approssimativamente $\mathbb{P}(T_n \leq 324)$.
(c) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(T_n > 252)$.
(d) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(252 < T_n \leq 324)$.
suggerimento: usare il fatto che la somma di variabili aleatorie indipendenti e tutte con distribuzione $Geom(p)$ ha distribuzione di Pascal di parametri n e p, e tenere presente l'Esercizio 16.4

Soluzione dell'Esercizio 16.1 Sia S_n il numero degli elettori che votano per il candidato, nel caso in cui il candidato spedisca n lettere.
(a) Calcolare approssimativamente la probabilità che egli riceva esattamente 71 voti se invia la lettera a 200 persone.
Per il TCL sappiamo che

$$\mathbb{P}(S_n = k) = \mathbb{P}(k - \tfrac{1}{2} < S_n \leq k + \tfrac{1}{2}) = \mathbb{P}\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}} < S_n^* \leq \frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) \approx \Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

In questo caso $n = 200$, $p = 40\% = 0,4$ e $k = 71$, per cui $np = 80$ e $\sqrt{np(1-p)} = \sqrt{80 \cdot 0,6} = \sqrt{48}$ per cui si ha che

$$\begin{aligned}\mathbb{P}(S_{200} = 71) &\approx \Phi\left(\frac{71,5 - 80}{\sqrt{48}}\right) - \Phi\left(\frac{70,5 - 80}{\sqrt{48}}\right) \approx \Phi(-1,22) - \Phi(-1,37) = 1 - \Phi(1,22) - \big(1 - \Phi(1,37)\big) \\ &= \Phi(1,37) - \Phi(1,22) \approx 0,9147 - 0,8888 = 0,025\end{aligned}$$

(b) Calcolare approssimativamente la probabilità di essere eletto se invia la lettera a 100 persone.
In generale

$$\mathbb{P}(S_n > k) = \mathbb{P}\left(S_n^* > \frac{k - np}{\sqrt{np(1-p)}}\right) \approx 1 - \Phi\left(\frac{k - np}{\sqrt{np(1-p)}}\right)$$

Nel nostro caso $n = 100$, $p = 40\% = 0,4$ e $k = 49$, per cui $np = 40$ e $\sqrt{np(1-p)} = \sqrt{40 \cdot 0,6} = \sqrt{24}$, e quindi

$$\mathbb{P}(S_{100} \geq 50) = \mathbb{P}(S_{100} > 49) \approx 1 - \Phi\left(\frac{49 - 40}{\sqrt{24}}\right) \approx 1 - \Phi(1,83) \approx 1 - 0,9664 = 0,0336$$

(c) Calcolare approssimativamente il numero minimo di persone alle quali deve inviare copia della lettera affinché la probabilità di essere eletto sia superiore all'80%
Stiamo cercando n tale che

$$\mathbb{P}(S_n > 49) > 80/100 = 0,8$$

Come visto al punto (b), sappiamo che

$$\mathbb{P}(S_n > k) \approx 1 - \Phi\left(\frac{k - np}{\sqrt{np(1-p)}}\right)$$

e quindi, per $k = 49$ e $p = 0,4$, per cui $p(1-p) = 0,4 \cdot 0,6 = 0,24$ cerchiamo n tale che

$$1 - \Phi\left(\frac{49 - np}{\sqrt{np(1-p)}}\right) > 0,8 \quad \Leftrightarrow \quad \Phi\left(\frac{n \cdot 0,4 - 49}{\sqrt{n \cdot 0,24}}\right) > 0,8$$

Dalla tavola della gaussiana troviamo che il primo valore $\bar{x}$ per il quale $\Phi(\bar{x}) > 0,8$ è il valore $\bar{x} = 0,85$. Essendo la funzione $\Phi(x)$ crescente basta prendere n tale che

$$\Phi\left(\frac{n \cdot 0,4 - 49}{\sqrt{n \cdot 0,24}}\right) > \Phi(\bar{x}) = \Phi(0,85) = 0,8023 > 0,8 \quad \Leftrightarrow \quad \frac{n \cdot 0,4 - 49}{\sqrt{n \cdot 0,24}} \geq \bar{x} = 0,85$$

ovvero, posto $y = \sqrt{n}$

$$\Updownarrow$$

$$y^2 \cdot 0,4 - 49 \geq 0,85 \cdot y \cdot \sqrt{0,24} \quad \Leftrightarrow \quad 0,4 \cdot y^2 - 0,85 \cdot \sqrt{0,24} \cdot y - 49 \geq 0$$

$$\Updownarrow$$

$$y \leq \frac{17 - \sqrt{31649}}{16} \text{ oppure } y \geq \frac{17 + \sqrt{31649}}{16} \quad \Leftrightarrow \quad y \leq -10,056 \text{ oppure } y \geq 12,181$$

Ovviamente, essendo $y = \sqrt{n}$ possiamo accettare solo le soluzioni del tipo

$$y = \sqrt{n} \geq 12,181 \quad \Leftrightarrow \quad n \geq (12,181)^2 = 148,376761 \quad \Leftrightarrow \quad n \geq 149$$

Con WolframAlpha si ottiene invece che la soluzione esatta è $n = 136$: con l'istruzione
`probability of at least 50 successes in 136 trials with p=.4`
si trova che tale probabilità vale 0,803999, mentre con 135 al posto di 136 si ottiene invece una probabilità uguale a 0,784705.

Soluzione dell'Esercizio 16.2 Siano X_i variabili aleatorie indipendenti e uniformi in $\{1,2,3\}$.
Posto $S_n = \sum_{i=1}^n X_i$, e prendendo $n = 54$

(a) Calcolare $\mathbb{E}[S_{54}]$ e $Var(S_{54})$.

Essendo le variabili aleatorie indipendenti e uniformi in $\{1,2,3\}$ si ha che $\mathbb{E}(X_i) = 2$ e $Var(X_i) = \frac{2}{3}$ e quindi

$$\mathbb{E}[S_{54}] = 54 \cdot 2 = 108, \quad Var(S_{54}) = 54 \cdot \frac{2}{3} = 36$$

(b) Calcolare approssimativamente $\mathbb{P}(S_{54} \leq 116)$.

Di nuovo per il TCL sappiamo che

$$\mathbb{P}(S_{54} \leq 116) = \mathbb{P}\left(S_{54}^* \leq \frac{116 - \mathbb{E}(S_{54})}{\sqrt{Var(S_{54})}}\right) = \mathbb{P}\left(S_{54}^* \leq \frac{116-108}{\sqrt{36}}\right) = \mathbb{P}\left(S_{54}^* \leq \frac{8}{6}\right) \approx \Phi(4/3)$$
$$\approx \Phi(1,33) \approx 0,9082$$

(c) Calcolare approssimativamente $\mathbb{P}(S_{54} > 100)$.

$$\mathbb{P}(S_{54} > 100) = \mathbb{P}\left(S_{54}^* > \frac{100 - \mathbb{E}(S_{54})}{\sqrt{Var(S_{54})}}\right) = \mathbb{P}\left(S_{54}^* > \frac{100-108}{\sqrt{36}}\right) = \mathbb{P}\left(S_{54}^* > -\frac{8}{6}\right) \approx 1 - \Phi(-4/3)$$
$$= 1 - \big(1 - \Phi(4/3)\big) = \Phi(4/3) \approx \Phi(1,33) \approx 0,9082.$$

(d) Calcolare approssimativamente $\mathbb{P}(100 < S_n \leq 116)$.

$$\mathbb{P}(100 < S_{54} \leq 116) = \mathbb{P}\left(\frac{100-108}{\sqrt{36}} < S_{54}^* \leq \frac{116 - \mathbb{E}(S_{54})}{\sqrt{Var(S_{54})}}\right) = \mathbb{P}\left(-\frac{4}{3} < S_{54}^* \leq \frac{4}{3}\right) \approx \Phi(4/3) - \Phi(-4/3)$$
$$= \Phi(4/3) - \big(1 - \Phi(4/3)\big) = 2\Phi(4/3) - 1 \approx 2\Phi(1,33) - 1 \approx 2 \cdot 0,9082 - 1 = 0,8164$$

(e) (domanda aggiuntiva)
Si può usare la disuguaglianza di Chebyshev per ottenere una minorazione di $\mathbb{P}(100 \leq S_{54} \leq 116)$?
in altre parole: potete trovare un valore $\alpha > 0$ tale che $\mathbb{P}(100 \leq S_{54} \leq 116) > \alpha$?

Osserviamo che

$$\{100 \leq S_{54} \leq 116\} = \{-8 \leq S_{54} - 108 \leq 8\} = \{-8 \leq S_{54} - \mathbb{E}(S_{54}) \leq 8\} = \{|S_{54} - 108| \leq 8\}$$

Con la disuguaglianza di Chebyshev otteniamo quindi che

$$\mathbb{P}(100 \leq S_n \leq 116) = \mathbb{P}(|S_{54} - 108| \leq 8) \geq 1 - \frac{Var(S_{54})}{8^2} = 1 - \frac{36}{64} = 1 - \frac{9}{16} = \frac{7}{16}.$$

Soluzione dell'Esercizio 16.3 Siano X_i variabili aleatorie indipendenti e tutte con distribuzione di Poisson di parametro λ, e con valore atteso $\mathbb{E}(X_i) = 3$.
Dopo aver trovato il valore di λ e di $Var(X_i)$, posto $S_n = \sum_{i=1}^n X_i$, dire che tipo di distribuzione ha S_n e scrivere la formula per calcolare $\mathbb{P}(S_n = k)$, specificando, come usuale e necessario, per quali valori di k, $\mathbb{P}(S_n = k) > 0$.
Prendere $n = 108$, e rispondere anche alle seguenti domande
(a) Calcolare $\mathbb{E}[S_n]$ e $Var(S_n)$.

Sappiamo che se $X \sim Poisson(\lambda)$ allora $\mathbb{E}(X) = \lambda$ e $Var(X) = \lambda$ e quindi $\mathbb{E}(X_i) = \lambda = 3 = Var(X_i)$, per ogni $i \geq 1$. Di conseguenza $E(S_n) = 108 \cdot \mathbb{E}(X_1) = 108 \cdot 3 = 324$, e $Var(S_n) = \mathbb{E}(S_n) = 324 = 4 \cdot 81 = 18^2$. Inoltre sappiamo che S_n è una variabile di Poisson di parametro $n\lambda = \mathbb{E}(S_n)$.

(b) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(S_n \leq 350)$, sempre per $n = 108$.

Sappiamo che la somma di variabili aleatorie indipendenti di Poisson ha ancora distribuzione di Poisson, inoltre il parametro della somma è la somma dei parametri delle variabili aleatorie (del resto il paramentro coincide con il valore atteso). Di conseguenza S_{108} ha distribuzione di Poisson di parametro 324, e quindi

$$\mathbb{P}(S_{108} \leq 350) = \sum_{k=0}^{350} \frac{324^k}{k!} e^{-324}.$$

D'altra parte, poiché $\mathbb{E}(S_{108}) = Var(S_{108}) = 324$, possiamo usare l'approssimazione normale per ottenere che

$$\mathbb{P}(S_{108} \leq 350) = \mathbb{P}\left(S^*_{108} \leq \frac{350-324}{\sqrt{324}}\right) \approx \Phi\left(\frac{26}{18}\right) \approx \Phi(1,44) \approx 0,9251$$

(c) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(S_{108} > 288)$.

Essendo $\mathbb{P}(S_{108} > 288) = 1 - \mathbb{P}(S_{108} \leq 288)$, in modo del tutto analogo al punto precedente si ottiene

$$\mathbb{P}(S_{108} > 288) = 1 - \mathbb{P}\left(S^*_{108} \leq \frac{288-324}{\sqrt{324}}\right) \approx 1 - \Phi\left(-\frac{36}{18}\right) \approx 1 - \Phi(-2) = \Phi(2) \approx 0,9772$$

(d) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(288 < S_{108} \leq 350)$.

$$\begin{aligned}\mathbb{P}(288 < S_{108} \leq 350) &= \mathbb{P}\left(\frac{288-324}{\sqrt{324}} < S^*_{108} \leq \frac{350-324}{\sqrt{324}}\right) \approx \Phi\left(\frac{26}{18}\right) - \Phi(-2) \approx \Phi(1,44) - \big(1-\Phi(2)\big)\\ &\approx 0,9251 + 0,9772 - 1 = 0,9023\end{aligned}$$

Soluzione dell'Esercizio 16.4 Siano X_i variabili aleatorie indipendenti e tutte $Geom(p)$, e con valore atteso $\mathbb{E}(X_i) = 3$. Dopo aver trovato il valore di p e di $Var(X_i)$, posto $S_n = \sum_{i=1}^{n} X_i$, dire che tipo di distribuzione ha S_n e scrivere la formula per calcolare $\mathbb{P}(S_n = k)$, specificando, come usuale e necessario, per quali valori di k, $\mathbb{P}(S_n = k) > 0$. Prendendo $n = 96$

(a) Calcolare $\mathbb{E}[S_n]$ e $Var(S_n)$.

Una variabile aleatoria $X \sim Geom(p)$, ha valore atteso $\mathbb{E}(X) = \frac{1}{p}$ e $Var(X) = \frac{1-p}{p^2}$ e quindi $p = 1/3$ e $Var(X_i) = \frac{\frac{2}{3}}{\frac{1}{3^2}} = 6$, per ogni $i \geq 1$. Di conseguenza, per la propriet à di linearità del valore atteso

$$\mathbb{E}[S_{96}] = 96 \cdot 3 = 32 \cdot 9 = 2^5 \cdot 3^2 = 288$$

e per l'indipendenza delle X_i (che implica $Cov(X_i, X_j) = 0$ per ogni $i \neq j$) la varianza della somma è uguale alla somma delle varianze

$$Var(S_{96}) = 96 \cdot 6 = 2\mathbb{E}(S_{96}) = 2^6 \cdot 3^2 = (8 \cdot 3)^2 = 24^2$$

Inoltre sappiamo che la somma di n variabili aleatorie indipendenti $Geom(p)$ ha la stessa distribuzione del tempo di n-simo successo in prove ripetute e indipendenti, ossia ha distribuzione di Pascal di parametri n e p, ossia

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

per $n = 96$ e $p = 1/3$, ossia

$$\mathbb{P}(S_{96} = k) = \binom{k-1}{95} \left(\frac{1}{3}\right)^{96} \left(\frac{2}{3}\right)^{k-96}, \quad k \geq 96.$$

(b) Calcolare approssimativamente $\mathbb{P}(S_{96} \leq 324)$.

Di nuovo per il TCL si ha

$$\mathbb{P}(S_{96} \leq 324) = \mathbb{P}\left(\frac{S_{96} - \mathbb{E}(S_n)}{\sqrt{S_n}} \leq \frac{324 - \mathbb{E}(S_{96})}{\sqrt{Var(S_{96})}}\right) = \mathbb{P}\left(S_{96}^* \leq \frac{324 - 288}{\sqrt{24^2}}\right) \approx \Phi\left(\frac{36}{24}\right) \approx \Phi(1,5) \approx 0,9332$$

(c) Calcolare approssimativamente $\mathbb{P}(S_n > 252)$.

$$\mathbb{P}(S_{96} > 252) = \mathbb{P}\left(\frac{S_{96} - \mathbb{E}(S_{96})}{\sqrt{Var(S_{96})}} > \frac{252 - \mathbb{E}(S_{96})}{\sqrt{Var(S_{96})}}\right) = \mathbb{P}\left(S_{96}^* > \frac{252 - 288}{\sqrt{24^2}}\right) \approx 1 - \Phi\left(-\frac{36}{24}\right) = 1 - \Phi(-1,5) = \Phi(1,5) \approx 0,9332$$

(d) Calcolare approssimativamente $\mathbb{P}(252 < S_n \leq 324)$.

$$\begin{aligned}\mathbb{P}(252 < S_n \leq 324) &= \mathbb{P}\Big(-1,5 < S_{96}^* \leq 1,5\Big) \approx \Phi(1,5) - \Phi(-1,5) = \Phi(1,5) - \big(1 - \Phi(1,5)\big)\\ &= 2 \cdot 0,9332 - 1 = 0,8664\end{aligned}$$