1 Variabili aleatorie in casi più generali: indipendenza, LGN e TCL.

1.1 Variabili aleatorie indipendenti

Molte delle definizioni e delle proprietà delle variabili aleatorie in spazi finiti valgono anche per le variabili aleatorie generali. Ad esempio si ha ancora che il valore atteso della somma di variabili aleatorie è la somma dei valori attesi e la regola per il calcolo della varianza della somma rimane identica.

In questo paragrafo ci chiediamo come si deve definire l'indipendenza per due variabili aleatorie X ed Y, nel caso generale, e daremo anche un'ulteriore definizione di indipendenza globale o completa per più di due variabili aleatorie.

Tra le varie caratterizzazioni di indipendenza, sicuramente non possiamo generalizzare[1] quella per cui $P(X = x, Y = y) = P(X = x)P(Y = y)$, in quanto, ad esempio, per le variabili aleatorie con funzione di distribuzione continua, la precedente relazione sarebbe solo una banalità: infatti si ridurrebbe alla relazione[2] $0 = 0$. Possiamo invece generalizzare quella data in ***Proposizione 1*** della Lezione 8, nel seguente modo.

Definizione 1.1 (Indipendenza di due variabili aleatorie). *Due variabili aleatorie X ed Y si dicono* **indipendenti** *se e solo se comunque scelti due intervalli I e J, limitati o illimitati,*

$$P(X \in I, Y \in J) = P(X \in I) \cdot P(Y \in J).$$

Come nel caso discreto, anche nel caso generale vale il risultato che l'indipendenza di due variabili aleatorie implica[3] la non correlazione, mentre non è vero il viceversa.

Strettamente collegata alla precedente definizione, c'è la seguente

Definizione 1.2 (indipendenza a due a due di n variabili aleatorie). *Siano X_1, X_2, ..., X_n n variabili aleatorie definite tutte sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, P)$. Esse si dicono* **indipendenti a due a due** *se comunque scelti $i \neq j$, con $i, j \in \{1, 2, \ldots, n\}$, le due variabili aleatorie X_i ed X_j sono indipendenti, ovvero comunque scelti I, J, intervalli (limitati o illimitati) di $\mathbb{R}$, si ha:*

$$P(X_i \in I,\, X_j \in J) = P(X_i \in I) \cdot P(X_j \in J).$$

Una condizione più forte dell'indipendenza a due a due è l'indipendenza globale

Un caso particolarmente interessante è quello in cui le variabili aleatorie X_i sono indipendenti (globalmente o completamente) tra loro, ovvero

Definizione 1.3 (indipendenza di n variabili aleatorie). *Siano X_1, X_2, ..., X_n n variabili aleatorie definite tutte sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, P)$. Esse si dicono* **indipendenti tra loro** *se comunque scelti J_1, J_2, ..., J_n, intervalli (limitati o illimitati) di $\mathbb{R}$, si ha:*

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

La precedente definizione implica l'indipendenza a due a due.

Proposizione 1 Se le n variabili aleatorie X_1, X_2, ..., X_n sono indipendenti (globalmente o completamente) fra loro, allora lo sono anche a due a due.

---

[1] Tuttavia nel caso delle variabili aleatorie discrete questa caratterizzazione rimane valida, infatti le dimostrazioni della equivalenza delle caratterizzazioni rimangono sostanzialmente invariate, pur di sostituire a somme finite somme infinite, per cui ad esempio due variabili aleatorie X ed Y con $X(\Omega) = \{x_1, x_2, \ldots\}$ ed $Y(\Omega) = \{y_1, y_2, \ldots\}$ sono indipendenti se e solo se $P(X = x_h, Y = y_k) = P(X = x_k)P(Y = y_h)$ per ogni h e k.

[2] Se $P(X = x) = 0$ per ogni $x \in \mathbb{R}$, allora anche $P(X = x, Y = y) = 0$, in quanto $\{X = x, Y = y\} \subseteq \{X = x\}$.

[3] Ovviamente è necessario che le variabili aleatorie ammettano valore atteso finito.

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

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

e quindi

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

Quanto visto per le variabili aleatorie discrete vale anche per le variabili aleatorie in generale, In particolare se le variabili aleatorie sono indipendenti a due a due, allora la varianza della somma è la somma delle varianze. Alla luce della precedente proposizione, lo stesso vale nel caso in cui le variabili aleatorie sono (globalmente o completamente) indipendenti tra loro.

1.1.1 Legge dei Grandi Numeri

Il risultato più importante di questa Lezione è noto come la ***Legge debole dei grandi numeri***. Tale risultato è enunciato alla fine (***Proposizione 2***) e riguarda le successioni di variabili aleatorie indipendenti a due a due.

Prima di arrivare ad enunciare e dimostrare la legge debole dei grandi numeri, riprendiamo quanto visto utilizzando la diseguaglianza di Chebyshev nella ***Proposizione 9*** della Lezione 10, ma allargando un poco la prospettiva.

Prima di tutto va detto che la diseguaglianza di Chebyshev continua a valere anche nel caso di variabili aleatorie generali, con l'unica accortezza che ***nel caso generale bisogna ipotizzare che esistano finiti valore atteso e varianza***[4] ***della variabile aleatoria*** X. Per cui, indicando come al solito $\mu = \mathbb{E}(X)$ e $\sigma^2 = Var(X)$ si ha

$$P(\{|X - \mu| > \varepsilon\}) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Se X_1, X_2, ..., X_n sono variabili aleatorie indipendenti a due a due, e con la stessa distribuzione, nell'***Osservazione 6*** della Lezione 10 abbiamo visto come trovare il numero n di prove per cui la probabilità dell'evento "*il valore atteso* μ *e la media aritmetica* $Y_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ *differiscono meno di una quantità prefissata* ε sia almeno $1 - \delta$", e nell'Esempio 10.6 ciò è stato applicato al caso di variabili binarie.

Anche qui la ***Proposizione 9*** continua a valere, ***pur di assumere che esistano finiti valore atteso*** $\mathbb{E}(X_i)$ ***e varianza*** $Var(X_i)$, che come al solito poniamo uguali rispettivamente a μ e σ^2.

In questi casi sappiamo che, se

$$n \geq \frac{\sigma^2}{\delta\, \varepsilon^2} \tag{1}$$

allora

$$P(\{|Y_n - \mu| > \varepsilon\}) \leq \frac{1}{n}\frac{\sigma^2}{\varepsilon^2} \leq \delta \quad \Leftrightarrow \quad P(\{-\varepsilon \leq Y_n - \mu \leq \varepsilon\}) \geq 1 - \delta.$$

Nel caso particolare in cui le variabili X_i siano variabili binarie, con $P(X_i = 1) = \theta$ e $P(X_i = 0) = 1 - \theta$, allora $\mu = \theta$, $\sigma^2 = \theta(1 - \theta)$ e basta prendere $n \geq \frac{\theta(1-\theta)}{\delta\, \varepsilon^2}$ per ottenere che la probabilità dell'evento "*la frequenza relativa dei successi*[5] Y_n *differisce dalla probabilità di*

[4] Sappiamo che possono esistere variabili aleatorie per le quali valore atteso e/o varianza non esistono, o valgono infinito. Questo problema non si pone nel caso finito in quanto in quel caso il calcolo del valore atteso e della varianza si riduce ad una somma finita e non presenta quindi nessun tipo di problema.

[5] Successo all'i-esima prova significa $X_i = 1$.

successo θ meno di ε" sia maggiore di $1-\delta$.

Ovvero

$$n \geq \frac{\theta(1-\theta)}{\delta\,\varepsilon^2} \quad \Rightarrow \quad P(\{-\varepsilon \leq Y_n - \theta \leq \varepsilon\}) \geq 1-\delta. \tag{2}$$

Nelle applicazioni si usa la frequenza relativa per "*stimare*" la probabilità θ: ovvero possiamo considerare il caso in cui possiamo osservare gli esiti di diversi esperimenti di uno stesso fenomeno, gli esperimenti sono condotti nelle stesse condizioni, per cui la probabilità di successo dell'esperimento è la stessa in tutte le prove, e infine ***si assume che le prove siano stocasticamente indipendenti tra loro***, tuttavia ***non si assume che sia noto esattamente il valore della probabilità di successo*** θ.

In questo contesto la misura di probabilità dipende dal parametro θ ed è quindi più opportuno indicarla con P_θ, invece che con P.

Riprendendo quanto detto nell'***Osservazione 6*** della Lezione 10 in questo contesto possiamo riscrivere

$$P_\theta(\{\theta-\varepsilon \leq Y_n \leq \theta+\varepsilon\}) \geq 1 - \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2},$$

ma anche

$$P_\theta(\{Y_n-\varepsilon \leq \theta \leq Y_n+\varepsilon\}) \geq 1 - \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2}.$$

Questa secondo modo di scrivere è più interessante, in quanto, in questo contesto, mentre possiamo osservare Y_n, invece non conosciamo affatto θ. L'idea è che vorremmo poter "*valutare*" la probabilità θ con Y_n, con un errore al più di ε. Ovviamente in nessun caso, facendo degli esperimenti, avremo la garanzia che la frequenza relativa Y_n e la probabilità di successo θ differiscano meno di ε, tuttavia la disuguaglianza di Chebyshev ci permette di affermare che ciò accade con probabilità elevata, e permette anche di trovare delle limitazioni inferiori a tale probabilità.

A prima vista però sorge una difficoltà: sembra che per adoperare la disuguaglianza di Chebyshev sia necessario conoscere θ. Ma a questo problema si può ovviare osservando che la funzione $h(x) = x(1-x)$ vale al massimo[6] $\frac{1}{4}$ e quindi si ha

$$P_\theta(|Y_n-\theta| > \varepsilon\}) \leq \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2} \leq \frac{1}{4\,\varepsilon^2\,n},$$

$$\Updownarrow$$

$$P_\theta(\{Y_n-\varepsilon \leq \theta \leq Y_n+\varepsilon\}) \geq 1 - \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2} \geq 1 - \frac{1}{4\,\varepsilon^2\,n}.$$

Ciò permette di affermare che, ***qualunque sia la probabilità di successo*** θ, la probabilità che θ e la frequenza relativa Y_n differiscano meno di ε vale almeno $1-\frac{1}{4\,\varepsilon^2\,n}$.

Più interessante ancora, dal punto di vista operativo, è tuttavia il fatto che siamo in grado di rispondere ***alla domanda***:

[6] La funzione $h(x) = x(1-x)$ ha il suo punto di massimo in $x = \frac{1}{2}$ come si vede subito, e quindi $h(x) \leq h(\frac{1}{2}) = \frac{1}{4}$.

Quante prove si devono effettuare, ovvero quanto si deve prendere grande n***, affinché, con probabilità almeno*** $1-\delta$***, la frequenza relativa differisca dalla probabilità di successo meno di*** ε***?***

La risposta alla precedente domanda è molto semplice: ***è sufficiente prendere***[7]

$$n \geq \frac{1}{4\,\delta\,\varepsilon^2},$$

in altre parole

$$n \geq \frac{1}{4\,\delta\,\varepsilon^2} \quad \Rightarrow \quad P(\{-\varepsilon \leq Y_n - \theta \leq \varepsilon\}) \geq 1-\delta. \tag{3}$$

Infatti in tale caso $\delta \geq \frac{1}{4\,\varepsilon^2\,n}$ e quindi, qualunque sia θ

$$P_\theta(|Y_n - \theta| > \varepsilon\}) \leq \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2} \leq \frac{1}{4\,\varepsilon^2\,n} \leq \delta,$$

$$\Updownarrow$$

$$P_\theta(\{Y_n - \varepsilon \leq \theta \leq Y_n + \varepsilon\}) \geq 1 - \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2} \geq 1 - \frac{1}{4\,\varepsilon^2\,n} \geq 1-\delta.$$

Esempio 1.1. *Sia* Y_n *la frequenza relativa dei successi in uno schema di Bernoulli di parametro* θ*. Si determini un* n *in modo che, qualunque sia il valore di* θ*, l'errore assoluto tra* Y_n *e* θ *sia minore di* 0.1*, con probabilita' almeno* 0.99*.*

Soluzione Siamo nel caso precedente con $\varepsilon = 0.1 = \frac{1}{10}$ e con $1-\delta = 0.99$, ovvero $\delta = \frac{1}{100}$. Quindi se

$$n \geq \frac{1}{4 \cdot \left(\frac{1}{10}\right)^2 \cdot \frac{1}{100}} = \frac{10000}{4} = 2500,$$

allora

$$P(-\varepsilon \leq Y_n - \theta \leq \varepsilon) \geq 0.99$$

E quindi, qualunque sia il valore di θ, è sufficiente prendere $n = 2500$.

Esempio 1.2. *Calcolare il minimo valore di* n *per il quale, in uno schema di Bernoulli di con probabilità* θ*,* **in base alla disuguaglianza di Chebyshev***, si possa scrivere*

$$P\left(\left\{\left|\frac{S_n}{n} - \theta\right| > \frac{1}{30}\right\}\right) \leq \frac{1}{10},$$

qualunque sia il valore di θ*.*

---

[7] Si deve prendere

$$n \geq n(\varepsilon, \delta)$$

dove

$$n(\varepsilon,\delta) := \left\lceil \frac{1}{4\,\varepsilon^2\,\delta} \right\rceil,$$

cioè la parte intera superiore di $\frac{1}{4\,\varepsilon^2\,\delta}$. Si ricordi che la parte intera superiore di un numero reale a è l'intero k tale che $k-1 < a \leq k$, ed è indicata appunto con $\lceil a \rceil$.

Soluzione Si può procedere considerando che

$$P\left(\left\{\left|\frac{S_n}{n}-\theta\right|>\frac{1}{30}\right\}\right)\leq\frac{\theta(1-\theta)}{n\,\left(\frac{1}{30}\right)^2}\leq\frac{1}{4\,n\,\left(\frac{1}{30}\right)^2}=\frac{900}{4\,n}\leq\frac{1}{10},$$

$$\Updownarrow$$

$$\frac{900}{4\,\frac{1}{10}}=\frac{9000}{4}=2250\leq n,$$

oppure direttamente utilizzando la formula (3)

$$n\geq\frac{1}{4\,\varepsilon^2\,\delta}=\frac{1}{4\,\left(\frac{1}{30}\right)^2\,\frac{1}{10}}=\frac{900}{4\,\frac{1}{10}}=\frac{9000}{4}=2250.$$

Si suggerisce di confrontare il risultato con quello dell'Esempio 10.6 in cui invece il valore di θ era calcolabile, e quindi si era ottenuto, sempre utilizzando la disuguaglianza di Chebyshev[8], che bastava prendere $n = 2223$.

Osservazione 1 Si faccia attenzione al fatto che queste limitazioni inferiori sono date in base alla disuguaglianza di Chebyshev. I valori ottenuti per n sono sicuramente validi, ma sono eccessivamente grandi ed in genere piu' elevati del necessario. In realtà bastano valori di n più piccoli (daremo un'idea del motivo per cui i valori trovati sono eccessivi nella Lezione sul Teorema centrale del limite).

Osservazione 2 **(Errore relativo)** Va anche sottolineato che finora abbiamo valutato solo l'errore assoluto, tra Y_n e θ, mentre avrebbe più interesse l'errore relativo, ovvero $\left|\frac{Y_n-\theta}{\theta}\right|$: infatti se θ fosse dell'ordine di un centesimo, stimare θ con un errore assoluto dell'ordine di un decimo non sarebbe molto ragionevole[9]. In questo caso la maggiorazione della disuguaglianza di Chebyshev permette di affermare che, per ogni θ

$$P_\theta\left(\left|\frac{Y_n-\theta}{\theta}\right|>\varepsilon\}\right)=P_\theta\left(|Y_n-\theta|>\varepsilon\theta\}\right)\leq\frac{1}{n}\frac{\theta(1-\theta)}{(\theta\varepsilon)^2}=\frac{1-\theta}{\theta}\,\frac{1}{n\,\varepsilon^2}\leq\delta,$$

per cui

$$n\geq\frac{1-\theta}{\theta}\frac{1}{\delta\,\varepsilon^2}\quad\Rightarrow\quad P_\theta\left(\left|\frac{Y_n-\theta}{\theta}\right|>\varepsilon\}\right)\leq\delta$$

Purtroppo, se θ non è noto, questa limitazione inferiore non è molto utile in quanto la funzione $h_1(x)=\frac{1-x}{x}=\frac{1}{x}-1$ converge ad infinito per $x\to 0^+$, ed è quindi impossibile[10] trovare un valore di n che sia valido qualunque sia θ.

Infine, nel formulare la domanda con la richiesta di scegliere n, c'è un punto che abbiamo volutamente trascurato fin qui. La possibilita' di scegliere n presuppone di avere a disposizione un numero di eventi indipendenti potenzialmente grande a piacere[11].

---

[8] In realtà nell'Esempio citato si è utilizzata la (2).

[9] Se nel misurare la distanza fra due città si commette un errore dell'ordine di un metro, ci possiamo dichiarare completamente soddisfatti, mentre certamente non lo saremmo se l'errore dell'ordine di un metro riguardasse la misura di un tavolo da mettere in cucina!!!

[10] Diverso è il caso in cui, pur non conoscendo esattamente θ si sappia che $\theta\geq\theta_0$ con $\theta_0>0$: allora basterà prendere $n\geq\frac{1-\theta_0}{\theta_0}\,\frac{1}{\delta\,\varepsilon^2}$.

[11] Si potrebbe ovviare al problema supponendo di avere una successione di spazi di probabilità $(\Omega_n,\mathcal{P}(\Omega_n),P^{(n)})$ e su ciascuno spazio n eventi $E_1^{(n)}$, $E_2^{(n)}$,..., $E_n^{(n)}$ che formano uno schema di Bernoulli con probabilità $\theta^{(n)}=\theta$ per ogni n.

Dal punto di vista matematico è più comodo poter affermare direttamente di avere a disposizione una successione di eventi indipendenti e tutti con la stessa probabilità θ. Ciò presuppone uno spazio di probabilità Ω infinito, e quindi solo dopo aver introdotto gli spazi di probabilità generali e la nozione di successioni di variabili aleatorie, riformuliamo la ***Proposizione 9*** per successioni di variabili aleatorie. Tale formulazione è nota con il nome di Legge Debole dei Grandi Numeri.

Proposizione 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[12], per le quali esistano finiti valore atteso e varianza. Posto $\mathbb{E}(X_i) = \mu$ e $Var(X_i) = \sigma^2$, $S_n = \sum_{i=1}^{n} X_i$ e $Y_n = \frac{S_n}{n}$, si ha, qualunque sia $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = \lim_{n\to\infty} P\left(|Y_n - \mu| > \varepsilon\right) = 0$$

Dimostrazione. Basta osservare che

$$0 \leq P\left(|Y_n - \mu| > \varepsilon\right) \leq \frac{1}{n}\frac{\sigma^2}{\varepsilon^2},$$

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

$$0 \leq \lim_{n\to\infty} P\left(|Y_n - \mu| > \varepsilon\right) \leq \lim_{n\to\infty} \frac{1}{n}\frac{\sigma^2}{\varepsilon^2} = 0.$$

Osservazione ?? Dalla ***Proposizione 1*** appare immediato che se $\{X_i,\ i \geq 1\}$ è una successione di variabili aleatorie completamente indipendenti, allora la Legge Debole dei Grandi Numeri continua a valere. Sotto questa ulteriore ipotesi vale anche il così detto Teorema centrale del limite che è oggetto del prossimo paragrafo. Nel prossimo paragrafo vedremo anche alcune relazioni tra questi due importantissimi risultati.

1.2 Teorema centrale del limite

Come abbiamo detto la disuguaglianza di Chebyshev permette di trovare delle limitazioni inferiori alle probabilità del tipo

$$P\left(\left|\frac{S_n}{n} - \mu\right| \leq \varepsilon\right)$$

che a loro volta permettono di dedurre la legge dei grandi numeri. Tuttavia se si conoscesse la funzione di distribuzione $F_{S_n}(x)$ della variabile aleatoria S_n, tale probabilità si potrebbe calcolare esattamente come

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - \mu\right| \leq \varepsilon\right) &= P\left(n(\mu-\varepsilon) \leq S_n \leq n(\mu+\varepsilon)\right) = F_{S_n}\left(n(\mu-\varepsilon)\right) - F_{S_n}\left(-\left(n(\mu-\varepsilon)\right)^-\right) \\ &= F_{S_n}\left(n(\mu-\varepsilon)\right) - F_{S_n}\left(-n(\mu-\varepsilon)\right) + P\big(\{S_n = -n(\mu-\varepsilon)\}\big) \end{aligned}$$

Appare quindi chiaro che calcolare la distribuzione della somma di variabili aleatorie $S_n = X_1 + X_2 + \ldots + X_n$ sia un problema interessante è, oltre che di per sé, anche per le connessioni con la legge dei grandi numeri e delle relazioni tra media aritmetica e valore atteso.

Sappiamo calcolare esattamente la distribuzione della somma di variabili aleatorie in alcuni casi specifici. Ad esempio quando le X_i sono le indicatrici di eventi E_i che formano uno schema di Bernoulli di parametro θ, sappiamo che la distribuzione della somma è la distribuzione binomiale $b(n;\theta)$.

[12]Poiché le variabili aleatorie X_n hanno tutte la stessa distribuzione, si ha che se esistono finiti valore atteso e varianza di X_1, allora esistono finiti valore atteso e varianza di X_i e coincidono con quelli di X_1.

Esempio 1.3. *Ancora sappiamo che se due variabili aleatorie X_1 ed X_2 sono indipendenti e hanno distribuzione binomiale di parametri n_i e θ (attenzione n_1 può essere diverso da n_2, ma θ è lo stesso per $i = 1, 2$), allora la somma $X_1 + X_2$ ha distribuzione binomiale di parametri $n_1 + n_2$ e θ (confrontare lo svolgimento dell'Esercizio 8.4). Questo risultato si estende anche al caso di n variabili aleatorie (globalmente) indipendenti tra loro: in particolare se le variabili aleatorie X_i hanno tutte la stessa distribuzione $bin(m; \theta)$ allora S_n ha distribuzione $bin(n \cdot m; \theta)$.*

Esempio 1.4. *Siano X_1 ed X_2 variabili aleatorie di Poisson di parametro $\lambda_1 (> 0)$ e $\lambda_2 (> 0)$ rispettivamente, ovvero*

$$P(X_i = k) = \frac{\lambda_i^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

Si assuma che le variabili siano indipendenti, ovvero che

$$P(X_1 = h, X_2 = k) = P(X_1 = h) P(X_2 = k), \qquad \textit{per ogni } h,\ k \in \{0,\ 1\ \ldots\}$$

Si vede facilmente che la variabile aleatoria $X_1 + X_2$ ha distribuzione di Poisson di parametro $\lambda_1 + \lambda_2$, ovvero che

$$P(X_1 + X_2 = m) = \frac{(\lambda_1 + \lambda_2)^m}{m!}, \qquad m = 0, 1, 2, \ldots$$

Infatti, per $m = 0, 1, \ldots$ l'evento

$$\{X_1 + X_2 = m\} = \bigcup_{k=0}^{\infty} \{X_1 = k,\ X_2 = m - k\} = \bigcup_{k=0}^{m} \{X_1 = k,\ X_2 = m - k\},$$

in quanto $\{X_2 = m - k\} = \emptyset$ per $k = m + 1, m + 2, \ldots$. Per cui

$$\begin{aligned} P(X_1 + X_2 = m) &= \sum_{k=0}^{m} P(X_1 = k,\ X_2 = m - k) = \sum_{k=0}^{m} P(X_1 = k) P(X_2 = m - k) \\ &= \sum_{k=0}^{m} \frac{\lambda_1^k}{k!} \frac{\lambda_2^{(m-k)}}{(m-k)!} = \frac{1}{m!} \sum_{k=0}^{m} m! \frac{1}{k!} \frac{1}{(m-k)!} \lambda_1^k \lambda_2^{(m-k)} \end{aligned}$$

dalla formula della potenza del binomio si ottiene la tesi:

$$\frac{1}{m!} \sum_{k=0}^{m} \binom{m}{k} \lambda_1^k \lambda_2^{(m-k)} = \frac{(\lambda_1 + \lambda_2)^m}{m!}, \qquad m = 0, 1, 2, \ldots$$

Più complesso risulta il caso della somma per altre variabili aleatorie, tuttavia si puo' innanzi tutto osservare come calcolare la funzione di distribuzione di S_n sia equivalente a calcolare la distribuzione di una sua trasformata ovvero:
se a_n e b_n sono numeri reali, con $b_n > 0$, allora[13]

$$\{S_n \le x\} = \left\{ \frac{S_n - a_n}{b_n} \le \frac{x - a_n}{b_n} \right\}$$

Una scelta naturale per a_n e per b_n è quella che trasforma S_n in una variabile aleatoria standard, ovvero quella di prendere $a_n = \mathbb{E}(S_n)$ e $b_n = \sqrt{Var(S_n)}$.

[13]Infatti

$$\omega \in \{S_n \le x\} \Leftrightarrow S_n(\omega) \le x \Leftrightarrow \frac{S_n(\omega) - a_n}{b_n} \le \frac{x - a_n}{b_n} \Leftrightarrow \omega \in \left\{ \frac{S_n - a_n}{b_n} \le \frac{x - a_n}{b_n} \right\}$$

In questo modo infatti, per la disuguaglianza di Chebyshev, sappiamo che

$$P(-\alpha \leq \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \leq \alpha) \geq 1 - \frac{1}{\alpha}.$$

Alla luce della seguente ***Proposizione 3***, nota come Teorema Centrale del Limite (o anche Teorema del Limite Centrale), si può dimostrare che se le variabili aleatorie sono (globalmente o completamente) indipendenti, hanno la stessa distribuzione, valore atteso finito μ, varianza finita σ^2 e non nulla, per cui $\mathbb{E}(S_n) = n\mu$ e $Var(S_n) = n\sigma^2 > 0$, allora

$$F_{S_n}(x) = P(S_n \leq x) = P\left(\frac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \leq \frac{x - n\mu}{\sqrt{n\,\sigma^2}}\right) \simeq \Phi\left(\frac{x - n\mu}{\sqrt{n\,\sigma^2}}\right), \tag{4}$$

dove $\Phi(x)$ è la funzione di distribuzione di una variabile aleatoria gaussiana standard.

La dimostrazione della precedente affermazione si basa sul seguente risultato basilare e che svolge un ruolo "centrale" nel Calcolo delle Probabilità.

***Proposizione 3* (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 $\mathbb{E}(X_i) = \mu$ e $Var(X_i) = \sigma^2$, si assuma che $\sigma^2 > 0$. Allora indicando con S_n^* variabile aleatoria standardizzata di S_n, si ha

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

e, indicando con $F_{S_n^*}(x)$ la funzione di distribuzione di S_n^*, si ha

$$\lim_{n\to\infty} F_{S_n^*}(x) = \lim_{n\to\infty} P\left(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \leq x\right) = \Phi(x), \tag{6}$$

dove Φ è la funzione di distribuzione di una variabile aleatoria Gaussiana standard: in altre parole

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{y^2}{2}}\, dy. \tag{7}$$

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

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

Non diamo la dimostrazione di questo risultato, ma notiamo solo che la (5) si dimostra tenendo conto che $\mathbb{E}(S_n) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n\mu$ e che per l'indipendenza dalle variabili aleatorie X_i, si ha $Var(S_n) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) = n\sigma^2$. La precedente relazione sarebbe valida anche nel caso in cui le variabili aleatorie fossero solo indipendenti a due a due(o addirittura solo non correlate), ma <u>sottolineiamo</u> il fatto che, mentre la Legge Debole dei Grandi Numeri, vale sotto l'ipotesi di indipendenza a due a due, e non è necessario supporre $\sigma^2 > 0$, invece ***per il Teorema Centrale del Limite, serve la condizione di indipendenza globale*** e ovviamente ***è necessario supporre*** $\sigma^2 > 0$, altrimenti non si potrebbe nemmeno formulare la tesi.

Fondamentale per dimostrare l'approssimazione (4) della funzione di distribuzione della somma S_n è il fatto che la convergenza sia uniforme: infatti, posto $E_n(x) = F_{S_n^*}(x) - \Phi(x)$, e $x_n = \frac{x-n\mu}{\sqrt{n\,\sigma^2}}$, si ha

$$F_{S_n}(x) = P(S_n \le x) = P\left(\frac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \le \frac{x-n\mu}{\sqrt{n\,\sigma^2}}\right) = F_{S_n^*}(x_n) = \Phi(x_n) + E_n(x_n),$$

per cui

$$|F_{S_n}(x) - \Phi(x_n)| = |E_n(x_n)| \le \sup_{x\in\mathbb{R}} |E_n(x)|.$$

Basta solo osservare che (7) garantisce che $\sup_{x\in\mathbb{R}} |E_n(x)|$ converge a zero[14] per n che tende all'infinito.

Si osservi che il Teorema Centrale del Limite implica che

$$\lim_{n\to\infty} P(a < \frac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \le b) = \Phi(b) - \Phi(a),$$

come si vede subito applicando la proprietà che per ogni variabile aleatoria X, con funzione di distribuzione $F(x)$, si ha $P(a < X \le b) = F(b) - F(a)$.

Il Teorema Centrale del Limite implica anche che

$$\lim_{n\to\infty} P\left(a \le \frac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \le b\right) = \Phi(b) - \Phi(a),$$

infatti, come si vede facilmente[15],

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n\,\sigma^2}} = a\right) = 0.$$

Dopo questa osservazione possiamo tornare indietro alle relazioni tra Legge dei Grandi Numeri e Teorema Centrale del Limite.

Indicando, come al solito, con Y_n la media aritmetica $\frac{S_n}{n}$, si ha

$$Y_n - \mu = \frac{S_n - n\mu}{n},$$

[14] Pur essendo assolutamente al di fuori dell'ambito di un corso elementare di probabilità, vale la pena di ricordare che esistono delle maggiorazioni per $\sup_{x\in\mathbb{R}} |E_n(x)|$, nel caso in cui si supponga che il valore atteso $\mathbb{E}(|X|^3)$ esista e sia finito. In particolare è stato dimostrato che

$$\sup_{x\in\mathbb{R}} |E_n(x)| \le \frac{C}{n} \frac{\mathbb{E}(|X|^3)}{\sigma^3},$$

con C costante. Il valore di C non è noto esattamente ma è noto che $0.4097 \le C \le 0.7975$, in particolare quindi vale

$$\sup_{x\in\mathbb{R}} |E_n(x)| \le \frac{1}{n} \frac{\mathbb{E}(|X|^3)}{\sigma^3},$$

I primi a fornire maggiorazioni in questa direzione sono stati Berry ed Eessen all'inizio degli anni 40 dello scorso XX secolo.

[15] Si osservi che

$$P\left(\frac{S_n - n\mu}{\sqrt{n\,\sigma^2}} = a\right) \le P\left(a - \frac{1}{n} \le \frac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \le a\right) = \Phi(a) - \Phi(a - \frac{1}{n}) \underset{n\to\infty}{\longrightarrow} 0$$

$$\sqrt{\frac{n}{\sigma^2}}\,\frac{S_n - n\mu}{n} = \frac{S_n - n\mu}{\sqrt{n\,\sigma^2}}$$

e quindi

$$\{|Y_n - \mu| \le \varepsilon\} = \left\{-\varepsilon \le \frac{S_n - n\mu}{n} \le \varepsilon\right\} = \left\{-\sqrt{\frac{n}{\sigma^2}}\,\varepsilon \le \frac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \le \sqrt{\frac{n}{\sigma^2}}\,\varepsilon\right\}.$$

Di conseguenza

$$P(\{|Y_n - \mu| \le \varepsilon\}) = P\left(-\varepsilon \le \frac{S_n - n\mu}{n} \le \varepsilon\right) = P\left(-\sqrt{\frac{n}{\sigma^2}}\,\varepsilon \le \frac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \le \sqrt{\frac{n}{\sigma^2}}\,\varepsilon\right)$$
$$\simeq \Phi\left(\sqrt{\frac{n}{\sigma^2}}\,\varepsilon\right) - \Phi\left(-\sqrt{\frac{n}{\sigma^2}}\,\varepsilon\right) = 2\Phi\left(\sqrt{\frac{n}{\sigma^2}}\,\varepsilon\right) - 1.$$

Esempio 1.5. *Sia X_1 una variabile aleatoria che può assumere i valori* 0, $\frac{1}{2}$, 1, $\frac{3}{2}$ *e con*

$$p_{X_1}(0) = P(X_1 = 0) = \frac{1}{10}, \qquad p_{X_1}(\frac{1}{2}) = P(X_1 = \frac{1}{2}) = \frac{1}{10},$$

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

Si ponga il valore atteso di X_1 uguale a μ e la sua varianza uguale a σ^2.

Siano X_1, X_2, $X_3, \ldots, X_{100}$ delle variabili aleatorie con la stessa distribuzione di X_1 e (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

$$P\left(\left\{\mu - \frac{1}{10} \le Y_{100} \le \mu + \frac{1}{10}\right\}\right),$$

Soluzione Innanzi tutto come si trova facilmente si ha $\mu = \frac{21}{20}$ e $\sigma^2 = \frac{89}{400}$. Quindi la probabilita' cercata è approssimata con

$$P\left(\left\{\mu - \frac{1}{10} \le Y_{100} \le \mu + \frac{1}{10}\right\}\right) \simeq 2\Phi\left(\sqrt{\frac{n}{\sigma^2}}\,\varepsilon\right) - 1 = 2\Phi\left(\sqrt{\frac{100}{\frac{89}{400}}}\,\frac{1}{10}\right) - 1$$
$$= 2\Phi\left(\sqrt{\frac{400}{89}}\right) - 1 = 2\Phi(\frac{20}{\sqrt{89}}) - 1$$
$$\simeq 2\Phi(2,1199) - 1 \simeq 2 \cdot 0.9826 - 1 = 1,9652 - 1 = 0.9652$$

Supponiamo ora di voler rispondere ***in modo approssimato*** alla domanda: ***sia*** n ***fissato, per quale valore di*** $\varepsilon = \varepsilon(n, \delta)$ ***posso affermare che***

$$P(\{|Y_n - \mu| \le \varepsilon\}) \simeq 1 - \delta?$$

Il seguente procedimento non è del tutto rigoroso, perché trascura l'errore di approssimazione E_n tra $F_{S_n^*}$ e Φ. Tuttavia permette di dare una buona valutazione del tipo di comportamento di ε: (trascurando E_n) andiamo a mostrare che $\varepsilon = \varepsilon(n, \delta)$ è un infinitesimo dell'ordine di $\frac{1}{\sqrt{n}}$.

Prima di tutto invece di valutare

$$P(\{|Y_n - \mu| \le \varepsilon\}) \simeq 1 - \delta$$

considerando che, per n sufficientemente grande,

$$P(\{|Y_n - \mu| \leq \varepsilon\}) \simeq 2\Phi\left(\sqrt{\frac{n}{\sigma^2}}\,\varepsilon\right) - 1$$

cerchiamo invece

$$2\Phi(\sqrt{\frac{n}{\sigma^2}}\,\varepsilon) - 1 = 1 - \delta \quad \Leftrightarrow \quad \Phi\left(\sqrt{\frac{n}{\sigma^2}}\,\varepsilon\right) = \frac{2-\delta}{2} = 1 - \frac{\delta}{2}$$

Sicuramente esiste un valore $x_{1-\ delta/2}$ per cui

$$\Phi\left(x_{1-\ delta/2}\right) = 1 - \frac{\delta}{2};$$

Inoltre possiamo trovarer un valore approssimato di $x_{1-\ delta/2}$ utilizzando le tavole della gaussiana.

A questo punto basta porre

$$\sqrt{\frac{n}{\sigma^2}}\,\varepsilon = x_{1-\delta/2} \quad \Leftrightarrow \quad \varepsilon = \varepsilon(n,\delta) = x_{1-\delta/2} \cdot \sqrt{\frac{\sigma^2}{n}} = \frac{x_{1-\delta/2}\cdot\sigma}{\sqrt{n}},$$

per ottenere il risultato desiderato.

Supponiamo ora di voler rispondere ***in modo approssimato*** alla domanda: ***per quali*** n ***posso affermare che***

$$P(\{|Y_n - \mu| \leq \varepsilon\}) \geq 1 - \delta?$$

Anche il seguente procedimento non è del tutto rigoroso, perché trascura l'errore di approssimazione E_n tra $F_{S_n^*}$ e Φ. Tuttavia permette di dare una buona valutazione del tipo di richiesta vada fatta su n per ottenere la limitazione inferiore richiesta.

Anche in questo problema, invece di cercare una limitazione inferiore esatta

$$P(\{|Y_n - \mu| \leq \varepsilon\}) \geq 1 - \delta$$

sempre considerando che, per n sufficientemente grande,

$$P(\{|Y_n - \mu| \leq \varepsilon\}) \simeq 2\Phi\left(\sqrt{\frac{n}{\sigma^2}}\,\varepsilon\right) - 1$$

cerchiamo invece una limitazione inferiore

$$2\Phi(\sqrt{\frac{n}{\sigma^2}}\,\varepsilon) - 1 \geq 1 - \delta \quad \Leftrightarrow \quad \Phi\left(\sqrt{\frac{n}{\sigma^2}}\,\varepsilon\right) \geq \frac{2-\delta}{2} = 1 - \frac{\delta}{2}$$

Come nel caso precedente possiamo trovare un valore $x_{1-\delta/2}$ per cui

$$\Phi\left(x_{1-\delta/2}\right) = 1 - \frac{\delta}{2}.$$

Si osservi che, essendo Φ una funzione non decrescente [16],

$$\Phi\left(x\right) \geq \Phi\left(x_{1-\delta/2}\right) = 1 - \frac{\delta}{2}, \qquad \text{per ogni } x \geq x_{1-\delta/2},$$

[16] In realta' basta trovare sulla tavola della gaussiana standard un valore $\overline{x}_{1-\delta/2}$ tale che

$$\Phi\left(\overline{x}_{1-\delta/2}\right) \geq 1 - \frac{\delta}{2}.$$

Il ragionamento fatto con $x_{1-\delta/2}$ si può ripetere mettendo $\overline{x}_{1-\delta/2}$ al posto di $x_{1-\delta/2}$.

A questo punto basta imporre che

$$\sqrt{\frac{n}{\sigma^2}}\,\varepsilon \geq x_{1-\delta/2} \quad \Leftrightarrow \quad \sqrt{n} \geq \frac{x_{1-\delta/2}\sqrt{\sigma^2}}{\varepsilon} \quad \Leftrightarrow \quad n \geq n_{TCL}(\varepsilon,\delta) := \frac{x^2_{1-\delta/2}\,\sigma^2}{\varepsilon^2} \tag{9}$$

per ottenere il risultato desiderato.

Osservazione ?? Si confrontino tra loro (1) e (9): come si vede (1) e (9) sono molto simili, la seconda si ottiene sostituendo al posto di $\frac{1}{\delta}$, il valore $x^2_{1-\delta/2}$.

Quindi a parità di valori di *varepsilon* e σ^2 si ottiene che la limitazione inferiore con la disuguaglianza di Chebyshev, pur essendo esatta, chiede

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

Per capire quindi la differenza si osservi che se $\delta = 0.01$, allora $\frac{1}{\delta} = 100$, mentre, essendo $x_{1-\delta/2} = x_{1-0.005} = x_0.995 = 2.576$ (come si può trovare dalle tavole) si ha che $x^2_{1-\delta/2} = 6.635776$.

In questo caso

$$n_{Ch}(\varepsilon,\delta) = \frac{1}{\delta\, x^2_{1-\delta/2}} n_{TCL}(\varepsilon,\delta) = \frac{16}{0.01 \cdot 6.63577}\, n_{TCL}(\varepsilon,\delta) \simeq 15.0698\, n_{TCL}(\varepsilon,\delta).$$

Se invece $\delta = 0.001$, allora $\frac{1}{\delta} = 1000$, mentre, essendo $x_{1-\delta/2} = x_{1-0.0005} = x_0.9995 = 3.291$ (come si può trovare dalle tavole) si ha che $x^2_{1-\delta/2} = 10,830681$.

In questo caso

$$n_{Ch}(\varepsilon,\delta) = \frac{x^2_{1-\delta/2}}{\delta} n_{TCL}(\varepsilon,\delta) = \frac{1}{0.001 \cdot 10,830681}\, n_{TCL}(\varepsilon,\delta) \simeq 92,3302\, n_{TCL}(\varepsilon,\delta).$$