1 Variabili Continue come limiti di variabili discrete

Supponiamo che n sia un numero intero grande, e di avere una variabile aleatoria U^n uniforme sull'insieme $\{x_i^{(n)} = \frac{i}{n}, \quad i = 1, 2, \dots, n\}$, ossia

$$P(U^n = \frac{i}{n}) = \frac{1}{n}, \qquad i \in \{1, 2, \dots, n\}$$

e vogliamo calcolare il valore della probabilità che U^n sia in un intervallo, ossia

$$P(a < U^n \leq b), \qquad \text{con } a < b, \tag{$*$}$$

e, per una funzione g continua,

$$E[g(U^n)]. \tag{$**$}$$

Il calcolo di (*) è relativamente semplice, infatti, prima di tutto, per $a < b$,

$$P(a < U^n \leq b) = P(U^n \leq b) - P(U^n \leq a),$$

come segue subito osservando che, per ogni variabile aleatoria X vale

$$\{X \leq b\} = \{a < X \leq b\} \cup \{X \leq a\}, \quad \text{da cui} \quad P(X \leq b) = P(a < X \leq b) + P(X \leq a).$$

Per il calcolo della probabilità che X sia nell'intervallo $(a, b]$, per ogni a, b, con $a < b$, basta quindi conoscere la funzione

$$x \mapsto F_X(x) := P(X \leq x).$$

La funzione $F_X : \mathbb{R} \to [0, 1]$, così definita ha un ruolo molto importante nella definizione delle variabili aleatorie continue (e non solo continue) viene detta ***funzione di distribuzione*** (o ***di ripartizione***) della variabile aleatoria X.

Nel nostro caso $\{U^n \leq b\} = \{a < U^n \leq b\} \cup \{U^n \leq a\}$ e $P(U^n \leq b) = P(a < U^n \leq b) + P(U^n \leq a)$, quindi, invece di calcolare direttamente

$$P(a < U^n \leq b) = \frac{\#\{i : \ 0 < i \leq n, \ a < \frac{i}{n} \leq b\}}{n} = \frac{\#\{i : \ 0 < i \leq n, \ an < i \leq bn\}}{n}$$

si può anche più semplicemente osservare che

$$\begin{aligned} P(a < U^n \leq b) &= P(U^n \leq b) - P(U^n \leq a) \\ &= \frac{\#\{i : \ 0 < i \leq n, \ \frac{i}{n} \leq b\}}{n} - \frac{\#\{i : \ 0 < i \leq n, \ \frac{i}{n} \leq a\}}{n} \\ &= \frac{\#\{i : \ 0 < i \leq n, \ i \leq bn\}}{n} - \frac{\#\{i : \ 0 < i \leq n, \ i \leq an\}}{n}. \end{aligned}$$

Occupiamoci quindi, al variare di x, della probabilità di

$$P(U^n \leq x) = \frac{\#\{i : \ 0 < i \leq n, \ \frac{i}{n} \leq x\}}{n} = \frac{\#\{i : \ 0 < i \leq n, \quad i \leq xn\}}{n} \tag{$\sharp$}$$

Tale probabilità è chiaramente nulla per $x \leq 0$ ed è chiaramente uguale ad 1, quando invece $x \geq 1$, in formule

$$\begin{aligned} P(U^n \leq x) &= 0 \qquad \text{per } x \leq 0, \\ P(U^n \leq x) &= 1 \qquad \text{per } x \geq 1, \end{aligned}$$

Il caso interessante è quindi il caso in cui $x \in (0,1)$. Per tali valori di x, dall'espressione ($\sharp$) si vede subito[1] che, denotando con $\lfloor \alpha \rfloor$ la parte intera inferiore di α,

$$P(U^n \le x) = \frac{\#\{i:\ 0 < i \le xn\}}{n} = \frac{\#\{i:\ 0 < i \le \lfloor xn \rfloor\}}{n} = \frac{\lfloor xn \rfloor}{n}, \qquad \text{per } x \in (0,1).$$

Abbiamo detto che ci interessa il calcolo per n grande, e quindi vogliamo vedere se, per n che tende ad infinito, c'è un limite per tale probabilità. Si vede facilmente[2] che

$$\lim_{n\to\infty} P(U^n \le x) = \lim_{n\to\infty} \frac{\lfloor xn \rfloor}{n} = x, \qquad \text{per } x \in (0,1).$$

Inoltre, vale chiaramente che

$$\lim_{n\to\infty} P(U^n \le x) = \lim_{n\to\infty} 0 = 0, \qquad \text{per } x \le 0.$$

e

$$\lim_{n\to\infty} P(U^n \le x) = \lim_{n\to\infty} 1 = 1, \qquad \text{per } x \ge 1.$$

Anche il calcolo di (**) è relativamente semplice, infatti,

$$E[g(U^n)] = \sum_{i=1}^{n} g(\tfrac{i}{n}) P(U^n = \tfrac{i}{n}) = \sum_{i=1}^{n} g(\tfrac{i}{n})\, \frac{1}{n}.$$

Anche qui cerchiamo l'espressione per n grande ossia vogliamo vedere se esiste finito (e quanto vale) il limite per n che tende ad infinito di tale espressione.

Il lettore più attento e smaliziato si sarà già accorto che vale

$$\lim_{n\to\infty} E[g(U^n)] = \lim_{n\to\infty} \sum_{i=1}^{n} g(\tfrac{i}{n})\, \frac{1}{n} = \int_0^1 g(x)\, dx,$$

tuttavia vale la pena di ricordare come si può arrivare a tale conclusione.

Poniamo $x_i^{(n)} := \frac{i}{n}$, per $i = 0, 1, \dots, n$ e notiamo che $\frac{1}{n} = \frac{i}{n} - \frac{i-1}{n} = x_i^{(n)} - x_{i-1}^{(n)}$, e riscriviamo

$$\sum_{i=1}^{n} g(\tfrac{i}{n})\, \frac{1}{n} = \sum_{i=1}^{n} g(x_i^{(n)})\, \big(x_i^{(n)} - x_{i-1}^{(n)}\big).$$

[1] La prima uguaglianza è ovvia, in quanto $nx < n$, per $x \in (0,1)$. La terza uguaglianza è una banalità, mentre un pochino meno ovvia è la seconda uguaglianza. Per questo motivo invitiamo il lettore a considerare un esempio

$$\{i:\ 0 < i \le \sqrt{5}\} = \{i:\ 0 < i \le 2\} \qquad (\text{e } \lfloor \sqrt{5} \rfloor = 2)$$

e a ricordare che la definizione di parte intera inferiore di α è il massimo degli interi i tali che $i \le \alpha$.

[2] Qualunque sia $x \in \mathbb{R}$ si ha

$$\lim_{n\to\infty} \frac{\lfloor xn \rfloor}{n} = x. \qquad (\natural)$$

Infatti, per definizione della parte intera inferiore, si ha

$$\lfloor nx \rfloor \le x < \lfloor nx \rfloor + 1.$$

Quindi valgono le seguenti disuguaglianze

$$0 \le nx - \lfloor nx \rfloor \le 1 \quad 0 \le \frac{nx - \lfloor nx \rfloor}{n} = x - \frac{\lfloor nx \rfloor}{n} \le \frac{1}{n}$$

da cui immediatamente segue il limite ($\natural$)

A questo punto, il lettore avrà riconosciuto che la precedente somma è una somma di Riemann, che serve per ottenere/definire il valore dell'integrale di $g(x)$ sull'intervallo $(0,1)$. Finiamo con l'osservare che il limite della precedente espressione esiste e vale appunto $\int_0^1 g(x)\,dx$, perché abbiamo assunto che g sia una funzione continua.

In generale si dirà che una variabile aleatoria continua U ha distribuzione uniforme in $(0,1)$ se

$$\begin{aligned} &P(U \leq x) = 0 \quad \text{per } x \leq 0 \\ &P(U \leq x) = x \quad \text{per } 0 < x < 1 \\ &P(U \leq x) = 1 \quad \text{per } x \geq 1 \end{aligned}$$

e, per questa variabile aleatoria, qualunque sia g continua, definiremo

$$E\big[g(U)\big] = \int_0^1 g(x)\,dx.$$

Consideriamo ora un altro esempio. Sia $T^{(n)}$ una variabile aleatoria Geometrica di parametro $p = p_n = \lambda/n$, cioè

$$P(T^{(n)} = k) = p(1-p)^{k-1} = \frac{\lambda}{n}\big(1 - \frac{\lambda}{n}\big)^{k-1}, \quad k = 1, 2, \ldots.$$

Come sappiamo $T^{(n)}$ si ottiene come tempo (calcolato in numero di prove) di primo successo in una successione di prove ripetute (ossia uno schema di Bernoulli infinito) e quindi

$$P(T^{(n)} > k) = (1-p)^k = \big(1 - \frac{\lambda}{n}\big)^k, \qquad k = 0, 1, 2, \ldots.$$

Consideriamo ora la variabile aleatoria[3]

$$X^{(n)} = \frac{T^{(n)}}{n}.$$

Vogliamo calcolare analogamente al caso precedente la sua funzione di distribuzione e vedere se ammette limite per n che tende ad infinito.

$$P(X^{(n)} \leq x) = 0 \qquad \text{per } x \leq 0$$

mentre

$$P(X^{(n)} \leq x) = 1 - P(X^{(n)} > x) \qquad \text{per } x > 0.$$

Inoltre, sempre per $x > 0$,

$$P(X^{(n)} > x) = P(T^{(n)} > nx) = P(T^{(n)} > \lfloor nx \rfloor) = (1 - \frac{\lambda}{n})^{\lfloor nx \rfloor} = (1 - \frac{\lambda}{n})^{n \frac{\lfloor nx \rfloor}{n}} \underset{n\to\infty}{\longrightarrow} (e^{-\lambda})^x = e^{-\lambda\, x}.$$

Riassumendo, abbiamo dimostrato che

$$\lim_{n\to\infty} P(X^{(n)} \leq x) = 0 \qquad \text{per } x \leq 0 \tag{1}$$

$$\lim_{n\to\infty} P(X^{(n)} \leq x) = 1 - e^{-\lambda x} \qquad \text{per } x > 0. \tag{2}$$

[3]Possiamo pensare sempre $X^{(n)}$ come tempo di primo successo, nella situazione in cui invece di fare una prova in ciascuna unità di misura in ogni unità di misura si effettuano n prove e il tempo è un tempo fisico e NON il numero di prove effettuate. Ad esempio se in ogni ora si effettuano 60 prove (una al minuto) $X^{(n)}$ calcola il tempo che si deve aspettare fino al primo successo incluso, con il tempo calcolato in ore. Mentre $T^{(n)}$ indica il numero di prove effettuate e quindi il tempo calcolato in minuti....

In generale si dice exponenziale di parametro λ, una variabile aleatoria tale che

$$\begin{aligned} P(X \le x) &= 0 \qquad\qquad \text{per } x \le 0,\\ P(X \le x) &= 1 - e^{-\lambda\, x} \qquad \text{per } x > 0, \end{aligned}$$

e quindi possiamo enunciare il risultato precedente dicendo che la variabile aleatoria $X^{(n)}$, che è una Geometrica di parametro λ/n riscalata converge ad una variabile esponenziale.

Può essere anche interessante cercare di calcolare approssimativamente

$$P(X^{(n)} = x_k^{(n)}) = \frac{\lambda}{n}\,(1-\tfrac{\lambda}{n})^{k-1}, \qquad \text{per } x_k^{(n)} = k/n.$$

Chiaramente tale probabilità converge a zero, ma se opportunamente riscalata, o meglio se moltiplicata per n possiamo fare un conto analogo al precedente ed ottenere che, per $x_k^{(n)} = k/n$,

$$n\,P(X^{(n)} = x_k^{(n)}) = \lambda\,(1-\tfrac{\lambda}{n})^{k-1} = \lambda\,(1-\tfrac{\lambda}{n})^{n\frac{k-1}{n}} = \lambda\,\big((1-\tfrac{\lambda}{n})^n\big)^{x_k^{(n)}-\frac{1}{n}}$$

e quindi che, se $k = k_n$ tende ad infinito in modo che $x_k^{(n)}$ rimanga in un insieme limitato[4], allora

$$\frac{n\,P(X^{(n)} = x_k^{(n)})}{\lambda\,e^{-\lambda\,x_k^{(n)}}} \underset{n\to\infty}{\longrightarrow} 1$$

che in altre parole significa che

$$P(X^{(n)} = x_k^{(n)}) \approx \frac{1}{n}\,\lambda\,e^{-\lambda\,x_k^{(n)}}.$$

Infine vediamo come si può approssimare il valore atteso

$$\begin{aligned} E[g(X^{(n)})] &= \sum_{k=0}^{\infty} g(k/n) P(X^{(n)} = k/n) = \sum_{k=0}^{\infty} g(k/n)\,\frac{\lambda}{n}\,(1-\tfrac{\lambda}{n})^{k-1}\\ &= \sum_{k=0}^{\infty} g(x_k^{(n)})\,\lambda\,\big((1-\tfrac{\lambda}{n})^n\big)^{x_k^{(n)}}\;(x_k^{(n)} - x_{k-1}^{(n)})\,\big(1-\tfrac{\lambda}{n}\big), \end{aligned}$$

[4]NOTA
Cerchiamo di capire il ruolo della condizione che $x_k^{(n)}$ rimanga in un insieme limitato. Sia $k = k_n$ tale che $0 \le x_k^{(n)} \le B$, allora, mostrare che

$$\frac{n\,P(X^{(n)} = x_k^{(n)})}{\lambda\,e^{-\lambda\,x_k^{(n)}}} = \frac{n\,\frac{\lambda}{n}\,\big((1-\frac{\lambda}{n})^n\big)^{x_k^{(n)}-\frac{1}{n}}}{\lambda\,e^{-\lambda\,x_k^{(n)}}} = \frac{\big((1-\frac{\lambda}{n})^n\big)^{x_k^{(n)}}}{e^{-\lambda\,x_k^{(n)}}}\,(1-\tfrac{\lambda}{n})^{-1} = \left(\frac{(1-\frac{\lambda}{n})^n}{e^{-\lambda}}\right)^{x_k^{(n)}}(1-\tfrac{\lambda}{n})^{-1} \underset{n\to\infty}{\longrightarrow} 1$$

equivale a mostrare che

$$\log\left(\frac{n\,P(X^{(n)} = x_k^{(n)})}{\lambda\,e^{-\lambda\,x_k^{(n)}}}\right) = x_k^{(n)}\,\log\left(\frac{(1-\frac{\lambda}{n})^n}{e^{-\lambda}}\right) - \log(1-\tfrac{\lambda}{n}) \underset{n\to\infty}{\longrightarrow} 0(=\log 1).$$

In questo modo si vede bene il ruolo della condizione che $x_k^{(n)}$ rimanga limitato.

per una funzione g continua e limitata. Pur non facendo alcune precisazioni doverose[5], possiamo intuire che la precedente somma può essere approssimata con

$$\int_0^\infty g(x)\,\lambda\,e^{-\lambda x}\,dx.$$

Si noti che in realtà, se prendiamo la funzione $g = \chi_{(0,a]}$ uguale alla funzione costante a tratti che vale 1 sull'intervallo $(0, a]$ e 0 altrove, sappiamo già che ciò è vero: infatti, da una parte

$$\int_0^\infty g(x)\,\lambda\,e^{-\lambda x}\,dx = \int_0^a \lambda\,e^{-\lambda x}\,dx = -\int_0^a d(e^{-\lambda x}) = -(e^{-\lambda x})\Big|_0^a = 1 - e^{-\lambda a},$$

mentra dall'altra parte, per tale funzione $g = \chi_{(0,a]}$,

$$E[g(X^{(n)})] = P(X^{(n)} \in (0, a]) \simeq 1 - e^{-\lambda a},$$

e abbiamo dimostrato l'approssimazione precedentemente .

Osservazione 1.1. *Si osservi che sia le variabili aleatorie Geometriche, che le variabili aleatorie Esponenziali hanno la proprietà della MANCANZA DI MEMORIA, ossia godono della seguente proprietà:*
Sia T una variabile aleatoria Geometrica di parametro p $\big(p \in (0, 1)\big)$. Si ha

$$P(T > h + k) = P(T > h)P(T > k) \qquad \forall\, h,\, k \in \{1,\, 2,\, \dots\},$$

(infatti $P(T > h + k) = (1 - p)^{h+k} = (1 - p)^h(1 - p)^k = P(T > h)P(T > k) \qquad \forall\, h,\, k \in \{0,\, 1,\, 2,\, \dots\},$)

ossia, $\forall\, h,\, k \in \{0\,1,\, 2,\, \dots\}$, (si noti che $\{T > h + k, T > k\} = \{T > h + k\}$, in quanto $h \geq 0$)

$$P(T > h + k | T > k) = \frac{P(T > h + k, T > k)}{P(T > k)} = \frac{P(T > h + k)}{P(T > k)} = \frac{(1 - p)^{h+k}}{(1 - p)^k} = (1 - p)^h = P(T > h)$$

[5]Solo per il lettore più curioso diamo, ad esempio, un'idea di come l'ipotesi di limitatezza di g sia importante. Infatti permette di dare una maggiorazione (qui solo approssimata) al resto

$$\sum_{k \geq Ln}^{\infty} g(x_k^{(n)})\,\lambda\,\big((1 - \tfrac{\lambda}{n})^n\big)^{x_k^{(n)}}\,(x_k^{(n)} - x_{k-1}^{(n)})\,(1 - \tfrac{\lambda}{n}) \leq \|g\|\lambda \sum_{k \geq Ln}^{\infty} (1 - \tfrac{\lambda}{n})^k$$
$$= \|g\|\,\lambda\,(1 - \tfrac{\lambda}{n})^{Ln}\,\frac{1}{n}\,\frac{1}{1 - (1 - \frac{\lambda}{n})} = \|g\|\,\lambda\,(1 - \tfrac{\lambda}{n})^{Ln}\,\frac{1}{\lambda} \simeq \|g\|\,e^{-L}.$$

La precedente approssimazione può essere utilizzata come segue

$$\sum_{k=0}^{\infty} g(x_k^{(n)})\,\lambda\,\big((1 - \tfrac{\lambda}{n})^n\big)^{x_k^{(n)}}\,(x_k^{(n)} - x_{k-1}^{(n)})\,(1 - \tfrac{\lambda}{n}) \approx$$
$$= \sum_{k=0}^{Ln} g(x_k^{(n)})\,\lambda\,\big((1 - \tfrac{\lambda}{n})^n\big)^{x_k^{(n)}}\,(x_k^{(n)} - x_{k-1}^{(n)}) + \sum_{k \geq Ln}^{\infty} g(x_k^{(n)})\,\lambda\,(1 - \tfrac{\lambda}{n})^n\big)^{x_k^{(n)}}\,(x_k^{(n)} - x_{k-1}^{(n)})$$
$$\simeq \int_0^L g(x)\lambda\,e^{-\lambda x}\,dx + ERR(n, L)$$

dove $ERR(n, L) \simeq \|g\|\,e^{-L}$ e quindi, mandando L all'infinito, si ottiene l'approssimazione

$$E[g(X^{(n)})] \simeq \int_0^\infty g(x)\lambda\,e^{-\lambda x}\,dx$$

cioè

$$P(T-k>h|T>k)=P(T>h) \qquad \forall\, h,\, k\in\{0,\,1,\,2,\,\dots\}.$$

Quest'ultima formulazione mette in evidenza perché tale proprietà è detta di mancanza di memoria: infatti , se sappiamo che $T>k$ (cioè che non ho avuto successi fino alla k-sima prova inclusa) allora sappiamo che $T-k$ rappresenta il tempo che devo ancora aspettare per ottenere il primo successo, e $P(T-k>h|T>k)$ rappresenta la probabilità di dover aspettare più di altre *h prove prima del primo successo, sapendo che ho già effettuato k prove e che sono tutte fallite.*
Il fatto che $P(T-k>h|T>k)=P(T>h)$ ci dice che la funzione distribuzione (condizionata a $T>k$) del numero di prove che devo ancora aspettare è la stessa funzione di distribuzione della variabile aleatoria tempo di primo successo (cioè è come se iniziassi a contare da capo il numero di prove, senza avere alcuna informazione su quanto è accaduto nelle prove precedenti).

Analogamente se X ha distribuzione esponenziale di parametro λ $\big(\lambda>0\big)$ si ha

$$P(X>s+t)=P(X>s)P(X>t) \qquad \forall\, s,\, t>0,$$

(infatti $P(X>s+t)=e^{-\lambda(s+t)}=e^{-\lambda s}e^{-\lambda t}=P(X>s)P(X>t) \qquad \forall\, s,\, t>0,$)

ossia, $\forall\ s,t>0$ (si noti che $\{X>s+t,X>t\}=\{X>s+t\}$, in quanto $s\geq 0$ e quindi $\{X>s+t\}\subseteq\{X>t\}$)

$$P(X>s+t|X>t)=\frac{P(X>s+t,X>t)}{P(X>t)}=\frac{P(X>s+t)}{P(X>t)}=\frac{e^{-\lambda(s+t)}}{e^{-\lambda t}}=e^{-\lambda s}=P(X>s)$$

cioè

$$P(X-t>s|X>t)=P(X>s) \qquad \forall\, s,\, t>0.$$

L'interpretazione è analoga se si pensa ad X come il tempo di guasto di un'apparecchiatura (o ad un tempo di attesa, ad esempio, di una prima telefonata, etc.).

2 TEOREMI LIMITE DI DE MOIVRE-LAPLACE

Iniziamo con un esempio

Esempio 2.1. *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

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

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) $\mathbb{P}(S_{10000} = 450) = \binom{10000}{450}(0.05)^{450}(0.95)^{9950}$

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.

Questa lezione è dedicata appunto al problema dell'approssimazione della legge binomiale.

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

$$\text{per } k = 0, 1, \cdots, n \quad \mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}. \tag{3}$$

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

$$\text{per } k = 0, 1, \cdots, n \quad \mathbb{P}(S_n \leq k) = \sum_{h=0}^{k} \binom{n}{h} p^h (1-p)^{n-h} = \frac{n!}{h!(n-h)!} p^h (1-p)^{n-h}. \tag{4}$$

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

$$\mathbb{P}(S_n = 0) = (1-p)^n \tag{5}$$

$$\mathbb{P}(S_n = k+1) = \frac{n-k}{k+1} \frac{p}{1-p} \mathbb{P}(S_n = k) \tag{6}$$

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 (6) 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.

NOTA DI APPROFONDIMENTO
ANDAMENTO A CAMPANA DELLA DISTRIBUZIONE BINOMIALE

Dalla formula (6), per k e $k-1$, si ha

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

e quindi

$$\mathbb{P}(S_n = k) \geq \mathbb{P}(S_n = k-1) \quad \text{se e solo se} \quad \frac{n-k+1}{k}\frac{p}{1-p} \geq 1$$

ovvero se e solo se

$$(n-k+1)p \geq (1-p)k, \quad \text{o meglio} \quad np + p \geq k.$$

Ciò significa che

$$\mathbb{P}(S_n = k) \geq \mathbb{P}(S_n = k-1)$$

fino a $k = \lfloor np + p \rfloor$, ovvero fino a k uguale alla parte intera inferiore di $np + p = \mathbb{E}[S_n] + p$. In generale c'è un solo punto di massimo, tranne nel caso in cui $np + p$ sia un intero, in tale caso allora $\mathbb{P}(S_n = k) = \mathbb{P}(S_n = k-1)$ per $k = np + p$ e $k - 1 = np - (1-p)$, cioè per i due valori interi più vicini a $np = \mathbb{E}[S_n]$.
Riassumendo il massimo di $\mathbb{P}(S_n = k)$ viene raggiunto per $k \in [np-(1-p), np+p]$.

Abbiamo visto che se consideriamo $p = p_n$ che tende a zero ed n che tende ad infinito, ma in modo che $np = \lambda$, si può usare l'approssimazione di Poisson, ma 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 $n \geq 100$, si abbia $\lambda = np \leq 10$. 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 $\lambda = np = 10000\frac{5}{100} = 500$. Riprendiamo il problema di calcolare $\mathbb{P}(S_n = k)$. Siamo interessati almeno ad ottenere un valore approssimato. Lo stesso problema se lo ponevano già, e a maggior ragione nel diciottesimo secolo.

Una prima risposta fu data da De Moivre per il caso $p = 1/2$, e poi il caso generale fu risolto da Laplace. Il risultato ottenuto da Laplace può essere riscritto in termini moderni nel seguente modo.

Teorema 2.1 (TEOREMA LOCALE di DE MOIVRE-LAPLACE).
Per ogni numero reale y, si definisca

$$\varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}},$$

e, per ogni k intero con $k = 0, 1, \cdots, n$, si definisca

$$x = x_k^{(n)} := \frac{k - np}{\sqrt{np(1-p)}}.$$

Se, per ogni n, S_n è una variabile aleatoria binomiale di parametri n e p, allora si ha che

$$\lim_{n\to\infty} \frac{\mathbb{P}(S_n = k)}{\frac{\varphi(x_k^{(n)})}{\sqrt{np(1-p)}}} = 1$$

uniformemente per k tale che $x = x_k^{(n)}$ varia in un insieme limitato.

In altre parole, per ogni $a < b$ si ha

$$\lim_{n\to\infty} \sup_{k:\, x_k^{(n)} \in [a,b]} \left| \frac{\sqrt{np(1-p)}\, \mathbb{P}(S_n = k)}{\varphi(x_k^{(n)})} - 1 \right| = 0$$

Rimandiamo la dimostrazione alla sezione successiva e facciamo alcune osservazioni.

Osservazione 2.1. *Sappiano che $\mathbb{P}(S_n = k)$ tende a zero per n che tende ad infinito, tuttavia, se opportunamente riscalato (in questo caso moltiplicato per $\sqrt{np(1-p)}$) si ha che $\mathbb{P}(S_n = k)$ converge ad un limite finito, nel senso precisato dal teorema locale di De Moivre e Laplace.*

Inoltre il teorema può essere usato per calcolare in modo approssimato $\mathbb{P}(S_n = k)$, ovvero

$$\mathbb{P}(S_n = k) \approx \frac{1}{\sqrt{np(1-p)}} \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2} \tfrac{(k-np)^2}{np(1-p)}\}$$

dove la notazione $a_n \approx b_n$ significa che b_n è definitivamente diverso da zero e $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

Il teorema di De Moivre e Laplace garantisce quindi che

$$\frac{\sqrt{np(1-p)}\, \mathbb{P}(S_n = k)}{\varphi(x_k^{(n)})} = \frac{\mathbb{P}(S_n = k)}{\frac{\varphi(x_k^{(n)})}{\sqrt{np(1-p)}}} = \frac{\mathbb{P}(S_n = k)}{\frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{np(1-p)}} \exp\{-\frac{1}{2} \frac{(k-np)^2}{np(1-p)}\}} \underset{n\mapsto\infty}{\longrightarrow} 1$$

sotto la condizione che $k = k_n$ dipenda da n in modo che $\frac{k-np}{\sqrt{np(1-p)}} = x_k^{(n)}$ si mantenga in un insieme limitato.

Rimane da capire, almeno intuitivamente come mai viene in mente proprio di guardare a questo rapporto. Il motivo è semplice, infatti da una parte la variabile aleatoria standardizzata di S_n è

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

e quindi, ovviamente,

$$\{S_n = k\} = \{ \tfrac{S_n - np}{\sqrt{np(1-p)}} = \tfrac{k-np}{\sqrt{np(1-p)}} \} = \{S_n^* = x_k^{(n)}\},$$

e d'altra parte

$$x_k^{(n)} - x_{k-1}^{(n)} = \frac{k - np - (k-1-np)}{\sqrt{np(1-p)}} = \frac{1}{\sqrt{np(1-p)}}.$$

Quindi il teorema limite locale di De Moivre-Laplace si può riformulare dicendo che, per ogni $a < b$,

$$\lim_{n\to\infty} \sup_{k:\, x_k^{(n)} \in [a,b]} \left| \frac{\mathbb{P}(S_n^* = x_k^{(n)})}{\varphi(x_k^{(n)})\,(x_k^{(n)} - x_{k-1}^{(n)})} - 1 \right| = 0. \tag{7}$$

Questa osservazione è alla base di un importante risultato, noto come teorema limite integrale di De Moivre-Laplace, ossia il Teorema 2.2.

Prima di enunciare questo risultato fondamentale possiamo spiegare euristicamente in che cosa consiste e come ci si arriva a partire dal teorema locale.

Prima di tutto si osservi che

$$\mathbb{P}(S_n \le k) = \mathbb{P}(S_n^* \le x_k^{(n)}) = \sum_{i\in\{0,1,\ldots,n\}:\, i\le k} \mathbb{P}(S_n = i) = \sum_{i\in\{0,1,\ldots,n\}:\, i\le k} \mathbb{P}(S_n^* = x_i^{(n)})$$

e quindi, dalla (7), possiamo scrivere

$$\simeq \sum_{i\in\{0,1,\ldots,n\}:\, x_i^{(n)}\le x_k^{(n)}} \varphi(x_i^n)\,(x_i^{(n)} - x_{i-1}^{(n)}).$$

A sua volta, tenendo conto del fatto che $x_0^{(n)} = \frac{0-np}{\sqrt{np(1-p)}} = -\sqrt{\frac{np}{1-p}} \to -\infty$, la precedente somma è una approssimazione del seguente integrale e quindi

$$\mathbb{P}(S_n \le k) \simeq \int_{-\infty}^{x_k^{(n)}} \varphi(x)\,dx = \int_{-\infty}^{x_k^{(n)}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\,dx.$$

Il Teorema Centrale del Limite, nella sua forma integrale asserisce proprio questo, e in modo più preciso.

Teorema 2.2. *Sia S_n (per ogni n) una variabile aleatoria $Bin(n,p)$. Allora, uniformemente in a e b, con $a < b$, vale:*

$$\lim_{n\to\infty} \mathbb{P}(a < \frac{S_n - np}{\sqrt{np(1-p)}} \le b) = \int_a^b \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\,dx. \tag{8}$$

In altre parole

$$\lim_{n\to\infty} \Big\{ \sup_{a<b} \Big| \mathbb{P}(a < \frac{S_n - np}{\sqrt{np(1-p)}} \le b) - \int_a^b \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\,dx \Big| \Big\} = 0 \tag{9}$$

Come conseguenza si ha anche che per ogni $x \in \mathbb{R}$

$$\lim_{n\to\infty} \Big\{ \sup_{x} \Big| \mathbb{P}(\frac{S_n - np}{\sqrt{np(1-p)}} \le x) - \int_{-\infty}^x \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}}\,dy \Big| \Big\} = 0 \tag{10}$$

Non daremo in queste note la dimostrazione di tale risultato. Tuttavia vogliamo illustrare alcune conseguenze dei due teoremi di De Moivre-Laplace.

Esempio 2.2 (Esempio 2.1 - seconda parte**).**
A questo punto possiamo utilizzare i due risultati di De Moivre-Laplace per poter calcolare in modo approssimato le probabilità dei punto a) e b) dell'esempio 2.1.

In particolare, essendo $np = 500$ ed $np(1-p) = 500 \cdot \frac{95}{100} = 5 \cdot 95 = 475$, possiamo affermare che

a) $\mathbb{P}(S_{10000} = 450) = \mathbb{P}(\frac{S_{10000}-500}{\sqrt{475}} = \frac{450-500}{\sqrt{475}}) \simeq \varphi(\frac{-50}{\sqrt{475}})\,\frac{1}{\sqrt{475}} = \frac{e^{-\frac{50^2}{2\cdot 475}}}{\sqrt{2\pi}}\,\frac{1}{\sqrt{475}} \simeq 3,002 \cdot 10^{-3}.$

b) $\mathbb{P}(S_{10000} \le 450) = \mathbb{P}(\frac{S_{10000}-500}{\sqrt{475}} \le \frac{450-500}{\sqrt{475}}) \simeq \Phi(\frac{-50}{\sqrt{475}}) = \Phi(-2,2941) = 1 - \Phi(2,2941) \simeq 0,062.$

Un'altra conseguenza importante consiste nell'ottenere una dimostrazione alternativa della legge dei grandi numeri, almeno nel caso della frequenza dei successi in prove ripetute.

Riportiamo qui di seguito una nuova dimostrazione della Legge dei Grandi Numeri per questo caso.

Teorema 2.3 (LEGGE DEI GRANDI NUMERI, per il caso di eventi)**.** *Sia S_n il numero di successi in uno schema di prove ripetute, allora qualunque sia $\eta > 0$*

$$\lim_{n\to\infty} \mathbb{P}(\{|\ \frac{S_n}{n} - p\ | \leq \eta\}) = 1.$$

Dimostrazione. **[attraverso il Teorema integrale di De Moivre - Laplace]**

Prima di tutto osserviamo che

$$\begin{aligned}
|\ \frac{S_n}{n} - p\ | \leq \eta \Leftrightarrow |\ S_n - np\ | \leq n\eta \Leftrightarrow -n\eta \leq S_n - np \leq n\eta \\
\Leftrightarrow \frac{-n\,\eta}{\sqrt{np(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{n\,\eta}{\sqrt{np(1-p)}} \\
\Leftrightarrow -\sqrt{n}\frac{\eta}{\sqrt{p(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \sqrt{n}\frac{\eta}{\sqrt{p(1-p)}}
\end{aligned}$$

di conseguenza

$$\mathbb{P}(\{|\ \frac{S_n}{n} - p\ | \leq \eta\}) = \mathbb{P}\Big(-\sqrt{n}\,\frac{\eta}{\sqrt{p(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \sqrt{n}\frac{\eta}{\sqrt{p(1-p)}}\Big)$$

e quindi,posto $\gamma = \gamma(n,p) := \frac{\eta}{\sqrt{p(1-p)}}$, si ha

$$\mathbb{P}(\{|\ \frac{S_n}{n} - p\ | \leq \eta\}) \simeq \int_{-\sqrt{n}\,\gamma}^{\sqrt{n}\,\gamma} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx \quad \underset{n\mapsto\infty}{\longrightarrow} 1$$

dove l'ultimo limite discende dal Teorema integrale 2.2. □

2.1 Dimostrazione del Teorema locale di De Moivre-Laplace

La dimostrazione del teorema locale di De Moivre-Laplace (Teorema 2.1) si basa sulla formula di Stirling

$$n! = \sqrt{2\pi} n^{n+1/2}\, e^{-n}\, e^{\theta_n} \qquad \text{con } 0 < \theta_n < \frac{1}{12n} \tag{11}$$

Per la dimostrazione di questa affermazione rimandiamo il lettore alla sezione 2.2, dove ne diamo una dimostrazione[6]. Iniziamo subito utilizzando la precedente formula per riscrivere il valore di

[6]La dimostrazione, per la verità, è incompleta perché viene solo dimostrato che esiste una costante C tale che

$$n! = C\, n^{n+1/2}\, e^{-n}\, e^{\theta_n} \qquad \text{con } 0 < \theta_n < \frac{1}{12n},$$

ma C non viene calcolata esplicitamente (confrontare Lemma 2.3.1).

$\mathbb{P}(S_n = k)$:

$$
\begin{aligned}
\mathbb{P}(S_n = k) &= \frac{n!}{k!(n-k)!}\, p^k (1-p)^{n-k} \\
&= \frac{\sqrt{2\pi}\, n^{n+1/2}\, e^{-n}\, e^{\theta_n}}{\sqrt{2\pi}\, k^{k+1/2}\, e^{-k}\, e^{\theta_k}\, \sqrt{2\pi}\, (n-k)^{n-k+1/2}\, e^{-(n-k)}\, e^{\theta_{n-k}}}\, p^k (1-p)^{n-k} \\
&= \frac{n^{n+1/2}\, e^{\theta_n}\, e^{-\theta_k}\, e^{-\theta_{n-k}}}{\sqrt{2\pi}\, k^{k+1/2} (n-k)^{n-k+1/2}}\, p^k (1-p)^{n-k} \\
&= \frac{n^{n+1/2}}{\sqrt{2\pi}\, k^{k+1/2}\, (n-k)^{n-k+1/2}} p^k (1-p)^{n-k} e^{\theta_n - \theta_k - \theta_{n-k}} \\
&= \frac{1}{\sqrt{2\pi}}\, \frac{n^{1/2}}{k^{1/2}(n-k)^{1/2}} \Big(\frac{np}{k}\Big)^k \Big(\frac{n(1-p)}{n-k}\Big)^{n-k} e^{\theta_n - \theta_k - \theta_{n-k}}
\end{aligned}
$$

da cui

$$\mathbb{P}(S_n = k) = F_n\, G_n\, H_n,$$

dove

$$
\begin{aligned}
F_n &= \frac{n^{1/2}}{k^{1/2}\, (n-k)^{1/2}} = \sqrt{\frac{n}{k(n-k)}} \\
G_n &= \frac{1}{\sqrt{2\pi}} \Big(\frac{np}{k}\Big)^k \Big(\frac{n(1-p)}{n-k}\Big)^{n-k} \\
H_n &= e^{\theta_n - \theta_k - \theta_{n-k}}
\end{aligned}
$$

Con queste posizioni si ha che, con notazioni ovvie,

$$\frac{\sqrt{np(1-p)}\, \mathbb{P}(S_n = k)}{\varphi(x_k^{(n)})} = \Big(\sqrt{np(1-p)}\, F_n\Big) \frac{G_n}{\varphi(x_k^{(n)})} H_n = \widetilde{F}_n \widetilde{G}_n H_n$$

Considereremo i tre fattori separatamente. Passando ai logaritmi

$$\widetilde{F}_n\, \widetilde{G}_n\, H_n = \exp\big\{ \log(\widetilde{F}_n) + \log(\widetilde{G}_n) + \log(H_n) \big\},$$

si vede subito che, per mostrare quindi la convergenza a 1 di $\frac{\sqrt{np(1-p)}\, \mathbb{P}(S_n=k)}{\varphi(x_k^{(n)})}$, basta mostrare che ciascuno dei tre logaritmi della formula precedente converge a zero.

Per ottenere la convergenza uniforme basta mostrare poi che ciascuno dei logaritmi converge uniformemente a zero. In un primo momento diamo solo la dimostrazione informale della convergenza, senza porre attenzione alla convergenza uniforme. La convergenza uniforme viene trattata nelle note di approfondimento, dedicate solo ai lettori più curiosi. Inoltre, in tali note, senza perdere in generalità, supporremo che esista un $B > 0$ tale che $a = -B$ e $b = B$, ossia supporremo che k sia tale che $|x_k^{(n)}| \leq B$.

In tutti e tre i casi è fondamentale l'osservazione che, poiché $x = x_k^{(n)} := \frac{k-np}{\sqrt{np(1-p)}}$ si ha che

$$k = np + \sqrt{np(1-p)}x_k^{(n)} \quad \text{e} \quad n - k = n(1-p) - \sqrt{np(1-p)}x_k^{(n)}. \tag{12}$$

Inoltre, tenendo conto di (12) si ha che

$$\frac{k}{np} = \frac{np + \sqrt{np(1-p)}\, x_k^{(n)}}{np} = 1 + \sqrt{\frac{1-p}{np}}\, x_k^{(n)} \tag{13}$$

e che

$$\frac{n-k}{n(1-p)} = \frac{n(1-p) - \sqrt{np(1-p)}\, x_k^{(n)}}{n(1-p)} = 1 - \sqrt{\frac{p}{n(1-p)}}\, x_k^{(n)}, \tag{14}$$

CASO DI $\widetilde{F}_n = \sqrt{np(1-p)}\, F_n$

Come sappiamo

$$\widetilde{F}_n = \sqrt{np(1-p)}\; F_n = \sqrt{np(1-p)}\sqrt{\frac{n}{k(n-k)}} = \sqrt{\frac{np}{k}\,\frac{n(1-p)}{n-k}}$$

e quindi

$$\log(\widetilde{F}_n) = \frac{1}{2}\left(\log(\frac{np}{k}) + \log(\frac{n(1-p)}{n-k})\right)$$

Tenendo conto di (13) e di (14) conviene scrivere

$$\log(\widetilde{F}_n) = -\frac{1}{2}\left(\log(\frac{k}{np}) + \log(\frac{n-k}{n(1-p)})\right)$$

$$= -\frac{1}{2}\left[\log\left(1 + \sqrt{\frac{1-p}{np}}\, x_k^{(n)}\right) + \log\left(1 - \sqrt{\frac{p}{n(1-p)}}\, x_k^{(n)}\right)\right]$$

Da cui immediatamente si vede che $\log(\widetilde{F}_n)$ tende a zero quando n tende all'infinito, se $x_k^{(n)}$ rimane in un limitato.

Solo per il lettore più curioso

La convergenza è uniforme: infatti, basta tenere conto del fatto che $|\log(1-r)| \leq |r|\frac{1}{2(1-|r|)}$, per $|r| < 1$, da cui

$$|\log(1-r)| \leq |r|, \qquad \text{per } |r| < 1/2,$$

e che, per ogni k tale che $|x_k^{(n)}| \leq B$ e per n sufficientemente grande, entrambi

$$\sqrt{\frac{1-p}{np}}\,|x_k^{(n)}| \leq \sqrt{\frac{1-p}{np}}\, B \quad \text{e} \quad \sqrt{\frac{p}{n(1-p)}}\,|x_k^{(n)}| \leq \sqrt{\frac{p}{n(1-p)}}\, B$$

sono minori o uguali di $1/2$.

CASO DI H_n

Grazie alla Formula di Stirling (si veda Lemma 2.3.1)

$$
\begin{aligned}
|\log(H_n)| = |\theta_n - \theta_k - \theta_{n-k}| &< \frac{1}{12n} + \frac{1}{12k} + \frac{1}{12(n-k)} \\
&= \frac{1}{12}\left[\frac{1}{n} + \frac{1}{np + \sqrt{np(1-p)}\,x_k^{(n)}} + \frac{1}{n - np - \sqrt{np(1-p)}\,x_k^{(n)}}\right] \\
&= \frac{1}{12n}\left[1 + \frac{1}{p + \sqrt{\frac{p(1-p)}{n}}\,x_k^{(n)}} + \frac{1}{(1-p) - \sqrt{\frac{p(1-p)}{n}}\,x_k^{(n)}}\right]
\end{aligned}
$$

Da cui immediatamente si vede che $\log(H_n)$ tende a zero quando n tende all'infinito, se $x_k^{(n)}$ rimane in un limitato.

Solo per il lettore più curioso

La convergenza di $\log H_n$ a zero è uniforme, infatti dalla precedente disuguaglianza si ha che, se $-B \le x_k^{(n)} \le B$, allora

$$
|\log(H_n)| \le \frac{1}{12n}\left[1 + \frac{1}{p - \sqrt{\frac{p(1-p)}{n}}\,B} + \frac{1}{(1-p) - \sqrt{\frac{p(1-p)}{n}}\,B}\right].
$$

Si osservi che, per n sufficientemente grande sia $p - \sqrt{\frac{p(1-p)}{n}}\,B$ che $(1-p) - \sqrt{\frac{p(1-p)}{n}}\,B$ sono strettamente positivi.

CASO DI $\widetilde{G}_n = \frac{G_n}{\varphi(x_k^{(n)})}$

Cominciamo con l'osservare che

$$\widetilde{G}_n = \frac{G_n}{\varphi(x_k^{(n)})} = \Big(\frac{np}{k}\Big)^k \Big(\frac{n(1-p)}{n-k}\Big)^{n-k} e^{\frac{(x_k^{(n)})^2}{2}}$$

per cui

$$\log(\widetilde{G}_n) = \log\left[\Big(\frac{np}{k}\Big)^k \Big(\frac{n(1-p)}{n-k}\Big)^{n-k}\right] + \frac{(x_k^{(n)})^2}{2} = k\log\Big(\frac{np}{k}\Big) + (n-k)\log\Big(\frac{n(1-p)}{n-k}\Big) + \frac{(x_k^{(n)})^2}{2}.$$

Inoltre, ricordando le espressioni (13) e (14), rispettivamente di $\frac{k}{np}$ e di $\frac{n-k}{n(1-p)}$, conviene scrivere

$$\begin{aligned}
\log(\widetilde{G}_n) &= -k\log\Big(\tfrac{k}{np}\Big) - (n-k)\log\Big(\tfrac{n-k}{n(1-p)}\Big) + \frac{(x_k^{(n)})^2}{2}\\
&= -np\,\frac{k}{np}\,\log\Big(\tfrac{k}{np}\Big) - n(1-p)\,\frac{n-k}{n(1-p)}\,\log\Big(\tfrac{n-k}{n(1-p)}\Big) + \frac{(x_k^{(n)})^2}{2}\\
&= -np\big[1+\sqrt{\tfrac{1-p}{np}}\,x_k^{(n)}\big]\log\Big(1+\sqrt{\tfrac{1-p}{np}}\,x_k^{(n)}\Big)\\
&\qquad - n(1-p)\big[1-\sqrt{\tfrac{p}{n(1-p)}}x_k^{(n)}\big]\log\Big(1-\sqrt{\tfrac{p}{n(1-p)}}x_k^{(n)}\Big) + \frac{(x_k^{(n)})^2}{2}
\end{aligned}$$

A questo punto poniamo

$$s = s_n = \sqrt{\tfrac{1-p}{np}}\,x_k^{(n)} \qquad \text{e} \qquad t = t_n = \sqrt{\tfrac{p}{n(1-p)}}\,x_k^{(n)}$$

per cui si ottiene

$$\log(\widetilde{G}_n) = -np\,(1+s)\,\log(1+s) - n(1-p)\,(1-t)\,\log(1-t) + \frac{(x_k^{(n)})^2}{2}.$$

Ricordiamo che (si veda la sezione 2.3)

$$-(1-r)\,\log(1-r) = (1-r)\big(r+\tfrac{r^2}{2}+O(r^3)\big) = r+\tfrac{r^2}{2}-r^2-\tfrac{r^3}{2}+(1-r)O(r^3) = r-\tfrac{r^2}{2}+O(r^3),$$

e osserviamo che, se $x_k^{(n)}$ si mantiene limitato allora

$$s = s_n = \sqrt{\tfrac{1-p}{np}}\,x_k^{(n)} = O(\tfrac{1}{\sqrt{n}}) \qquad \text{e} \qquad t = t_n = \sqrt{\tfrac{p}{n(1-p)}}\,x_k^{(n)} = O(\tfrac{1}{\sqrt{n}}).$$

Si ottiene allora che

$$\begin{aligned}
\log(\widetilde{G}_n) &= np\big(-s-\tfrac{s^2}{2}+O(s^3)\big) + n(1-p)\,\big(t-\tfrac{t^2}{2}+O(t^3)\big) + \frac{(x_k^{(n)})^2}{2}\\
&= -nps + n(1-p)t - \frac{1}{2}\Big(np\,s^2 + n(1-p)\,t^2\Big) + \Big(np\,O(s^3) + n(1-p)\,O(t^3)\Big) + \frac{(x_k^{(n)})^2}{2}\\
&= 0 - \frac{(x_k^{(n)})^2}{2} + np\,O\Big(\big(\tfrac{1}{\sqrt{n}}\big)^3\Big) + n(1-p)\,O\Big(\big(\tfrac{1}{\sqrt{n}}\big)^3\Big) + \frac{(x_k^{(n)})^2}{2} = O(\tfrac{1}{\sqrt{n}})
\end{aligned}$$

dove la penultima uguaglianza dipende dal fatto che, per definizione di s e t si ha

$$-nps + n(1-p)t = -np\sqrt{\tfrac{1-p}{np}}\,x_k^{(n)} + n(1-p)\sqrt{\tfrac{p}{n(1-p)}}\,x_k^{(n)} = x_k^{(n)}\Big(-\sqrt{np(1-p)}+\sqrt{pn(1-p)}\Big) = 0,$$

e inoltre

$$np\,s^2 + n(1-p)\,t^2 = np\,\frac{1-p}{np}\,(x_k^{(n)})^2 + n(1-p)\,\frac{p}{n(1-p)}\,(x_k^{(n)})^2 = (1-p+p)\,(x_k^{(n)})^2 = (x_k^{(n)})^2$$

Solo per il lettore più curioso

Anche in questo caso la convergenza di $\log \widetilde{G}_n$ a zero è uniforme. Infatti basta dimostrare che esiste una constante L tale che, per n sufficientemente grande e per ogni $|x_k^{(n)}| \leq B$

$$\left|np\, O\Big(\big(\tfrac{1}{\sqrt{n}}\big)^3\Big) + n(1-p)\, O\Big(\big(\tfrac{1}{\sqrt{n}}\big)^3\Big)\right| \leq L\frac{1}{\sqrt{n}}.$$

Prima di tutto, è noto che (si veda la disugualgianza (27) nella sezione 2.3)

$$|R_3(r)| := \Big| -\log(1-r) - \big[r + \frac{r^2}{2}\big]\Big| \leq \frac{|\,r\,|^3}{3(1-|\,r\,|)} \qquad \text{per } |\,r\,| < 1,$$

e quindi, dall'uguaglianza

$$-(1-r)\,\log(1-r) = (1-r)\big(r + \tfrac{r^2}{2} + R_3(r)\big) = r + \tfrac{r^2}{2} - r^2 - \tfrac{r^3}{2} + (1-r)R_3(r),$$

si ottiene che, sempre per $|r| < 1$,

$$\Big| -(1-r)\,\log(1-r) - r - \tfrac{r^2}{2}\Big| \leq \Big| -\frac{r^3}{2}\Big| + |1-r|\,\frac{|\,r\,|^3}{3(1-|\,r\,|)},$$

ossia, per $|r| < 1/2$

$$|O(t^3)| = \Big| -(1-r)\,\log(1-r) - r - \tfrac{r^2}{2}\Big| \leq |r|^3\big(\frac{1}{2} + \frac{3}{2}\,\frac{2}{3}\big) = \frac{3}{2}\,|r|^3.$$

Di conseguenza, se $|\,x_k^{(n)}\,| \leq B$, allora vale la disuguaglianza

$$|\,t\,| = \Big|\sqrt{\frac{p}{n(1-p)}}\,x_k^{(n)}\Big| \leq \sqrt{\frac{p}{n(1-p)}}B < \frac{1}{2} < 1, \qquad \text{per } n > 2\,\frac{p}{(1-p)}\,B^2$$

e analogamente

$$|\,s\,| = \Big|\sqrt{\frac{1-p}{np}}\,x_k^{(n)}\Big| \leq \sqrt{\frac{1-p}{np}}\,B < \frac{1}{2} < 1, \qquad \text{per } n > 2\,\frac{(1-p)}{p}\,B^2,$$

da cui, per $n > 2MB^2$, con $M = \max\Big(\frac{p}{(1-p)}, \frac{(1-p)}{p}\Big)$, si ha

$$\begin{aligned}&\Big|np\, O\Big(\big(\tfrac{1}{\sqrt{n}}\big)^3\Big) + n(1-p)\, O\Big(\big(\tfrac{1}{\sqrt{n}}\big)^3\Big)\\ &\Big| \leq np\,\frac{3}{2}\,\Big(\sqrt{\frac{1-p}{np}}\,B\Big)^3 + n(1-p)\,\frac{3}{2}\,\Big(\sqrt{\frac{p}{n(1-p)}}B\Big)^3 \leq L\frac{1}{\sqrt{n}}.\end{aligned}$$

Solo per il lettore più curioso

Perché la convergenza uniforme dei logaritmi a zero implica la convergenza uniforme della funzione

Vogliamo dimostrare che, se $\log g_n(x)$ converge a zero uniformemente, allora $g_n(x)$ converge a d 1 uniformemente.
Iniziamo osservando che, dallo sviluppo di Taylor in 0 per la funzione esponenziale, cioè per $y \mapsto \exp(y)$. Ricordiamo che, per una funzione $x \mapsto f(x)$ si ha

$$f(x) = f(0) + f'(0)x + \frac{1}{2} f''(\xi)x^2,$$

dove ξ è un numero tra 0 e x, se $x > 0$, ovvero tra x e 0, se $x < 0$, o meglio $\xi = x\theta(x)$, con $\theta(x) \in (0, 1)$.
Per la funzione esponenziale si ha

$$\exp\{\alpha\} = 1 + \alpha + \frac{1}{2}\,\alpha^2\, exp\{\theta(\alpha)\alpha\}$$

e quindi per ogni α tale che $\mid \alpha \mid \leq A$

$$\begin{aligned}\mid \exp\{\alpha\} - 1 \mid &= \left|\alpha + \frac{1}{2}\,\alpha^2 \,\exp\{\theta(\alpha)\alpha\}\right| \leq |\alpha| + \frac{1}{2}\,\alpha^2 \,\exp\{\theta(\alpha)\alpha\} \\ &\leq |\alpha| + \frac{1}{2} \mid \alpha \mid^2 \,\exp\{|\alpha|\} \leq A + \frac{1}{2}\,A^2 \exp\{A\}\end{aligned}$$

Quindi poiché

$$g_n(x) = \exp\{\log g_n(x)\}$$

se

$$\sup_{x \in I_n} |\log g_n(x)| \leq A_n \quad \text{e} \quad \lim_{n \to \infty} A_n = 0,$$

allora

$$\sup_{x \in I_n} |g_n(x) - 1| = \sup_{x \in I_n} |\exp\{\log g_n(x)\} - 1| \leq A_n + \frac{1}{2}\,A_n^2 \exp\{A_n\} \underset{n\to\infty}{\longrightarrow} 0.$$

2.2 FORMULA DI STIRLING

La formula di Stirling permette di calcolare in modo approssimato $n!$:

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi}\, n^{n+1/2}\, e^{-n}} = 1 \tag{15}$$

In realtà vale una forma più precisa della formula di Stirling, ossia,

$$n! = \sqrt{2\pi}\, n^{n+1/2}\, e^{-n}\, e^{\theta_n} \qquad \text{con } 0 < \theta_n < \frac{1}{12n} \tag{16}$$

In queste note non dimostreremo le precendenti formule (15) e (16), ma solo che esiste una costante $C \in \mathbb{R}$ per cui

$$\lim_{n\to\infty} \frac{n!}{n^{n+1/2}\, e^{-n}} = e^{C}, \tag{17}$$

anzi in realtà mostreremo la forma più precisa della formula di Stirling, ma sempre con la costante da determinare, ossia non dimostreremo invece il fatto che e^C coincide con $\sqrt{2\pi}$. Tuttavia, dopo aver enunciato il teorema limite centrale di De Moivre e Laplace nella sua forma integrale, ne daremo una giustificazione euristica.

Lemma 2.3.1. *Esiste una costante $C \in \mathbb{R}$ tale che, per ogni $n \geq 1$, esiste un numero θ_n per cui vale la seguente uguaglianza*

$$n! = e^{C}\, n^{n+1/2}\, e^{-n}\, e^{\theta_n} \qquad \textit{con } 0 < \theta_n < \frac{1}{12n} \tag{18}$$

Va detto che è possibile dimostrare che $C = \log(\sqrt{2\pi})$. di conseguenza si ha

$$n! = \sqrt{2\pi}\, n^{n+1/2}\, e^{-n}\, e^{\theta_n} \qquad \text{con } 0 < \theta_n < \frac{1}{12n} \tag{19}$$

Dimostrazione. La dimostrazione che daremo qui segue quella del libro di W. Feller: An Introduction to Probability Theory and its Applications Vol. I

NOTA DI APPROFONDIMENTO
MOTIVAZIONE EURISTICA della Formula di Stirling:
(non è necessaria per la comprensione del seguito)

Vediamo qui sotto come può venire in mente la formula di Stirling
Per il calcolo di $n!$ si può procedere nel seguente modo utilizzando le tavole dei logaritmi (ed era l'unico modo prima dell'avvento dei computer):
prima si osserva che

$$n! = \exp\{\log(n!)\},$$

e successivamente che

$$\log(n!) = \sum_{k=1}^{n} \log(k),$$

Quindi si usa la tavola dei logaritmi per calcolare $\log(k)$, si esegue la somma e poi si usa la tavola in senso inverso per calcolare in modo approssimato $n!$.
Per tale calcolo si può utilizzare il seguente criterio integrale, per ottenere con una certa approssimazione la somma dei logaritmi in $\log(n!) = \sum_{k=1}^{n} \log(k)$:
La funzione $\log(x)$ è crescente e quindi

$$\int_{k-1}^{k} \log(x)\,dx < \log(k) < \int_{k}^{k+1} \log(x)\,dx$$

Sommando per $k = 1, 2, \cdots, n$, si ottiene

$$\int_{0}^{n} \log(x)dx < \log(n!) < \int_{1}^{n+1} \log(x)dx$$

e quindi, ricordando che

$$\int \log(x)\,dx = x\log(x) - x + costante$$

(come si vede subito derivando o integrando per parti), si ottiene che

$$\alpha_n := n\log(n) - n < \log(n!) < (n+1)\log(n+1) - (n+1) + 1 =: \beta_n$$

Quindi $\log(n!)$ è in un intervallo (α_n, β_n) di ampiezza

$$(n+1)\log(n+1) - n\log(n) = (n+1)\log(\frac{n+1}{n}) + \log(n) \leq \frac{n+1}{n} + \log(n)$$

che è ancora insoddisfacente, visto che cresce con n.
Tuttavia si può sperare che il punto di mezzo dell'intervallo (α_n, β_n), ovvero la semisomma

$$\frac{1}{2}(\alpha_n + \beta_n) = \frac{1}{2}\{(n+1)\log(n+1) + n\log(n) - 2n\}$$

sia più vicina a $\log(n!)$. Si noti che (aggiungnendo e togliendo $(n+1)\log(n)$ tale semisomma si può anche esprimere come

$$\frac{1}{2}(\alpha_n + \beta_n) = (n+1/2)\log(n) - n + \frac{1}{2}(n+1)\log(\frac{n+1}{n})$$

e che $\frac{1}{2}(n+1)\log(\frac{n+1}{n}) = \frac{1}{2}(n+1)\log(1-\frac{1}{n})$ si mantiene limitato.
Ciò giustifica l'idea di fare il confronto tra $\log(n!)$ e $(n+1/2)\log(n) - n$.

Passando ai logaritmi nella formula (18), dobbiamo dimostrare che esistono una costante C e un $\theta_n \in (0, \frac{1}{12n})$ tali che

$$\log(n!) - (n + 1/2)\log(n) + n + \theta_n = C. \tag{20}$$

A questo scopo mostreremo che

$$\lim_{n\to\infty} \log(n!) - (n + 1/2)\log(n) + n = C, \tag{21}$$

e, posto

$$d_n :\, = \log(n!) - (n + 1/2)\; \log(n) + n, \tag{22}$$

mostreremo che le successioni $\{d_n\}$ e $\{d_n - \frac{1}{12n}\}$ sono monotone, e più precisamente che $\{d_n\}$ è monotona non crescente e $\{d_n - \frac{1}{12n}\}$ è monotona non decrescente. Di conseguenza entrambe le successioni sono convergenti e entrambe allo stesso numero C ($d_n \searrow C$ e $d_n - \frac{1}{12n} \nearrow C$) e inoltre

$$d_n - \frac{1}{12n} < C < d_n. \tag{23}$$

Grazie a quest'ultima disuguaglianza e all'evidente identità

$$n! = n^{n+1/2}\, e^{-n}\, e^{d_n},$$

e posto $\theta_n :\, = d_n - C$, si ha immediatamente la tesi, ossia che

$$n! = e^C\, n^{n+1/2}\, e^{-n}\, e^{\theta_n} \qquad \text{con } 0 < \theta_n < \frac{1}{12n}.$$

cioè, appunto, la formula cercata.

Non rimane quindi che mostrare la monotonia delle due successioni $\{d_n\}$ e $\{d_n - \frac{1}{12n}\}$.
Prima di tutto notiamo che dalla definizione (23) si ha

$$d_n - d_{n+1} = \log(n!) - (n + 1/2)\; \log(n) + n - \log((n + 1)!) + (n + 1 + 1/2)\; \log(n + 1) - (n + 1),$$

tenendo conto che $\log((n + 1)!) = \log(n!) + \log(n + 1)$ si ha quindi

$$\begin{aligned} d_n - d_{n+1} &= -(n + 1/2)\log(n) - \log(n + 1) + (n + 1 + 1/2)\log(n + 1) - 1 \\ &= -(n + 1/2)\log(n) + (n + 1/2)\log(n + 1) - 1 \\ &= (n + 1/2)\log(\tfrac{n+1}{n}) - 1 \end{aligned}$$

Posto

$$t = \frac{1}{2n + 1},$$

si ha

$$n + 1/2 = \frac{1}{2}\,(2n + 1) = \frac{1}{2t}, \qquad \text{e} \qquad \frac{n + 1}{n} = \frac{1 + t}{1 - t}.$$

e quindi

$$d_n - d_{n+1} = (n + 1/2)\; \log(\frac{n + 1}{n}) - 1 = \frac{1}{2t}\; \log\left(\tfrac{1+t}{1-t}\right) - 1.$$

Da un noto risultato di analisi (si veda la formula (28) della sezione sui richiami di analisi)

$$d_n - d_{n+1} = \sum_{h=0}^{\infty} \frac{t^{2h}}{2h + 1} - 1 = \sum_{h=1}^{\infty} \frac{t^{2h}}{2h + 1}. \tag{24}$$

Ciò mostra che $d_n - d_{n+1}$ è chiaramente positivo e quindi la monotonia della prima successione (e più precisamente $d_{n+1} \leq d_n$).

Inoltre, sempre per $t = \frac{1}{2n+1}$, dalla equazione (24) si ottiene che

$$\begin{aligned} d_n - d_{n+1} &= \sum_{h=1}^{\infty} \frac{t^{2h}}{2h+1} \leq \sum_{h=1}^{\infty} \frac{t^{2h}}{3} = \frac{1}{3} \sum_{h=1}^{\infty} (t^2)^h = \frac{1}{3} \frac{t^2}{1-t^2} = \frac{1}{3} \frac{1}{t^{-2}-1} = \frac{1}{3} \frac{1}{(2n+1)^2 - 1} \\ &= \frac{1}{3} \frac{1}{4n^2+4n} = \frac{1}{12} \left(\frac{1}{n} - \frac{1}{(n+1)} \right) \end{aligned}$$

da cui la monotonia della seconda successione:

$$d_n - \frac{1}{12n} \leq d_{n+1} - \frac{1}{12(n+1)}$$

e ciò conclude la prova. □

Solo per il lettore più curioso

Una osservazione matematico-probabilistica su come ottenere la costante $\sqrt{2\pi}$ nella formula di Stirling

In realtà la dimostrazione del teorema locale (Teorema 2.1) è inficiata dal fatto che abbiamo utilizzato la formula di Stirling, mentre abbiamo dimostrato solo la formula Stirling con e^C al posto di $\sqrt{2\pi}$.
La dimostrazione del Teorema locale 2.1 non è completa e di conseguenza anche quella del teorema integrale 2.2(che è basata sul teorema locale), non sarebbe completa. In realtà abbiamo dimostrato solo che

$$\lim_{n\to\infty} \frac{\mathbb{P}(S_n = k)}{\frac{e^{-\frac{(x_k^{(n)})^2}{2}}}{e^C\sqrt{np(1-p)}}} = 1$$

e avremmo potuto dimostrare solo che

$$\lim_{n\to\infty} \left\{ \sup_{a<b} \left| \mathbb{P}(a < \frac{S_n - np}{\sqrt{np(1-p)}} \leq b) - \int_a^b \frac{1}{e^C} e^{-\frac{x^2}{2}}\, dx \right| \right\} = 0$$

Tuttavia, siccome, qualunque sia n, $\mathbb{P}(a < \frac{S_n - np}{\sqrt{np(1-p)}} \leq b)= 1$, pur di scegliere $a < \frac{0-np}{\sqrt{np(1-p)}} = -\sqrt{n}\sqrt{p/(1-p)}$ e $b > \frac{n-np}{\sqrt{np(1-p)}} = \sqrt{n}\sqrt{(1-p)/p}$
ad esempio $a < -\sqrt{nM}$ e $b > \sqrt{nM}$ dove $M := \max\{p/(1-p), (1-p)/p\}$.
Necessariamente deve valere che

$$\lim_{n\to\infty} \left| 1 - \int_{-\sqrt{nM}}^{\sqrt{nM}} \frac{1}{e^C} e^{-\frac{x^2}{2}}\, dx \right| = 0$$

ovvero che

$$\lim_{n\to\infty} \frac{1}{e^C} \int_{-\sqrt{nM}}^{\sqrt{nM}} e^{-\frac{x^2}{2}}\, dx = 1.$$

Una strategia per trovare la costante C potrebbe esssere quella di trovare quanto vale il limite (per $n \to \infty$) dell'integrale di $\exp\{-\frac{x^2}{2}\}$ sugli intervalli $[-\sqrt{nM}, \sqrt{nM}]$.
Poiché si può dimostrare che

$$\lim_{n\to\infty} \int_{-\sqrt{nM}}^{\sqrt{nM}} e^{-\frac{x^2}{2}}\, dx = \sqrt{2\pi},$$

si ottiene così la costante cercata, ovvero

$$e^C = \sqrt{2\pi}.$$

Solo per il lettore più curioso

Cenno di dimostrazione del fatto che $\varphi(x)$ è una densità.
ATTENZIONE SERVONO DELLE NOZIONI SUGLI INTEGRALI DOPPI

Per la dimostrazione del fatto che

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\,dx = \sqrt{2\pi},$$

si può procedere come segue: è equivalente mostrare che

$$\Big(\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\,dx\Big)^2 = 2\pi,$$

ossia che

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\,dx \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\,dy = 2\pi.$$

Ma il prodotto dei due integrali qui sopra si può interpretare come un integrale doppio

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}}\,dx\,dy$$

Ponendo $x^2+y^2=r^2$ e facendo variare $\theta\in(0,2\pi]$ si ottiene che il precedente integrale si può riscrivere come

$$\int_0^{+\infty}\int_0^{2\pi} e^{-\frac{r^2}{2}}\,rdr\,d\theta = \int_0^{2\pi} d\theta \int_0^{+\infty} e^{-\frac{r^2}{2}}\,rdr\; = 2\pi\; d(-e^{-r^2/2})\Big|_0^{\infty} = 2\pi.$$

2.3 Richiami di analisi

SOMMA RIDOTTA della serie GEOMETRICA:

$$\sum_{k=0}^{m} \alpha^k = \begin{cases} \frac{1-\alpha^{m+1}}{1-\alpha}, & \text{per } \alpha \neq 1 \\ m+1, & \text{per } \alpha = 1. \end{cases}$$

Sia α un numero reale e sia

$$s_m = s_m(\alpha) := \sum_{k=0}^{m} \alpha^k$$

Ovviamente se $\alpha = 1$ si ha $s_m = m + 1$. Passiamo quindi al caso $\alpha \neq 1$. Si noti che

$$s_{m+1} = \sum_{k=0}^{m+1} \alpha^k = \sum_{k=0}^{m} \alpha^k + \alpha^{m+1} = s_m + \alpha^{m+1}$$

e che

$$s_{m+1} = \sum_{k=0}^{m+1} \alpha^k = 1 + \sum_{k=1}^{m+1} \alpha^k = 1 + \sum_{h=0}^{m} \alpha^{h+1} = 1 + \alpha \sum_{h=0}^{m} \alpha^h = 1 + \alpha s_m,$$

quindi, dalle due uguaglianze precedenti, si ha

$$s_m + \alpha^{m+1} = 1 + \alpha s_m \qquad \text{ovvero} \qquad s_m - \alpha s_m = 1 - \alpha^{m+1},$$

da cui

$$(1-\alpha) s_m = 1 - \alpha^{m+1}$$

e quindi

$$s_m = s_m(\alpha) := \sum_{k=0}^{m} \alpha^k = \frac{1-\alpha^{m+1}}{1-\alpha}, \qquad \text{per } \alpha \neq 1$$

SOMMA DELLE SERIE GEOMETRICA PER $|\alpha| < 1$

Dall'espressione precedente pert la somma ridotta e tenendo conto che $|\alpha|^{m+1}$ tende a zero se $|\alpha| < 1$ si ha immediatamente che la serie geometrica è assolutamente convergente e che

$$\sum_{k=0}^{\infty} \alpha^k = \frac{1}{1-\alpha}, \qquad \text{per } |\alpha| < 1.$$

Infine è facile vedere[7] che

$$\sum_{k=n}^{\infty} \alpha^k = \frac{\alpha^n}{1-\alpha}, \qquad \text{per } |\alpha| < 1$$

[7] Infatti

$$\sum_{k=n}^{\infty} \alpha^k = \alpha^n \sum_{k=n}^{\infty} \alpha^{k-n} = \alpha^n \sum_{h=0}^{\infty} \alpha^h = \frac{\alpha^n}{1-\alpha}, \qquad \text{per } |\alpha| < 1.$$

SVILUPPO IN SERIE DEL LOGARITMO:
Una conseguenza immediata delle espressioni precedenti per la serie geometrica e del fatto che, per $|t| < 1$

$$-\log(1-t) = \int_0^t \frac{1}{1-s}\,ds = \int_0^t \sum_{k=0}^{\infty} s^k\,ds = \sum_{k=0}^{\infty} \int_0^t s^k\,ds = \sum_{k=0}^{\infty} \frac{s^{k+1}}{k+1},$$

è la seguente:

$$\log(1-t) = -\sum_{k=0}^{\infty} \frac{t^{k+1}}{k+1}, \qquad \text{per } |\, t \,| < 1, \tag{25}$$

che a sua volta, osservando che $\log(1+t) = \log(1-(-t))$, porta immediatamente a

$$\log(1+t) = \sum_{k=0}^{\infty} (-1)^k \frac{t^{k+1}}{k+1}, \qquad \text{per } |\, t \,| < 1. \tag{26}$$

Inoltre le due serie precedenti sono assolutamente convergenti.

Si noti che dalla (25) segue immediatamente che

$$-\log(1-t) = t + \frac{t^2}{2} + R_3(t)$$

con $R_3(t) = O(t^3)$, infatti

$$|R_3(t)| = \big| -\log(1-t) - [t + \tfrac{t^2}{2}]\big| \leq \sum_{k=2}^{\infty} \frac{|\, t \,|^{k+1}}{k+1} |\leq \sum_{k=2}^{\infty} \frac{1}{3} \,|\, t \,|^{k+1} = \frac{|\, t \,|^3}{3(1-|\, t \,|)} \qquad \text{per } |\, t \,| < 1. \tag{27}$$

UN ALTRO SVILUPPO IN SERIE

Nella dimostrazione della Formula di STIRLING si utilizza la seguente identità, che deriva dallo sviluppo in serie del logaritmo.

$$\frac{1}{2t} \log\left(\tfrac{1+t}{1-t}\right) = \sum_{h=0}^{\infty} \frac{t^{2h}}{2h+1}, \qquad \text{per } |\, t \,| < 1. \tag{28}$$

Infatti

$$\frac{1}{2t} \log\left(\tfrac{1+t}{1-t}\right) = \frac{1}{2t}\Big[\log(1+t) - \log(1-t)\Big] = \frac{1}{2t}\Big[\sum_{k=0}^{\infty}(-1)^k \frac{t^{k+1}}{k+1} - \Big(-\sum_{k=0}^{\infty} \frac{t^{k+1}}{k+1}\Big)\Big]$$

$$= \frac{1}{2t}\Big\{\sum_{k=0}^{\infty}(-1)^k \frac{t^{k+1}}{k+1} + \sum_{k=0}^{\infty} \frac{t^{k+1}}{k+1}\Big\} = \frac{1}{2t}\Big\{\sum_{k=0}^{\infty}\Big[(-1)^k \frac{t^{k+1}}{k+1} + \frac{t^{k+1}}{k+1}\Big]\Big\}$$

in quanto le due serie sono assolutamente convergenti. Di conseguenza

$$\frac{1}{2t} \log\left(\tfrac{1+t}{1-t}\right) = \frac{1}{2t} \sum_{\substack{k=0\\k=2h}}^{\infty} 2\frac{t^{k+1}}{k+1} = \frac{1}{t} \sum_{h=0}^{\infty} \frac{t^{2h+1}}{2h+1} = \sum_{h=0}^{\infty} \frac{t^{2h}}{2h+1}.$$