Valore atteso

Sia $X : \Omega \equiv \{\omega_1, \ldots, \omega_N\} \to X(\Omega) \equiv \{x_1, \ldots, x_n\} \subseteq \mathbb{R}$ una variabile aleatoria, dove $X(\Omega)$ è l'insieme dei valori che può assumere la variabile aleatoria X. Se $\mathbb{P}$ è una probabilità su $(\Omega, \mathcal{P}(\Omega))$, con $\mathbb{P}(A) = \sum_{\omega_i \in A} p(\omega_i)$, allora il valore atteso di X (rispetto a $\mathbb{P}$) è dato dalla media di $X(\omega_i)$ pesata con $p(\omega_i)$, cioè

$$\mathbb{E}(X) := \sum_{\omega_i \in \Omega} p(\omega_i) X(\omega_i)$$

La densità discreta di X è data da $p_X : X(\Omega) \to [0,1];\ x \mapsto p_X(x) := \mathbb{P}(X = x)$, dove $\mathbb{P}(X = x)$ è la probabilità di $H_x^X := \{\omega \in \Omega : X(\omega) = x\}$ e quindi

$$p_X(x) := \mathbb{P}(X = x) = \sum_{\omega_i \in H_x^X} p(\omega_i) = \sum_{\omega_i :\, X(\omega_i) = x} p(\omega_i)$$

Il valore atteso di X dipende solo dalla sua densità discreta, ossia vale

$$\mathbb{E}(X) := \sum_{\omega_i \in \Omega} p(\omega_i) X(\omega_i) = \sum_{k=1}^{n} x_k \, p_X(x_k)$$

Una dimostrazione del fatto che $\mathbb{E}(X)$ dipende solo dalla distribuzione è basata sul calcolo della densità discreta di X: infatti da

$$p_X(x) := \mathbb{P}(X = x) = \sum_{\omega_i \in H_x^X} p(\omega_i) = \sum_{\omega_i :\, X(\omega_i)=x} p(\omega_i)$$

si vede subito che si può riscrivere il valore atteso sommando prima sugli $\omega_i \in H_{x_1}^X$, poi sugli $\omega_i \in H_{x_2}^X$, e così via, in quanto la famiglia $\{H_{x_1}^X, H_{x_2}^X, \cdots, H_{x_n}^X\}$ è una partizione (ossia gli eventi $H_{x_k}^X$ sono incompatibili a due a due, ed esaustivi) per cui, da

$$\mathbb{E}(X) = \sum_{\omega_i \in \Omega} p(\omega_i) X(\omega_i)$$

si ottiene

$$\mathbb{E}(X) = \sum_{\omega_i \in H_{x_1}^X} \underbrace{X(\omega_i)}_{=x_1} p(\omega_i) + \sum_{\omega_i \in H_{x_2}^X} \underbrace{X(\omega_i)}_{=x_2} p(\omega_i) + \cdots + \sum_{\omega_i \in H_{x_n}^X} \underbrace{X(\omega_i)}_{=x_n} p(\omega_i)$$

$$= \sum_{\omega_i \in H_{x_1}^X} x_1\, p(\omega_i) + \sum_{\omega_i \in H_{x_2}^X} x_2\, p(\omega_i) + \cdots + \sum_{\omega_i \in H_{x_n}^X} x_n\, p(\omega_i) = \sum_{k=1}^{n} x_k \underbrace{\sum_{\omega_i \in H_{x_k}^X} p(\omega_i)}_{\mathbb{P}(H_{x_k}^X) =: p_X(x_k)}$$

e quindi

$$\mathbb{E}(X) = \sum_{k=1}^{n} x_k\, p_X(x_k) = \sum_{k=1}^{n} x_k\, \mathbb{P}(X = x_k)$$

Varianza

Sia $X : \Omega \equiv \{\omega_1, \dots, \omega_N\} \to X(\Omega) \equiv \{x_1, \dots, x_n\}$ una variabile aleatoria ed indichiamo brevemente con μ il valore atteso di X. La ***varianza*** di X viene definita come il valore atteso di $(X-\mu)^2$ e si indica brevemente con il simbolo $Var(X)$, cioè

$$Var(X) = \mathbb{E}\left((X-\mathbb{E}(X))^2\right) \equiv \sum_{\omega_i \in \Omega} p(\omega_i)\,(X(\omega_i)-\mu)^2 .$$

Come il valore atteso $\mathbb{E}(X)$, anche la varianza $Var(X)$ dipende soltanto dalla distribuzione di probabilità di X, ossia, in sostanza, dalla densità discreta di X: $p_X(x_k) := \mathbb{P}(X = x_k)$, per $x_k \in X(\Omega)$.
Si veda a questo proposito anche la ***Proposizione* 9.7 di [SN]** e la formula $\mathbb{E}[h(X)] = \sum_{k=1}^n h(x_k)\,p_X(x_k)$, con $h(x) = (x-\mu_X)^2$, dove $\mu_X := \mathbb{E}[X]$), per cui equivalentemente vale

$$Var(X) = \sum_{k=1}^{n} (x_k - \mu_X)^2\, p_X(x_k)$$

Proprietà della Varianza

$\boxed{Var(X) \geq 0}$ e inoltre $Var(X) = 0 \Leftrightarrow$ esiste $\hat{x}$ tale che $\mathbb{P}(X = \hat{x}) = 1$

$$\boxed{Var(X) = \mathbb{E}(X^2) - \left(\mathbb{E}(X)\right)^2}$$

$$\boxed{Var(aX + b) = a^2\, Var(X), \text{ per ogni } a, b \in \mathbb{R}}$$

$$\boxed{Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)}$$

dove $Cov(X, Y)$ è la **covarianza di X e Y**

$$\boxed{Cov(X, Y) := \mathbb{E}\left[\left(X - \mathbb{E}(X)\right)\left(Y - \mathbb{E}(Y)\right)\right] = \ldots = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}$$

A che serve la varianza? (i)

In generale il valore atteso di una variabile aleatoria X non basta per dare un'idea, anche se "grossolana" della distribuzione di una variabile aleatoria. La varianza ci fornisce un elemento in più e di permette di farci un'idea di quanto la distribuzione di X sia "vicina" (o "lontana") dal suo valore atteso.

Ad esempio se $X \sim Unif(\{-1,0,1\})$, allora $\mathbb{E}(X) = 0$ e $Var(X) = \frac{2}{3}$

$\mathbb{E}(X) = -1\mathbb{P}(X=-1) + 0\mathbb{P}(X=0) + 1\mathbb{P}(X=1) = -\frac{1}{3} + 0 + \frac{1}{3}$

e

$$Var(X) = \mathbb{E}\left[(X-0)^2\right] = (-1-0)^2\mathbb{P}(X=-1) + (0-0)^2\mathbb{P}(X=0) + (1-0)^2\mathbb{P}(X=1) =$$
$$= (-1)^2\frac{1}{3} + 0 + \frac{1}{3} = \frac{2}{3}$$

Se invece $Y \sim Unif(\{-2,-1,0,1,2\})$,
allora $\mathbb{E}(Y) = \mathbb{E}(X) = 0$, ma $Var(Y) = 2 > Var(X) = \frac{2}{3}$

$$\mathbb{E}(Y) = -2\mathbb{P}(X=-2) - 1\mathbb{P}(Y=-1) + 0\mathbb{P}(Y=0) + 1\mathbb{P}(Y=1) + 2\mathbb{P}(Y=2)$$
$$= -2\,\frac{1}{5} - \frac{1}{5} + 0 + \frac{1}{5} + 2\,\frac{1}{5} = 0$$

e

$$Var(Y) = \mathbb{E}\left[(Y-0)^2\right] = \sum_{k=-2}^{2}(k-0)^2\mathbb{P}(Y=k) = (-2)^2\,\frac{1}{5} + (-1)^2 + 0^2\frac{1}{5} + 1^2\frac{1}{5} + 2^2\,\frac{1}{5} =$$
$$= \frac{4+1+0+1+4}{5} = 2$$

La disuguaglianza di Chebyshev

Per rendere più preciso quanto appena detto si ricorre alla disuguaglianza di Chebyshev.

Proposizione [Disuguaglianza di Chebyshev]

Sia X una variabile aleatoria, con valore atteso $\mathbb{E}(X) = \mu_X$ e varianza $Var(X) = \sigma_X^2$. Allora **per ogni** $\varepsilon > \mathbf{0}$

$$\mathbb{P}(\{|X - \mu_X| > \varepsilon\}) \leq \frac{\sigma_X^2}{\varepsilon^2}, \quad \Leftrightarrow \quad \mathbb{P}\big(\{|X - \mathbb{E}(X)| > \varepsilon\}\big) \leq \frac{Var(X)}{\varepsilon^2}.$$

Si osservi che ovviamente nella diseguaglianza di Chebyshev si prende $\varepsilon > 0$ in quanto, per $\varepsilon < 0$ si ha $\mathbb{P}(\{|X - \mu_X| > \varepsilon\}) = \mathbb{P}(\Omega) = 1$, mentre per $\varepsilon = 0$ non avrebbe senso il secondo membro della diseguaglianza.
Infine va osservato che la diseguaglianza di Chebyshev è interessante solo se $\frac{\sigma_X^2}{\varepsilon^2} < 1$, in quanto in caso contrario si ottiene solo una banalità, ovvero che $\mathbb{P}(\{|X - \mu_X| > \varepsilon\})$ è minore o uguale di un numero maggiore o uguale di 1.

La dimostrazione segue immediatamente dalla definizione stessa di varianza, infatti possiamo scrivere

$$
\begin{aligned}
\boxed{\sigma_X^2} &= \mathbb{E}\big((X-\mu_X)^2\big) = \sum_{\omega_i \in \Omega} p(\omega_i)\cdot(X(\omega_i)-\mu_X)^2 \\
&= \sum_{\omega_i : |X(\omega_i)-\mu_X|>\varepsilon} p(\omega_i)\cdot(X(\omega_i)-\mu_X)^2 + \sum_{\omega_i : |X(\omega_i)-\mu_X|\leq\varepsilon} \underbrace{p(\omega_i)\cdot(X(\omega_i)-\mu_X)^2}_{\geq 0} \\
&\geq \sum_{\omega_i : |X(\omega_i)-\mu_X|>\varepsilon} p(\omega_i)\cdot(X(\omega_i)-\mu_X)^2 \geq \sum_{\omega_i : |X(\omega_i)-\mu_X|>\varepsilon} p(\omega_i)\cdot\varepsilon^2 \\
&= \varepsilon^2 \cdot \sum_{\omega_i : |X(\omega_i)-\mu_X|>\varepsilon} p(\omega_i) = \boxed{\varepsilon^2 \cdot \mathbb{P}(\{|X-\mu_X| > \varepsilon\})}
\end{aligned}
$$

La dimostrazione è quindi terminata:

$$
\sigma_X^2 \geq \varepsilon^2 \cdot \mathbb{P}(\{|X-\mu_X| > \varepsilon\}) \quad \Leftrightarrow \quad \mathbb{P}(\{|X-\mu_X| > \varepsilon\}) \leq \frac{\sigma_X^2}{\varepsilon^2}
$$

Si osservi che la dimostrazione della Disuguaglianza di Chebyshev rimane pressoché invariata se invece di considerare l'evento l'evento $\{|X-\mu| \geq \varepsilon\}$, per cui vale anche

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

e che, essendo $\{|X-\mu| > \varepsilon\} \subseteq \{|X-\mu| \geq \varepsilon\}$, la precedente disuguaglianza è anche più forte della disuguaglianza appena dimostrata.

A che serve la varianza? (ii)

Il valore atteso è un parametro interessante, che a volte caratterizza addirittura una distribuzione (ad esempio il parametro di una variabile aleatoria di Poisson coincide con il suo valore atteso). Tuttavia spesso non è noto e siamo interessati a cercare un modo per ottenerlo. In questo senso ci aiuta la Legge dei Grandi Numeri, nella cui dimostrazione la varianza gioca un ruolo fondamentale.

Legge debole dei Grandi Numeri(LGN) "informale": La media aritmetica di n variabili aleatorie "simili" e "indipendenti" è, con grande probabilità, vicina al loro valore atteso comune, **purché *n* sia "abbastanza grande"**.
IMPORTANTE: come vedremo in seguito, per sapere quanto basta che n sia "grande" è sufficiente conoscere la varianza comune delle n variabili aleatorie.
In particolare la frequenza relativa del numero dei successi è, con grande probabilità, vicina alla comune probabilità di successo delle n prove, **purché *n* sia "abbastanza grande"**.

Per rendere la Legge dei Grandi Numeri un vero teorema, come vedremo, entra in gioco la varianza, insieme alla nozione di indipendenza di variabili aleatorie[1].

Teorema (verso la Legge debole dei Grandi Numeri: il caso di un numero finito di variabili aleatorie)
Siano date n variabili aleatorie X_1, X_2, ..., X_n, tutte con **la stessa distribuzione e indipendenti**. Allora, posto

$$S_n := \sum_{i=1}^{n} X_i, \quad \mu_X := \mathbb{E}(X_i) \quad \text{e} \quad \sigma_X^2 := Var(X_i),$$

per ogni numero reale $\varepsilon > 0$ si ha che

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

[1]Nel caso Ω finito chiedere che X_1, X_2, ..., X_n abbiano la stessa distribuzione significa chiedere che abbiano la stessa densità: per ogni i si ha $X_i(\Omega) = \{x_1, x_2, ..., x_\ell\}$ e $\mathbb{P}(X_i = x) = \mathbb{P}(X_1 = x)$, per ogni $x \in X(\Omega)$. Inoltre chiedere che siano indipendenti equivale a chiedere che siano indipendenti le partizioni $\mathcal{P}_i$, $i = 1, 2, \cdots, n$, dove, per ogni i,

$$\mathcal{P}_i := \left\{H_1^{X_i} = \{X_i = x_1\}, H_2^{X_i} = \{X_i = x_2\}, \cdots, H_n^{X_i} = \{X = x_n\}\right\}$$

è la partizione generata dalla variabile aleatoria X_i.

La dimostrazione/verifica è basata su

$$\boxed{\mathbb{E}\left(\frac{S_n}{n}\right)} = \frac{1}{n}\mathbb{E}(S_n) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\left(\sum_{i=1}^{n} \underbrace{\mathbb{E}(X_i)}_{=\mu_X}\right) = \frac{1}{n} n\mu_X \boxed{= \mu_X}$$

la formula per calcolare la varianza di una somma di variabili aleatorie

$$Var\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} Var(X_k) + \sum\sum_{h\neq k}^{1\leq h,k\leq n} Cov(X_h, X_k)$$

e il fatto che

$$\boxed{X_1, \ldots, X_n \text{ indipendenti}} \Rightarrow \boxed{\forall\, i \neq j\ X_i \text{ e } X_j \text{ indipendenti}} \Rightarrow \boxed{\forall\, i \neq j\ Cov(X_i, X_j) = 0}$$

da cui $\boxed{Var\left(\frac{S_n}{n}\right)} = Var\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \left(\frac{1}{n}\right)^2 Var\left(\sum_{k=1}^{n} X_k\right)$

$$= \frac{1}{n^2}\left[\sum_{k=1}^{n} \overbrace{Var(X_k)}^{=\sigma_X^2} + \sum\sum_{h\neq k}^{1\leq h,k\leq n} \overbrace{Cov(X_h, X_k)}^{=0}\right] = \frac{1}{n^2} n\sigma_X^2 \boxed{= \frac{\sigma_X^2}{n}}$$

e sulla disuguaglianza di Chebyshev applicata alla variabile aleatoria $\frac{S_n}{n}$:

$$\boxed{\mathbb{P}\left(\left|\frac{S_n}{n}-\mu_X\right|>\varepsilon\right)}=\mathbb{P}\left(\left|\frac{S_n}{n}-\mathbb{E}\left(\frac{S_n}{n}\right)\right|>\varepsilon\right)\leq\frac{Var\left(\frac{S_n}{n}\right)}{\varepsilon^2}=\boxed{\frac{\sigma_X^2}{n\cdot\varepsilon^2}}$$

Possiamo pensare che X_1, X_2, ..., X_n rappresentino i valori di una stessa quantità (ad esempio temperatura, o pressione, o altro) che osserveremo in n esperimenti effettuati con le stesse modalità.
La legge dei grandi numeri assicura che, fissata una soglia $\varepsilon > 0$ (ovviamente $\varepsilon > 0$ è un numero "piccolo") purché n sia abbastanza grande, possiamo affermare che con probabilità grande, e precisamente almeno $1 - \frac{\sigma_X^2}{n \cdot \varepsilon^2}$, la media aritmetica di X_1, X_2, ..., X_n differirà di meno di ε dal comune valore atteso μ_X:

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

intervallo di fiducia

L'intervallo (con estremi aleatori) $\left[\frac{S_n}{n} - \varepsilon, \frac{S_n}{n} + \varepsilon \right]$ è detto intervallo di fiducia del parametro μ_X, ossia al valore atteso della distribuzione comune a tutte le variabili aleatorie X_1, X_2, ..., X_n

Un caso particolarmente interessante è quello in cui $X_i = \mathbf{1}_{A_i}$ dove $\{A_i, 1 \leq i \leq n\}$ sono eventi indipendenti, tutti con la stessa probabilità, ossia $\mathbb{P}(A_i) = \mathbb{P}(A_1)$, per ogni $i = 1, 2, \dots, n$, ma non conosciamo il valore esatto di $p := \mathbb{P}(A) = \mathbb{E}\big(\mathbf{1}_{A_i}\big) = \mathbb{E}(X_i) = \mu_X$, di modo che

$$\frac{S_n}{n} = \text{frequenza relativa dei successi è una “stima” di } p$$

Ovviamente possiamo affermare che, con probabilità almeno $1 - \frac{\sigma_X^2}{n \cdot \varepsilon^2}$, si ha che

$$\left|\frac{S_n}{n} - p\right| \leq \varepsilon \quad \Leftrightarrow \quad -\varepsilon \leq p - \frac{S_n}{n} \leq \varepsilon \quad \Leftrightarrow \quad \frac{S_n}{n} - \varepsilon \leq p \leq \frac{S_n}{n} + \varepsilon.$$

Se volessimo essere sicuri che tale probabilità sia almeno $1 - \delta$ (valori tipici sono $\delta = 5\% = \frac{5}{100}$ oppure $\delta = 1\% = \frac{1}{100}$)

Se fissiamo il numero di esperimenti n **e il valore** δ basta uguagliare

$$1 - \frac{\sigma_X^2}{n \cdot \varepsilon^2} = 1 - \delta \quad \Leftrightarrow \quad \frac{\sigma_X^2}{n \cdot \varepsilon^2} = \delta$$

per ottenere che $\quad \varepsilon^2 = \frac{\sigma_X^2}{n \cdot \delta} \quad \Leftrightarrow \quad \varepsilon = \frac{\sigma_X}{\sqrt{n} \cdot \sqrt{\delta}} \quad$ e quindi

$$\mathbb{P}\left(\frac{S_n}{n} - \frac{\sigma_X}{\sqrt{n} \cdot \sqrt{\delta}} \leq p \leq \frac{S_n}{n} + \frac{\sigma_X}{\sqrt{n} \cdot \sqrt{\delta}}\right) \geq 1 - \delta.$$

ATTENZIONE: nel ragionamento precedente sorge immediatamente un problema:

$$\text{NON CONOSCIAMO } \mu_X = p,$$

e quindi

$$\text{NON CONOSCIAMO NEMMENO } \sigma_X^2 = Var(X_i) = Var\left(\mathbf{1}_{A_i}\right) = p(1-p)$$

COSA POSSIAMO FARE?

Sappiamo però che, per $x \in [0,1]$ la funzione $x \mapsto x(1-x)$ ammette valore massimo uguale a $1/4$ (il punto di massimo è in $x_{MAX} = \frac{1}{2}$), per cui:

$$\sigma_X^2 := p(1-p) \leq \frac{1}{4}$$

e quindi l'intervallo

$$\left[\frac{S_n}{n} - \frac{1}{4 \cdot \sqrt{n} \cdot \sqrt{\delta}}, \frac{S_n}{n} + \frac{1}{4 \cdot \sqrt{n} \cdot \sqrt{\delta}}\right] \supseteq \left[\frac{S_n}{n} - \frac{\sigma_X^2}{\sqrt{n} \cdot \sqrt{\delta}}, \frac{S_n}{n} + \frac{\sigma_X^2}{\sqrt{n} \cdot \sqrt{\delta}},\right]$$

e di conseguenza

$$\begin{aligned} &\mathbb{P}\left(\frac{S_n}{n} - \frac{1}{4 \cdot \sqrt{n} \cdot \sqrt{\delta}} \leq p \leq \frac{S_n}{n} + \frac{1}{4 \cdot \sqrt{n} \cdot \sqrt{\delta}}\right) \\ \geq &\mathbb{P}\left(\frac{S_n}{n} - \frac{\sigma_X}{\sqrt{n} \cdot \sqrt{\delta}} \leq p \leq \frac{S_n}{n} + \frac{\sigma_X}{\sqrt{n} \cdot \sqrt{\delta}}\right) \geq 1 - \delta. \end{aligned}$$

Se invece fissiamo l'ampiezza dell'intervallo (ossia fissiamo ε) e δ, allora, per trovare n abbastanza grande in modo che la probabilità che *la frequenza relativa* $\frac{S_n}{n}$ *e il valore p differiscano (in valore assoluto) meno di* ε sia almeno $1-\delta$, è sufficiente richiedere che

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

In altre parole, posto $n(\varepsilon,\delta):=\lceil\frac{\sigma_X^2}{\delta\cdot\varepsilon^2}\rceil$, (dove $\lceil x\rceil$ denota la parte intera superiore di x), si ha che

$$\mathbb{P}\Big(\frac{S_n}{n}-\varepsilon\leq p\leq\frac{S_n}{n}+\varepsilon\Big)\geq 1-\delta,\qquad \text{per ogni } n\geq n(\varepsilon,\delta)$$

L'enunciato della Legge di Grandi Numeri viene più spesso formulato nel seguente modo, a partire da una successione di variabili aleatorie

Teorema (Legge debole dei Grandi Numeri: per una successione di variabili aleatorie) Sia data una successione di variabili aleatorie $\{X_n;\, n \geq 1\}$, tutte con **la stessa distribuzione e indipendenti.** Allora, posto

$$S_n := \sum_{i=1}^{n} X_i, \quad \mu_X := \mathbb{E}(X_i) \quad \text{e} \quad \sigma_X^2 := Var(X_i),$$

per ogni numero reale $\varepsilon > 0$ si ha che

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

La dimostrazione è immediata perché basta osservare che per ogni n

$$1 \geq \mathbb{P}\left(\left|\frac{S_n}{n} - \mu_X\right| \leq \varepsilon\right) \geq \underbrace{1 - \frac{\sigma_X^2}{n \cdot \varepsilon^2}}_{\to_{n\to\infty} 1} \quad \Leftrightarrow \quad 0 \leq \mathbb{P}\left(\left|\frac{S_n}{n} - \mu_X\right| > \varepsilon\right) \leq \underbrace{\frac{\sigma_X^2}{n \cdot \varepsilon^2}}_{\to_{n\to\infty} 0}$$

e usare il Teroema del confronto (anche detto "*teorema dei carabinieri*")

Abbiamo già calcolato[2] il valore atteso e la varianza di una variabile aleatoria $Bin(n, \theta)$ e di una variabile aleatoria $Hyp(M, m_1; n)$ usando il fatto che hanno lo stesso valore atteso e la stessa varianza delle rispettive variabili aleatorie "prototipo", e la **proprietà di linearità** del valore atteso e la **formula della varianza della somma** di variabili aleatorie.

Nella prossima slide vedremo come ottenere il valore atteso e la varianza usando solo la densità discreta: si tratta di calcoli abbastanza lunghi, che fanno apprezzare la "fatica" fatta per mostrare le proprietà appena citate.

[2] vedere gli Appunti [SN] e il file SCHEMA-VARIABILI-ALEATORIE-16-novembre-2024.pdf

Valore atteso di $X \sim Bin(n, \theta)$, con la densità discreta

Per ottenere che $\mathbb{E}(X) = n\theta$, si può procedere anche usando la densità discreta, ma a costo di un po' di calcoli, che invece la proprietà di linearità del valore atteso ci permette di risparmiare:

$$\mathbb{E}(X) = \sum_{k=0}^{n} k\, p_X(k) = \sum_{k=0}^{n} k\, \mathbb{P}(X = k) = \sum_{k=0}^{n} k \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

$$= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} \theta^k (1-\theta)^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} \theta^k (1-\theta)^{n-k},$$

e quindi, ponendo $h = k-1$, e tenendo conto che $1 \leq k \leq n \Leftrightarrow 0 \leq h \leq n-1$, che $n-k = n-1-(k-1) = n-1-h$, $n! = n \cdot (n-1)!$, $k = h+1$, si ha

$$\mathbb{E}(X) = \sum_{h=0}^{n-1} n \frac{(n-1)!}{h!(n-1-h)!} \theta^{h+1} (1-\theta)^{n-1-h}$$

$$= n\theta \sum_{h=0}^{n-1} \frac{(n-1)!}{h!(n-1-h)!} \theta^h (1-\theta)^{n-1-h} = n\theta \sum_{h=0}^{n-1} \binom{n-1}{h} \theta^h (1-\theta)^{n-1-h},$$

infine, tenendo conto dello sviluppo della potenza del binomio, otteniamo

$$\mathbb{E}(X) = n\theta(\theta + (1-\theta))^{n-1} = n\theta \cdot 1^{n-1} = n\theta.$$

Varianza di $X \sim Bin(n, \theta)$, con la densità discreta

Per ottenere che $Var(X) = n\theta(1-\theta)$, si può procedere anche usando la densità discreta, ma a costo di un po' di calcoli, che invece la formula della varianza di una somma ci permette di risparmiare. Prima di tutto si usa il fatto che

$$Var(X) := \mathbb{E}\left[\left(X - \mathbb{E}(X)\right)^2\right] \left(= \mathbb{E}\left[X^2 - 2\mathbb{E}(X) \cdot X + (\mathbb{E}(X))^2\right]\right.$$

$$\left. = \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + (\mathbb{E}(X))^2 \right) = \mathbb{E}(X^2) - \left(\mathbb{E}(X)\right)^2$$

Inoltre, da $X^2 = X^2 - X + X = X(X-1) + X$ e dalla proprietà di linearità del valore atteso si ottiene

$$Var(X) = \overbrace{\mathbb{E}\left(X(X-1)\right)}^{=n\cdot(n-1)\theta^2} + \overbrace{\mathbb{E}(X)}^{=n\theta} - \overbrace{\left(\mathbb{E}(X)\right)^2}^{=n^2\theta^2} = n\theta\,\overbrace{\left((n-1)\theta + 1 - n\theta\right)}^{=1-\theta} = n\theta(1-\theta)$$

(per il calcolo di $\mathbb{E}\left(X(X-1)\right)$ vedere la prossima pagina)

$$\mathbb{E}\big(X(X-1)\big) = \sum_{k=0}^{n} k(k-1)\, p_X(k) = \sum_{k=0}^{n} k(k-1) \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

$$= \sum_{k=2}^{n} k(k-1) \frac{n!}{k!(n-k)!} \theta^k (1-\theta)^{n-k} = \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} \theta^k (1-\theta)^{n-k},$$

e quindi, ponendo $h = k-2$, e tenendo conto che $2 \leq k \leq n \Leftrightarrow 0 \leq h \leq n-2$, che $n-k = n-2-(k-2) = n-2-h$, $n! = n \cdot (n-1) \cdot (n-2)!$, $k = h+2$, si ha

$$\mathbb{E}\big[X(X-1)\big] = \sum_{h=0}^{n-2} n \cdot n(-1) \frac{(n-2)!}{h!(n-2-h)!} \theta^{h+2} (1-\theta)^{n-2-h}$$

$$= n \cdot (n-1)\, \theta^2 \sum_{h=0}^{n-2} \frac{(n-2)!}{h!(n-2-h)!} \theta^h (1-\theta)^{n-2-h}$$

$$= n \cdot (n-1)\, \theta^2 \sum_{h=0}^{n-2} \binom{n-2}{h} \theta^h (1-\theta)^{n-2-h},$$

infine, tenendo conto dello sviluppo della potenza del binomio, otteniamo

$$\mathbb{E}\big[X(X-1)\big] = n \cdot (n-1)\, \theta^2 \big(\theta + (1-\theta)\big)^{n-2} = n \cdot (n-1)\, \theta^2 \cdot 1^{n-2} = n \cdot (n-1)\, \theta^2.$$

Valore atteso di $X \sim Hyp(M, m_1; n)$, con la densità discreta

Poniamo, come al solito $m_2 = M - m_1$,

$$\mathbb{E}(X) = \sum_{k=0}^{n} k\, p_X(k) = \sum_{k=1}^{n} k\, \mathbb{P}(X=k) = \sum_{k\geq 1}^{*} k\, \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{m_1+m_2}{n}},$$

dove la somma con $*$ è estesa ai valori k per cui l'addendo ha senso ed è diverso da zero: $k \leq m_1$, $n-k \leq m_2$ e $1 < n \leq M = m_1 + m_2$. Ricordando che per i coefficienti binomiali vale la seguente indentità *(pensare all'esempio della giunta con ℓ membri di cui uno è il presidente)*

$$\binom{r}{\ell} \cdot \ell = r \cdot \binom{r-1}{\ell-1}, \quad \text{e quindi} \quad \binom{r}{\ell} = \frac{r}{\ell} \cdot \binom{r-1}{\ell-1}$$

si ottiene allora

$$\mathbb{E}(X) = \sum_{k\geq 1}^{*} \frac{\cancel{k} \cdot \frac{m_1}{\cancel{k}} \binom{m_1-1}{k-1}\binom{m_2}{n-k}}{\frac{m_1+m_2}{n}\binom{m_1+m_2-1}{n-1}} = n\, \frac{m_1}{m_1+m_2} \overbrace{\sum_{k\geq 1}^{*} \frac{\binom{m_1-1}{k-1}\binom{m_2}{n-k}}{\binom{m_1+m_2-1}{n-1}}}^{=1} = n\,\frac{m_1}{M},$$

in quanto, ponendo $h = k-1$, da cui $0 \leq h \leq m_1 - 1$ e $0 \leq n-1-h = n-k \leq m_2$,

$$\sum_{k\geq 1}^{*} \frac{\binom{m_1-1}{k-1}\binom{m_2}{n-k}}{\binom{m_1+m_2-1}{n-1}} = \sum_{h\geq 0}^{**} \frac{\binom{m_1-1}{h}\binom{m_2}{n-1-h}}{\binom{m_1+m_2-1}{n-1}} = \sum_{h\geq 0}^{**} p_{X'}(h) = 1 \quad (X' \sim Hyp(M-1, m_1-1; n-1))$$

Varianza di $X \sim Hyp(M, m_1; n)$, con la densità discreta

Per ottenere che $Var(X) = n \cdot \frac{m_1}{M} \cdot \left(1 - \frac{m_1}{M}\right) \left[1 - \frac{n-1}{M-1}\right]$ procederemo in modo analogo a quanto fatto nel caso di una variabile aleatoria con distribuzione $Bin(n, \theta)$: sappiamo già che

$$Var(X) = \mathbb{E}\big(X(X-1)\big) + \mathbb{E}(X) - \big(\mathbb{E}(X)\big)^2$$

Basterà quindi mostrare che $\mathbb{E}\big(X(X-1)\big) = n \cdot (n-1)\, \frac{m_1}{M}\, \frac{m_1-1}{M-1}$, per ottenere che

$$Var(X) = \overbrace{\mathbb{E}\big(X(X-1)\big)}^{= n\cdot(n-1)\frac{m_1}{M}\frac{m_1-1}{M-1}} + \overbrace{\mathbb{E}(X)}^{= n\frac{m_1}{M}} - \overbrace{\big(\mathbb{E}(X)\big)^2}^{= n^2\left(\frac{m_1}{M}\right)^2} = n\cdot(n-1)\,\frac{m_1}{M}\,\frac{m_1-1}{M-1} + n\,\frac{m_1}{M} - \left(n\,\frac{m_1}{M}\right)^2$$

$$= n\,\frac{m_1}{M}\left[(n-1)\,\frac{m_1-1}{M-1} + 1 - n\,\frac{m_1}{M}\right] = n\,\frac{m_1}{M}\left(1 - \frac{m_1}{M}\right)\left[1 - \frac{n-1}{M-1}\right]$$

in quanto $\left[(n-1)\,\frac{m_1-1}{M-1} + 1 - n\,\frac{m_1}{M}\right] = \frac{(n-1)(m_1-1)M + M(M-1) - n\,m_1\,(M-1)}{M(M-1)} =$

$$= \frac{(nm_1 - n - m_1 + 1)M + M^2 - M - n\,m_1\,M + nm_1}{M(M-1)} = \frac{-nM - m_1 M + M^2 + nm_1}{M(M-1)}$$

coincide con $\left(1 - \frac{m_1}{M}\right)\left[1 - \frac{n-1}{M-1}\right] = \frac{(M-m_1)(M-\cancel{1}-(n-\cancel{1}))}{M(M-1)} = \frac{M^2 - m_1 M - nM + nm_1}{M(M-1)}$

Per il calcolo di $\mathbb{E}\big[X(X-1)\big]$ poniamo, come al solito, $m_2 := M - m_1$,

$$\mathbb{E}\big[X(X-1)\big] = \sum_{k=0}^{n} k(k-1)\,p_X(k) = \sum_{k\geq 2}^{*} k(k-1)\,\frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{m_1+m_2}{n}},$$

dove la somma con $*$ è estesa ai valori k per cui l'addendo ha senso ed è diverso da zero: $k \leq m_1$, $n-k \leq m_2$ e $1 < n \leq M = m_1 + m_2$.

Estendendo l'identità già usata per il calcolo del valore atteso, si ha

$$\binom{r}{\ell} = \frac{r}{\ell}\cdot\binom{r-1}{\ell-1} \quad \text{e quindi} \quad \binom{r}{\ell} = \frac{r}{\ell}\cdot\frac{r-1}{\ell-1}\cdot\binom{r-2}{\ell-2}$$

e quindi si ottiene

$$\mathbb{E}\big[X(X-1)\big] = \sum_{k\geq 1}^{*} \frac{\cancel{k(k-1)}\cdot\frac{m_1(m-1)}{\cancel{k(k-1)}}\binom{m_1-2}{k-2}\binom{m_2}{n-k}}{\frac{(m_1+m_2)(m_1+m_2-1)}{n(n-1)}\binom{m_1+m_2-2}{n-2}}$$

$$= n(n-1)\,\frac{m_1}{m_1+m_2}\,\frac{m_1-1}{m_1+m_2-1}\overbrace{\sum_{k\geq 2}^{*}\frac{\binom{m_1-2}{k-2}\binom{m_2}{n-k}}{\binom{m_1+m_2-1}{n-1}}}^{=1} = n(n-1)\,\frac{m_1}{M}\,\frac{m_1-1}{M-1},$$

in quanto, ponendo $h = k-2$, da cui $0 \leq h \leq m_1 - 2$ e $0 \leq n-2-h = n-k \leq m_2$,

$$\sum_{k\geq 2}^{*}\frac{\binom{m_1-2}{k-2}\binom{m_2}{n-k}}{\binom{m_1+m_2-2}{n-2}} = \sum_{h\geq 0}^{**}\frac{\binom{m_1-2}{h}\binom{m_2}{n-2-h}}{\binom{m_1+m_2-2}{n-2}} = \sum_{h\geq 0}^{**} p_{X''}(h) = 1 \quad (X'' \sim Hyp(M-2, m_1-2; n-2))$$

Esempio 3.4 rivisitato ("Paradosso del Cavalier De Méré").
Il numero di volte in cui si ottiene il risultato "asso" in quattro lanci di un dado è una variabile aleatoria con distribuzione $\mathit{Bin}(4, 1/6)$ e il suo valore atteso è dato da $4 \cdot 1/6 = 2/3$.

Anche il valore atteso della variabile aleatoria numero di volte in cui si presenta il doppio asso in ventiquattro lanci di una coppia di dadi vale $2/3$: infatti ha distribuzione $\mathit{Bin}(24, 1/36)$ e il cui valore atteso è dato da $24 \cdot 1/36 = 2/3$.

Possiamo concludere quindi che in entrambi i tipi di gioco d'azzardo si ha uguale valore atteso del numero dei successi.

Tuttavia, come avevamo visto nell'**Esempio 3.4**, sono diverse le probabilità di ottenere almeno un successo nelle due diverse scommesse, e questo spiega il motivo per cui questo esempio viene ricordato come "paradosso".