1. Integrazione ad alcuni argomenti del corso

Definizione. *Una variabile aleatoria X si dice di Bernoulli di parametro $p \in (0,1)$ se ha densità discreta*

$$p_X(a) = \begin{cases} p & \text{se } a = 1\,, \\ 1-p & \text{se } a = 0\,, \\ 0 & \text{altrimenti.} \end{cases}$$

Definizione. *Una variabile aleatoria X si dice binomiale di parametri n, p (dove $n \in \{1,2,3,\dots\}$ e $p \in (0,1)$) se ha densità discreta*

$$p_X(a) = \begin{cases} \binom{n}{a} p^a (1-p)^{n-a} & \text{se } a \in \{0,1,2,\dots,n-1,n\}\,, \\ 0 & \text{altrimenti.} \end{cases}$$

Definizione. *Una variabile aleatoria X si dice geometrica di parametro $p \in (0,1)$ se ha densità discreta*

$$p_X(a) = \begin{cases} (1-p)^{a-1} p & \text{se } a \in \{1,2,3,\dots\}\,, \\ 0 & \text{altrimenti.} \end{cases}$$

La v.a. geometrica gode della seguente proprietà detta *assenza di memoria*:

Proposizione. *Se X è v.a. geometrica di parametro p, allora dati n_1, n_2 interi positivi vale*

$$P(X = n_1 + n_2 | X > n_1) = P(X = n_2)\,.$$

Dim.

$$\begin{aligned} P(X = n_1 + n_2 | X > n_1) &= \frac{P(X = n_1 + n_2,\, X > n_1)}{P(X > n_1)} = \frac{P(X = n_1 + n_2)}{P(X > n_1)} \\ &= \frac{p(1-p)^{n_1+n_2-1}}{(1-p)^{n_1}} = p(1-p)^{n_2-1} = P(X = n_2)\,. \end{aligned}$$

□

Definizione. *Una variabile aleatoria X si dice di Poisson di parametro $\lambda > 0$ se ha densità discreta*

$$p_X(a) = \begin{cases} e^{-\lambda} \lambda^a / a! & \text{se } a \in \{0,1,2,3,\dots\}\,, \\ 0 & \text{altrimenti.} \end{cases}$$

La v.a. di Poisson, in certe condizioni, approssima la v.a. binomiale. Infatti, vale la seguente proprietà detta anche *legge dei piccoli numeri*:

Proposizione. *Per ogni $n \in \{1,2,3,\dots\}$ sia $p_n \in (0,1)$ tale che il limite $\lim_{n\uparrow\infty} np_n$ esiste finito e positivo. Allora, posto $X_n = Bin(n, p_n)$ e $Y = Poisson(\lambda)$ dove $\lambda = \lim_{n\uparrow\infty} np_n$, vale*

$$\lim_{n\uparrow\infty} P(X_n = k) = P(Y = \lambda)\,, \qquad \forall k \in \{0,1,2,\dots\}\,.$$

Dim. Dobbiamo dimostrare che per ogni k intero non negativo vale

$$\lim_{n\uparrow\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda} \frac{\lambda^n}{n!}\,.$$

A tal fine riscriviamo il primo membro come

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} = (n\,p_n)^k \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)(1-p_n)^n/(1-p_n)^k\,.$$

Siccome $n\,p_n \to \lambda$ abbiamo che $(n\,p_n)^k \to \lambda^k$. Inoltre, vale

$$\lim_{n\uparrow\infty}\left(1-\frac{1}{n}\right) = \lim_{n\uparrow\infty}\left(1-\frac{2}{n}\right) = \cdots = \lim_{n\uparrow\infty}\left(1-\frac{k-1}{n}\right) = 1\,.$$

Siccome $n\,p_n \to \lambda$ abbiamo che $p_n \to 0$. Quindi $(1-p_n)^k \to 1$. Infine, abbiamo che

$$\lim_{n\uparrow\infty}(1-p_n)^n = \lim_{n\uparrow\infty} e^{n\log(1-p_n)} = e^{\lim_{n\uparrow\infty}(np_n)\frac{\log(1-p_n)}{p_n}}\,.$$

L'ultimo limite si calcola ricordando che $np_n \to \lambda$, mentre per De L'Hospital vale

$$\lim_{x\downarrow 0}\frac{\log(1-x)}{x} = -\lim_{x\downarrow 0}\frac{1}{1-x} = -1\,.$$

Quindi

$$e^{\lim_{n\uparrow\infty}(np_n)\frac{\log(1-p_n)}{p_n}} = e^{-\lambda}\,.$$

Mettendo insieme tutti i limiti calcolati finora otteniamo la tesi. □

Proposizione. *Siano X, Y v.a. di Possoin indipendenti rispettivamente di parametro λ e μ, definite sullo stesso spazio di probabilità. Allora $X+Y$ è v.a. di Poisson di parametro $\lambda+\mu$.*

Dim. $X+Y$ assume con probabilità positiva solo valori interi non negativi. Sia k un numero intero non negativo. Allora

$$P(X+Y=k) = \sum_{j=0}^{k} P(X=j, Y=k-j) = \sum_{j=0}^{k} e^{-\lambda}\frac{\lambda^j}{j!}e^{-\mu}\frac{\mu^{k-j}}{(k-j)!} =$$

$$e^{-(\lambda+\mu)}\sum_{j=0}^{k}\frac{\lambda^j}{j!}\frac{\mu^{k-j}}{(k-j)!} = \frac{e^{-(\lambda+\mu)}}{k!}\sum_{j=0}^{k}\binom{k}{j}\lambda^j\mu^{k-j} = e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^k}{k!}$$

(l'ultima uguaglianza segue dal teorema del binomio). Siccome l'ultima espressione coincide con la probabilità per una v.a. di Poisson di parametro $\lambda+\mu$ di essere pari a k, abbiamo la tesi. □

2. Leggi dei grandi numeri

Supponiamo che su uno stesso spazio di probabilità sia definita una successione di variabili aleatorie X_1, X_2,... indipendenti e identicamente distribuite (scriveremo più brevemente "i.i.d."). Ricordiamo che indipendenti significa che per ogni famiglia finita di indici $i_1, i_2, \dots, i_n$ e di sottinsiemi di $\mathbb{R}$ $A_1, A_2, \dots, A_n$ vale

$$P(X_{i_1}\in A_1, X_{i_2}\in A_2, \dots, X_{i_n}\in A_n) = P(X_{i_1}\in A_1)P(X_{i_2}\in A_2)\cdots P(X_{i_n}\in A_n).$$

Identicamente distribuite significa che

$$P(X_i\in A) = P(X_j\in A) \qquad \forall i,j,\ \forall A\subset\mathbb{R}.$$

Se X_i sono v.a. discrete, allora la suddetta proprietà equivale al fatto che $p_{X_i} = p_{X_j}$ $\forall i,j$. Si noti che v.a. identicamente distribuite hanno stessa media e stessa varianza.

I teoremi limite che studieremo riguardano il comportamento asintotico di $S_n = X_1 + X_2 + \cdots + X_n$ per grandi n. Si noti che se X_i ha media finita μ allora $E(S_n) = n\mu$, mentre se X_i ha anche varianza finita pari a σ^2 allora $Var(S_n) = n\sigma^2$.

Esempio 1)
Consideriamo una successione di esperimenti identici e indipendenti. X_i vale 1 se l'evento E si è verificato all'i–esimo esperimento altrimenti X_i vale 0. Allora S_n coincide con il numero di volte in cui l'evento E si è verificato nei primi n esperimenti. Inoltre abbiamo $E(S_n) = nP(E)$ e $Var(S_n) = nP(E)(1 - P(E))$.
Esempio 2)
Se le variabili $X_1, X_2, \ldots$ sono i.i.d. con X_i Bernoulli di parametro p, allora $S_n = Bin(n, p)$.

Teorema (legge debole dei grandi numeri) *Sia $X_1, X_2, \ldots$ una successione di variabili i.i.d. definite sullo stesso spazio di probabilità aventi media finita μ. Allora, per ogni $\epsilon > 0$,*

$$\lim_{n\uparrow\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0$$

Teorema (legge forte dei grandi numeri) *Sia $X_1, X_2, \ldots$ una successione di variabili i.i.d. definite sullo stesso spazio di probabilità aventi media finita μ. Allora*

$$P\left(\lim_{n\uparrow\infty} \frac{S_n}{n} = \mu\right) = 1$$

Si può provare che la legge forte dei grandi numeri implica quella debole (da cui gli aggettivi forte, debole).

Si noti che

$$E(S_n/n) = \mu, \qquad Var(S_n/n) = Var(S_n)/n^2 = \sigma^2/n$$

Poichè la varianza di S_n/n tende a zero, ci si aspetta che S_n/n sia sempre più piccata attorno al suo valor medio μ. Quindi le suddette leggi dei grandi numeri non dovrebbero sorprendere.

Consideriamo l'esempio 1) descritto precedentemente. Allora S_n/n è la frequenza con cui l'evento E si è verificato nei primi n esperimenti. Per la legge forte dei grandi numeri otteniamo che

$$\lim_{n\uparrow\infty} \frac{S_n}{n} = P(E)$$

con probabilità 1. La legge forte dei grandi numeri corrisponde all'interpretazione della probabilità nell'approccio frequentistico, solo che qui essa è un teorema che deriva dagli assiomi di base mentre nell'approccio frequentistico viene assunta come definizione stessa di probabilità di un evento.