

Probabilità e statistica

Andrea Chelini

12 febbraio 2021

Indice

1	Esperimenti, eventi e Probabilità	3
1.1	Probabilità	3
1.1.1	Spazio campionario continuo	4
1.2	Densità	5
1.3	Probabilità condizionata	6
1.4	Distribuzioni congiunte	8
1.5	Funzione di ripartizione	9
1.6	Prodotto di convoluzione	11
1.7	Calcolo dei momenti	13
1.7.1	Media	13
1.7.2	Varianza e Covarianza	15
2	Coefficiente di correlazione	19
2.1	Disuguaglianze	19
2.2	Coefficiente di correlazione	20
2.3	Funzione generatrice dei momenti	21
3	Distribuzioni	25
3.1	Distribuzioni discrete	25
3.1.1	Distribuzione uniforme (discreta)	25
3.1.2	Binomiale	26
3.1.3	Poisson	26
3.1.4	Distribuzione geometrica	28
3.1.5	Bernulli	29
3.2	Variabili aleatorie continue	29
3.2.1	Uniforme continua	29
3.2.2	Esponenziale	30
3.2.3	Legge normale (o gaussiane)	31
3.2.4	Distribuzione gamma	32
3.2.5	χ^2	34
3.2.6	t -student	35
3.2.7	Fisher-Snedecor	35
3.2.8	Distribuzione Beta	36
3.3	Famiglia esponenziali	37
3.3.1	Vettori gaussiani	37
3.4	Proprietà variabili aleatorie continue	40
3.4.1	Calcolo dei momenti per v.a. continua	41
3.5	Vettori Aleatori	41
3.6	Convergenza di V.A	46
4	Esempi	49
4.0.1	esercizio 1	49
4.0.2	La raccolta delle figurine	50

5	Legge (debole) dei grandi numeri	52
5.1	Convergenza in Distribuzione	53
5.1.1	Approssimazione di Poisson alla binomiale	56
5.2	Teorema centrale del limite	57
5.2.1	Correzione di continuità	59
5.3	Passeggiata casuale	61
5.3.1	Passeggiata casuale simmetrica semplice	61
5.3.2	Principio di riflessione	62
5.3.3	Ritorni all'origine	62
6	Entropia	68
6.1	Entropia	68
6.1.1	Dado di Jaynes	70
6.2	Entropia congiunta	71
6.2.1	Entropia della somma di due variabili aleatorie indipendenti	75

Capitolo 1

Esperimenti, eventi e Probabilità

L'insieme di tutti i possibili esiti è contenuto nello **Spazio campionario** Ω .

Definizione 1.0.1. (Algebra)

Sia F una famiglia di insiemi, si definisce algebra se sono verificate queste 3 proprietà:

- $\Omega \in F$.
- Se $A \in F$ allora $A^c \in F$.
- Se $A, B \in F$ allora $A \cup B \in F$.

Proprietà 1.0.1. Data un'algebra F allora valgono le seguenti proprietà:

- $\phi \in F$.
- Siano $A, B \in F$ vale che
 - $A \cap B \in F$.
 - $A \setminus B \in F$
- Data una sottofamiglia $\{A_i\} \subset F$ allora
 - $(\bigcup_{i=1}^n A_i) \in F$.
 - $(\bigcap_{i=1}^n A_i) \in F$

L'intersezione tra algebre è un'algebra mentre l'unione non lo è necessariamente.

Definizione 1.0.2. Data un'algebra F e sia $A \in F$ allora si definisce atomo se $\exists B \subset A, B \neq \phi$ allora $B = A$.

1.1 Probabilità

Definizione 1.1.1 (Probabilità).

Una misura di probabilità è una funzione $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ definita su di una σ -algebra con le seguenti proprietà

1. **Normalizzazione:** $\mathbb{P}(\Omega) = 1$.
2. σ -**additiva:** sia $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A} | A_n \cap A_m = \phi, \forall m \neq n$ ho che

$$\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n)$$

Proprietà 1.1.1. Data un'algebra F e una misura μ su essa definita allora valgono le seguenti proprietà:

- $\mu(\phi) = 0$.
- $\forall A, B \in \mathcal{F}, B \subset A$ allora $\mu(B) \leq \mu(A)$ e $\mu(A \setminus B) = \mu(A) - \mu(B)$
- $\forall A, B \in \mathcal{F}$ vale che $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.

Definizione 1.1.2 (Variabile aleatoria).

Data uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$, una variabile aleatoria è una funzione

$$X : \Omega \rightarrow E$$

Che associa ad ogni evento un valore.

Esempio 1.1.1. Sia X una variabile aleatoria che descrive l'esito del lancio di una moneta, dunque si avrà che $\Omega = \{T, C\}$ e che

$$X : \Omega \rightarrow \mathbb{R}$$

$$x \mapsto X(x) = \begin{cases} 0 & x = C \\ 1 & x = T \end{cases}$$

Osservazione 1.1.1. Il termine aleatorio è dovuto al fatto per cui ci occupiamo degli esiti possibili di un esperimento il cui esito è incerto.

Definizione 1.1.3. (Eventi generati da un variabile aleatoria)

Dato uno spazio misurabile (Ω, \mathcal{A}) allora se $X : \Omega \rightarrow E$ è una variabile aleatoria, $\forall A \subset E$ definisco

$$\{X \in A\} := X^{-1} = \{\omega \in \Omega | X(\omega) \in A\}$$

Dunque $\{X \in A\} \subset \Omega$ rappresenta l'evento costituito da tutti e soli gli esiti ω dell'esperimento aleatorio e inoltre $\{X \in A\} \in \mathcal{A}$.

Esempio 1.1.2. Un esempio di una variabile aleatoria è la funzione indicatrice di eventi tale per cui $\forall A \subset \Omega$ tale funzione è definita

$$\mathbb{1}_A : \Omega \rightarrow \mathbb{R}$$

$$x \mapsto \mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

Esempio 1.1.3. Nel caso di una dado dire che esce un numero pari viene tradotto con $\{X \in \{2, 4, 6\}\}$ che definisce un sottoinsieme di Ω .

1.1.1 Spazio campionario continuo

Definizione 1.1.4. (Variabile aleatoria generale)

Sia $(\Omega, \mathcal{A}, \mathbb{P})$ uno spazio di probabilità, (E, \mathcal{E}) uno spazio misurabile e $X : \Omega \rightarrow E$ una funzione. Allora si definisce variabile aleatoria se $\forall C \in \mathcal{E}$ si ha che

$$\{X \in C\} = X^{-1}(C) \in \mathcal{A}$$

Definizione 1.1.5. (Distribuzione di una variabile aleatoria)

Sia X una variabile aleatoria definita su uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ a valori in (E, \mathcal{E}) . Si dice distribuzione di X

$$\mu_X : \mathcal{E} \rightarrow [0, 1]$$

$$C \mapsto \mu_X(C) = \mathbb{P}(X \in C)$$

Osservazione 1.1.2. La distribuzione di μ_X di una variabile aleatoria X a valori in (E, \mathcal{E}) è una probabilità su (E, \mathcal{E}) .

Osservazione 1.1.3. La condizione di una funzione di essere σ -finita permette di lavorare con unioni numerabili e inoltre ho che su \mathbb{R} la σ -algebra più piccola è costituita dai boreliani $B(\mathbb{R})$.

Se considero uno spazio campionario Ω e un'algebra \mathcal{A} tale per cui $\mathcal{A} := P(\Omega)$ allora $\{i\} | i \in \Omega$ sono atomi. Dunque posso definire una misura di probabilità sugli atomi come segue

$$\begin{aligned}\mathbb{P}: \mathcal{A} &\rightarrow \mathbb{R} \\ \{i\} &\mapsto \mathbb{P}(\{i\}) = \alpha_i\end{aligned}$$

Dunque dato che ogni insieme appartenente a quest'algebra è costituito da atomi allora posso dire che $\forall A \in \mathcal{A}, A = \bigcup_{i=1}^n \{x_i\}$ dunque dato che tutti questi atomi sono per definizione disgiunti allora vale che

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^n \{x_i\}\right) = \sum_{i=1}^n \mathbb{P}(\{x_i\})$$

Dunque se fisso $\mathbb{P}(\{x_i\}) = \alpha_i$ allora posso estendere tale misura di probabilità a una misura σ -additiva su \mathcal{A} .

1.2 Densità

Definizione 1.2.1. (*Densità discreta*)

Sia $X: \Omega \rightarrow E$ una v.a. definita su uno spazio di probabilità discreto (Ω, \mathbb{P}) a valori in un insieme arbitrario E , si definisce:

- **Distribuzione di X l'applicazione**

$$\begin{aligned}\mu_X: \mathcal{P}(X) &\rightarrow [0, 1] \\ A &\mapsto \mu_X(A) := \mathbb{P}(X \in A)\end{aligned}$$

- **Densità discreta l'applicazione**

$$\begin{aligned}p_X: E &\rightarrow [0, 1] \\ x &\mapsto p_X(x) := \mathbb{P}(X = x) = \mu_X(\{x\})\end{aligned}$$

Inoltre $\sum_{\omega \in \Omega} p(\omega) = 1$.

Proposizione 1.2.1. Sia p una densità discreta su un insieme arbitrario Ω . La funzione

$$\begin{aligned}P: \mathcal{P}(\Omega) &\rightarrow \mathbb{R} \\ A &\mapsto P(A) = \sum_{\omega \in A} p(\omega)\end{aligned}$$

È una misura di probabilità.

Dimostrazione. Per dimostrare che è una probabilità devo dimostrare 2 cose:

1. Per costruzione ho che $P(\Omega) = \sum_{\omega \in \Omega} p(\omega)$. Dato che per ipotesi p è una densità su Ω allora per definizione ho che $\sum_{\omega \in \Omega} p(\omega) = 1$ e quindi $P(\Omega) = 1$.
2. Sia $A \subset \Omega$ e sia $\{A_i\}_{i \in \mathbb{N}}$ una sua partizione allora per costruzione ho che $P(A) = \sum_{\omega \in A} p(\omega)$ dato che $\forall i \neq j, A_i \cap A_j = \emptyset$ allora ho che tale sommatoria può essere riscritta come

$$\sum_{k \in \mathbb{N}} \left(\sum_{\omega \in A_k} p(\omega) \right) = \sum_{k \in \mathbb{N}} p(A_k)$$

Quindi per definizione P è una probabilità. □

Teorema 1.2.1. Sia f_x una densità allora la funzione definita

$$\begin{aligned}\mathbb{P}: B(\mathbb{R}) &\rightarrow \mathbb{R} \\ A &\mapsto \mathbb{P}(A) = \int_A f d\mu\end{aligned}$$

È una misura di probabilità.

Dimostrazione. Per dimostrare che è una probabilità devo dimostrare 2 cose:

1. Questo punto è banale in quanto per definizione ho che

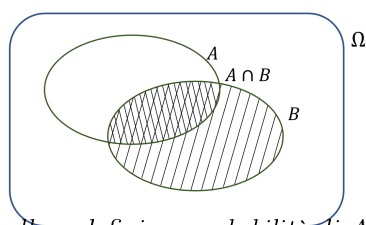
$$\mathbb{P}(\mathbb{R}) = \int_{\mathbb{R}} f(x) dx = 1$$

2. Sia $A \subset \mathbb{R} | A = \bigcup_{i=1}^n A_i, A_i \cap A_j = \emptyset, \forall i \neq j$ allora

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \int_{\bigcup_{i=1}^n A_i} f(x) dx = \sum_{i=1}^n \int_{A_i} f(x) dx = \sum_{i=1}^n \mathbb{P}(A_i)$$

Dunque per definizione \mathbb{P} è una probabilità. □

1.3 Probabilità condizionata



Data una misura di probabilità \mathbb{P} allora se si vuole conoscere la probabilità di un determinato evento A sapendo che un altro evento B si è verificato precedentemente, bisogna costruire un nuovo giudizio di probabilità

$$\mathbb{P}_B : \mathcal{A} \rightarrow [0, 1]$$

Definizione 1.3.1 (probabilità condizionata).

Dato uno spazio di probabilità (Ω, \mathbb{P}) due eventi $A, B \in \Omega | \mathbb{P}(B) > 0$ allora definisco probabilità di A condizionata da B

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Osservazione 1.3.1. Ciò che sta dietro alla proprietà condizionata è che si rivaluta la probabilità che un determinato evento si verifichi sapendo che in precedenza se ne è verificato un altro e quindi restringere Ω ad un suo sottoinsieme B .

Teorema 1.3.1. Sia $\{B_i\}$ una partizione di Ω finita o numerabile allora vale:

- Formula di **disintegrazione**:

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i)$$

- Formula delle **proprietà totali**:

$$\mathbb{P}(A) = \sum_i \mathbb{P}(B_i) \mathbb{P}_{B_i}(A)$$

In particolare

$$\mathbb{P}(A) = \mathbb{P}(B) \mathbb{P}_B(A) + \mathbb{P}(B^c) \mathbb{P}_{B^c}(A)$$

Definizione 1.3.2. (Eventi indipendenti) Dato uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ diciamo che due eventi $A, B \subset \Omega$ sono indipendenti se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B)$$

Definizione 1.3.3. Dati gli insiemi $\{A_i\}_{i \in \mathbb{N}}$, diremo che sono globalmente indipendenti se $\forall I \subset \mathbb{N}$ finito si ha che

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i)$$

Esempio 1.3.1. Sia $\Omega := \{n \in \mathbb{N} | 1 \leq n \leq 24\}$ e dati due eventi $A, B | A = \text{numeri pari e } B = \text{numeri dispari}$ allora

$$\mathbb{P}(A \cap B) = \frac{6}{24} = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B)$$

Quindi per definizione questi due eventi sono indipendenti.

Proposizione 1.3.1. Siano A, B due eventi indipendenti allora anche A^c e B sono indipendenti.

Dimostrazione. Dato che

$$B = (A \cap B) \cup (A^c \cap B)$$

Allora per definizione posso dire che

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$$

$$\mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(A^c \cap B)$$

Ora dato che per ipotesi so che A e B sono eventi indipendenti allora ho per definizione che $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ dunque ottengo che

$$\mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(A^c \cap B)$$

Ma dato che per definizione di probabilità ho che $1 - \mathbb{P}(A) = \mathbb{P}(A^c)$, allora la mia tesi è dimostrata. \square

Teorema 1.3.2. (Formula di Bayes)

Dati due eventi $A, B \in \Omega | \mathbb{P}(A) > 0, \mathbb{P}(B) > 0$ allora vale la **formula di Bayes**:

$$\mathbb{P}_A(B) = \frac{\mathbb{P}(B)\mathbb{P}_B(A)}{\mathbb{P}(A)}$$

Inoltre presa una partizione $\{B_i\}$ di Ω allora la formula di Bayes può essere riscritta come:

$$\mathbb{P}_A(B) = \frac{\mathbb{P}(B_i)\mathbb{P}_{B_i}(A)}{\sum_i \mathbb{P}(B_i)\mathbb{P}_{B_i}(A)}$$

Dimostrazione. Assunto per ipotesi che $A, B \in \Omega | \mathbb{P}(A) > 0, \mathbb{P}(B) > 0$.

Punto 1

Per definizione di proprietà condizionata posso calcolare in due modi del tutto equivalenti $\mathbb{P}(A \cap B)$ infatti

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}_A(B) = \mathbb{P}(B)\mathbb{P}_B(A)$$

Da qui la mia tesi

$$\mathbb{P}_A(B) = \frac{\mathbb{P}(B)\mathbb{P}_B(A)}{\mathbb{P}(A)}$$

Punto 2

Dato che per ipotesi abbiamo una partizione ho che

$$A = A \cap \Omega = A \cap \bigcup_i B_i$$

Allora per le proprietà degli operatori booleani ho che

$$A = \bigcup_i (A \cap B_i)$$

Dunque dato che quella famiglia rappresenta per ipotesi una partizione allora

$$(A \cap B_i) \cap (A \cap B_j) = A \cap (B_i \cap B_j) = A \cap \emptyset \quad \forall i \neq j$$

Quindi per definizione di funzione di probabilità ho che

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_i (A \cap B_i)\right) = \sum_i \mathbb{P}(A \cap B_i)$$

A questo punto ho per il primo punto che $\mathbb{P}(A \cap B_i) = \mathbb{P}_{B_i}(A)\mathbb{P}(B_i)$, dunque

$$\mathbb{P}_A(B) = \frac{\mathbb{P}(B)\mathbb{P}_B(A)}{\mathbb{P}(A)} = \mathbb{P}_A(B) = \frac{\mathbb{P}(B)\mathbb{P}_B(A)}{\sum_i \mathbb{P}_{B_i}(A)\mathbb{P}(B_i)}$$

Perciò la mia tesi è dimostrata. □

Teorema 1.3.3. (Regola della catena)

Siano $\{A_i\}$ eventi, allora vale la regola della catena:

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \mathbb{P}(A_1)\mathbb{P}_{A_1}(A_2)\mathbb{P}_{A_1 \cap A_2}(A_3) \cdots \mathbb{P}_{A_1 \cap \cdots \cap A_{n-1}}(A_n)$$

Dimostrazione. Dimostro per induzione:

- $n = 2$, per Bayes ho che

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}_{A_1}(A_2)$$

- $n = 3$

$$\mathbb{P}\left(\bigcap_{i=1}^3 A_i\right) = \mathbb{P}(A_1)\mathbb{P}_{A_1}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}_{A_1}(A_2)\mathbb{P}_{A_1 \cap A_2}(A_3)$$

- Facendo il passo induttivo e supponendo che vale per n allora lo devo provare per $n + 1$ dunque considero $\{A_i\}_{i=1}^n, A_i \subset \Omega, \forall i$. A questo punto considero A_{n+1} . A questo punto ci sono due casi:

1. $\exists i_0 | A_{i_0} = A_{n+1}$ in quel caso la dimostrazione è banale.
2. $\nexists i_0 | A_{i_0} = A_{n+1}$ allora considero $(\bigcap_{i=1}^n A_i) \cap A_{n+1}$ allora per Bayes ho che

$$\mathbb{P}\left(\left(\bigcap_{i=1}^n A_i\right) \cap A_{n+1}\right) = \mathbb{P}\left(\bigcap_{i=1}^n A_i\right) \mathbb{P}_{\bigcap_{i=1}^n A_i}(A_{n+1})$$

Dunque dato che $\mathbb{P}(\bigcap_{i=1}^n A_i) = \mathbb{P}(A_1)\mathbb{P}_{A_1}(A_2)\mathbb{P}_{A_1 \cap A_2}(A_3) \cdots \mathbb{P}_{A_1 \cap \cdots \cap A_{n-1}}(A_n)$ allora la mia tesi è dimostrata. □

1.4 Distribuzioni congiunte

Definizione 1.4.1. Dato uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ e sia $X = (X_1, \dots, X_n)$ tale per cui $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, B(\mathbb{R}))$, si definisce distribuzione di probabilità di X la funzione

$$\begin{aligned} \mu_X : B(\mathbb{R}) &\rightarrow [0, 1] \\ I &\mapsto \mu_X(I) = \mathbb{P}(X^{-1}(I)) \end{aligned}$$

Definizione 1.4.2. Dato uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ e sia $X = (X_1, \dots, X_n)$ tale per cui $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, B(\mathbb{R}))$, si definisce distribuzione marginale di μ_X la funzione

$$\begin{aligned} \mu_i : B(\mathbb{R}) &\rightarrow [0, 1] \\ E &\mapsto \mu_X(I) = \mathbb{P}(X_i^{-1}(E)) \end{aligned}$$

Definizione 1.4.3. (Indipendenza delle variabili aleatorie)

n variabili aleatorie X_1, \dots, X_n su uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ sono dette indipendenti se lo sono le rispettive σ -algebre generate: $\Sigma(X_i)$

1.5 Funzione di ripartizione

Definizione 1.5.1. (*Funzione di ripartizione*)

Si chiama *funzione di ripartizione* della variabile aleatoria X la funzione $F_X : \mathbb{R} \rightarrow \mathbb{R}$ tale per cui $\forall t \in \mathbb{R}$

$$F_X(t) = \mu_X((-\infty, t]) = \mathbb{P}(X \leq t)$$

Osservazione 1.5.1. F_X determina completamente la distribuzione di una variabile aleatoria reale discreta μ_X .

Proposizione 1.5.1. Sia X una v.a. reale discreta con funzione di distribuzione p_X allora vale che

$$p_X(x) = F_X(x) - F_X(x^-)$$

Questo mi permette di vedere che F_X è discontinua in un punto $x \iff p_X(x) > 0$.

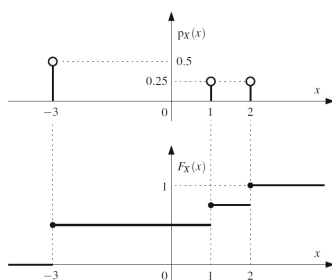
Definizione 1.5.2. (*Funzione di ripartizione congiunta della coppia*)

Sia $X : \Omega \rightarrow \mathbb{R}^n$, $X = (X_1, \dots, X_n)$ e posto $(X_1 \leq t_1, \dots, X_n \leq t_n) = (X \leq t) \in \mathcal{A}$ allora $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ con $\mathbb{P}(X \leq t, Y \leq s) = F(t, s)$ si definisce *funzione di ripartizione congiunta della coppia* (X, Y) .

Proprietà 1.5.1. Di questa funzione si possono notare alcune cose:

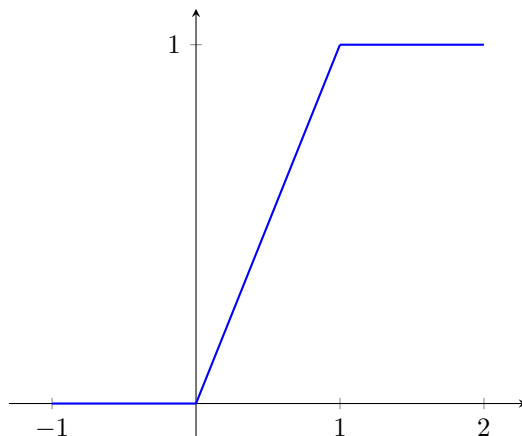
- $\lim_{x \rightarrow +\infty} F_X(x) = \lim_{t \rightarrow +\infty} \mathbb{P}(X \leq t) = 1$.
- F_X è crescente e continua da destra.
- Vale che:

$$\begin{aligned} \lim_{s \rightarrow +\infty} \mathbb{P}(X \leq t, Y \leq s) &= \mathbb{P}(X \leq t) = F_X(t) = \lim_{s \rightarrow +\infty} F(t, s) \\ \lim_{t \rightarrow +\infty} \mathbb{P}(X \leq t, Y \leq s) &= \mathbb{P}(Y \leq s) = F_Y(s) = \lim_{t \rightarrow +\infty} F(t, s) \end{aligned}$$



Teorema 1.5.1. Data una funzione di ripartizione $F : \mathbb{R} \rightarrow [0, 1]$ tale per cui $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$ continua da destra e crescente allora esiste uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ e una variabile aleatoria $X : \Omega \rightarrow \mathbb{R} | F_X = F$.

Esempio 1.5.1. Considero la funzione $F(x) := \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$



Soddisfa tutte le ipotesi del teorema e quindi $\exists(\Omega, \mathcal{A}, \lambda) | \lambda((-\infty, x]) = F(x)$ dunque $\lambda((a, b]) = b - a, \forall [a, b] \subset [0, 1]$ e in questo caso ho che λ è la misura di Lebesgue.

Proposizione 1.5.2. (indipendenza delle variabili aleatorie)

Se X, Y , variabili aleatorie, hanno la stessa funzione di ripartizione $F_x(t) = F_y(t), \forall t \in \mathbb{R}$ allora (X_1, \dots, X_n) sono indipendenti se e solo se lo sono (Y_1, \dots, Y_n) .

Se X, Y sono variabili aleatorie indipendenti e se f, g sono funzioni reali misurabili allora $f(X), g(Y)$ sono variabili aleatorie indipendenti.

Se X, Y sono variabili aleatorie indipendenti allora $X_1 + \dots + X_n$ e $Y_1 + \dots + Y_n$ sono indipendenti.

Osservazione 1.5.2. Sia $X = (X_1, \dots, X_n)$ una v.a. e X_i la v.a. che mi informa se dal lancio di una moneta è uscita testa allora

$$Z = \sum_{i=1}^n X_i$$

Rappresenta la v.a. che indica quante volte è uscita testa inseguito ad n -lanci di moneta.

Osservazione 1.5.3. Date due variabili aleatorie X, Y e sia $Z = (X, Y)$ dunque ho che $Z \sim \{(z_{i,j}, p_{ij})\}_{ij}$ esiste un legame tra la distribuzione congiunta e quella marginale:

$$\begin{aligned} \mathbb{P}(X = x_i) &= \sum_j \mathbb{P}(X = x_i, Y = y_j) = \sum_j p_{ij} \\ \mathbb{P}(Y = y_j) &= \sum_i \mathbb{P}(X = x_i, Y = y_j) = \sum_i p_{ij} \end{aligned}$$

1.6 Prodotto di convoluzione

Definizione 1.6.1. (Prodotto di convoluzione)

Dati p, q due densità tali per cui $p, q : \mathbb{N} \rightarrow [0, 1]$ (dunque $\sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty} q_n = 1$) allora definisco prodotto di convoluzione di p, q

$$p \star q : \mathbb{N} \rightarrow [0, 1]$$

$$z \mapsto (p \star q)(z) = \sum_{n=0}^z p_n q_{z-n}$$

Teorema 1.6.1. $p \star q = q \star p$ è una distribuzione di probabilità.

Dimostrazione. Per definizione posso scrivere che

$$\sum_{z \in \mathbb{N}} (p \star q)(z) = \sum_{z \in \mathbb{N}} \sum_{n=0}^z p_n q_{z-n} = \sum_{n=0}^{\infty} \sum_{z=n}^{\infty} p_n q_{z-n}$$

Dunque ponendo $x = z - n$ ottengo

$$\sum_{n=0}^{\infty} \sum_{x=0}^{\infty} p_n q_x = \sum_{n=0}^{\infty} p_n \left(\sum_{x=0}^{\infty} q_x \right)$$

Dato che per ipotesi ho che $\sum_{x=0}^{\infty} q_x = 1$ allora il tutto è equivalente a $\sum_{n=0}^{\infty} p_n = 1$ per ipotesi e quindi la mia tesi è dimostrata. \square

Proposizione 1.6.1. Siano $X : \Omega \rightarrow E$ e $Y : \Omega \rightarrow F$ variabili aleatorie definite sullo stesso spazio di probabilità con densità marginali p_X, p_Y e densità congiunte $p_{X,Y}$, allora

$$X \amalg Y \iff p_{X,Y}(x, y) = p_X(x)p_Y(y)$$

Dimostrazione. Procedendo per ordine.

$$"X \amalg Y \Rightarrow p_{X,Y}(x, y) = p_X(x)p_Y(y)"$$

Per definizione di variabile aleatoria indipendenti ho che $\forall x \in E, \forall y \in F$ vale

$$p_{X,Y}(X = x, Y = y) = p_X(X = x)p_Y(Y = y)$$

$$"X \amalg Y \Leftarrow p_{X,Y}(x, y) = p_X(x)p_Y(y)"$$

Sia $A \subset E, B \subset F$ allora so che

$$\mathbb{P}(X \in A, Y \in B) = \sum_{\substack{x \in A \\ y \in B}} p_{X,Y}(x, y)$$

Dunque per ipotesi so che il tutto è equivalente a

$$\sum_{y \in B} \sum_{x \in A} p_X(x)p_Y(y) = \sum_{y \in B} p_Y(y) \sum_{x \in A} p_X(x) =$$

$$\sum_{y \in B} p_Y(y)p_X(A) = p_X(A) \sum_{y \in B} p_Y(y) = p_X(X \in A)p_Y(Y \in B)$$

Perciò per definizione ho che $X \amalg Y$ e quindi la mia tesi è dimostrata. \square

Teorema 1.6.2. Sia (X, Y) un vettore aleatorio e sia $S = X + Y, X \geq 0, Y \geq 0$ allora $p_{X,Y}(x, y)$ distribuzione congiunta $p_X(x), p_Y(y)$ distribuzioni marginali allora vale che

1.

$$p_{X+Y}(s) = \sum_x p_{X,Y}(x, s-x) = \sum_y p_{X,Y}(s-y, y)$$

2. Se X, Y sono indipendenti allora $p_S = p_X * p_Y$.

Dimostrazione. Procedendo con ordine.

Punto 1

Sapendo che $S = X + Y$ allora logicamente se dico che $S = s$ significa che $X + Y = s$ dunque facendo variare X in tutto il suo dominio trovo tutti quei valori per cui $X + Y = s$

$$(S = s) = \bigcup_x (Y = s - x, X = x)$$

Dunque

$$P(S = s) = \sum_x P(X = x, Y = s - x) = \sum_x P_{X,Y}(x, s - x)$$

Punto 2

Se $X \perp Y$ allora ho per definizione che

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = s - x)$$

Perciò

$$P(S = s) = \sum_x \mathbb{P}(X = x)\mathbb{P}(Y = s - x)$$

Ma questa sempre per definizione rappresenta $p_X * p_Y$, quindi la mia tesi è dimostrata. \square

Esempio 1.6.1. Supponiamo che $U = \{1, 2, 3, 4, 5, 6\}$ e che X sia l'esito della prima estrazione e Y della seconda estrazione allora ci possono essere due casi in entrambi i casi posso definire $Z = (X, Y)$ come

$$Z \sim \{(Z_{i,j}, p_{i,j}), z_{i,j}(i, j), i, j = 1, 2, 3, 4, 5, 6\}$$

- **Estrazione con reimmissione**

Dato che l'estrazione è equilibrata e che $|U \times U| = 36$ allora $p_{i,j} = \frac{1}{36}$ dunque si vede che sto simulando il lancio di due dati in quanto

$$P(X = i) = \sum_{j=1}^6 p_{i,j} = \frac{1}{36} 6 = \frac{1}{6}$$

Inoltre vedo che

$$\frac{1}{36} = \frac{1}{6} \frac{1}{6} = p_i p_j$$

Dunque per definizione queste sono due variabili indipendenti.

- **Estrazione senza reimmissione**

Si vede che $p_{i,i} = 0$ in quanto se alla prima estrazione pesco x_i alla seconda estrazione x_i non è presente e quindi non potrà mai esserci tale coppia. dunque

$$p_{i,j} = \frac{1}{30}$$

Inoltre noto che

$$P(X = i) = \sum_{j=1}^6 p_{i,j} = \frac{1}{30} \sum_{i \neq j} = \frac{1}{6}$$

Ma dato che $0 = \frac{1}{6} \frac{1}{6}$ allora queste due variabili non sono indipendenti.
(Ho preso 0 perché l'uguaglianza deve vale $\forall z \in Z$)

Esempio 1.6.2. Supponiamo di avere un dado ed una moneta equilibrati dunque definisco $Z = (D, X)$ come la v.a. che mi indica, lanciato il dado D ed ottenuto $i = 1, \dots, 6$, quante volte ottengo testa lanciando la moneta i volte.

$D \setminus N$	0	1	2	3	4	5	6	
1	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	0	0	$= \frac{1}{6}$
2	$\frac{1}{24}$	$\frac{2}{24}$	$\frac{1}{24}$	0	0	0	0	$= \frac{1}{6}$
3	$\frac{1}{48}$	$\frac{3}{48}$	$\frac{3}{48}$	$\frac{1}{48}$	0	0	0	$= \frac{1}{6}$
4	$\frac{1}{96}$	$\frac{4}{96}$	$\frac{6}{96}$	$\frac{4}{96}$	$\frac{1}{96}$	0	0	$= \frac{1}{6}$
5	$\frac{1}{192}$	$\frac{5}{192}$	$\frac{10}{192}$	$\frac{10}{192}$	$\frac{5}{192}$	$\frac{1}{192}$	0	$= \frac{1}{6}$
6	$\frac{1}{384}$	$\frac{6}{384}$	$\frac{15}{384}$	$\frac{20}{384}$	$\frac{15}{384}$	$\frac{6}{384}$	$\frac{1}{384}$	$= \frac{1}{6}$
$= \frac{63}{384} \mid = \frac{120}{384} \mid = \frac{72}{384} \mid = \frac{69}{384} \mid = \frac{29}{384} \mid = \frac{8}{384} \mid = \frac{1}{384}$								

Dunque chiedere

$$\mathbb{P}(D=4, N=3) = \mathbb{P}(D=4)\mathbb{P}_{D=4}(N=3) = \frac{1}{6} \binom{4}{3} \frac{1}{2^3} \frac{1}{2} = \frac{4}{96}$$

1.7 Calcolo dei momenti

1.7.1 Media

Definizione 1.7.1. Sia $k \in \mathbb{N}$ e $\mu \in \mathbb{R}$ e data una variabile aleatoria X definisco momento di ordine k e origine m :

- Se X è discreta $X \sim \{(x_i, p_i)\}$

$$\mu_{m,k} = \sum_{i=1}^n (x_i - m)^k p_i$$

- Se X è continua allora

$$\mu_{m,k} = \int_{-\infty}^{\infty} (x - m)^k p_X(x) dx$$

Preso un insieme X con cardinalità finita N allora se i suoi elementi sono indicati da x_i allora la media di tale insieme sarà determinata da

$$\mu = \frac{1}{N} \sum_i x_i$$

Dunque se ogni elemento x_i viene ripetuto m_i volte allora posso riscrivere il tutto come

$$\mu = \frac{1}{N} \sum_{i=1}^n x_i m_i = \sum_{i=1}^n x_i \frac{m_i}{N}$$

A questo punto definisco peso di un elemento $p_i = \frac{m_i}{N}$.

Definizione 1.7.2. (Media)

Sia X una variabile aleatoria discreta tale per cui $X \sim \{(x_k, p_k), k \in \mathbb{N}\}$, se $X \geq 0$ dirò **media (valore atteso)** di X il valore

$$\mathbb{E}[X] = \sum_{k \in \mathbb{N}} x_k p_k \in [0, +\infty]$$

Se $X \leq 0$ allora

$$\mathbb{E}[X] = \sum_{k \in \mathbb{N}} x_k p_k \in [-\infty, 0]$$

Nel caso generale diremo che X ammette media se almeno uno tra X^+, X^- , con $\begin{cases} X^+ = X\mathbb{1}_{(X \geq 0)} \\ X^- = X\mathbb{1}_{(X \leq 0)} \end{cases}$, ammette media finita, altrimenti X non ammette media, inoltre ho che

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} xp_X$$

Osservazione 1.7.1. Una variabile aleatoria reale X ammette valore medio se e solo se almeno uno tra i valori medi di X^+, X^- è finito, in tal caso si ha che

$$\mathbb{E}[X] = \mathbb{E}[X^+] + \mathbb{E}[X^-]$$

Proposizione 1.7.1. Sia X una variabile aleatoria reale, definita su uno spazio di probabilità discreto (Ω, \mathbb{P}) . X ammette valore medio se e solo se la famiglia di numeri reali $\{X(\omega)\mathbb{P}(\{\omega\})\}_{\omega \in \Omega}$ ammette somma. In questo caso si ha

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\})$$

Dimostrazione. Per definizione ho che

$$\mathbb{E}[X] = \sum_{x \in \mathbb{R}} xp_X(x) = \sum_{x \in \mathbb{R}} x\mathbb{P}(X = x) = \sum_{x \in \mathbb{R}} x \sum_{\omega \in \{X=x\}} \mathbb{P}(\{\omega\})$$

Dunque dato che $X^{-1}(x)$ è un insieme di elementi dello spazio campionario tali per cui $X(\omega) = x$ allora posso riscrivere il tutto come

$$\sum_{x \in \mathbb{R}} \sum_{\omega \in \{X=x\}} X(\omega)\mathbb{P}(\{\omega\})$$

Dato che X ha immagine $X(\Omega) \subset \mathbb{R}$ e che $\Omega = \bigcup_{x \in \mathbb{R}} \{X = x\}$ allora ottengo la mia tesi in quanto

$$\sum_{x \in X(\Omega)} \sum_{\omega \in \{X=x\}} X(\omega)\mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\})$$

□

Proprietà 1.7.1. La media dipende solo dalla distribuzione di X inoltre vale che

$$\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]$$

Teorema 1.7.1. Sia X una variabile aleatoria discreta, sia $g: \mathbb{R} \rightarrow \mathbb{R}$ allora $Y = g(X)$ ha media pari a

$$\mathbb{E}[Y] = \sum_{k \in \mathbb{N}} g(x_k)p_k$$

\iff almeno uno tra $\mathbb{E}[Y^+], \mathbb{E}[Y^-]$ sia finito. In tal caso ho che

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$$

Dimostrazione. Il mio obiettivo è provare che

$$\sum_{j \in \mathbb{N}} y_j q_j = \sum_{k \in \mathbb{N}} g(x_k)p_k$$

Dunque per definizione so che $\mathbb{E}[Y] = \sum_{j \in \mathbb{N}} y_j q_j$ perciò supposto di lavorare con $g(x) \geq 0, \forall x \in \mathbb{R}$ ho che

$$q_j = \mathbb{P}(Y = y_j) = \mathbb{P}(X \in g^{-1}(y_j)) = \sum_{x_k | g(x_k) = y_j} p_k$$

Quindi moltiplicando ambo i membri per una costante ottengo

$$y_j q_j = y_j \sum_{x_k | g(x_k) = y_j} p_k = \sum_{x_k | g(x_k) = y_j} g(x_k)p_k$$

Perciò ho che

$$\sum_j y_j p_j = \sum_j \sum_{x_k | g(x_k) = y_j} g(x_k) p_k$$

Perciò per l'osservazione precedente la mia tesi è dimostrata. \square

Osservazione 1.7.2. La **media** è un momento del primo ordine mentre la **varianza** è un momento del secondo ordine centrato nella media μ .

Teorema 1.7.2. (Proprietà della media)

1. Se X è una variabile aleatoria che ammette media allora $|\mathbb{E}[x]| \leq \mathbb{E}[|x|]$.
2. (**Linearità**) Se X, Y hanno media finita (X, Y sono positive) e se $a, b \in \mathbb{R}$ allora

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

Esiste ed è finito.

Dimostrazione. Prima di tutto devo dimostra che $aX + bY$ ha media finita ma questo è ovvio in quanto è sufficiente mostrare che

$$|\mathbb{E}[aX + bY]| \leq \mathbb{E}[|aX + bY|] = \sum_{i,j} |ax_i + by_j| p_{ij} \leq |a| \sum_{i,j} |x_i| p_{ij} + |b| \sum_{i,j} |y_j| p_{ij}$$

Ma questo per definizione corrisponde a $|a|\mathbb{E}[|X|] + |b|\mathbb{E}[|Y|]$ che per ipotesi è finito.

A questo punto in maniera del tutto analoga ho che

$$\mathbb{E}[aX + bY] = \sum_{i,j} (ax_i + by_j) p_{ij} = a \sum_{i,j} x_i p_{ij} + b \sum_{i,j} y_j p_{ij} = a \sum_i x_i \sum_j p_{ij} + b \sum_j y_j \sum_i p_{ij}$$

Dunque per quanto calcolato in precedenza ho che

$$\sum_i p_{ij} = p_j \quad \sum_j p_{ij} = p_i$$

Perciò ho che il tutto è equivalente a

$$a \sum_i x_i p_i + b \sum_j y_j p_j = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

Quindi la mia tesi è dimostrata. \square

1.7.2 Varianza e Covarianza

Definizione 1.7.3. Definisco l'insieme delle variabili aleatorie reali definiti su Ω che ammette valore medio di ordine k finito come

$$L^k(\Omega, \mathbb{P}) := \{X : \Omega \rightarrow \mathbb{R} | \mathbb{E}[X^k] < +\infty\}$$

Osservazione 1.7.3. (X, Y) variabile aleatoria con $g(x, y) = xy$ allora

$$\mathbb{E}[g(X, Y)] = \mathbb{E}[XY] = \sum_{k,j} g(x_k, y_j) p(x_k, y_j)$$

Se $x, y > 0$ allora $g(x_k, y_j) > 0$ e quindi esiste sempre e al limite è infinito inoltre p è la distribuzione congiunta di (X, Y) .

Proposizione 1.7.2. Siano $p, q \in \mathbb{N} | 0 < p < q < \infty$ e sia $X \in L^q(\Omega, \mathbb{P})$ continua allora $X \in L^p(\Omega, \mathbb{P})$.

Dimostrazione. Prima di tutto si può osservare che

$$\begin{cases} |x|^p < 1 & |x| < 1 \\ |x|^p < |x|^q & |x| > 1 \end{cases}$$

Dunque è ovvio affermare che $|x|^p < 1 + |x|^q$. A questo punto per definizione e ricordando che essendo X continua per ipotesi ho per definizione che $\exists f_X(x)$, funzione densità, che risulta essere ≥ 0 dunque ho che

$$\begin{aligned} \mathbb{E}[X^p] &= \int_{-\infty}^{+\infty} |x|^p f_X(x) dx \\ &\leq \int_{-\infty}^{+\infty} (1 + |x|^q) f_X(x) dx \\ &= \int_{-\infty}^{+\infty} f_X(x) dx + \int_{-\infty}^{+\infty} |x|^q f_X(x) dx \\ &= 1 + \mathbb{E}[X^q] < +\infty \end{aligned}$$

Dunque per definizione ho che $X \in L^p(\Omega, \mathbb{P}), \forall p < q$. □

Definizione 1.7.4. (Varianza e Covarianza)

Sia (Ω, \mathbb{P}) uno spazio di probabilità discreta e siano X, Y v.a. reali definite su Ω allora

- Se $X, Y, XY \in L^1(\Omega, \mathbb{P})$ (in particolare $X, Y \in L^2(\Omega, \mathbb{P})$) si definisce covarianza di X, Y

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Se $\text{Cov}(X, Y) = 0$ allora X, Y si dicono essere **scorrelate**.

- Se $X \in L^2$ allora si definisce varianza di X

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Osservazione 1.7.4. Da tutto questo si può osservare che:

- **Media:** rappresenta il baricentro della distribuzione ed è indicata come

$$\mathbb{E}[X] = \sum_{k \in \mathbb{N}} x_k p_k = \mu$$

- **Varianza:** mi informa quanto lontano dalla media ci troviamo quindi più è piccola e più mi avvicino ad ogni risultato alla media.

$$\text{var}(X) = \sigma^2 = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \mathbb{E}[\mu^2]$$

Infatti questa rappresenta una media degli scarti del valore medio.

Proposizione 1.7.3. Se X, Y sono v.a. sullo stesso spazio di probabilità e hanno momento secondo finito

$$X \in L^2(\Omega, \mathbb{P}) \iff \mathbb{E}[|X|^2] < +\infty$$

$$Y \in L^p(\Omega, \mathbb{P}) \iff \mathbb{E}[|Y|^p] < +\infty$$

Allora $XY \in L^1(\Omega, \mathbb{P}) \iff \mathbb{E}[XY] < +\infty$.

$$\mathbb{E}[XY] = \sum_{k,j} g(x_k, y_j) p(x_k, y_j)$$

Definizione 1.7.5. Se X, Y sono v.a. sullo stesso spazio di probabilità e hanno momento secondo finito con

$$\begin{aligned} (X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \\ &= \left(\sum_{k,j} x_k y_j p(x_k, y_j) \right) - \mathbb{E}[X]\mathbb{E}[Y] = \sum_{k,j} (x_k y_j - \mu_x \mu_y) p(x_k, y_j) \end{aligned}$$

Proprietà 1.7.2. *Da tutto questo si possono osservare alcune proprietà:*

- $cov(X, X) = var(X)$.
- $cov(X, Y) = cov(Y, X)$.
- $var(c) = 0$.
- $var(cX) = c^2 var(X)$.
- $cov(\alpha X + \beta Y, Z) = \alpha cov(X, Z) + \beta cov(Y, Z)$.
- $var(X + Y) = var(X) + var(Y) + cov(X, Y)$.

Proposizione 1.7.4. *Siano $X_1, \dots, X_n \in L^2$ variabili aleatorie indipendenti allora vale che*

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i)$$

Dimostrazione. Per definizione di varianza ho che

$$Var\left(\sum_{i=1}^n X_i\right) = Cov\left(\left(\sum_{i=1}^n X_i\right), \left(\sum_{i=1}^n X_i\right)\right)$$

Dunque dalle proprietà della covarianza ho che

$$Cov\left(\left(\sum_{i=1}^n X_i\right), \left(\sum_{i=1}^n X_i\right)\right) = \sum_{i=1}^n \sum_{j=1}^n Cov(X_i, X_j) = \sum_{i=1}^n var(X_i) + \sum_{\substack{1 \leq j \leq n \\ 1 \leq i \leq n \\ i \neq j}} Cov(X_i, X_j)$$

Ma dato che per ipotesi X_1, \dots, X_n sono indipendenti allora per le proprietà della covarianza ho che

$$\sum_{\substack{1 \leq j \leq n \\ 1 \leq i \leq n \\ i \neq j}} Cov(X_i, X_j) = 0$$

Quindi ottengo la mia tesi

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i)$$

□

Teorema 1.7.3. *Se X, Y sono v.a. su $(\Omega, \mathcal{A}, \mathbb{P})$ indipendenti e in $L^1(\Omega, \mathbb{P})$ allora $XY \in L^1(\Omega, \mathbb{P})$ e vale inoltre che*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

.

Dimostrazione. Per definizione ho che

$$\mathbb{E}[|XY|] = \sum_{k,j} |x_k| |y_j| p(x_k, y_j) = \sum_{k,j} |x_k| |y_j| p_X(x_k) p_Y(y_j) = \sum_k \left(\sum_j |y_j| p_Y(y_j) \right) |x_k| p_X(x_k)$$

Dato che $\sum_j |y_j| p_Y(y_j) = \mathbb{E}[|Y|] < +\infty$ per ipotesi allora ho che il tutto è equivalente a

$$\mathbb{E}[|Y|] \sum_k |x_k| p_X(x_k) = \mathbb{E}[|Y|] \mathbb{E}[|X|]$$

Inoltre

$$\mathbb{E}[XY] = \left(\sum_k x_k p_X(x_k) \right) \left(\sum_j y_j p_Y(y_j) \right) = \mathbb{E}[X] \mathbb{E}[Y]$$

□

Proposizione 1.7.5. *Siano X, Y v.a. se sono indipendenti allora $Cov(X, Y) = 0$.*

Dimostrazione. Banalmente applicando la definizione ho che

$$Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Ma per ipotesi essendo le due variabili aleatorie indipendenti, ho che $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ dunque $Cov(X, Y) = 0$. \square

Osservazione 1.7.5. *Non vale il viceversa in quanto se ho due variabili aleatorie scorrelate non è detto che siano indipendenti, come in questo caso: $\Omega := \{0, 1, 2\}$*

$$X(\omega) = \begin{cases} 1 & w = 0 \\ 0 & w = 1 \\ -1 & w = 2 \end{cases} \quad Y(\omega) = \begin{cases} 0 & w = 0 \\ 1 & w = 1 \\ 0 & w = 2 \end{cases}$$

Si può osservare banalmente che

$$Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$$

E che

$$\mathbb{P}(X = 1, Y = 0) = \mathbb{P}(w = 0) = \frac{1}{3} \neq \mathbb{P}(X = 1)\mathbb{P}(Y = 0) = \frac{2}{9}$$

Capitolo 2

Coefficiente di correlazione

2.1 Disuguaglianze

Teorema 2.1.1. *Sia X una v.a. reale definito in uno spazio di probabilità discreto (Ω, \mathbb{P}) allora valgono le seguenti disuguaglianze:*

- **Markov**

Se X è a valori positivi e $\mu = \mathbb{E}[X]$ allora $\forall \epsilon > 0$

$$\mathbb{P}(X \geq \epsilon) \leq \frac{\mu}{\epsilon}$$

- **Chebyshev**

Se $X \in L^2(\Omega, \mathbb{P})$, $\mu = \mathbb{E}[X]$ e $\sigma^2 = \text{Var}(X)$ allora $\forall \epsilon > 0$

$$\mathbb{P}(|X - \mu| > \epsilon) = \mathbb{P}((X - \mu)^2 > \epsilon^2) \leq \frac{\sigma^2}{\epsilon^2}$$

- **Cauchy-Schwarz**

Data inoltre Y v.a. reale e supposto che $X, Y \in L^2$ allora $XY \in L^1$ e vale che

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$$

Proposizione 2.1.1. $|\mathbb{E}[XY]| = \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]} \iff \exists c \in \mathbb{R} | Y = cX$.

Dimostrazione. Consideriamo il caso non banale per cui $X, Y \neq 0$ dunque chiamo

$$\xi = \frac{X}{\sqrt{\mathbb{E}[X^2]}} \quad \mu = \frac{Y}{\sqrt{\mathbb{E}[Y^2]}}$$

Dunque ho che $(\xi - \mu)^2 \geq 0$ perciò $\mathbb{E}[(\xi - \mu)^2] \geq 0$ perciò dato che $\mathbb{E}[(\xi - \mu)^2] = \mathbb{E}[\xi^2] + \mathbb{E}[\mu^2] - 2\mathbb{E}[\xi\mu] = 1 + 1 - 2\mathbb{E}[\xi\mu] \geq 0$ allora ho che

$$1 \geq \mathbb{E}[\xi\mu]$$

Se $Y = cX$ allora $XY = cX^2$ e quindi ho che

$$|\mathbb{E}[XY]| = |c|\mathbb{E}[X^2]$$

Viceversa se $\mathbb{E}[(\xi - \mu)^2] = 0$ allora

$$\mathbb{P}(\xi - \mu = 0) = 1 = \mathbb{P}\left(\frac{X}{\sqrt{\mathbb{E}[X^2]}} = \frac{Y}{\sqrt{\mathbb{E}[Y^2]}}\right) = \mathbb{P}(Y = cX)$$

□

Proposizione 2.1.2. *Sia Y una v.a. positiva avente momento di ordine k finito allora*

$$\mathbb{P}(Y > \epsilon) \leq \frac{\mathbb{E}[Y^k]}{\epsilon^k}$$

Inoltre

$$\mathbb{P}(Y > \epsilon) \leq \frac{\mathbb{E}[e^{tX}]}{e^{t\epsilon}}$$

Dimostrazione. Molto banalmente ho che

$$\{Y > \epsilon\} = \{Y^k > \epsilon^k\}$$

Dunque per la disuguaglianza di Markov ho che

$$\mathbb{P}(Y > \epsilon) = \mathbb{P}(Y^k > \epsilon^k) \leq \frac{\mathbb{E}[Y^k]}{\epsilon^k}$$

Inoltre

$$\{Y > \epsilon\} = \{e^{tX} > e^{t\epsilon}\}$$

Dunque sempre per Markov ho che

$$\mathbb{P}(Y > \epsilon) \leq \frac{\mathbb{E}[e^{tX}]}{e^{t\epsilon}}$$

Dunque la tesi è dimostrata. □

2.2 Coefficiente di correlazione

Definizione 2.2.1. *(Coefficiente di correlazione)*

Siano X, Y v.a. tali per cui $\text{Var}(X), \text{Var}(Y) > 0$ allora definisco coefficiente di correlazione

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}}$$

Osservazione 2.2.1. *Il coefficiente di correlazione indica il grado di correlazione lineare tra le variabili X, Y ossia di quanto bene Y possa essere approssimata da funzioni lineari affini di X .*

Osservazione 2.2.2. *Da ciò posso osservare alcune cose:*

- *Tale coefficiente rimane costante per un cambiamento di scala (per questo motivo è migliore della covarianza).*
- *Per Cauchy-Schwarz vale la seguente disuguaglianza*

$$|\text{cov}(X, Y)| \leq \sqrt{\text{var}(X) \text{var}(Y)}$$

- *Verifica che*

$$|\rho(X, Y)| \leq 1$$

Inoltre ho che se tali v.a. sono indipendenti tra loro $\rho(X, Y) = 0$ (non vale necessariamente il viceversa).

Proposizione 2.2.1. $|\rho(X, Y)| = 1 \iff Y = aX + b$ con $a \in \mathbb{R} \setminus \{0\}$ e $b \in \mathbb{R}$ e $\mathbb{P}(Y = aX + b) = 1$

Dimostrazione. Data la seguente funzione

$$\begin{aligned} \varphi: \mathbb{R} \setminus \{0\} \times \mathbb{R} &\rightarrow \mathbb{R} \\ (a, b) &\mapsto \varphi(a, b) = \|Y - (aX + b)\|_2^2 \end{aligned}$$

Dunque è ovvio affermare che

$$\varphi(a, b) = \mathbb{E}[(y - aX - b)^2] = \text{Var}(Y - aX) + (\mathbb{E}[Y] - a\mathbb{E}[X] - b)^2$$

Dunque

$$\nabla\varphi(a, b) = (0, 0) \iff a = \frac{Cov(Y, X)}{Var(X)} \quad b = \mathbb{E}[Y] - \frac{\mathbb{E}[X]Cov(X, Y)}{Var(X)}$$

Sotto queste ipotesi ho che

$$\varphi(a, b) = Var(Y) (1 - \rho^2(X, Y))$$

Quindi se $\rho = 1$ allora Y, X sono approssimabili. □

2.3 Funzione generatrice dei momenti

(questa è sempre una funzione di trasformazione di variabile aleatoria)

Definizione 2.3.1. (*Funzione generatrice dei momenti*)

Sia X una v.a. reale allora definisco funzione generatrice dei momenti la funzione così definita:

$$M_X: \mathbb{R} \rightarrow \overline{\mathbb{R}} \\ t \mapsto M_X(t) = \mathbb{E}[e^{tx}] = \sum_{k \in \mathbb{N}} e^{tx_k} p_k$$

Osservazione 2.3.1. A questo punto si possono vedere alcune cose:

- Questa funzione non è detto che sia ben definita su tutto \mathbb{R} allora ne definisco il dominio come $\{t \in \mathbb{R} | \mathbb{E}[e^{tx}] < +\infty\} \neq \emptyset$ in quanto 0 appartiene al dominio in quanto $\mathbb{E}[1] < +\infty$.
- Se $X \sim Be(p) \Rightarrow M_X(t) = (1 - p) + pe^t$ e questo si vede che è definita $t \in \mathbb{R}$.

Teorema 2.3.1. Siano X, Y v.a. indipendenti sullo stesso spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ allora $\forall \alpha, \beta \in \mathbb{R}, t \in \mathbb{R}$ vale che

$$M_{\alpha X + \beta Y}(t) = M_X(\alpha t) M_Y(\beta t)$$

Dimostrazione. Assumo per ipotesi che $X \perp Y$ allora $f(X) \perp g(Y)$ quando f, g sono funzioni misurabili in particolare se considero $f(x) = e^{\alpha x}, g(y) = e^{\beta y}$ dunque ho che $e^{\alpha x} \perp e^{\beta y}$, quindi ho che

$$\mathbb{E}[e^{\alpha x} e^{\beta y}] = \mathbb{E}[e^{\alpha x}] \mathbb{E}[e^{\beta y}]$$

Dunque dato che $e^{\alpha x} e^{\beta y} = e^{\alpha x + \beta y}$ ho per definizione che

$$\mathbb{E}[e^{\alpha x + \beta y}] = M_{\alpha X + \beta Y}(t) = \mathbb{E}[e^{\alpha x}] \mathbb{E}[e^{\beta y}] = M_X(\alpha t) M_Y(\beta t)$$

Dunque la mia tesi è dimostrata. □

Teorema 2.3.2. Sia X una v.a. reale e sia $M_X(t)$ la sua funzione generatrice dei momenti. Inoltre se $\exists a > 0 |M_X(t) < +\infty, \forall |t| < a$ allora

1. X ha momenti finiti di ogni ordine cioè $\mathbb{E}|x|^p < +\infty \forall p \geq 1$ e il momento di X può essere scritta come serie di Taylor:

$$M_X(t) = \sum_{n=0}^{\infty} \mathbb{E}[x^n] \frac{t^n}{n!}$$

2. $M_X(t) \in C^\infty((-a, a))$ e vale che

$$\frac{d^n}{dt^n} M_X(t) = \mathbb{E}[x^n]$$

Dimostrazione. (1)

Se $M_{|X|}(t) < +\infty$ allora $\mathbb{E}[e^{t|x|}] < +\infty, \forall |t| < a$.

Dallo sviluppo della serie di Taylor della funzione esponenziale ho che

$$e^\rho = \sum_{k=0}^{\infty} \frac{\rho^k}{k!} \Rightarrow e^\rho \geq \frac{\rho^k}{k!}, \forall k \in \mathbb{N}$$

Inoltre posso constatare anche

$$n \sum_{k=n}^{\infty} \frac{\rho^k}{k!} \leq \sum_{k=n}^{\infty} k \frac{\rho^k}{k!} = \rho \sum_{k=n}^{\infty} \frac{\rho^{k-1}}{(k-1)!} \leq \rho \sum_{k=0}^{\infty} \frac{\rho^k}{k!} = \rho e^\rho \quad (2.1)$$

Dunque ho che

$$n \sum_{k=n}^{\infty} \frac{\rho^k}{k!} \leq \rho e^\rho \Rightarrow \sum_{k=n}^{\infty} \frac{\rho^k}{k!} \leq \frac{\rho}{n} e^\rho \quad (2.2)$$

A questo considerando $h \in \mathbb{R} |h + |t| < a$ e riprendendo la disuguaglianza (2.1) posso scrivere che

$$|x|^n e^{t|x|} = \frac{(h|x|)^n}{h^n} e^{t|x|} \leq \frac{n!}{h^n} e^{h|x|} e^{|t||x|} = \frac{n!}{h^n} e^{(h+|t|)|x|} = \frac{n!}{h^n} M_{|X|}(h + |t|)$$

Quindi dato che $h + |t| < a$ per costruzione ho per ipotesi che $M_{|X|}(h + |t|) < +\infty$ dunque

$$\mathbb{E}[|x|^n e^{t|x|}] \leq \mathbb{E}\left[\frac{n!}{h^n} e^{(h+|t|)|x|}\right] = \frac{n!}{h^n} \mathbb{E}[e^{(h+|t|)|x|}] < +\infty$$

A questo punto allora scrivendo che

$$\left| M_x(t) - \sum_{k=0}^n \mathbb{E}[x^k] \frac{t^k}{k!} \right| = \left| \mathbb{E}\left(e^{tx} - \sum_{k=0}^n x^k \frac{t^k}{k!}\right) \right| \leq \mathbb{E}\left| e^{tx} - \sum_{k=0}^n x^k \frac{t^k}{k!} \right|$$

Quindi dato che $e^{tx} = \sum_{k=0}^{+\infty} x^k \frac{t^k}{k!}$ allora posso riscrivere tale termine come

$$\mathbb{E}\left| \sum_{k \geq n+1} x^k \frac{t^k}{k!} \right| \leq \mathbb{E}\left[\sum_{k \geq n+1} |x|^k \frac{|t|^k}{k!} \right]$$

Per la disequazione (2.2) ho che

$$\mathbb{E}\left[\sum_{k \geq n+1} \frac{|x||t|^k}{k!} \right] \leq \mathbb{E}\left[\frac{|x||t|}{n+1} e^{|x||t|} \right] = \frac{|t|}{n+1} \mathbb{E}[|x| e^{|x||t|}]$$

Ma dato che ho dimostrato precedentemente che $\mathbb{E}[|x| e^{t|x|}] < +\infty$ allora

$$\lim_{n \rightarrow +\infty} \left| M_x(t) - \sum_{k=0}^n \mathbb{E}[x^k] \frac{t^k}{k!} \right| \leq \lim_{n \rightarrow +\infty} \frac{|t|}{n+1} \mathbb{E}[|x| e^{|x||t|}] = 0$$

E quindi ho la mia tesi:

$$\lim_{n \rightarrow +\infty} \sum_{k=0}^n \mathbb{E}[x^k] \frac{t^k}{k!} = M_x(t)$$

□

Dimostrazione. (2)

Dunque dato che per definizione ho mostrato il fatto che

$$M_x(t) = \sum_{k=0}^{\infty} \mathbb{E}[x^k] \frac{t^k}{k!}$$

Allora, sapendo che tutti i termini di grado inferiore a n si annullano se derivo per t^n , ottengo

$$\frac{d^n}{dt^n} M_X(t) = \sum_{k=n}^{\infty} \mathbb{E}[x^k] \frac{t^{k-n}}{(k-n)!}$$

Facendo un cambio di variabili: $l = k - n$ allora ottengo

$$\sum_{l=0}^{\infty} \mathbb{E}[x^{l+n}] \frac{t^l}{l!} = \mathbb{E} \left[x^n \sum_{l=0}^{\infty} x^l \frac{t^l}{l!} \right] = \mathbb{E}[x^n e^{tx}]$$

Dunque

$$\left. \frac{d^n}{dt^n} M_X(t) \right|_{t=0} = \mathbb{E}[x^n]$$

Quindi è infinitamente derivabile. □

Esercizio. Dimostrare che:

- Determinare la funzione generatrice dei momenti per $X \sim \text{Bin}(n, p)$
- Se X_1, \dots, X_n sono v.a. indipendenti $B(p) \Rightarrow M_{X_1+\dots+X_n}(t) = M_X(t)$.
- Analogamente per $X \sim \text{Po}(\lambda)$: calcolare la funzione generatrice dei momenti e mostrare che se X_1, \dots, X_n sono v.a. indipendenti $\text{Po}(\lambda_i)$, $\lambda_1 + \dots + \lambda_n = \lambda \Rightarrow M_{X_1+\dots+X_n}(t) = M_X(t)$.
- Calcolare media e varianza di una v.a. $X \sim \text{Po}(\lambda)$.

Dimostrazione. (1)

Assunto per ipotesi che $X \sim \text{Bin}(n, p)$, $x_k = k$ dunque $p_k = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, \dots, n$ allora ho per definizione che

$$M_X(t) = \mathbb{E}[e^{tx}] = \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n \binom{n}{k} (pe^t)^k (1-p)^{n-k} = (1-p+pe^t)^n$$

□

Dimostrazione. (2)

Ora se $X_1 \sim \text{Be}(p) \Rightarrow M_{X_1}(t) = (1-p+pe^t)$ dunque dato che X_1, \dots, X_n sono variabili indipendenti avrei per il teorema precedente che

$$M_{X_1+\dots+X_n}(t) = \prod_{k=1}^n M_{X_i}(t) = \prod_{k=1}^n (1-p+pe^t) = (1-p+pe^t)^n$$

Dunque dato che questa equazione è uguale a quella precedente allora anche il secondo punto è dimostrato. □

Dimostrazione. (3)

Se assumo che $X \sim \text{Po}(\lambda)$, $x_k = k$, $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$, $k \geq 0$ allora ho che

$$M_x(t) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{\lambda(e^t-1)}$$

E che

$$M_{X_1+\dots+X_n}(t) = \prod_{k=1}^{\infty} e^{\lambda_k(e^t-1)} = e^{\sum_{k=1}^{\infty} \lambda_k(e^t-1)} = e^{\lambda(e^t-1)}$$

Dunque anche il terzo punto è dimostrato. □

Dimostrazione. (4)

A questo punto considero $M_X(t) = e^{\lambda(e^t-1)}$ dunque ho che

$$\frac{d}{dt}M_x(t) = M_x(t)\lambda e^t \Big|_{t=0} = \lambda = \mathbb{E}[X]$$

$$\frac{d^2}{dt^2}M_x(t) = M_x(t)\lambda^2 e^{2t} + M_x(t)\lambda \Big|_{t=0} = \lambda^2 + \lambda = \mathbb{E}[X^2]$$

Perciò ho che $\text{var}(X) = \mathbb{E}[x^2] - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

□

Capitolo 3

Distribuzioni

3.1 Distribuzioni discrete

3.1.1 Distribuzione uniforme (discreta)

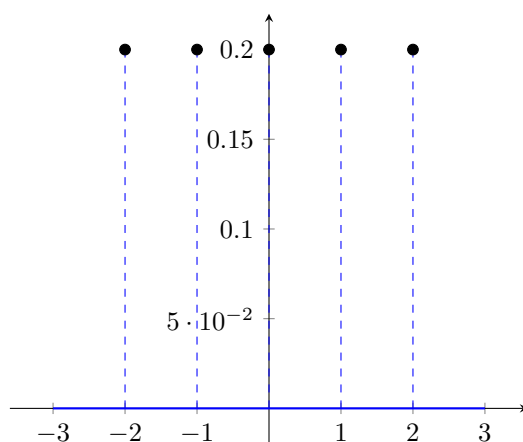
Definizione 3.1.1. (*Distribuzione uniforme (discreta)*)

Sia X una v.a. e sia E un insieme arbitrario finito, diremo che X è una variabile aleatoria uniforme discreta a valori in E , indicata come $X \sim \text{Unif}\{E\}$, se assume ugual probabilità $\forall x \in E$:

$$p_X(x) = \frac{1}{|E|}$$

Analogamente si può scrivere che $X \sim \{(k, p_k), k = 1, \dots, N, p_k = 1/N\}$

Supponiamo che $E := \{-2, -1, 0, 1, 2\}$ allora



Proposizione 3.1.1. Valgono i seguenti fatti:

- **Media**

$$\mathbb{E}[X] = \frac{n+1}{2}$$

- **Varianza**

$$\text{Var}(X) = \frac{(n+1)(5n-1)}{12}$$

- **Funzione generatrice dei momenti**

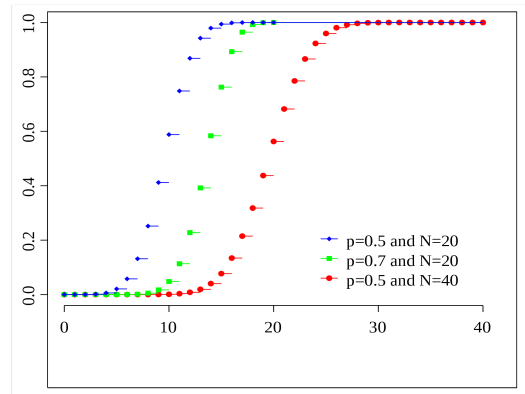
$$M_X(t) = \mathbb{E}(e^{tX}) = \sum e^{tx} p_X = \frac{1}{n} \sum_{k=1}^n e^{tk} = \frac{e^t}{n} \left(\frac{e^{tn} - 1}{e^t - 1} \right)$$

3.1.2 Binomiale

Definizione 3.1.2. Sia X una v.a. si definisce binomiale e si indica con $X \sim \text{Bin}(n, p)$ dove n è il numero di prove effettuate e p è il successo della singola prova di Bernulli X_i . Inoltre possiede distribuzione

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Quindi la probabilità è data da una funzione del tipo:



Proposizione 3.1.2. Valgono i seguenti fatti:

- **Media**

$$\mathbb{E}[X] = np$$

- **Varianza**

$$\text{Var}(X) = np(1-p)$$

- **Funzione generatrice dei momenti**

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k} = (e^t p + 1 - p)^n$$

Osservazione 3.1.1.

$$\text{Bin}(1, p) = \text{Be}(p)$$

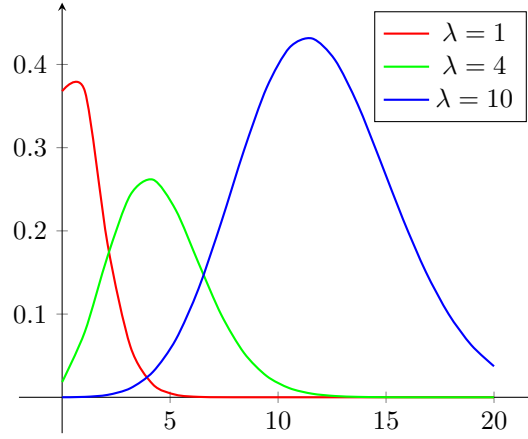
3.1.3 Poisson

Definizione 3.1.3. (Poisson)

Sia X v.a. allora si dice di Poisson e si indica con $X \sim \text{Po}(\lambda)$ se

$$\mathbb{P}(x) = e^{-\lambda} \frac{\lambda^x}{x!} \mathbb{1}_{\mathbb{N}}(x)$$

Dunque ammette un grafico del tipo



Proposizione 3.1.3. Valgono i seguenti fatti:

- *Media*

$$\mathbb{E}[X] = \lambda$$

- *Varianza*

$$\text{Var}(X) = \lambda$$

- *Funzione generatrice dei momenti*

$$M_X(t) = \exp(\lambda(e^t - 1))$$

Proposizione 3.1.4. Siano X, Y due v.a. indipendenti tali per cui $X \sim Po(\lambda_1)$ $Y \sim Po(\lambda_2)$ allora

$$X + Y \sim Po(\lambda_1 + \lambda_2)$$

Dimostrazione. Per definizione e ricordando che tali variabili sono definite positivamente ho che

$$\begin{aligned} \mathbb{P}(X + Y = x) &= \mathbb{P}\left(Y = x - k, \bigcup_{k \in \mathbb{N}} \{X = k\}\right) \\ &= \sum_{k \in \mathbb{N}} \mathbb{P}(Y = x - k, X = k) \\ &= \sum_{k \in \mathbb{N}} \mathbb{P}(Y = x - k) \mathbb{P}(X = k) \\ &= \sum_{k=0}^x \mathbb{P}(Y = x - k) \mathbb{P}(X = k) \\ &= \sum_{k=0}^x e^{-\lambda_2} \frac{\lambda_2^{x-k}}{(x-k)!} \mathbb{1}_{\mathbb{N}}(x-k) e^{-\lambda_1} \frac{\lambda_1^k}{k!} \mathbb{1}_{\mathbb{N}}(x) \\ &= e^{-\lambda_2 - \lambda_1} \lambda_2^x \mathbb{1}_{\mathbb{N}}(x) \sum_{k=0}^x \frac{1}{(x-k)! k!} \left(\frac{\lambda_1}{\lambda_2}\right)^k \\ &= e^{-\lambda_2 - \lambda_1} \lambda_2^x \mathbb{1}_{\mathbb{N}}(x) \frac{\left(\frac{\lambda_1 + \lambda_2}{\lambda_2}\right)^x}{x!} \\ &= e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^x}{x!} \end{aligned}$$

Dunque per definizione questa rappresenta una distribuzione di Poisson di parametro $\lambda_1 + \lambda_2$. □

Lemma 3.1.1. Se una tra queste:

- La funzione di ripartizione congiunta $F(x, y)$

- La distribuzione congiunta $p(x, y)$
- La densità $f(x, y)$.

Si fattorizza come prodotto di due funzioni che dipendono solo da x, y allora le variabili aleatorie marginali sono indipendenti e hanno le rispettive funzioni di ripartizione date dalla decomposizione precedente.

Teorema 3.1.1. Siano X, Y due v.a. indipendenti tali per cui $X \sim Po(\lambda)$ e $Y \sim Po(\mu)$ allora

$$X|(X+Y=b) \sim Bin\left(b, \frac{\lambda}{\lambda+\mu}\right)$$

Dimostrazione.

$$\begin{aligned} \mathbb{P}_{(X+Y=b)}(X=a) &= \frac{\mathbb{P}(X=a, X+Y=b)}{\mathbb{P}(X+Y=b)} = \frac{\mathbb{P}(X=a, Y=b-a)}{\mathbb{P}(X+Y=b)} = \frac{\mathbb{P}(X=a)\mathbb{P}(Y=b-a)}{\mathbb{P}(X+Y=b)} \\ &= e^{-\lambda} \frac{\lambda^a}{a!} e^{-\mu} \frac{\mu^{b-a}}{(b-a)!} e^{\lambda+\mu} \frac{b!}{(\lambda+\mu)^b} \\ &= \frac{b!}{a!(b-a)!} \left(\frac{\lambda}{\lambda+\mu}\right)^a \left(1 - \frac{\lambda}{\lambda+\mu}\right)^{b-a} \\ &= \binom{b}{a} \left(\frac{\lambda}{\lambda+\mu}\right)^a \left(1 - \frac{\lambda}{\lambda+\mu}\right)^{b-a} \end{aligned}$$

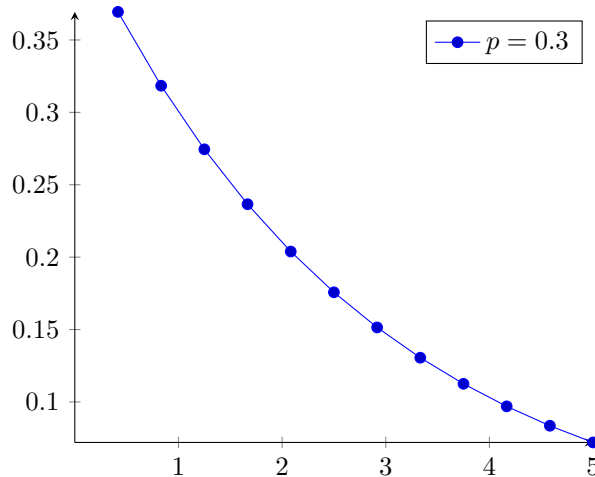
Dunque questa per definizione rappresenta proprio la nostra tesi. □

3.1.4 Distribuzione geometrica

Definizione 3.1.4. Una v.a X si definisce geometrica di parametro p e si indica con $N \sim Geo(p)$ se la sua densità discreta è

$$p_X(k) = p(1-p)^{k-1} \mathbb{1}_{\mathbb{N}}(k)$$

Dunque ammette un grafico del tipo



Proposizione 3.1.5. Valgono i seguenti fatti:

- **Media**

$$\mathbb{E}[X] = \frac{1}{p}$$

- **Varianza**

$$Var(X) = \frac{q}{p^2}$$

- *Funzione generatrice dei momenti*

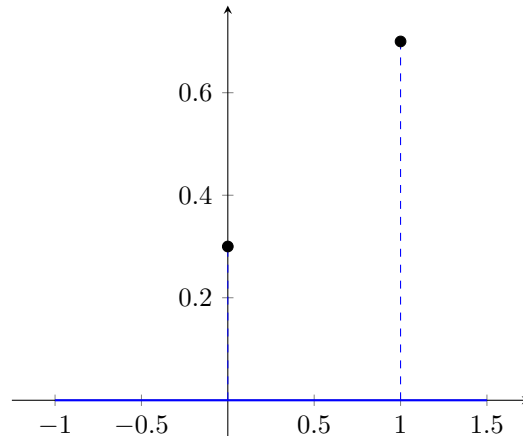
$$M_X(t) = \begin{cases} \frac{p}{e^t - (1-p)} & t < \log \frac{1}{1-p} \\ +\infty & t \geq \log \frac{1}{1-p} \end{cases}$$

3.1.5 Bernulli

Definizione 3.1.5. Una v.a. X si definisce di Bernulli di parametro p e si indica con $X \sim Be(p)$ se $S_X = \{0, 1\}$ e la funzione di probabilità è data da

$$\mathbb{P}(X = k) = (1-p)\mathbb{1}_{\{0\}}(k) + p\mathbb{1}_{\{1\}}(k)$$

Dunque



Proposizione 3.1.6. Valgono i seguenti fatti:

- *Media*

$$\mathbb{E}[X] = p$$

- *Varianza*

$$\text{Var}(X) = p(1-p)$$

- *Funzione generatrice dei momenti*

$$M_X(t) = \mathbb{E}[e^{tX}] = e^{1t}p + e^{0t}(1-p) = e^t p + 1 - p$$

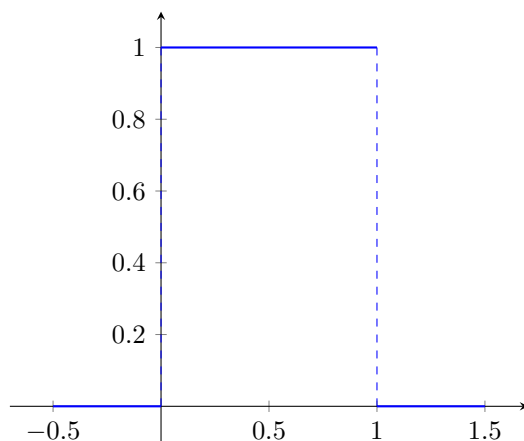
3.2 Variabili aleatorie continue

3.2.1 Uniforme continua

Definizione 3.2.1. Sia X una variabile aleatoria continua, si definisce uniforme se $\exists f \geq 0$, $\int_{\mathbb{R}} f(x) dx = 1$ funzione densità tale per cui

$$f(x) = \mathbb{1}_{(a,b)}(x) \frac{1}{b-a}$$

Dunque se suppongo che $[a, b] = [0, 1]$ allora



Proposizione 3.2.1. *Valgono i seguenti fatti:*

- *Media*

$$\mathbb{E}[X] = \frac{b+a}{2}$$

- *Varianza*

$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

- *Funzione generatrice dei momenti*

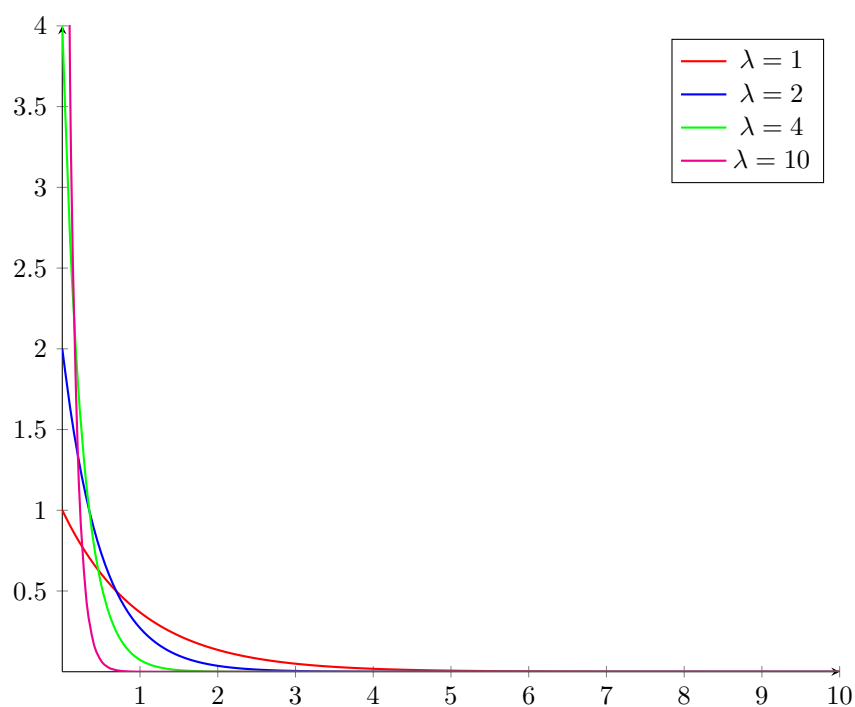
$$M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

3.2.2 Esponenziale

Definizione 3.2.2. *Sia X una variabile aleatoria continua, si definisce di tipo esponenziale di parametro λ e si indica come $\text{Exp}(\lambda) \sim X$ se possiede funzione di densità*

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{(0,+\infty)}(x)$$

Dunque



Proposizione 3.2.2. Valgono i seguenti fatti:

- **Media**

$$\mathbb{E}[X] = \frac{1}{\lambda}$$

- **Varianza**

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

- **Funzione generatrice dei momenti**

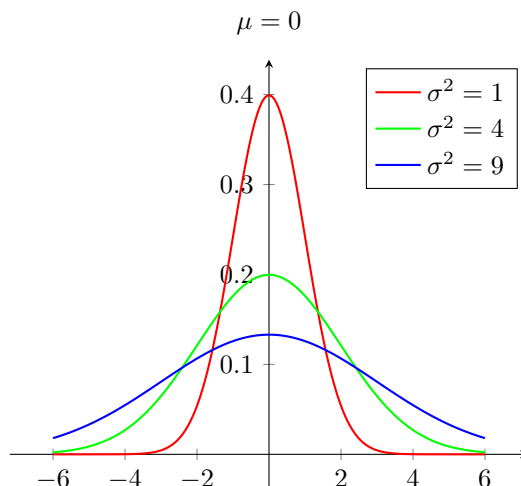
$$M_X(t) = \begin{cases} \frac{\lambda}{\lambda - t} & t < \lambda \\ +\infty & t \geq \lambda \end{cases}$$

3.2.3 Legge normale (o gaussiana)

Definizione 3.2.3. Sia $X : \Omega \rightarrow \mathbb{R}$ una variabile aleatoria continua, si dice che X segue una legge normale (o gaussiana) di parametri $\mu \in \mathbb{R}$ e $\sigma > 0$, $X \sim N(\mu, \sigma^2)$, se la funzione densità è tale per cui

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}$$

Se $Z \sim N(0, 1)$ allora la chiamo legge normale standard inoltre il grafico della densità è dato (a parità di media in quanto rappresenta solo una traslazione) da



Proposizione 3.2.3. Sia $X \sim N(\mu, \sigma^2)$ allora valgono i seguenti fatti:

- **Funzione generatrice dei momenti**

$$M_X(t) = \exp \left\{ \mu t + \frac{\sigma^2 t^2}{2} \right\}$$

Dimostrazione. Per definizione ho che

$$\begin{aligned} M_X(t) &= \mathbb{E} [e^{tX}] = \int_{-\infty}^{+\infty} \frac{e^{xt}}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right) dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{2\sigma^2 tx - x^2 - \mu^2 + 2x\mu}{2\sigma^2} dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{((x + \mu + 2\sigma^2 t)^2}{2\sigma^2} + \frac{\sigma^2 t^2}{2} + t\mu}{2\sigma^2} dx \\ &= e^{\frac{\sigma^2 t^2}{2} + t\mu} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(x + \mu + 2\sigma^2 t)^2}{2\sigma^2} dx \\ &= \exp \left\{ \frac{\sigma^2 t^2}{2} + t\mu \right\} \end{aligned}$$

□

- **Media**

$$\mathbb{E}[X] = \mu$$

Dimostrazione. Per quanto abbiamo visto prima abbiamo che

$$\mathbb{E}[X] = \left. \frac{d}{dt} M_X(t) \right|_{t=0}$$

Dunque poiché

$$\frac{d}{dt} M_X(t) = (\mu + t\sigma^2) \exp \left\{ \frac{\sigma^2 t^2}{2} + t\mu \right\}$$

Allora

$$\mathbb{E}[X] = \mu$$

□

- **Varianza**

$$\text{Var}(X) = \sigma^2$$

Dimostrazione. Per quanto abbiamo visto prima abbiamo che

$$\mathbb{E}[X^2] = \left. \frac{d^2}{dt^2} M_X(t) \right|_{t=0}$$

Dunque poiché

$$\frac{d^2}{dt^2} M_X(t) = \sigma^2 \exp \left\{ \frac{\sigma^2 t^2}{2} + t\mu \right\} + (\mu + t\sigma^2)^2 \exp \left\{ \frac{\sigma^2 t^2}{2} + t\mu \right\}$$

Allora

$$\mathbb{E}[X^2] = \sigma^2 + \mu^2$$

Quindi

$$\text{Var}(X) = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

□

Proposizione 3.2.4. Siano X_1, \dots, X_n v.a. $|X_i \sim N(\mu_i, \sigma_i^2), \forall i$ allora

$$Y = \sum_{i=1}^n a_i X_i \sim N \left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2 \right)$$

3.2.4 Distribuzione gamma

Definizione 3.2.4. Una v.a X si dice che verifica la funzione gamma e si indica con $X \sim \Gamma(\alpha, \beta)$ se possiede densità

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp^{-\beta x} \mathbb{1}_{(0,+\infty)}(x)$$

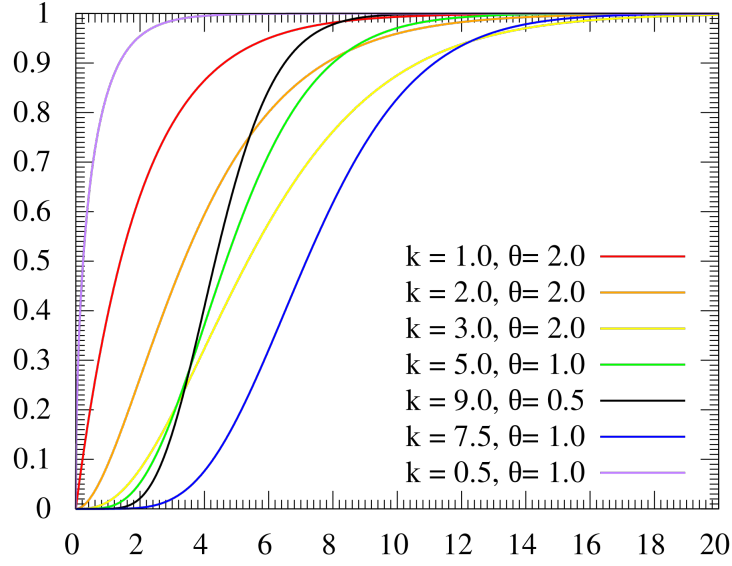
O equivalentemente $X \sim \Gamma(\lambda, k)$

$$f_X(x) = \frac{1}{\Gamma(\alpha) k^\lambda} x^{\lambda-1} \exp^{-\frac{x}{k}} \mathbb{1}_{(0,+\infty)}(x)$$

Con

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

Questa funzione è ben definita $\forall k \in \mathbb{R}, k > 0$ e il grafico della funzione è dato da



Proprietà 3.2.1. Le proprietà più importanti da osservare di questa funzione sono:

- $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$.
- Se $k \in \mathbb{N} \Rightarrow \Gamma(k) = (k - 1)!$.
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Proposizione 3.2.5. Sia $X \sim \Gamma(\alpha, \beta)$ allora valgono i seguenti fatti:

- **Momento di ordine n**

$$\mathbb{E}[X] = \frac{\alpha}{\beta}$$

- **Varianza**

$$\text{Var}(X) = \frac{\alpha}{\beta^2}$$

- **Funzione generatrice dei momenti**

$$M_X(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha} \quad |t| < \beta$$

Osservazione 3.2.1. Sia X una v.a. allora valgono le seguenti affermazioni:

•

$$X \sim \exp(\lambda) \iff X \sim \Gamma(1, \lambda)$$

•

$$X \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right) \iff X \sim \chi_n^2$$

Proposizione 3.2.6. Siano X_1, \dots, X_n v.a. indipendenti tali per cui $X_i \sim \Gamma(\alpha_i, \beta)$ allora

$$Y = \sum_{i=1}^n X_i \sim \Gamma\left(\sum_{i=1}^n \alpha_i, \beta\right)$$

Inoltre sia $a \in \mathbb{R}$ allora se $X \sim \Gamma(\alpha, \beta)$ allora

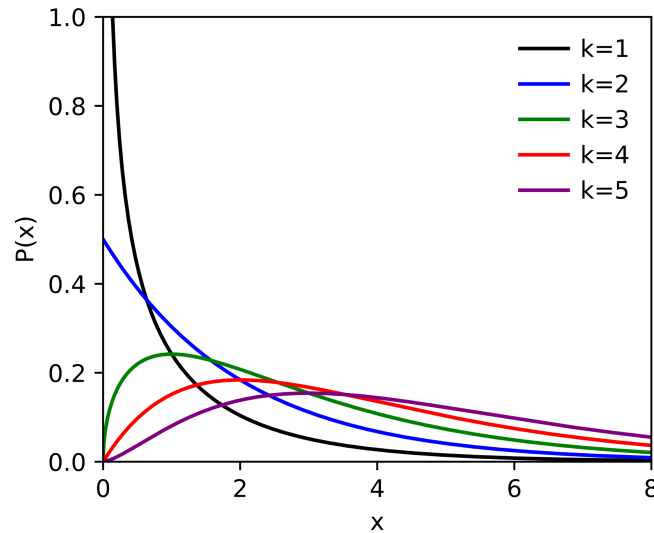
$$aX \sim \Gamma\left(\frac{\alpha}{a}, \beta\right)$$

3.2.5 χ^2

Definizione 3.2.5. Una v.a. X si dice seguire una distribuzione χ^2 a k gradi di libertà e si indica con $X \sim \chi_k^2$ se presenta come funzione di densità

$$f_k(x) = \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-x/2} \quad x > 0$$

Dunque la rappresentazione di tale funzione è



Proposizione 3.2.7. Sia $X \sim \chi_k^2$ allora valgono i seguenti fatti:

- **Media**

$$\mathbb{E}[X] = k$$

- **Varianza**

$$\text{Var}(X) = 2k$$

- **Funzione generatrice dei momenti**

$$M_X(t) = (1 - 2t)^{-\frac{k}{2}} \quad -\frac{1}{2} \leq t \leq \frac{1}{2}$$

- **Funzione caratteristica**

$$\phi_X(t) = (1 - 2it)^{-\frac{k}{2}}$$

Proposizione 3.2.8. Siano X_1, \dots, X_n v.a. indipendenti che seguono una distribuzione normale standard allora

$$Y = \sum_{k=1}^n X_k^2 \sim \chi_n^2$$

Proposizione 3.2.9. Siano X_1, X_2 v.a. indipendenti tali per cui $X_1 \sim \chi_{n_1}^2$ e $X_2 \sim \chi_{n_2}^2$ allora

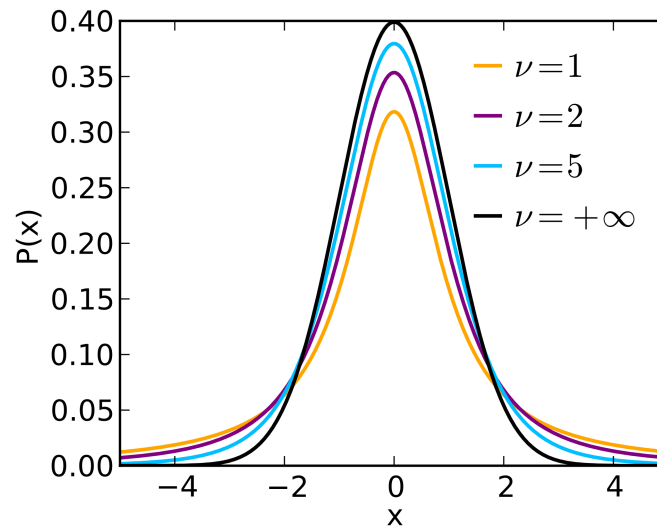
$$X_1 + X_2 \sim \chi_{n_1+n_2}^2$$

3.2.6 t -student

Definizione 3.2.6. Una v.a. X si dice seguire una distribuzione t di student a n gradi di libertà e si indica con $X \sim t_k$ se possiede come funzione di densità

$$f_k(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\left(\frac{n+1}{n}\right)}$$

Dunque la rappresentazione di tale funzione è



Proposizione 3.2.10. Sia $X \sim t_k$ allora valgono i seguenti fatti:

- **Media**

$$\mathbb{E}[X] = 0$$

- **Varianza**

$$\text{Var}(X) = \frac{n}{n-2} \quad n > 2$$

- **Funzione caratteristica**

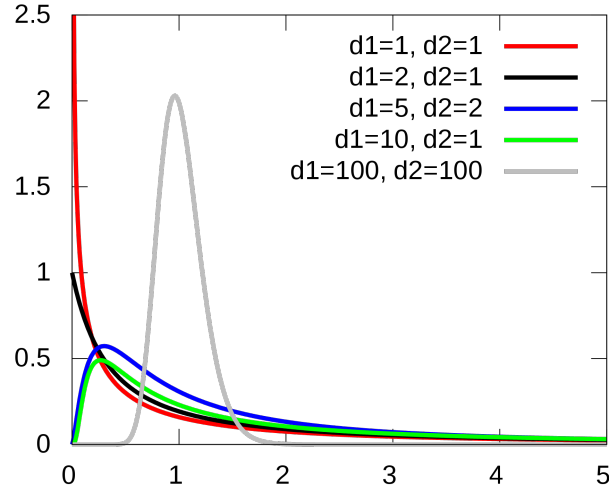
$$\phi_X(t) = \frac{K_{n/2}(\sqrt{n}|t|)(\sqrt{n}|t|)^{n/2}}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}-1}}$$

3.2.7 Fisher-Snedecor

Definizione 3.2.7. Una v.a. X si dice seguire una distribuzione di Fisher-Snedecor a (m, n) gradi di libertà e si indica con $X \sim F_{m,n}$ se possiede come funzione di densità

$$f_k(x) = \frac{1}{x} \left(B\left(\frac{m}{2}, \frac{n}{2}\right)\right)^{-1} \left(\frac{(mx)^m n^n}{(nx+n)^{n+m}}\right)^{\frac{1}{2}}$$

Dunque la rappresentazione di tale funzione è



Proposizione 3.2.11. Sia $X \sim \chi_k^2$ allora valgono i seguenti fatti:

- **Media**

$$\mathbb{E}[X] = \frac{n}{n-2} \quad n > 2$$

- **Varianza**

$$\text{Var}(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$$

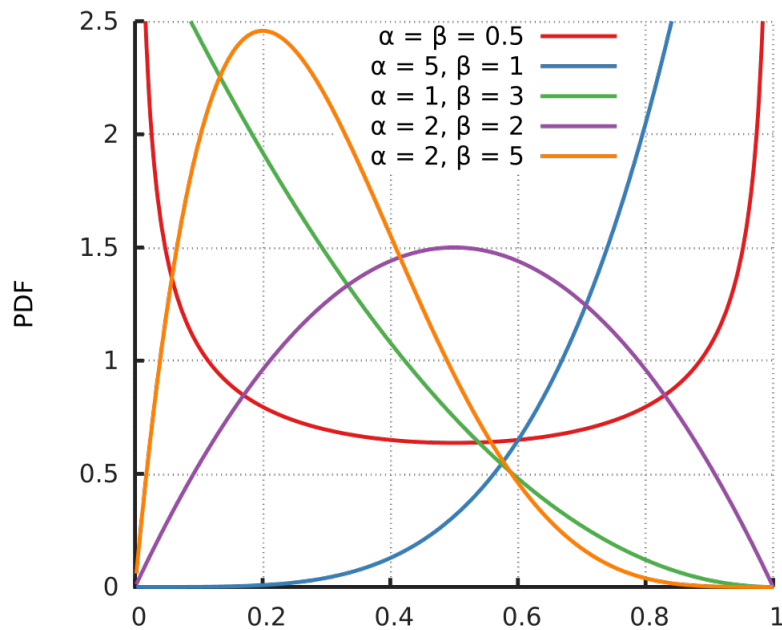
3.2.8 Distribuzione Beta

Definizione 3.2.8. Sia X una v.a. e $\alpha, \beta \in \mathbb{R}^+$ allora dico che X segue una distribuzione beta di parametri α, β , $X \sim B(\alpha, \beta)$ se la funzione di densità è

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \mathbb{1}_{[0,1]}(x)$$

Con

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}$$



Proposizione 3.2.12. Sia X una v.a. tale per cui $X \sim B(\alpha, \beta)$ allora valgono i seguenti fatti:

- **Media**

$$\mathbb{E}[X] = \frac{\alpha}{\alpha + \beta}$$

- **Varianza**

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

3.3 Famiglia esponenziali

Le seguenti distribuzioni rappresentano distribuzioni esponenziali:

Distribuzione	Forma esponenziale	numero parametri
Binomiale $b(n, p)$	$\binom{n}{x} \mathbb{1}_{\mathbb{N}}(x)(1-p)^n \exp \left\{ x \log \left(\frac{p}{1-p} \right) \right\}$	1
Esponenziale $\exp(\theta)$	$\frac{\theta}{1-\theta} \exp \{ x \log(1-\theta) \}$	1
Normale $N(\mu, \sigma^2)$	$\frac{e^{-\frac{\mu^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2}x^2 + \frac{2\mu}{\sigma^2}x \right\}$	2

3.3.1 Vettori gaussiani

Definizione 3.3.1. (*Vettori gaussiani/Distribuzione gaussiana multivariata*)

Sia $X : \Omega \rightarrow \mathbb{R}^n$ un vettore aleatorio del tipo $X = (X_1, \dots, X_n)^T$ lo definisco vettore gaussiano multivariato se ogni combinazione lineare delle sue componenti è una legge gaussiana con funzione densità

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det(A^2)}} \exp \left(-\frac{1}{2} \langle A^2(x - \mu), (x - \mu) \rangle \right)$$

Equivalentemente sia $a \in \mathbb{R}^n$ allora

$$Y = \langle a, X \rangle = \sum_{i=1}^n a_i X_i$$

Deve essere una legge gaussiana $N(\mu_Y, \sigma_Y^2)$.

Definizione 3.3.2. La funzione generatrice dei momenti per un vettore aleatorio X è

$$M_X(a) = [e^{\langle a, X \rangle}]$$

Con $M_X : \mathbb{R}^n \rightarrow \mathbb{R}$ e $M_X(0) = 1$ come quella unidimensionale può non essere definita su tutto \mathbb{R} .

Sia $Z \sim N(0, 1) | Z = (Z_1, \dots, Z_n)^T$ con Z_1, \dots, Z_n sono variabili aleatorie gaussiane standard indipendenti allora abbiamo che

$$Y = \sum_{i=1}^n a_i Z_i \sim N \left(0, \sum_{i=1}^n a_i^2 \right)$$

$\forall a \in \mathbb{R}^n \Rightarrow Z \sim N(0, Id)$.

$$\text{Cov}(Z_i, Z_j) = \mathbb{E}[Z_i, Z_j] = 0$$

In quanto le variabili aleatorie sono supposte indipendenti inoltre dato che $\text{Cov}(Z_i, Z_i) = \text{Var}(Z_i) = 1$ ecco quindi che la matrice di covarianza C di Z è Id quindi la matrice identità.

Sia A una matrice $n \times n$ simmetrica, ≥ 0 con elementi $A = (\sigma_{ij})_{i,j=1}^n$ allora posto $X = AZ + \mu$ ho che

$$\text{Cov}(X) = AA^T \text{Cov}(Z) = AA^T = A^2$$

E

$$\mathbb{E}[X] = \mu$$

Dunque diremo che $X \sim N(\mu, A^2)$, a questo punto dovremmo vedere cosa succede al prodotto scalare

$$\begin{aligned} \langle a, X \rangle &= \sum a_i X_i = \sum a_i (AZ + \mu)_i = \sum a_i \left(\sum_j \sigma_{ij} Z_j \right) + \sum a_i \mu_i \\ &= \sum_{j=1}^n \left(\sum_{i=1}^n \sigma_{ij} a_i \right) Z_j + \langle a, \mu \rangle \\ &= \langle Z, Aa \rangle + \langle a, \mu \rangle \sim N(\langle a, \mu \rangle, |Aa|^2) \end{aligned}$$

Dunque per definizione segue che X p una legge gaussiana multivariata.

Questo mi permette di calcolare subito

$$M_X(y) = \mathbb{E} [e^{\langle X, y \rangle}] = \mathbb{E} [e^{\langle \langle u, \mu \rangle + \langle Z, Au \rangle \rangle}] = e^{\langle u, \mu \rangle} \mathbb{E} [e^{\langle Z, Au \rangle}] = e^{\langle u, \mu \rangle} e^{\frac{1}{2} |Au|^2}$$

Teorema 3.3.1. *Sia X una variabile aleatoria ha legge gaussiana multivariata*

$$N(\mu, A^2) \iff M_X(u) = e^{\langle u, \mu \rangle + \frac{1}{2} \langle A^2 u, u \rangle}$$

Definizione 3.3.3. *Sia $X \sim N(\mu, A^2)$ è una gaussiana degenera se $\det(A) = 0$*

Nel caso in cui ho $X = (X_1, \frac{1}{2}X_1 + 1)$ e

$$\text{Cov}(X) = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} \end{bmatrix} \Rightarrow \det \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} \end{bmatrix} = 0$$

Teorema 3.3.2. *Se $X \sim N(\mu, A^2)$ e se $\det(A^2) > 0$ allora $\exists f_X(x)$ tale per cui*

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det(A^2)}} e^{-\frac{1}{2} \langle A^2(x-\mu), (x-\mu) \rangle}$$

Dimostrazione. Vale se Z è la gaussiana standard quindi la densità congiunta è il prodotto delle densità marginali e quindi vale tale formula.

Se $X = AZ + \mu$ la tesi segue dalla formula di trasformazione delle densità . □

Teorema 3.3.3. *Se le variabili aleatorie sono non correlate allora sono indipendenti.*

Dimostrazione. Se X, Y sono indipendenti allora

$$M_{(X,Y)}(u, v) = \mathbb{E} [e^{uX+vY}] = \mathbb{E} [e^{uX} e^{vY}]$$

Avendo supposto essere indipendenti ho che

$$\mathbb{E} [e^{uX} e^{vY}] = \mathbb{E} [e^{uX}] \mathbb{E} [e^{vY}] = M_X(u) M_Y(v)$$

Supponiamo ora (senza perdita di generalità) che X, Y hanno legge gaussiana centrata, quindi a questo punto la tesi diventa che $\mathbb{E}[X, Y] = 0$ dunque

$$M_{(X,Y)}(u, v) = e^{\langle A^2(u,v), (u,v) \rangle}$$

Nel nostro caso ho che

$$A = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_Y^2 \end{bmatrix}$$

Perciò il tutto è equivalente a

$$\exp(\langle (\sigma_x^2 u, \sigma_Y^2 v), (u, v) \rangle) = e^{\sigma_x^2 u^2 + \sigma_Y^2 v^2} = M_X(u) M_Y(v)$$

X, Y sono indipendenti perché

$$M_{(X,Y)}(u) = M_X(u) M_Y(v)$$

□

Precedentemente ho già dimostrato questo risultato:

Teorema 3.3.4. *Se X, Y sono variabili aleatorie tali per cui $M_{(X,Y)}(u) = M_X(u)M_Y(v)$ allora X, Y sono indipendenti.*

Osservazione 3.3.1. *Una caratteristica peculiare delle variabili aleatorie gaussiane è che sono indipendenti se e solo se non sono correlate.*

Teorema 3.3.5. *Se X_1, \dots, X_n sono variabili aleatorie gaussiane indipendenti allora $X = (X_1, \dots, X_n)$ è una variabile aleatoria gaussiana multivariata $X \sim N(\mu, A^2)$, $\mu = (\mu_{X_1}, \dots, \mu_{X_n})$ e $A = \text{diag}(\sigma_{X_1}^2, \dots, \sigma_{X_n}^2)$*

controesempio

Se X_1, \dots, X_n sono variabili aleatorie gaussiane univariate in \mathbb{R} allora non è detto che $X = (X_1, \dots, X_n)$ sia un vettore gaussiano.

Sia $X \sim N(0, 1)$ e $X_2 = \begin{cases} X_1 & |X_1| \leq 1 \\ -X_1 & |X_1| > 1 \end{cases}$ dunque $X = (X_1, X_2)$.

In questo caso X non è un vettore Gaussiano in quanto se lo fosse per definizione avrei che $X_1 + X_2 \sim N(\mu, \sigma^2)$ ma

$$\mathbb{P}(|X_1 + X_2| > 2) = 0$$

E questo non è caratteristica di una variabile aleatoria gaussiana.

Proposizione 3.3.1. *Se $X = (X_1, \dots, X_n)$ è un vettore aleatorio gaussiano multivariato allora $\forall i$ ho che $X_i \sim N(\mu_i, \sigma_i^2)$ dove μ_i è l'elemento i -esimo del vettore μ e $\sigma_i^2 = (A^2)_{ii}$.*

Dimostrazione. Per definizione sappiamo che $(\alpha, X) \sim N(\mu, \sigma^2)$, $\forall \alpha \in \mathbb{R}$ dunque basterà prendere $\alpha = e_i = (0, \dots, 0, 1, 0, \dots, 0)$. \square

3.4 Proprietà variabili aleatorie continue

Teorema 3.4.1. *Sia X una v.a. continua, $X \geq 0$ con densità f allora vale che*

$$\mathbb{E} = \int_0^{\infty} xf(x) dx$$

Dimostrazione. Per semplicità supponiamo che $\int_0^{\infty} xf(x) dx < +\infty$ dunque so che

$$\exists \lim_{L \rightarrow +\infty} \int_L^{+\infty} xf(x) dx = 0$$

Inoltre

$$\forall \epsilon > 0, \exists L_0 > 0 \mid \int_L^{+\infty} xf(x) dx < \epsilon, \forall L > L_0$$

A questo punto basterà mostrare solo una disuguaglianza.

L'idea della dimostrazione è prendere una successione Y_n di v.a. semplici e mostriamo che vale

$$\mathbb{E}[Y_n] \geq \int_0^{+\infty} xf(x) dx - \epsilon$$

Da questo so che passando al sup in n e uso l'arbitrarietà di ϵ per concludere.

Scelgo $L \mid \int_L^{+\infty} xf(x) dx < \frac{\epsilon}{2}$ poi scelgo una partizione π tale per cui la sua ampiezza $\|\pi\| = \sup\{x_{j+1} - x_j\} \leq \frac{\epsilon}{2}$ e definisco

$$\mathbb{E}[Y_n] = \sum_{j=0}^N x_j \int_{x_j}^{x_{j+1}} f(x) dx + L\mathbb{P}(X \leq L)$$

E

$$\begin{aligned} \int_0^{\infty} xf(x) dx - \mathbb{E}[Y_n] &= \sum_{j=0}^N \int_{x_j}^{x_{j+1}} (x - x_j) f(x) dx + \int_L^{\infty} xf(x) dx - L\mathbb{P}(X \leq L) \\ &\leq \|\pi\| \int_0^L f(x) dx + \frac{\epsilon}{2} \leq \epsilon \end{aligned}$$

Quindi data l'arbitrarietà di ϵ ho vinto. □

Se X è una variabile aleatoria continua senza restrizioni sul segno allora scriveremo $X = X^+ - X^-$ dove $X^+, X^- \geq 0$ sono v.a. Diremo che H ha media $\mathbb{E}[X] \in [-\infty, +\infty]$ se almeno uno dei $\mathbb{E}[X^+], \mathbb{E}[X^-]$ è finito e nel caso poniamo la $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$. Diremo che X è in $L^p(\Omega, \mathbb{P})$ se $|x|^p$ ha media finita e infine varrà che se $X \in L^1(\Omega, \mathbb{P})$ allora $\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x) dx$.

Se X è una v.a. continua con densità f e se $g : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione sufficientemente regolare: continua a tratti, allora

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x) dx$$

3.4.1 Calcolo dei momenti per v.a. continua

Teorema 3.4.2. Sia X v.a. continua con densità f_X allora $Y = \alpha X + \beta$, $\alpha > 0$ è una v.a. assolutamente continua e inoltre la sua densità vale

$$f_Y(y) = \frac{1}{\alpha} f_X\left(\frac{y - \beta}{\alpha}\right)$$

Dimostrazione. $\forall t \in \mathbb{R}$ ho che

$$\mathbb{P}(Y \leq t) = \mathbb{P}(\alpha X + \beta \leq t) = \mathbb{P}\left(X \leq \frac{t - \beta}{\alpha}\right)$$

Dunque per definizione ho che

$$\mathbb{P}(Y \leq t) = F_X\left(\frac{t - \beta}{\alpha}\right)$$

Perciò dato che $f_Y(t) = F'_Y(t)$ allora

$$f_Y(t) = \frac{d}{dt} F_X\left(\frac{t - \beta}{\alpha}\right) = \frac{1}{\alpha} F'_X\left(\frac{t - \beta}{\alpha}\right) = \frac{1}{\alpha} f_X\left(\frac{t - \beta}{\alpha}\right)$$

A questo punto devo provare che tale variabile è assolutamente continua dunque

$$F_Y(y) = F_X\left(\frac{y - \beta}{\alpha}\right) = \int_{-\infty}^{\frac{y - \beta}{\alpha}} f_X(x) dx$$

Quindi facendo un cambio di variabile $x = \frac{t - \beta}{\alpha}$ allora ho che

$$F_Y(y) = \int_{-\infty}^y \frac{1}{\alpha} f_X\left(\frac{t - \beta}{\alpha}\right) dt$$

Questa per definizione mi mostra che Y è assolutamente continua. □

3.5 Vettori Aleatori

Definizione 3.5.1. Siano X_1, \dots, X_n v.a. allora definiamo densità congiunta di X_1, \dots, X_n come

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_n|X_{n-1}, \dots, X_1}(x_n|x_{n-1}, \dots, x_1) f_{X_{n-1}|X_{n-2}, \dots, X_1}(x_{n-1}|x_{n-2}, \dots, x_1) \cdots f_{X_1}(x_1)$$

Osservazione 3.5.1. Siano X_1, \dots, X_n v.a. indipendenti allora

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$$

Proprietà 3.5.1. Siano X, Y due v.a. con S_X, S_Y rispettivamente i loro domini e distribuzioni congiunta $f_{X,Y}(x, y)$ allora

$$f_X(x) = \int_{S_Y} f_{X,Y}(x, y) dy$$

$$f_Y(y) = \int_{S_X} f_{X,Y}(x, y) dx$$

Osservazione 3.5.2. Sia (X, Y) un vettore aleatorio bidimensionale con densità $f_{x,y}$ allora la v.a. $X+Y$ è continua con densità pari a

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_{X,Y}(x, z - x) dx = \int_{-\infty}^{+\infty} f_{X,Y}(z - y, y) dy$$

Prendiamo il quadrato $Q = [0, 1] \times [0, 1]$ scegliamo un punto a caso in Q e chiamiamo questa variabile aleatoria $P = (X, Y)$. La legge congiunta è

$$\mathbb{P}(P \in R) = \frac{|R \cap Q|}{|Q|}$$

Si vede che questo è equivalente a

$$\int \int_{R \cap Q} 1 \, dx \, dy$$

Dunque la densità della variabile aleatoria P è $f(x, y) = 1_Q(x, y)$. Con ciò posso anche calcolare le leggi marginali:

$$\mathbb{P}(X \leq t) = \int_0^t \int_0^1 f(x, y) \, dy \, dx = t$$

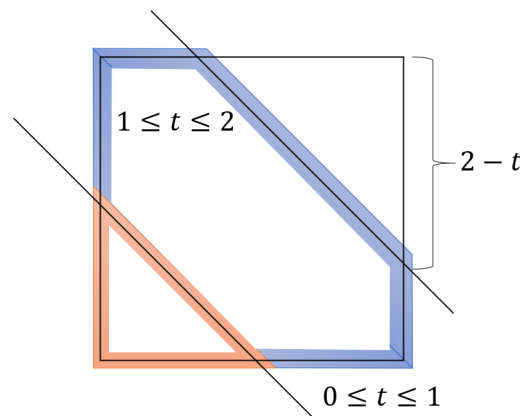
Dunque

$$\begin{aligned} X &\sim \text{Unif}[0, 1] & f_X(t) &= 1_{[0, 1]} \\ Y &\sim \text{Unif}[0, 1] & f_Y(t) &= 1_{[0, 1]} \end{aligned}$$

Dunque vale che $f(x, y) = f_X(x)f_Y(y) \Rightarrow$ le variabili aleatorie X, Y sono indipendenti. Questa uguaglianza vale $\forall x, y \in \mathbb{R}$.

Siano $S = X + Y$, $T = XY$, $V = \frac{X}{Y}$, $M = \max\{X, Y\}$, $N = \min\{X, Y\}$.
 $S = X + Y$

$$\mathbb{P}(S \leq t) = \mathbb{P}(X + Y \leq t) = \mathbb{P}((X, Y) \in R_t) = |R_t|$$



Si possono vedere due cose:

- Se $0 \leq t \leq 1$ allora

$$|R_t| = \frac{t^2}{2}$$

- Se $1 \leq t \leq 2$ allora l'area è data dall'area del complementare del quadratino

$$|R_t| = 1 - \frac{(2-t)^2}{2}$$

Perciò

$$f_S(t) = \begin{cases} t & 0 < t < 1 \\ 2-t & 1 < t < 2 \end{cases}$$

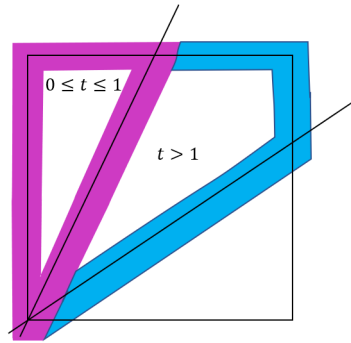
Per calcolare i momenti uso semplicemente la linearità quindi

$$\mathbb{E}[S] = \mathbb{E}[X] + \mathbb{E}[Y] = \frac{1}{2} + \frac{1}{2} = 1$$

$V = \frac{X}{Y}$
 Si osserva che $V \in (0, +\infty)$ e abbiamo

$$\mathbb{P}(V \leq t) = \mathbb{P}(X \leq tY) = \mathbb{P}\left(Y \geq \frac{1}{t}X\right)$$

Dunque il disegno è del tipo



Si possono vedere due cose:

- Se $0 \leq t \leq 1$ allora

$$|R_t| = \frac{t}{2}$$

- Se $t > 1$ allora

$$|R_t| = 1 - \frac{1}{2t}$$

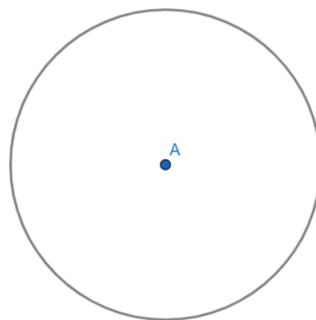
Dunque posso calcolare la densità

$$f_V(t) = \begin{cases} \frac{1}{2} & 0 < t \leq 1 \\ \frac{1}{2}t^2 & t > 1 \end{cases}$$

Quindi si vede che

$$\mathbb{E}[V] = \int_0^1 \frac{1}{2}t \, dt + \int_1^\infty \frac{1}{2t} \, dt = +\infty$$

Esempio 3.5.1. Dato il cerchio centrato nell'origine $C = B_0(1)$ suppongo che $(X, Y) \sim \text{Unif}(C)$



TH:

1. Dato un punto qualsiasi (x, y) , calcolare che $\mathbb{P}(X \leq x)$.
2. Calcolare media di X, Y .

Punto 1

Per definizione posso vedere che la densità congiunta è della forma

$$f(x, y) = \frac{1}{\pi} 1_C(x, y)$$

Dunque si può vedere per Fubini che fissato una x allora $y \in [-\sin(\arccos(x)), \sin(\arccos(x))]$ dunque per definizione posso calcolare che

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f(x, y) dy$$

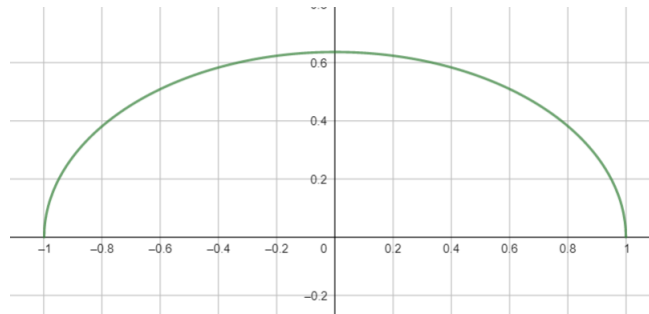
Dunque considero un punto all'interno di tale cerchio

$$\mathbb{P}(X \leq x) = \int_{-1}^x \int_{-\sqrt{1-t^2}}^{\sqrt{1-t^2}} f(t, s) ds dt = 1 - \frac{\arccos(x) - x\sqrt{1-x^2}}{\pi}$$

Dunque

$$f_X(x) = \frac{2}{\pi} \sqrt{1-x^2}$$

Che se rappresentata è della forma



Punto 2

Si può facilmente vedere che

$$f_Y(x) = f_X(x)$$

Dunque è ovvio che

$$f(x, y) = \frac{2}{\pi} 1_C(x, y) \neq f_X(x)f_Y(y) = \left(\frac{2}{\pi}\right)^2 \sqrt{1-x^2}\sqrt{1-y^2}$$

Allora le variabili aleatorie non sono indipendenti. Per quanto riguarda la media ho che

$$\mathbb{E}[X] = \int_{-1}^1 x f_X(x) dx = 0 = \mathbb{E}[Y]$$

Esercizio. Sia $X \sim \text{Unif}([0, 1])$ e supponiamo di prendere $Y \sim \text{Unif}([a, a+1])$ se $(X = a)$ dunque $f_{Y|X}(y|a) = 1_{(a, a+1)}(y)$ quindi la distribuzione congiunta è

$$f(x, y) = 1_{(0,1)}(x) 1_{(x, x+1)}(y)$$

Dunque

$$f_y(y) = \int_{\mathbb{R}} f(x, y) dx = \begin{cases} y & 0 < y \leq 1 \\ 2 - y & 1 < y < 2 \end{cases}$$

Quale è la probabilità che

$$\mathbb{P}(Y > 2X) = \int_0^1 \int_{2x}^{1+x} f(x, y) dy dx = \frac{1}{2}$$

Sia P una variabile aleatoria in \mathbb{R}^n con densità $f(x)$ e sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continua tranne in un insieme trascurabile di punti

Teorema 3.5.1. *Sia X un vettore aleatorio continuo n dimensionale con densità f_X e sia $g : \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione continua allora*

$$\mathbb{E}[g(X)] < +\infty \iff |g(x)|f_X(x) \text{ è integrabile}$$

In tal caso vale che

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^n} g(x)f_X(x) dx$$

Teorema 3.5.2. *Sia X un vettore aleatorio continuo n dimensionale e supponiamo che $\exists U \subset \mathbb{R}^n | \mathbb{P}(X \in U) = 1$. Sia $\varphi : U \rightarrow V$ un diffeomorfismo e sia $Y = \varphi(X)$ allora Y è un vettore aleatorio continuo e la sua densità vale*

$$f_Y(y) = \begin{cases} f_X(\varphi^{-1}(Y)) |\det(D\varphi^{-1})| & y \in V \\ 0 & y \notin V \end{cases}$$

Corollario 3.5.1. *se $\varphi = Ax + b$ allora*

$$f_Y(y) = \frac{1}{|\det(A)|} f_X(A^{-1}(y - b))$$

Definizione 3.5.2. *Sia $X = (X_1, \dots, X_n)^T$, definisco $V = Cov(X) = (V_{i,j})_{i,j=1}^n$ la matrice di covarianza. Definisco*

$$V_{i,j} = cov(X_i, X_j) = \mathbb{E}[X_i, X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]$$

Dunque V è una matrice $n \times n$ simmetrica ($V = V^T$)

Teorema 3.5.3. *Se le componenti del vettore X sono indipendenti allora V è diagonale.*

Teorema 3.5.4. *La matrice V è semidefinita positiva: $\forall x \in \mathbb{R}^n$*

$$\langle Vx, x \rangle \geq 0$$

Dimostrazione. Per definizione ho che

$$\langle Vx, x \rangle = \sum_{i,j} Cov(X_i, X_j) x_i x_j = Cov \left(\sum_i x_i X_i, \sum_j x_j X_j \right) = Var \left(\sum_i x_i X_i \right)$$

Ma questi numeri sono tutti positivi e quindi il tutto è positivo. \square

Se V è simmetrica e semidefinita positiva allora esiste un'unica matrice simmetrica $V^{\frac{1}{2}} | V = \left(V^{\frac{1}{2}}\right)^2$.

Teorema 3.5.5. *Data una matrice V simmetrica e semidefinita positiva allora esiste un vettore aleatorio $X | V = Cov(X)$.*

Dimostrazione. Sia Z un vettore aleatorio n -dimensionale tale per cui le sue componenti Z_i sono indipendenti e $\forall i$ ho che

$$V(Z_i) = 1 \iff Cov(Z) = I_{n \times n}$$

Definiamo $X = V^{\frac{1}{2}}Z$, perciò

$$X_i = \sum_{j=1}^n V_{ij}^{\frac{1}{2}} Z_j$$

inoltre

$$Cov(X)_{ij} = Cov(X_i, X_j) = Cov \left(\sum_{l=1}^n V_{il}^{\frac{1}{2}} Z_l, \sum_{m=1}^n V_{jm}^{\frac{1}{2}} Z_m \right) = \sum_l \sum_m V_{il}^{\frac{1}{2}} V_{jm}^{\frac{1}{2}} Cov(Z_l, Z_m) = \sum_l \sum_m V_{il}^{\frac{1}{2}} V_{jl}^{\frac{1}{2}} = V_{i,j}$$

Quindi segue che $Cov(X) = V$. \square

3.6 Convergenza di V.A

Definizione 3.6.1. Sia $\{X_n\}_{n \in \mathbb{N}}$ una successione di variabili aleatorie definite sullo stesso spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$. $\forall n, X_n$ si definisce **iid**= indipendenti e identicamente distribuite se comunque preso preso $(X_{n_1}, \dots, X_{n_j})$ un sottoinsieme finito abbiamo che

$$X_{n_i} \perp X_{n_j} \quad \forall i \neq j$$

E le leggi di X_{n_1}, \dots, X_{n_j} sono uguali.

Definizione 3.6.2 (Convergenza in senso quasi certo). $X_n : \Omega \rightarrow \mathbb{R}$ misurabili allora dico che X_n converge in senso quasi certo a X v.a. se

$$X_n(\omega) \rightarrow X(\omega) \quad \omega \in \Omega' | \mathbb{P}(\Omega') = 1$$

Ovvero che $\mathbb{P}(X_n \not\rightarrow X) = 0$.

Definizione 3.6.3 (Convergenza in probabilità). Sia $\{X_n\}_{n \in \mathbb{N}}$ una successione di v.a. in $(\Omega, \mathcal{A}, \mathbb{P})$ tali per cui $X_n : \Omega \rightarrow \mathbb{R}, \forall n \in \mathbb{N}$, si dice che X_n converge in probabilità alla v.a. X ($X_n \xrightarrow{\mathbb{P}} X$) se

$$\forall \epsilon > 0 \quad \mathbb{P}(|X_n - X| > \epsilon) \rightarrow 0$$

Controesempio

Se $X_n \xrightarrow{\mathbb{P}} X$ non è detto che $X_n \xrightarrow{q.c.} X$. Infatti se considero $X_n \sim B(p_n = \frac{1}{n})$ con $\mathbb{P}(X_n = 1) = \frac{1}{n}$ e $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$ indipendenti allora $\forall \epsilon > 0$ ho che

$$\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(X_n = 1) \rightarrow 0$$

Possiamo considerare l'evento $E = \{\omega | X_n(\omega) = 1 \text{ infinite volte}\}$ e mostriamo che questo evento ha probabilità 1.

$$\mathbb{P}(E^c) = \mathbb{P}(\{\omega | \exists k, \forall n \geq k, X_n(\omega) = 0\})$$

Dunque definisco $A_n = \{X_n(\omega) = 0\}$ e $B_k = \bigcap_{n \geq k} A_n$ Allora $E^c = \bigcup_k B_k$ dunque

$$\mathbb{P}\left(\bigcup_k B_k\right) \leq \sum_k \mathbb{P}(B_k)$$

$\forall k$ ho che

$$\mathbb{P}(B_k) = \mathbb{P}\left(\bigcap_{n \geq k} A_n\right) \leq \prod_{n \geq k} (1 - p_n) = \exp\left(\log\left(\prod_{n \geq k} (1 - p_n)\right)\right) = \exp\left(\sum_{n \geq k} \log(1 - p_n)\right)$$

Dunque dato che $\log(1 + x) \leq x$ allora

$$\mathbb{P}(B_k) \leq \exp\left(-\sum_{n \geq k} \frac{1}{n}\right) = e^{-\infty} = 0$$

Quindi

$$\mathbb{P}(E) \leq \sum_k \mathbb{P}(B_k) = 0$$

Esempio 3.6.1. $\Omega = [0, 1), \mathcal{A} = \mathcal{B}(\Omega)$ e $\mathbb{P} = \lambda$ la misura di Lebesgue.

Definisco la successione di variabili aleatorie $X_n = 1_{E_n}$ con

$$E_1 = \left[0, \frac{1}{2}\right), \quad E_2 = \left[\frac{1}{2}, 1\right) \\ E_3 = \left[0, \frac{1}{4}\right), \quad E_4 = \left[\frac{1}{4}, \frac{1}{2}\right), \quad E_5 = \left[\frac{1}{2}, \frac{3}{4}\right), \quad E_6 = \left[\frac{3}{4}, 1\right)$$

Dunque è facile vedere che $X_n \xrightarrow{\mathbb{P}} 0$ ma $X_n \not\xrightarrow{q.c.} 0$ perché $\forall \omega, \exists n_k \rightarrow +\infty | X_{n_k}(\omega) = 1$ costruisco una sotto successione

Proposizione 3.6.1. *Data una famiglia di variabili aleatorie $\{X_i\}_{i=1}^n$ sono iid e strettamente positive allora vale che*

$$\mathbb{E} \left[\frac{X_1 + \dots + X_k}{X_1 + \dots + X_n} \right] = \frac{k}{n}$$

Dimostrazione. Se X_1, \dots, X_n sono iid allora per definizione ho che

$$(X_1, \dots, X_n) \stackrel{\text{in distribuzione}}{\cong} (X_{\sigma(1)}, \dots, X_{\sigma(n)})$$

Se X, Y sono variabili aleatorie id a valori in (E, Ξ) e $f : (E, \Xi) \rightarrow (F, \mathcal{F})$ misurabile allora $f(X) = f(Y)$ in distribuzione.

Dunque prendendo

$$f(x_1, \dots, x_n) = \frac{x_1}{x_1 + \dots + x_n}$$

Avremo che ciò è uguale in distribuzione a $\frac{x_{\sigma(1)}}{x_1 + \dots + x_n}$ dunque le variabili aleatorie

$$\left\{ Y_i = \frac{x_1}{x_1 + \dots + x_n} \right\}_i$$

Sono id allora ho che $\mathbb{E}[Y_i] = \mu$ ma dato che $Y_1 + \dots + Y_n = 1$ allora

$$1 = \sum \left[\sum_{i=1}^n Y_i \right] = \sum_{i=1}^n \mathbb{E}[Y_i] = n\mu \Rightarrow \mu = \frac{1}{n}$$

Quindi posso vedere che

$$\mathbb{E} \left[\frac{X_1 + \dots + X_k}{X_1 + \dots + X_n} \right] = \sum_{i=1}^k \mathbb{E}[Y_i] = \frac{k}{n}$$

Dunque la mia tesi è dimostrata. □

Teorema 3.6.1. *Se X_n sono variabili aleatorie con media μ_n e varianza σ_n^2 e se $\mu_n \rightarrow \mu$ e $\sigma_n^2 \rightarrow 0$ allora $X_n \xrightarrow{\mathbb{P}} \mu$.*

Dimostrazione. Per ipotesi so che $\mu_n \rightarrow \mu$ allora $\forall \epsilon, \exists n_\epsilon \forall n > n_\epsilon$ ho che

$$|\mu_n - \mu| > \frac{\epsilon}{2}$$

Dunque dato che

$$B_{\mu_n} \left(\frac{\epsilon}{2} \right) \subset B_\mu(\epsilon)$$

Allora

$$\mathbb{P}(|X_n - \mu| > \epsilon) \leq \mathbb{P}(|X_n - \mu_n| > \frac{\epsilon}{2})$$

Adesso per la disuguaglianza di Chebishev ho che

$$\mathbb{P}(|X_n - \mu| > \epsilon) \geq \frac{4\sigma_n^2}{\epsilon^2}$$

Ma dato che per ipotesi $\sigma_n^2 \rightarrow 0$ allora il tutto tende a 0 e quindi per definizione questo è equivalente a dire che $X_n \xrightarrow{\mathbb{P}} \mu$. □

Sia $U_n \sim Unif([0, 1])$ variabili aleatorie iid e chiamiamo $X_n = \min\{U_k | k \leq n\}$ allora abbiamo che $X_n \xrightarrow{\mathbb{P}} 0$ in quanto

$$\mathbb{P}(X_n > \epsilon) \leq \mathbb{P}(U_k > \epsilon, \forall k = 1, \dots, n) = \prod_{k=1}^n (1 - \epsilon) = (1 - \epsilon)^n$$

Ma questa tende a 0 per $n \rightarrow +\infty$ quindi per definizione ho che $X_n \xrightarrow{\mathbb{P}} 0$. La velocità di convergenza è più alta di quella data da Chebichev in quanto lui mi avrebbe detto che

$$\mathbb{P}(X_n > \epsilon) \leq \frac{\mathbb{E}[X_n]}{\epsilon}$$

Ma dato che

$$\mathbb{E}[X_n] = \int_0^1 nx(1-x)^{n-1} dx = \int_0^1 n(1-x)x^{n-1} dx = 1 - \frac{n}{n+1} = \frac{1}{n+1}$$

Allora

$$\mathbb{P}(X_n > \epsilon) \leq \frac{\mathbb{E}[X_n]}{\epsilon} \leq \frac{1}{(n+1)\epsilon}$$

Capitolo 4

Esempi

4.0.1 esercizio 1

Data una moneta di parità p definisco:

- S_i il tempo per ottenere l' i -esima vincita.
- N_i il tempo di attesa per l' i -esimo successo a partire dal $(i - 1)$ -esimo successo.

Dunque bisogna provare che:

1. S_1, S_2 non sono indipendenti.
2. N_1, N_2 sono indipendenti.

S_1, S_2 non sono indipendenti

Per definizione ho che

$$\mathbb{P}((S_1 = N_1) \cap (S_2 = N_1 + N_2)) = \mathbb{P}(S_2 = N_1 + N_2 | S_1 = N_1) \mathbb{P}(S_1 = N_1)$$

Ora dato che

$$\mathbb{P}(S_2 = k_2) = \sum_{n=1}^{k_2-1} \mathbb{P}(S_2 = k_2 | S_1 = n) \mathbb{P}(S_1 = n) = (k_2 - 1)(1 - p)^{k_2-1} p^2$$

E che

$$\mathbb{P}(S_2 = N_1 + N_2 | S_1 = N_1) = (1 - p)^{N_2-1} p$$

Si può concludere banalmente che $\mathbb{P}(S_2 = N_1 + N_2 | S_1 = N_1) \neq \mathbb{P}(S_2 = N_1 + N_2)$ dunque questo punto è dimostrato.

N_1, N_2 sono indipendenti

Quale è la distribuzione di S_2 ? È data da $\mathbb{P}(S_2 = b) = \mathbb{P}(\text{nei primi } b - 1 \text{ lanci c'è stato un solo successo e } X_b = T)$

Questo è data dalla probabilità

$$\begin{aligned} \mathbb{P}(\text{nei primi } b - 1 \text{ lanci c'è stato un solo successo}) \mathbb{P}(X_b = T) \\ = \left[\binom{b-1}{1} q^{b-2} p \right] p = (b - 1) q^{b-2} p^2 \end{aligned}$$

Quindi posso calcolare a

$$\mathbb{P}(N_1 = a, N_2 = b - a) = \mathbb{P}(N_1 = a, S_2 = b) = q^{b-2} p^2$$

Se voglio calcolare

$$\mathbb{P}(N_2 = c) = \sum_a \mathbb{P}(N_1 = a, N_2 = c) = \sum_a \mathbb{P}(N_1 = a, S_2 = ac) = \sum_a q^{a+c-2} p^2 = q^{c-1} p^2 \sum_{a \geq 1} q^{a-1} = p q^{c-1}$$

Dunque si vede che $N_2 \sim \text{geom}(p)$

$$\mathbb{P}(N_1 = a, N_2 = b) = q^{a+b-2} p^2 = q^{a-1} p q^{b-1} p = \mathbb{P}(N_1 = a) \mathbb{P}(N_2 = b)$$

Dunque sono indipendenti

Teorema 4.0.1. (*Estrazioni lanci ripetuti*)

La probabilità

$$\mathbb{P}(S_k = a) = \binom{a-1}{k-1} p^k q^{a-k}$$

Dunque $N_k \sim \text{geom}(p)$. Inoltre le variabili aleatorie N_1, \dots, N_k sono indipendenti.

Esercizio. Se ho avuto il secondo successo al lancio b quindi ($S_2 = b$) come è distribuito $S_1 = N_1$?
 $\mathbb{P}_{S_2=b}(S_1 = a) = 0$ se $a \geq b$ ma per la probabilità condizionata ho che

$$\mathbb{P}_{S_2=b}(S_1 = a) = \frac{\mathbb{P}(S_1 = a, S_2 = b)}{\mathbb{P}(S_2 = b)} = \frac{q^{b-2} p^2}{(b-1) q^{b-2} p^2} = \frac{1}{b-1}$$

Dunque S_1 sotto la misura $\mathbb{P}_{S_2=b}$ ha distribuzione uniforme in $\{1, \dots, b-1\}$

4.0.2 La raccolta delle figurine

Abbiamo N figurine diverse e ogni volta compro un pacchetto che tiene una figurina e ciascuna di esse ha la stessa probabilità. Quanto tempo ci vuole per finire un album?

- Se $N = 1$ basta comprare un pacchetto.
- Se $N = 2$ ci sono due casi:
 - Bastano 2 pacchetti con probabilità $\frac{1}{2}$.
 - Ne servono più di 2. Dato che è la stessa situazione della moneta so che servono in media 3.

Definiamo:

- N_1, N_2, \dots, N_n i tempi di attesa per la prima, seconda, ..., figurina.
- S_1, S_2, \dots, S_n tempi di arrivo della figurina, con $S_n = \sum_{i=1}^n N_i$, dunque

$$\mathbb{E}[S_n] = \sum_{i=1}^n \mathbb{E}[N_i]$$

Se riusciamo a dimostrare che N_1, \dots, N_n sono indipendenti ($\mathbb{E}[N_1] = 1 \Rightarrow \text{Var}(N_1) = 0$) allora

$$\text{Var}(S_n) = \sum_{k=1}^n \text{Var}(N_k)$$

Dunque calcoliamo

$$\begin{aligned} \mathbb{P}(N_2 = b) &= \mathbb{P}(X_1 = \dots = X_b, X_{b+1} \neq X_b) = \sum_{j=1}^N \mathbb{P}(X_1 = \dots = X_b = j, X_{b+1} \neq j) = \\ &= \sum_{j=1}^N \left(\frac{1}{N}\right)^b \frac{N-1}{N} = N \left(\frac{1}{N}\right)^b \left(\frac{N-1}{N}\right) = \left(1 - \frac{1}{N}\right) \left(\frac{1}{N}\right)^{b-1} \end{aligned}$$

Dunque si osserva che $N_2 \sim \text{geom}\left(P = 1 - \frac{1}{N}\right)$.

A questo punto ho che

$$\mathbb{P}(N_2 = b, N_3 = c) = \mathbb{P}(X_1 = \dots = X_b, X_{b+1} \neq X_1, X_{b+1}, \dots, X_{b+c} \in \{X_1, X_b\}, X_{b+c+1} \notin \{X_1, X_b\})$$

$$\begin{aligned}
&= \sum_{X_1 \neq X_2} \mathbb{P}(X_1 = \dots = X_b = x_1, X_{b+1} = x_2, X_{b+2}, \dots, X_{b+c} \in \{x_1, x_2\}, X_{b+c+1} \notin \{x_1, x_2\}) \\
&= \sum_{X_1 = X_2} \left(\frac{1}{N}\right)^b \frac{1}{N} \left(\frac{2}{N}\right)^{c-1} \left(1 - \frac{2}{N}\right) = N(N-1) \left(\frac{1}{N}\right)^b \left(\frac{2}{N}\right)^{c-1} \left(1 - \frac{2}{N}\right) \\
&= \left(1 - \frac{1}{N}\right) \left(\frac{1}{N}\right)^{b-1} \left(1 - \frac{2}{N}\right) \left(\frac{2}{N}\right)^{c-1} \\
&= \mathbb{P}(N_2 = b) \mathbb{P}(N_3 = c)
\end{aligned}$$

Dunque $N_3 \sim \text{geom}\left(1 - \frac{2}{N}\right)$ dunque $N_2 \text{ II } N_3$ in effetti vale che $N_k \sim \left(1 - \frac{k-1}{N}\right)$ e che le variabili $\{N_k\}$ sono tra loro indipendenti.

Osservazione 4.0.1. *Posto $n =$ numero di figurine distinte abbiamo che $N_1 = 1$, $N_2 \sim \text{geom}\left(1 - \frac{1}{n}\right)$ in generale ho che $N_k \sim \left(1 - \frac{k-1}{n}\right)$ dunque*

$$\begin{aligned}
S_n &= \sum_{k=1}^n N_k \\
\mathbb{E}[S_n] &= \sum_{k=1}^n \frac{1}{1 - \frac{k-1}{n}} = \sum_{k=1}^n \frac{n}{n - (k-1)} = \sum_{j=1}^n \frac{n}{j}
\end{aligned}$$

Essendo una serie armonica ho che

$$\sum_{j=1}^n \frac{1}{j} = \log(n) + \gamma + \frac{1}{2n} + o\left(\frac{1}{n}\right)$$

Dunque

$$\sum_{j=1}^n \frac{n}{j} = n \log(n) + n\gamma + \frac{1}{2} + c$$

Questo mi dice che cresce in maniera più che lineare e inoltre mi dice quanti pacchetti devo comprare. Per $n = 5$ ho che

$$\mathbb{E}[S_5] \approx 11,4167$$

Teorema 4.0.2. *Possiamo stimare la probabilità di non essere vicini a $n \log(n)$, preso $c > 0$, con*

$$\mathbb{P}(S_n > n \log(n) + cn + 1) \leq e^{-c}$$

Analogamente per $n \rightarrow +\infty$

$$n \log(n) \pm cn = n \log(n) \left(1 \pm \frac{c}{\log(n)}\right)$$

Dunque

$$\mathbb{P}\left(\frac{|S_n - n \log(n)|}{n \log(n)} > \frac{c}{\log(n)}\right) \leq 1 - 2e^{-c}$$

Capitolo 5

Legge (debole) dei grandi numeri

Definizione 5.0.1. Sia data una successione di v.a. reali $\{X_i\}_{i \in I}$ definiti nello spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ tali per cui $\mathbb{E}[X_i] = \mu, \forall i$ allora posto

$$\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n}$$

Diremo che $\{X_i\}_{i \in I}$ soddisfa la legge debole dei grandi numeri se $\overline{X} \xrightarrow{\mathbb{P}} \mu$

Teorema 5.0.1. Sia data una successione di v.a. reali $\{X_i\}_{i \in I}$ iid definiti nello spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ tali per cui $\mathbb{E}[X_i] = \mu < +\infty$ e $\text{Var}(X_i) = \sigma^2, \forall i$ allora posto

$$\overline{X}_n = \frac{S_n}{n} = \frac{(X_1 + \dots + X_n)}{n}$$

Ho che

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu$$

Dimostrazione. Per definizione è facile vedere che

$$\mathbb{E} \left[\frac{S_n}{n} \right] = \frac{1}{n} \sum_{k=1}^n \mathbb{E}[X_k] = \mu$$

Inoltre

$$\text{Var} \left(\frac{S_n}{n} \right) = \frac{1}{n} \sum_{k=1}^n \text{Var}(X_k) = \sigma^2$$

Dunque a questo punto ho per la disuguaglianza di Chebichev che

$$\mathbb{P} \left(\left| \frac{S_n}{n} - \mu \right| > \epsilon \right) \leq \frac{\sigma^2}{n\epsilon^2}$$

Perciò ho che

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\left| \frac{S_n}{n} - \mu \right| > \epsilon \right) = 0$$

E per definizione ciò significa che $\overline{X} \xrightarrow{\mathbb{P}} \mu$ quindi la mia tesi è dimostrata. □

Osservazione 5.0.1. Basta che $\mu_n \rightarrow \mu$ dove $\mu_n = \mathbb{E}[X_n]$ e $\mu \in \mathbb{R}$

Se $X_n \sim B(p)$ esito del lancio di una moneta di parità p e se S_n conta il numero di successi (esce testa) in n lanci allora $\frac{S_n}{n}$ che è uguale alla percentuale di successi converge in probabilità al valore p quindi alla parità della moneta.

Possiamo dire che

$$\mathbb{P} \left(\left| \frac{S_n}{n} - p \right| > \epsilon \right) \rightarrow 0$$

Inoltre Chebichev ci dice che

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2}$$

Anche se questa stima è molto grezza.

Teorema 5.0.2. *Se X_n sono variabili aleatorie iid aventi media μ e con funzione generatrice dei momenti $M_X(s)$ definita in un intorno completo dell'origine allora $\forall \epsilon > 0 \exists \rho \in \mathbb{R}, \rho < 1$ tale per cui*

$$\mathbb{P}\left(\frac{S_n}{n} - \mu > \epsilon\right) \leq \rho^n$$

Osservazione 5.0.2. *Si può dimostrare che*

$$\rho = \inf \left\{ e^{-(\mu+\epsilon)s} M_X(s) \mid 0 < s < 1 \right\} < 1$$

Se $X_n \sim B(p) \Rightarrow p \sim 1 - 2\epsilon^2$ allora

$$\mathbb{P}\left(\frac{S_n}{n} - \mu > \epsilon\right) \leq (1 - 2\epsilon^2)^n$$

E rispetto alla stima precedente è migliore.

5.1 Convergenza in Distribuzione

Definizione 5.1.1. *Sia $\{X_n\}_{n \in \mathbb{N}}$ una successione di variabili aleatorie su $(\Omega, \mathcal{A}, \mathbb{P})$, diremo che $X_n \xrightarrow{D} X$ convergenza in distribuzione (o in legge) se e solo se*

$$\lim_{n \rightarrow +\infty} F_{X_n}(t) = F_X(t)$$

$\forall t$ punto di continuità di F_X .

Definizione 5.1.2. *Se definisco*

$$D_X := \{t \mid F_X(t) \text{ non è continua}\}$$

Allora so che D_X è un insieme discreto che contiene al più un insieme numerabile di punti.

$$t \in D_X \iff \mathbb{P}(X = t) = F_X(t) - F_X(t^-) > 0$$

Esempio 5.1.1. *Sia $\{X_n\}$ variabili aleatorie iid tali per cui $X_n \sim \exp(1)$ inoltre definiamo*

$$Z_n = \min\{X_k \mid k \leq n\}$$

Dunque ho che

$$\mathbb{P}(X_n > t) = (e^{-t})^n = e^{-nt} \rightarrow 0$$

Dunque questo per definizione significa che $Z_n \xrightarrow{\mathbb{P}} 0$.

A questo punto introduco $W_n = nZ_n$ quindi

$$\mathbb{P}(W_n > t) = \mathbb{P}\left(Z_n > \frac{t}{n}\right) = \left(e^{-\frac{t}{n}}\right)^n = e^{-t}$$

Quindi $W_n \sim \exp(1)$ dunque per definizione ho che $W_n \xrightarrow{D} W \sim \exp(1)$.

Adesso prendo in considerazione $X_n \sim \text{Unif}([0, 1])$ dunque

$$\mathbb{P}(Z_n > t) = (1 - t)^n$$

Ma questa converge a 0 per $n \rightarrow +\infty$ quindi per definizione ho che $Z_n \xrightarrow{\mathbb{P}} 0$. Analogamente a prima ho che

$$\mathbb{P}(W_n > t) = \mathbb{P}\left(Z_n > \frac{t}{n}\right) = \left(1 - \frac{t}{n}\right)^n$$

Ma

$$\lim_{n \rightarrow +\infty} \mathbb{P}(W_n > t) = e^{-t}$$

Perciò per definizione ho che $W_n \xrightarrow{D} W \sim \exp(1)$.

In questo caso considero $X_n \sim N(\mu, \sigma_n^2)$ $\sigma_n^2 \frac{1}{n}$, voglio vedere se questa successione converge in distribuzione quindi essendo una successione di v.a. gaussiane devo vedere che converga a μ .

Teorema 5.1.1. Sia $\{X_n\}$ una successione di variabili aleatoria su $(\Omega, \mathcal{A}, \mathbb{P})$ tali per cui $X_n \xrightarrow{\mathbb{P}} X$ allora

$$X_n \xrightarrow{D} X$$

Dimostrazione. Sia $z \notin D_x$ e $\epsilon > 0$ dunque

$$(X_n \leq z) \subset (X \leq z + \epsilon) \cup (|X_n - X| \geq \epsilon) \implies \mathbb{P}(X_n \leq z) \leq \mathbb{P}(X \leq z + \epsilon) + \mathbb{P}(|X_n - X| \geq \epsilon)$$

Ma dato che la successione converge in probabilità per ipotesi allora posso affermare che

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X \leq z + \epsilon) + \mathbb{P}(|X_n - X| \geq \epsilon) = \mathbb{P}(X \leq z + \epsilon)$$

Dunque riprendendo quanto visto in precedenza ho che

$$\limsup \mathbb{P}(X_n \leq z) \leq \mathbb{P}(X \leq z + \epsilon)$$

Data l'arbitrarietà di ϵ allora

$$\limsup \mathbb{P}(X_n \leq z) \leq \mathbb{P}(X \leq z)$$

Analogamente scambiando X con X_n e z con $z - \epsilon$ abbiamo che

$$(X \leq z - \epsilon) \subset (X_n \leq z) \cup (|X - X_n| \geq \epsilon) \implies \mathbb{P}(X \leq z - \epsilon) \leq \mathbb{P}(X_n \leq z) + \mathbb{P}(|X - X_n| \geq \epsilon)$$

Da cui posso dimostrare l'altra disuguaglianza

$$\mathbb{P}(X \leq z - \epsilon) \leq \mathbb{P}(X_n \leq z) + \mathbb{P}(|X_n - X| > \epsilon)$$

Allora data la continuità della funzione di ripartizione

$$\mathbb{P}(X \leq z - \epsilon) \leq \liminf \mathbb{P}(X \leq z)$$

Ma data sempre l'arbitrarietà di ϵ ho che

$$\mathbb{P}(X \leq z) \leq \liminf \mathbb{P}(X \leq z)$$

Dunque i due limiti coincidono perciò ho che

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_n \leq z) = \mathbb{P}(X \leq z)$$

Quindi per definizione ho che

$$\lim_{n \rightarrow +\infty} F_{X_n}(z) = F_X(z)$$

E quindi per definizione $X_n \xrightarrow{D} X$ perciò la mia tesi è dimostrata. □

Osservazione 5.1.1. *Il viceversa in generale non è vero in quanto se considero $(\Omega = [0, 1), \mathcal{A} = B(\Omega), \mathbb{P} = \lambda)$ e prendiamo le variabili aleatorie in questo modo:*

$$X_1 = X_3 = \dots = X_{2n+1} = 1_{[0, \frac{1}{2})}$$

Con $X_1 : \Omega \rightarrow \mathbb{R}, X_1 \sim \{(0, \frac{1}{2}), (\frac{1}{2}, 1)\} \sim B(\frac{1}{2})$. Mentre

$$X_2 = X_4 = \dots = X_{2n} = 1_{[\frac{1}{2}, 1)}$$

Con $X_2 : \Omega \rightarrow \mathbb{R}, X_2 \sim \{(0, \frac{1}{2}), (\frac{1}{2}, 1)\} \sim B(\frac{1}{2})$.

Si vede che $X_n \xrightarrow{D} X \sim B(p)$ dato che tutte le funzioni sono uguali, ma $X_n \not\xrightarrow{\mathbb{P}} X$ in quanto se per assurdo convergesse avrei che

$$\mathbb{P}(|X_{2n} - X| > \epsilon) \rightarrow 0$$

$$\mathbb{P}(|X_{2n+1} - X| > \epsilon) \rightarrow 0$$

Ma queste due sotto-successioni hanno concentrazioni in intervalli diversi dato che nel primo caso ho che $X(\omega) \approx 1$ e che $X(\omega) \approx 0$ nel secondo caso ma ciò non può accadere contemporaneamente.

Teorema 5.1.2. *Se $X_n \xrightarrow{D} X \equiv \mu$ allora $X_n \xrightarrow{\mathbb{P}} X$.*

Dimostrazione. Per definizione posso affermare che

$$\mathbb{P}(X_n > \mu + \epsilon) = 1 - \mathbb{P}(X_n \leq \mu + \epsilon)$$

Dato che per ipotesi la successione converge in distribuzione allora ho che il tutto tende a $1 - F_X(\mu + \epsilon) = 0$ dunque

$$\mathbb{P}(X_n < \mu - \epsilon) \rightarrow 1 - F_X(\mu - \epsilon) = 0$$

Dunque la somma di questi due numeri che corrisponde a

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - \mu| > \epsilon) = 0$$

□

Esempio 5.1.2. *Sia $X_n \sim Unif([0, \frac{1}{n}])$ dunque so che*

$$F_{X_n}(t) = \begin{cases} 0 & t \leq 0 \\ nt & 0 \leq t \leq \frac{1}{n} \\ 1 & t \geq \frac{1}{n} \end{cases}$$

Quindi abbiamo 2 casi:

- Se $t < 0$ $F_{X_n}(t) \rightarrow F_X(t) = 0$.
- Se $t > 0$ $\forall n > \frac{1}{t}, F_{X_n}(t) = 1 \rightarrow F_X(t) = 1$.
- $t = 0$ non ci interessa perché $0 \in D_X$.

Quindi si vede che abbiamo una successione che converge in distribuzione ma dato che

$$X_n \rightarrow 0$$

Allora per il risultato precedente ho che questa successione converge anche in probabilità.

Teorema 5.1.3. *Se $X_n \xrightarrow{D} X$ allora $\mathbb{P}(X_n \in A) \rightarrow \mathbb{P}(X \in A)$, per ogni boreliano A tale per cui $\mathbb{P}(X \in \partial A) = 0$.*

Lemma 5.1.1. Sia $X_n \xrightarrow{D} X$ (per semplicità supponiamo che $D_X = \phi$). Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ una funzione semplice a supporto compatto:

$$f(x) = \sum_{k=1}^K \varphi_k \mathbb{1}_{A_k}(x)$$

Dove $\bigcup_k A_k \subset \mathbb{R}$ è un compatto e $A_k \cap A_j = \emptyset$ allora

$$\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$$

Dimostrazione. per definizione ho che

$$\mathbb{E}[f(X_n)] = \sum \varphi_k \mathbb{P}(X_n \in A_k) \rightarrow \sum \varphi_k \mathbb{P}(X \in A_k)$$

Ma quest'ultimo per definizione è $\mathbb{E}[f(X)]$. □

Teorema 5.1.4. Se $f : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione continua a supporto compatto allora $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$.

Osservazione 5.1.2. Vale anche il viceversa se $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ allora $X_n \xrightarrow{D} X$.

5.1.1 Approssimazione di Poisson alla binomiale

Teorema 5.1.5. Siano $\{X_n\}_{n \in \mathbb{N}}$, X v.a. a valori in \mathbb{N} allora

$$X_n \xrightarrow{D} X \iff \forall n \in \mathbb{N}, p_n(x) = \mathbb{P}(X_n = x) \rightarrow p_X(x)$$

Dimostrazione. Per dimostrare questo risultato è necessario dimostrare due implicazioni.

$X_n \xrightarrow{D} X \iff \forall n \in \mathbb{N}, p_n(x) = \mathbb{P}(X_n = x) \rightarrow p_X(x)$
Prendiamo $y \notin \mathbb{N}$ allora per definizione di funzione di ripartizione per funzioni discrete ho che

$$F_{X_n}(y) = \sum_{x=0}^{\lfloor y \rfloor} p_n(x) \leq 1$$

Dunque è ovvio osservare che la serie converge quindi

$$\lim_{n \rightarrow +\infty} F_{X_n}(y) = \lim_{n \rightarrow +\infty} \sum_{x=0}^{\lfloor y \rfloor} p_n(x) = \sum_{x=0}^{\lfloor y \rfloor} \lim_{n \rightarrow +\infty} p_n(x)$$

Adesso sfruttando l'ipotesi ho che $\lim_{n \rightarrow +\infty} p_{X_n}(x) = p_X(x)$ dunque

$$\lim_{n \rightarrow +\infty} F_{X_n}(y) = \lim_{n \rightarrow +\infty} F_X(y)$$

$X_n \xrightarrow{D} X \Rightarrow \forall n \in \mathbb{N}, p_n(x) = \mathbb{P}(X_n = x) \rightarrow p_X(x)$
Per le proprietà della funzione di ripartizione ho che $\forall n \in \mathbb{N}$

$$p_n(x) = F_{X_n}(x + \epsilon) - F_{X_n}(x - \epsilon)$$

Poiché per ipotesi ho che $X_n \xrightarrow{D} X$ allora ho che

$$\lim_{n \rightarrow +\infty} p_n(x) = \lim_{n \rightarrow +\infty} F_{X_n}(x + \epsilon) - F_{X_n}(x - \epsilon) = F_X(x + \epsilon) - F_X(x - \epsilon) = p_X(x)$$

Quindi la tesi è dimostrata. □

Teorema 5.1.6. Sia $\{X_n\}$ una successione di v.a. con $X_n \sim B(n, p_n)$ dunque supposto che $p_n \rightarrow 0$ e che $np_n \rightarrow \lambda$ allora

$$X_n \xrightarrow{D} X \sim Po(\lambda)$$

Dimostrazione. Per definizione ho che

$$p_n(k) = \mathbb{P}(X_n = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{n!}{k!(n-k)!} p_n^k (1 - p_n)^{n-k}$$

Dato che

$$\frac{n!}{k!(n-k)!} = \frac{n(n-1) \cdots (n-k+1)}{k! n^k} n^k$$

E che

$$(1 - p_n)^{n-k} = \left(1 - \frac{np_n}{n}\right)^{n-k}$$

Dunque

$$\frac{n!}{k!(n-k)!} p_n^k (1 - p_n)^{n-k} = \frac{n(n-1) \cdots (n-k+1)}{k! n^k} (np_n)^k \left(1 - \frac{np_n}{n}\right)^{n-k}$$

Passando al limite e ricordando che $np_n \rightarrow \lambda$ allora ottengo che

$$\lim_{n \rightarrow +\infty} \frac{n(n-1) \cdots (n-k+1)}{k! n^k} (np_n)^k \left(1 - \frac{np_n}{n}\right)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!} = p_X(n)$$

Dunque per il teorema precedente ho che $X_n \xrightarrow{D} X$ allora per definizione ho che $X \sim Po(\lambda)$ □

5.2 Teorema centrale del limite

Dato uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ e sia $\{X_n\}_{n \in \mathbb{N}}$ l'esito del lancio di una moneta X_n iid $B(p)$, noi sappiamo che

$$S_n(X_1 + \cdots + X_n) \sim B(n, p)$$

Il numero di successi in n lanci e $\bar{X}_n = \frac{S_n}{n}$ media campionaria che è anche la percentuale di successi in n lanci, quindi per la legge dei grandi numeri mi dice che converge in probabilità alla media quindi

$$\bar{X} \xrightarrow{\mathbb{P}} \mathbb{E}[X_n]$$

Sappiamo che

$$V(\bar{X}_n) = \frac{1}{n^2} V(S_n) = \frac{1}{n^2} n V(X_n) = \frac{V(X_n)}{n} \xrightarrow{n \rightarrow +\infty} 0$$

Osservazione 5.2.1. Standardizzare S_n cioè

$$S_n^* = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{V(S_n)}} = \frac{S_n - np}{\sqrt{np(1-p)}}$$

Significa renderla una gaussiana standard, inoltre si nota che $\mathbb{E}[S_n^*] = 0$ e $V(S_n^*) = 1, \forall n \geq 1$ e

$$-\sqrt{\frac{np}{1-p}} \leq S_n^* \leq \sqrt{\frac{n(1-p)}{p}}$$

Esempio 5.2.1. Sia $\{X_n\}$ iid con $N(\mu, \sigma^2)$ dunque $S_n = (X_1 + \cdots + X_n) \sim N(n\mu, n\sigma^2)$ dunque

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad S_n^* = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$$

Dunque $S_n^* \xrightarrow{D} Z \sim N(0, 1)$ in quanto tutte lo sono.

Inoltre

$$M_Z(t) = e^{\frac{t^2}{2}} \quad M_{X_n}(t) = e^{\mu t + \frac{1}{2} \sigma^2 t^2}$$

Teorema 5.2.1. Se due v.a. X, Y hanno la stessa funzione generatrice dei momenti in un intervallo completo dell'origine allora $X \sim Y$.

Teorema 5.2.2 (Teorema centrale del limite). Sia $\{X_i\}_{i \in \mathbb{N}}$ una successione di v.a. iid che ammettono momento secondo finito e sia $\forall i$

$$\begin{aligned}\mu &= \mathbb{E}[X_i] \\ \sigma^2 &= \text{Var}(X_i) \neq 0\end{aligned}$$

Allora posto

$$S_n^* = \frac{S_n - \mathbb{E}[S]_n}{\sqrt{V(S_n)}}$$

E $Z \sim N(0,1)$ allora si ha che $\{S_n^*\} \xrightarrow{D} Z$ dunque

$$\lim_{n \rightarrow +\infty} \mathbb{P}(S_n^* \leq x) = \mathbb{P}(Z \leq x)$$

Dimostrazione. Basterà dimostrare che $M_{S_n^*} \rightarrow M_Z(t)$ in $(-b, b)$ possiamo supporre che

$$\mathbb{E}[X_n] = 0 \quad V(X_n) = \mathbb{E}[X_n^2] = 1$$

Questo lo possiamo supporre in quanto

$$S_n - n\mu = (X_1 - \mu) + \dots + (X_n - \mu)$$

Dunque standardizzando questa variabile aleatoria, ricordando che $\text{Var}(S_n) = n \text{Var}(X_i)$ essendo v.a. iid, allora ottengo che

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{(X_1 - \mu)}{\sqrt{n\sigma^2}} + \dots + \frac{(X_n - \mu)}{\sqrt{n\sigma^2}}$$

Per Taylor ho che

$$M_X(t) = 1 + tM'_X(0) + \frac{1}{2}t^2M''_X(0) + o(t^2) = 1 + \frac{1}{2}t^2 + o(t^2)$$

Dunque dato che $\sigma^2 = 1$ sfruttando le proprietà della funzione generatrice dei momenti ho che

$$M_{S_n^*} = M_{\frac{S_n}{\sqrt{n}}}(t) = M_{S_n}\left(\frac{t}{\sqrt{n}}\right) = \left(M_X\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + o\left(\left(\frac{t}{\sqrt{n}}\right)^2\right)\right)^n$$

Dunque

$$\log(M_{S_n^*}(t)) = n \log\left(1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)$$

Allora sviluppando mediante Taylor la funzione $\log(1+x)$ ho che il tutto è equivalente

$$n \left(\frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right)$$

Quindi

$$\lim_{n \rightarrow +\infty} \log(M_{S_n^*}(t)) = \frac{t^2}{2} \implies \lim_{n \rightarrow +\infty} M_{S_n^*} = e^{\frac{t^2}{2}} = M_Z(t)$$

□

Osservazione 5.2.2. Per poter calcolare il valore della gaussiana bisogna usare le tavole. non ci sono numeri negativi in quanto per simmetria

$$\Phi(-x) = \int_{-\infty}^{-x} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt = \int_x^{\infty} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt = 1 - \Phi(x)$$

5.2.1 Correzione di continuità

La correzione di continuità consiste nell'ampliare di $\frac{1}{2}$ gli estremi dell'intervallo sulla quale si integra la densità di probabilità usata per approssimare una distribuzione discreta.

Osservazione 5.2.3. Correzione di continuità

Se considero una successione di v.a. discrete $\{X_i\}_{i=1}^n$ iid, dal quale voglio sfruttare il TLC allora preso $L \in \mathbb{N}$ posso affermare che

$$\begin{aligned} \mathbb{P}(S_N \leq L) &= \mathbb{P}\left(S_N \leq L + \frac{1}{2}\right) = \mathbb{P}\left(S_N \leq L + \frac{9}{10}\right) \\ &\parallel \mathbb{P}\left(\frac{S_n - \mu}{\sqrt{\sigma^2}} \leq \frac{L - \mu}{\sqrt{\sigma^2}}\right) \parallel \mathbb{P}\left(\frac{S_n - \mu}{\sqrt{\sigma^2}} \leq \frac{L + \epsilon - \mu}{\sqrt{\sigma^2}}\right) \\ &\approx \Phi\left(\frac{L - \mu}{\sigma}\right) \approx \Phi\left(\frac{L + \epsilon - \mu}{\sigma}\right) \end{aligned}$$

In quanto essendo S_n una v.a. discreta essa potrà assumere solo valori in \mathbb{N} .

Osservazione 5.2.4. Supponiamo di avere una successione di v.a. discrete $\{X_i\}_{i=1}^n$ iid $X \sim Be(p)$ dunque definisco

$$S_n := \sum_{i=1}^n X_i \sim Bin(n, p)$$

Perciò per un corollario del TLC posso affermare che

$$S_n \sim_D N(np, np(1-p))$$

Dunque

$$\begin{aligned} \mathbb{P}(S_n \leq s) &= \mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{s - np}{\sqrt{np(1-p)}}\right) \\ &= \mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{s + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) \\ &\approx \mathbb{P}\left(Z \leq \frac{s + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) \\ &= \Phi\left(\frac{s + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

A questo punto ponendo $p = 0.5$ ho che $S_n \sim N(n0.5, n0.25)$ quindi

$$\mathbb{P}(S_N \leq \mu - k) \quad \mu = \mathbb{E}[S_N] = \frac{N}{2}$$

Per quanto detto prima, applicando la correzione di continuità, il tutto è equivalente a

$$\begin{aligned} \mathbb{P}(S_N \leq \mu - k) &\approx \mathbb{P}\left(Z \leq \frac{(Np - k) + \frac{1}{2} - Np}{\sqrt{Np(1-p)}}\right) \\ &= \mathbb{P}\left(Z \leq \frac{-k + \frac{1}{2}}{\sqrt{0.25N}}\right) \\ &= \mathbb{P}\left(Z \leq \frac{1 - 2k}{\sqrt{N}}\right) \end{aligned} \tag{5.1}$$

Osservazione 5.2.5. La correzione di continuità si fa solo la dove il campione statistico presenta un numero di osservazioni n alquanto basso, solitamente

$$n < 30$$

Inoltre possiamo anche evitare la correzione di continuità quando la varianza della v.a. continua che si usa per approssimare il campione discreto è molto alta, solitamente

$$\sigma > 100$$

In quanto avrei

$$\frac{1}{2\sigma} < 0,005$$

E questo non mi porterebbe contributo nella lettura delle tavole di distribuzione.

Esempio 5.2.2. Problema è se lancio una moneta equilibrata $N = 100$ volte quale è la probabilità di ottenere non più di $L = 45$ teste? Se $N = 1000$ e $L = 450$?...

Riprendendo l'esempio iniziale ho che se $N = 10$ ho che $X_N \sim B(N, 0.5)$ dunque

$$P(X \leq 4) = \sum_{k=0}^4 \binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{2^{10}} \left(\sum_{k=0}^4 \binom{n}{k} \right) = \frac{386}{1024} \approx 37,7\%$$

Mentre per quanto riguarda gli altri casi il ho che

- Se $N = 100$ allora $\text{Var}(S_N) = 100$ e quindi essendo $k = 50 - 45 = 5$ ho che

$$\mathbb{P}(S_{100} \leq 45) = \Phi\left(-\frac{2k-1}{\sqrt{100}}\right) = \Phi\left(\frac{-9}{10}\right) = 0.1841$$

- Se $N = 1000$ allora $\text{Var}(S_N) = 1000$ e quindi essendo $k = 500 - 450 = 50$ ho che

$$\mathbb{P}(S_{1000} \leq 450) = \Phi\left(-\frac{2k-1}{\sqrt{1000}}\right) = \Phi\left(\frac{-99}{\sqrt{1000}}\right) = 0.0009$$

Osservazione 5.2.6. Supponiamo di avere una successione $\{X_n\}$ v.a. iid in $L^2(\Omega, \mathbb{P})$ con $\mu = \mathbb{E}[X]$, $\sigma^2 = \text{Var}(X)$ e $X_N = \sum_{k=1}^N X_k \in \mathbb{N}$, dunque la domanda è $\mathbb{P}(S_N = L)$

$$(S_N = L) = \left(L - \frac{1}{2} < S_N \leq L + \frac{1}{2}\right) \Rightarrow \mathbb{P}(S_N = L) = \mathbb{P}\left(S_N \leq L + \frac{1}{2}\right) - \mathbb{P}\left(S_N \leq L - \frac{1}{2}\right)$$

Standardizzando ottengo che

$$\mathbb{P}\left(S_N^* \leq \frac{L + \frac{1}{2} - N\mu}{\sqrt{N}\sigma^2}\right) - \mathbb{P}\left(S_N^* \leq \frac{L - \frac{1}{2} - N\mu}{\sqrt{N}\sigma^2}\right) \approx \Phi(L_+^*) - \Phi(L_-^*) = \int_{L_-^*}^{L_+^*} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt$$

Ma dato che

$$L_+^* - L_-^* = \frac{1}{\sigma\sqrt{N}}$$

Dunque per N grande questo intervallo è piccolo, quindi posso usare un'approssimazione dell'integrale ottenendo quindi

$$(L_+^* - L_-^*) \varphi(L^*) = \frac{1}{\sigma\sqrt{N}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(L - N\mu)^2}{2N\sigma^2}\right)$$

Perciò l'approssimazione è

$$\mathbb{P}(S_N = L) \approx \frac{1}{\sigma\sqrt{2N\pi}} \exp\left(-\frac{(L - N\mu)^2}{2N\sigma^2}\right)$$

5.3 Passeggiata casuale

Date X_1, X_2, X_3, \dots variabili aleatorie indipendenti e equidistribuite dunque della forma

$$X_n = \begin{cases} 1 & p \\ -1 & q \end{cases} \quad p + q = 1$$

Che rappresentano l'esito del gioco n-esimo .

Definizione 5.3.1. Si definisce passeggiata casuale

$$\begin{aligned} S_0 &= a \\ S_1 &= S_0 + X_1 = \begin{cases} a+1 & p \\ a-1 & q \end{cases} \\ &\vdots \\ S_{n+1} &= S_n + X_n \end{aligned}$$

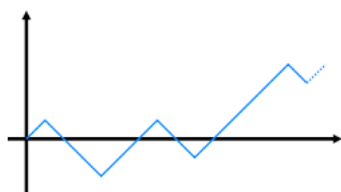
Osservazione 5.3.1. Se n è pari allora S_n ha la stessa parità di a altrimenti la posizione è impossibile.

5.3.1 Passeggiata casuale simmetrica semplice

Preso in riferimento una passeggiata casuale con $p = q = \frac{1}{2}$, $S_0 = 0$ e consideriamo

$$\Omega_n = \{(\epsilon_1, \epsilon_2, \dots, \epsilon_n) | \epsilon_k \in \{-1, 1\}\}$$

Dunque $\#\Omega_n = 2^n$ perciò $\mathbb{P}(\omega) = 2^{-n}$ mentre la rappresentazione grafica:



$$\mathbb{P}((0,0) \rightarrow (n,x)) = \mathbb{P}(S_n = x | S_0 = 0)$$

Da tutto ciò si può osservare che

- Se n e x non hanno la stessa parità allora $\mathbb{P}(S_n = x | S_0 = 0) = 0$
- Se $|x| > n$ allora $\mathbb{P}(S_n = x | S_0 = 0) = 0$
- Se voglio arrivare ad un punto x ho che

$$((0,0) \rightarrow (n,x)) = \underbrace{1, 1, \dots, 1}_{x\text{-volte}}, \underbrace{1, -1, 1, -1, \dots, 1}_{(n-x)\text{-volte}}$$

Osservazione 5.3.2. Da quanto osservato fino ad ora si può osservare che per passare da $(0,0)$ a (n,x) sono necessari $\frac{n+x}{2}$ "1" e $\frac{n-x}{2}$ "-1" perciò posso concludere che

$$\mathbb{P}(S_n = x | S_0 = 0) = \binom{n}{\frac{n+x}{2}} 2^{-n}$$

Inoltre è possibile affermare in maniera più generale che il numero di cammini possono essere calcolati come

$$N((0,a) \rightarrow (n,b)) = \binom{n}{\frac{n+(b-a)}{2}}$$

5.3.2 Principio di riflessione

Supponiamo di avere che $S_0 = a$ e $S_n = b$ dove la passeggiata $(0, a) \rightarrow (n, b)$ è una passeggiata possibile.

Definizione 5.3.2. Data una passeggiata aleatoria allora e un punto $A = (k, \alpha)$ definisco il suo riflesso come $A' = (k, -\alpha)$

Definizione 5.3.3. Definisco cammini positivi quei cammini $(0, a) \rightarrow (n, b)$ tali per cui $S_0 > 0, S_1 > 0, \dots, S_n > 0$ e lo indico come

$$N^+((0, a) \rightarrow (n, b))$$

Lemma 5.3.1. (Principio di riflessione)

Vale che

$$N^+((0, a) \rightarrow (n, b)) = N((0, a) \rightarrow (n, b)) - N((0, -a) \rightarrow (n, b))$$

Dimostrazione. Il numero di cammini che vanno $(0, a) \rightarrow (n, b)$ e che toccano/attraversano l'asse orizzontale sono esattamente tanti quanti quelli che $(0, -a) \rightarrow (n, b)$. Dunque tutti i cammini che toccano 0 in un tempo $k < n$ possono essere riflessi nella prima parte (fino al tempo k) costruendo un cammino che va da $(0, -a) \rightarrow (n, b)$ \square

Teorema 5.3.1. Sia $(n, b), b > 0$ un punto di arrivo ammissibile per una passeggiata casuale che parte da $(0, 0)$.

$$N^+((0, 0) \rightarrow (n, b)) = \frac{b}{n} N((0, 0) \rightarrow (n, b))$$

Dimostrazione. Prima di tutto si vede che

$$N^+((0, 0) \rightarrow (n, b)) = N^+((1, 1) \rightarrow (n, b)) = N((1, 1) \rightarrow (n, b)) - N((1, -1) \rightarrow (n, b))$$

Dunque per quanto visto in precedenza ho che il tutto è equivalente a

$$\binom{n-1}{\frac{n-1+(b-1)}{2}} - \binom{n-1}{\frac{n-1+(b+1)}{2}} = \binom{n-1}{\frac{n-b}{2} + 1} - \binom{n-1}{\frac{n+b}{2}} = \frac{b}{n} \binom{n}{\frac{n+b}{2}}$$

\square

Ritornando al problema originario ho che

$$\mathbb{P}(A \text{ fosse sempre in vantaggio durante lo spoglio}) = \frac{p-m}{p+m}$$

5.3.3 Ritorni all'origine

Per tornare all'origine ho che

$$u_{2n} = \mathbb{P}(S_{2n} = 0 | S_0 = 0) = \binom{2n}{n} 2^{-2n}$$

Osservazione 5.3.3. Per le formule di Stirling ho che

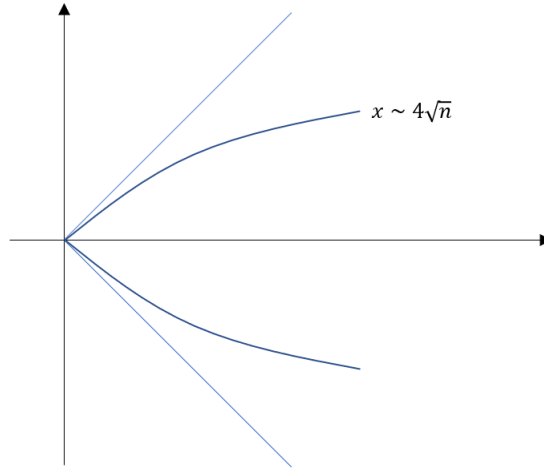
$$u_{2n} \sim \frac{1}{\sqrt{\pi n}}$$

Inoltre per il teorema centrale del limite ho che

$$2 \frac{S_n}{\sqrt{n}} \sim N(0, 1)$$

Quindi

$$\mathbb{P}(|S_n| > x\sqrt{n}) \approx 1 - \Phi(x)$$



Se al tempo $2n$ ho circa $c\sqrt{n}$ possibili valori allora mi aspetto che

$$u_{2n} \sim \frac{1}{c\sqrt{n}}$$

Per capire quando tornerò a toccare l'asse ho che

$$f_{2n} = \mathbb{P}(S_1 \neq 0, \dots, S_{2n-1} \neq 0, S_{2n} = 0)$$

Ma in tutto questo posso togliere le condizioni ai tempi dispari in quanto posso solo toccarlo per tempi pari.

$$u_{2n} = \mathbb{P}(S_{2n} = 0 | S_0 = 0) = \sum_{k=1}^n \mathbb{P}(S_1 \neq 0, \dots, S_{2k} = 0)$$

Questa è la probabilità di tornare a 0 per la prima volta al tempo $2k$

$$\mathbb{P}(S_{2k} = 0 \rightarrow S_{2n} = 0) = \mathbb{P}(S_0 = 0 \rightarrow S_{2n-2k} = 0)$$

Lemma 5.3.2. Vale che

$$\mathbb{P}(S_1 \neq 0, \dots, S_{2n} \neq 0) = \mathbb{P}(S_{2n} = 0) = u_{2n}$$

Dimostrazione.

$$\begin{aligned} \mathbb{P}(S_1 > 0, \dots, S_{2n} > 0) &= \sum_{r=1}^n \mathbb{P}(S_1 > 0, \dots, S_{2n} = 2r) = \sum_{r=1}^n N^+((0, 0) \rightarrow (2n, 2r)) 2^{-2r} \\ &= \frac{1}{2} \sum_{r=1}^n [N((1, 1) \rightarrow (2n, 2r)) - N((1, -1) \rightarrow (2n, 2r))] 2^{-(2n-1)} \\ &= \frac{1}{2} \sum_{r=1}^n \mathbb{P}(S_{2n} = 2r | S_1 = 1) - \mathbb{P}(S_{2n} = 2r + 2 | S_1 = 1) \\ &= \frac{1}{2} \mathbb{P}(S_{2n} = 2 | S_1 = 1) \\ &= \frac{1}{2} \mathbb{P}(S_{2n} = 1 | S_0 = 0) \end{aligned}$$

□

Esempio 5.3.1. (Votazione)

Supponiamo che in una elezione il candidato A abbia ricevuto p voti e il candidato B abbia ricevuto m voti con $p > m$. Quale è la probabilità che durante tutto lo spoglio A sia sempre in vantaggio rispetto a B?

Per quanto visto fino ad ora possiamo dire che tale probabilità equivale a $\frac{p-m}{p+m}$

$$\begin{aligned}
f_{2n} &= \mathbb{P}(S_2 \neq 0, \dots, S_{2n-2} \neq 0, S_{2n} = 0) \\
&= \mathbb{P}(S_2 \neq 0, \dots, S_{2n-2} \neq 0) - \mathbb{P}(S_2 \neq 0, \dots, S_{2n-2} \neq 0, S_{2n} \neq 0) \\
&= u_{2n-2} - u_{2n}
\end{aligned}$$

Lemma 5.3.3. *Quasi certamente la passeggiata casuale torna all'origine.*

Dimostrazione.

$$\mathbb{P}(\text{ritorno all'origine}) = \sum_k^\infty f_{2k} = \sum_k^\infty u_{2k-2} - u_{2k} = u_0 = 1$$

□

Mentre la probabilità di non essere mai ritornati a 0 in $2n$ lanci è

$$u_{2n} \sim \frac{1}{\sqrt{\pi n}}$$

Dunque la probabilità di non tornare a 0 in 100 lanci è dell'ordine di 8%

La passeggiata casuale con $S_0 = x$ e $S_{n+1} = S_n + X_{n+1}$. Diremo che

- Simmetrica se $\mathbb{E}[X_n] = 0$
- Semplice se $X_n \in \{-1, 0, 1\}$

$$r = \mathbb{P}(X_n = 0) < 1$$

Con $\mathbb{P}(X_n = 1) = p, \mathbb{P}(X_n = -1) = q$ e $p + q = 1$

$$\mathbb{E}[X_n] = p - q$$

E

$$\begin{aligned}
\mathbb{E}[S_n] &= \mathbb{E}[x + X_1 + \dots + X_n] = x + n(p - q) \\
V(X_n) &= (p + q) - (p - q)^2
\end{aligned}$$

Definizione 5.3.4. *Chiamiamo T il tempo di prima uscita. La distribuzione di T dipende da x .*

$$T = \inf\{n \geq 0 | S_n \leq a \vee S_n \geq b\}$$

Con $\inf \phi = +\infty$.

Proposizione 5.3.1. *Identità di Wald*

$$\mathbb{E}[S_T] = x + (p - q)\mathbb{E}[T]$$

Dimostrazione. Possiamo scrivere

$$S_T = x + \sum_{k=1}^{+\infty} X_k 1_{\{k \leq T\}}$$

Si osserva che l'evento

$$(T \geq k) = (T < k)^c$$

Dipende solo da X_1, \dots, X_{k-1} quindi è indipendente X_k quindi posso dire che

$$\mathbb{E}[S_T] = x + \sum_{k=1}^{+\infty} \mathbb{E}[X_k] \mathbb{E}[1_{\{k \leq T\}}] = x + \sum_{k=1}^{+\infty} \mathbb{E}[X_k] \mathbb{P}(T \geq k)$$

Se Y è una variabile aleatoria a valori interi positivi allora per definizione ho che

$$\mathbb{E}[Y] = \sum_{k=1}^{+\infty} k p_k = \sum_{k=1}^{\infty} \left(\sum_{j=1}^k 1 \right) p_k = \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} p_k = \sum_{j=1}^{+\infty} \mathbb{P}(Y \geq j)$$

Perciò posso dire che

$$\mathbb{E}[S_T] = x + (p - q)\mathbb{E}[T]$$

□

Se $p = q$ allora $\mathbb{E}[X_n] = 0$ e $V(X_n) = 1 - 2r = \sigma^2$ con $r = \mathbb{P}(X_n = 0)$.

Proposizione 5.3.2. Seconda identità di Wald

$$\text{Var}(S_T) = \text{Var}(X_n)\mathbb{E}[T]$$

Dimostrazione. Per quanto visto prima sappiamo che

$$S_T - x = \sum_{j=1}^{+\infty} X_j 1_{\{j \leq T\}}$$

Dunque è ovvio che

$$(S_T - x)^2 = \left(\sum_{j=1}^{+\infty} X_j 1_{\{j \leq T\}} \right) \left(\sum_{k=1}^{+\infty} X_k 1_{\{k \leq T\}} \right) = \sum_{j,k=1}^{+\infty} X_j X_k 1_{\{j \leq T\}} 1_{\{k \leq T\}}$$

Dunque per definizione posso dire che

$$\begin{aligned} \mathbb{E}[(S_T - x)^2] &= \sum_{j,k=1}^{+\infty} \mathbb{E}[X_j X_k 1_{\{j \leq T\}} 1_{\{k \leq T\}}] \\ &= \sum_{j=1}^{+\infty} \mathbb{E}[X_j^2 1_{\{j \leq T\}}] + 2 \sum_{j < k}^{+\infty} \mathbb{E}[X_j X_k 1_{\{j \leq T\}} 1_{\{k \leq T\}}] \end{aligned}$$

Dato che $(j \leq T) = (j > T)^c$ dipende da X_1, \dots, X_{j-1} è indipendente da X_k . X_j è indipendente da X_k , dunque tutto è equivalente a

$$\sum_{j=1}^{+\infty} \mathbb{E}[X_j^2] \mathbb{E}[1_{\{j \leq T\}}] + 2 \sum_{j < k} \mathbb{E}[X_k] \mathbb{E}[X_j 1_{\{k \leq T\}}] = \sigma^2 \sum_{j=1}^{+\infty} \mathbb{P}(T \geq j) = \sigma^2 \mathbb{E}[T]$$

□

Dato che tutto dipende dal dato di partenza x posso dire che

$$\mathbb{P}^x(S_t = a) + \mathbb{P}^x(S_t = b) = 1$$

Inoltre

$$\mathbb{E}[S_t] = a\mathbb{P}^x(S_t = a) + b\mathbb{P}^x(S_t = b)$$

Allora abbiamo due formule che dicono quanto vale la media si S_T dunque

$$x + \mu \mathbb{E}^x[T] = a\mathbb{P}^x(S_t = a) + b\mathbb{P}^x(S_t = b) = a\mathbb{P}^x(S_t = a) + b(1 - \mathbb{P}^x(S_t = a))$$

Dunque ho che

$$-b + x + \mu \mathbb{E}^x[T] = (a - b)\mathbb{P}^x(S_t = a)$$

Caso simmetrico

Supponiamo che $p = q$ allora

$$\mathbb{P}^x(S_T = a) = \frac{b - x}{b - a}$$

$$\mathbb{P}^x(S_T = b) = \frac{x-a}{b-a}$$

Applicando la seconda identità di Wald ho che

$$\begin{aligned} Var(S_T) &= \mathbb{E}[(S_t - x)^2] = (b-x)^2 \mathbb{P}^x(S_t = b) + (x-a)^2 \mathbb{P}^x(S_t = a) \\ &= \sigma^2 \mathbb{E}[T] = (b-x)^2 \frac{x-a}{b-a} + (x-a)^2 \frac{b-x}{b-a} = (x-a)(b-x) \end{aligned}$$

Dunque otteniamo che

$$\mathbb{E}^x[T] = \frac{(x-a)(x-b)}{1-r}$$

Allora

$$\mathbb{E}^x[T] = x$$

Caso non simmetrico

Supponiamo che $p \neq q$, in questo caso non posso utilizzare le identità di Wald dunque poniamo

$$\begin{aligned} f(x) &= \mathbb{P}^x(S_T = b) = \mathbb{P}^x(S_T = b, S_1 = x+1) + \mathbb{P}^x(S_T = b, S_1 = x) + \mathbb{P}^x(S_T = b, S_1 = x-1) \\ &= \mathbb{P}^x(S_1 = x+1) \mathbb{P}_{S_1=x+1}(S_T = b) + \mathbb{P}^x(S_1 = x) \mathbb{P}_{S_1=x}(S_T = b) + \mathbb{P}^x(S_1 = x-1) \mathbb{P}_{S_1=x-1}(S_T = b) \\ &= pf(x+1) + rf(x) + qf(x-1) \end{aligned}$$

Ricordando che $p+q=1-r$ ho che

$$\begin{aligned} (p+q)f(x) &= pf(x+1) + qf(x-1) \\ \begin{cases} p(f(x+1) - f(x)) &= q(f(x) - f(x-1)) \\ f(a) &= 0 \\ f(b) &= 1 \end{cases} \end{aligned}$$

Questo problema si può risolvere iterando il tutto e vedendo che

$$f(x+1) - f(x) = \frac{q}{p}(f(x) - f(x-1)) = \left(\frac{q}{p}\right)^2 (f(x-1) - f(x-2)) = \dots = \left(\frac{q}{p}\right)^{x-a} (f(a+1) - f(a)) = \left(\frac{q}{p}\right)^{x-a} f(a+1)$$

Dunque abbiamo che

$$f(x+1) = f(x) + \left(\frac{q}{p}\right)^{x-a} f(a+1)$$

Perciò ho che

$$\sum_{x=1}^{b-1} (f(x+1) - f(x)) = f(b) - f(a) = 1 = \left[\sum_{x=a}^{b-1} \left(\frac{q}{p}\right)^{x-a} \right] f(a+1)$$

Allora posto $c = \frac{q}{p}$ ho che

$$1 = f(a+1) \frac{1 - c^{b-a}}{1 - c} \implies f(a+1) = \frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^{b-a}}$$

Infine

$$f(x) = f(x) - f(a) = \sum_{y=a}^{x-1} (f(y+1) - f(y)) = \frac{1 - \left(\frac{q}{p}\right)^{x-a}}{1 - \left(\frac{q}{p}\right)^{b-a}}$$

Dunque

$$\begin{aligned} \mathbb{E}^x[S_T] &= (b-a) \frac{1 - \left(\frac{q}{p}\right)^{x-a}}{1 - \left(\frac{q}{p}\right)^{b-a}} + a \\ \mathbb{E}^x[T] &= \left(\frac{b-a}{p-q}\right) \frac{1 - \left(\frac{q}{p}\right)^{x-a}}{1 - \left(\frac{q}{p}\right)^{b-a}} - \frac{x-a}{p-q} \end{aligned}$$

Esercizio. Dato $p = \frac{2}{3}$ e $q = \frac{1}{3}$ dato $x = 1, a = 0, b \nearrow +\infty$

$$f(y) = \frac{1 - \left(\frac{q}{p}\right)^y}{1 - \left(\frac{q}{p}\right)^b} = \frac{1 - \frac{1}{2^y}}{1 - \frac{1}{2^b}}$$

Per quanto visto fino ad ora

$$f(y) = \mathbb{P}^x(S_n > 0, \forall n) = 1 - \frac{1}{2^y}$$

Dunque

$$\mathbb{P}^{n+1}(S_n > 0) = \frac{1}{2}$$

Per quanto riguarda il caso simmetrico allora è certo che l'ubriaco passerà da 0 in quanto

$$\mathbb{P}^x(S_T = a = 0) = \frac{b-x}{b} \xrightarrow{b \rightarrow +\infty} 1$$

Quindi

$$\mathbb{E}^x[T] = +\infty$$

Esercizio. Dati due giocatori con due fondi diversi che giocano facendo puntate unitarie fin tanto che uno non ha esaurito i fondi.

Supponiamo che i capitali iniziali siano m, n con $p = \frac{2}{3}, q = \frac{1}{3}$.

In questo caso le due frontiere sono 0: il giocatore è in bancarotta, e $m+n$: il giocatore avversario è in bancarotta.

$$\mathbb{P}^m(S_T = m+n) = \frac{1 - \left(\frac{q}{p}\right)^m}{1 - \left(\frac{q}{p}\right)^{m+n}} = \frac{1 - \left(\frac{1}{2}\right)^m}{1 - \left(\frac{1}{2}\right)^{m+n}}$$

Supposto $m = 1, n = 2$ allora

$$\mathbb{P}^m(S_T = m+n) = \frac{1 - \left(\frac{1}{2}\right)^1}{1 - \left(\frac{1}{2}\right)^3} = \frac{4}{7}$$

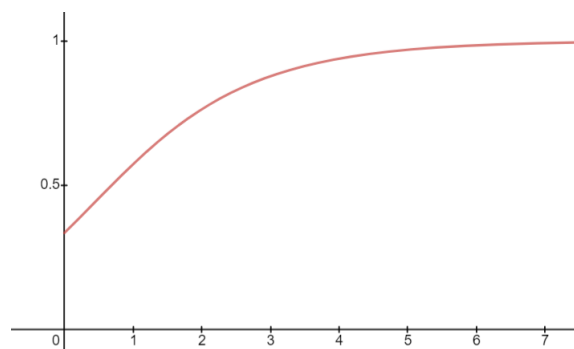
E

$$\mathbb{E}^1[T] = \frac{3}{1} \cdot \frac{1 - \left(\frac{1}{2}\right)^1}{1 - \left(\frac{1}{2}\right)^3} - \frac{1}{\frac{1}{3}} = \frac{15}{7} > 2$$

Da questo esempio si può osservare che posto $n = 2m$ ho che

$$f(m) = \frac{1 - \left(\frac{1}{2}\right)^m}{1 - \left(\frac{1}{2}\right)^{3m}}$$

È una funzione del tipo



Capitolo 6

Entropia

6.1 Entropia

Definizione 6.1.1 (Informazione).

Sia X una v.a. discreta definita su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e sia $S_X := \{x_1, \dots, x_n\}$ il supporto della v.a.

Dunque posto

$$E_i := X^{-1}(x_i) = \{\omega \in \Omega | X(\omega) = x_i\}$$

Allora definisco informazione associata ad un evento E_i

$$I(E_i) = -\log_2 \mathbb{P}(E_i)$$

Definizione 6.1.2 (Entropia).

Sia X una v.a. discreta definita su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e sia $S_X := \{x_1, \dots, x_n\}$ il supporto della v.a.

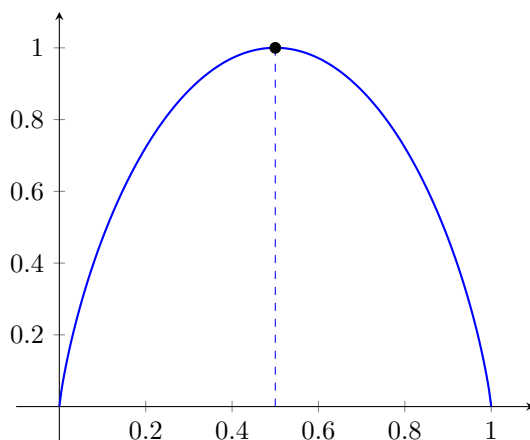
Allora definisco entropia di X

$$H(X) = -\sum_{i=1}^n \mathbb{P}(X = x_i) \log_2 (\mathbb{P}(X = x_i))$$

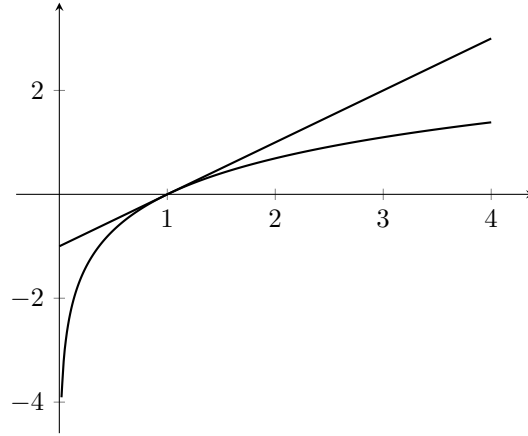
Esempio 6.1.1. Sia X una v.a. tale per cui $X \sim Be(p)$ dunque per definizione ho che $S_X = \{0, 1\}$ e quindi per definizione di entropia ho che

$$H(X) = -p \log_2(p) - (1-p) \log_2(1-p)$$

Il cui massimo si trova quando $p = \frac{1}{2}$.



Osservazione 6.1.1. La funzione $\ln(x) \leq x - 1$, $\forall x > 0$ in quanto



Teorema 6.1.1. Sia X una v.a. discreta definita su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e sia $S_X := \{x_1, \dots, x_n\}$ il supporto della v.a. (dunque $\#S_X = n$) allora valgono i seguenti risultati:

1. $H(X) \geq 0$.
2. $H(X) = 0 \iff \exists i | \mathbb{P}(X = x_i) = 1$.
3. Presenta un limite superiore

$$H(X) \leq \log_2 n$$

4. Raggiungimento limite superiore

$$H(X) = \log_2 n \iff X \sim \text{Unif}\{x_1, \dots, x_n\}$$

Dimostrazione.

Punto 1

Per definizione

$$H(X) = \sum p_k \log_2 \left(\frac{1}{p_k} \right)$$

Ma dato che

$$1 \geq p_k \geq 0$$

Allora vale che

$$1 \leq \frac{1}{p_k} \implies 0 \leq \log_2 \left(\frac{1}{p_k} \right)$$

Quindi

$$H(X) = \sum p_k \log_2 \left(\frac{1}{p_k} \right) \geq 0$$

Punto 2

Se assumo che $\exists i | \mathbb{P}(X = x_i) = 1$ allora è ovvio che $\forall j \neq i | j \in S_X$ vale che

$$\mathbb{P}(X = j) = 0$$

Quindi per definizione ho che

$$H(X) = -\mathbb{P}(X = i) \log_2(\mathbb{P}(X = i)) - \sum_{j \in S_X, j \neq i} \mathbb{P}(X = j) \log_2(\mathbb{P}(X = j)) = 0$$

Punto 3

Per definizione ho che

$$\begin{aligned}
H(X) - \log_2(n) &= \frac{1}{\ln 2} \sum p_k \log_2 \left(\frac{1}{p_k} \right) - \frac{1}{\ln 2} \sum p_k \ln n = \frac{1}{\ln 2} \sum p_k \ln \left(\frac{1}{np_k} \right) \\
&\leq \frac{1}{\ln 2} \sum p_k \log_2 \left(\frac{1}{np_k} - 1 \right) \\
&\leq \frac{1}{\ln 2} \sum_{k=1}^n \left(\frac{1}{n} - p_k \right) = 0
\end{aligned}$$

Inoltre vale l'uguaglianza $\iff \frac{1}{np_k} = 1 \quad \forall k \iff p_k = \frac{1}{n}$. \square

Osservazione 6.1.2. *Da quanto visto nel teorema precedente il fatto che l'entropia sia massima nel caso di una distribuzione di Bernulli solo se $p = 0.5$ rispecchia il fatto che in quel caso la distribuzione è uniforme.*

6.1.1 Dado di Jaynes

Dato un dado a 6 facce di cui è noto che il valore medio uscito è $B = 4.5$ quindi il nostro vincolo ci impedisce di scegliere la distribuzione uniforme per la variabile aleatoria X .

$$\begin{cases} \sum_{i=1}^6 p_x = 1 \\ \sum_{i=1}^6 xp_x = B \\ \max H(p_1, \dots, p_6) = \sum_{i=1}^6 p_x \log_2 \frac{1}{p_x} \end{cases}$$

Conviene massimizzare la funzione $\tilde{H}(p_1, \dots, p_6) = \sum_{i=1}^6 p_x \ln \frac{1}{p_x}$ in quanto è più semplice e pur avendo valore massimo diverso, il punto di massimo è lo stesso.

In questo caso con queste condizioni ho una funzione con 4 variabili e per questo è conveniente utilizzare i moltiplicatori di Lagrange.

Definizione 6.1.3. *Sia $f \in A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ con A aperto e siano $g_1(x_1, \dots, x_n) = \dots = g_n(x_1, \dots, x_n) = 0$ vincoli espressi sotto forma di luoghi geometrici. Supponiamo che $f, g_1, \dots, g_n \in C^1(A)$ allora definisco la lagrangiana del sistema come*

$$L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_n) = f(x_1, \dots, x_n) + \sum_{j=1}^m \lambda_j g_j(x_1, \dots, x_n)$$

Osservazione 6.1.3. *La ricerca dei punti stazionari di questa funzione mi permette di trovare i punti stazionari della funzione f vincolata agli m -vincoli imposti inizialmente, dunque in questo caso ho che la lagrangiana del sistema è*

$$L(p_1, \dots, p_6, \alpha, \beta) = \tilde{H} + (\alpha + 1) \left(\sum_{i=1}^6 p_x - 1 \right) + \beta \left(\sum_{i=1}^6 xp_x - B \right)$$

Se ho un punto di massimo allora

$$\frac{\partial L}{\partial \alpha} = 0 \Rightarrow \sum_{i=1}^6 p_x = 1$$

Dunque il vincolo è verificato, inoltre sempre se il punto è massimo ho che

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow \sum_{i=1}^6 xp_x = B$$

Dunque anche questo vincolo è verificato.

A questo punto per trovare tale massimo bisogna andare a trovare quando il suo gradiente si annulla

$$\frac{\partial L}{\partial p_x} = 0 = \frac{\partial \tilde{H}}{\partial p_x} + (\alpha + 1) + \beta x = -\ln(p_x) - 1 + \alpha + 1 + \beta x$$

Dunque a questo punto abbiamo un problema con 8 equazioni dato che ho

$$\begin{cases} \ln(p_x) = \alpha + \beta x & x = 1, \dots, 6 \\ \sum p_x = 1 \\ \sum xp_x = B \end{cases}$$

Esempio 6.1.2. Sia $X \sim Be(\frac{1}{2})$ lancio di una moneta ideale e prendiamo $Z \sim Be(p)$ indipendente da X e costruiamo la variabile aleatoria

$$Y = Z(1_X) + (1 - Z)X$$

Dunque se $Z = 1 \Rightarrow Y = 1 - X$ e se $Z = 0 \Rightarrow Y = 0$ dunque

$$\begin{aligned} \mathbb{P}(Y = 1) &= \mathbb{P}(Z(1_X) + (1 - Z)X = 1) \\ &= \mathbb{P}(X = 1, Z = 0) + \mathbb{P}(1 - X = 1, Z = 1) \\ &= \mathbb{P}(X = 1)\mathbb{P}(Z = 0) + \mathbb{P}(X = 0)\mathbb{P}(Z = 1) = \frac{1}{2} \end{aligned}$$

Dunque $Y \sim Be(\frac{1}{2})$ A questo punto per calcolare $H(X, Y)$ ho che

	$X = 0$	$X = 1$
$Y = 0$	$\mathbb{P}(X = 0, Z = 0) = \frac{1}{2}(1 - p)$	$\mathbb{P}(X = 1, Z = 0) = \frac{1}{2}p$
$Y = 1$	$\mathbb{P}(X = 0, Z = 1) = \frac{1}{2}p$	$\mathbb{P}(X = 1, Z = 1) = \frac{1}{2}(1 - p)$

Dunque dato che per definizione ho che

$$H(X, Y) = \sum_{x,y=0}^1 p_{xy} \log_2 \frac{1}{p_{xy}} = 2 \left[\frac{1}{2}p \log_2 \frac{2}{p} + \frac{1}{2}(1 - p) \log_2 \frac{2}{1 - p} \right]$$

X, Y sono indipendenti se e solo se $p = \frac{1}{2}$ $Z \sim Be(\frac{1}{2})$. In questo caso ho che

$$H(X, Y) = 2 = H(X) + H(Y)$$

Inoltre è il massimo valore possibile dell'entropia per una v.a. che assume quattro valori.

Se fosse $p = 1$ o $p = 0$ allora X, Y sono perfettamente correlate in quanto $Y = 1 - X$, allora $H(X, Y) = 1$ quindi tutta l'incertezza ricade sull'esito della prima moneta.

6.2 Entropia congiunta

Definizione 6.2.1 (Entropia congiunta).

Siano X, Y due v.a. discrete definite su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e siano rispettivamente i loro supporti

$$S_X := \{x_1, \dots, x_n\}$$

$$S_Y := \{y_1, \dots, y_m\}$$

Dunque definisco entropia congiunta

$$H(X, Y) = - \sum_{i=1}^n \sum_{j=1}^m \mathbb{P}(X = x_i, Y = y_j) \log_2 (\mathbb{P}(X = x_i, Y = y_j))$$

Definizione 6.2.2 (Entropia condizionata).

Siano X, Y due v.a. discrete definite su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e siano rispettivamente i loro supporti

$$S_X := \{x_1, \dots, x_n\}$$

$$S_Y := \{y_1, \dots, y_m\}$$

Dunque definisco entropia di Y condizionata da X

$$H_X(Y) = \sum_{i=1}^n \mathbb{P}(X = x_i) H_i(Y)$$

Con

$$H_i(Y) = - \sum_{j=1}^m \mathbb{P}(Y = y_j | X = x_i) \log_2(\mathbb{P}(Y = y_j | X = x_i))$$

Lemma 6.2.1 (Disuguaglianza di Gibbs).

Siano (p_1, \dots, p_n) e (q_1, \dots, q_n) tali per cui

$$\sum_i p_i = \sum_i q_i = 1$$

Allora vale che

$$- \sum_i p_i \log(p_i) \leq \sum_i p_i \log(q_i)$$

Dimostrazione. Escludendo i casi banali ho che $p_i, q_i > 0$ dunque è ovvio che

$$\frac{q_i}{p_i} > 0$$

Perciò posso affermare che

$$\log\left(\frac{q_i}{p_i}\right) \leq \frac{q_i}{p_i} - 1$$

Che è del tutto equivalente ad affermare che

$$p_i \log\left(\frac{q_i}{p_i}\right) \leq p_i \left[\frac{q_i}{p_i} - 1\right]$$

Dunque

$$\sum_i p_i \log\left(\frac{q_i}{p_i}\right) \leq \sum_i p_i \left[\frac{q_i}{p_i} - 1\right]$$

Poiché

$$\sum_i p_i \left[\frac{q_i}{p_i} - 1\right] = \sum_i q_i - \sum_i p_i = \sum_i q_i - \sum_i p_i = 0$$

Allora ho che

$$\begin{aligned} 0 &\geq \sum_i p_i \log\left(\frac{q_i}{p_i}\right) \\ &= \sum_i p_i \log(q_i) + \sum_i p_i \log\left(\frac{1}{p_i}\right) \\ &= \sum_i p_i \log(q_i) - \sum_i p_i \log(p_i) \end{aligned}$$

Da cui la tesi. □

Teorema 6.2.1. Siano X, Y due v.a. discrete definite su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ allora vale che

$$H(X, Y) = H(X) + H_Y(X)$$

Corollario 6.2.1. Siano X, Y due v.a. discrete definite su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ allora vale che

1. Se X, Y sono indipendenti

$$H(X, Y) = H(X) + H(Y)$$

2. Sussiste la seguente equivalenza

$$H_X(Y) = 0 \iff \rho(Y, X) = 1$$

3.

$$H_X(Y) \leq H(Y)$$

Dimostrazione.

Punto 1

Se le due variabili sono indipendenti allora ho per Bayes

$$\begin{aligned} \mathbb{P}(Y = y_j | X = x_i) &= \frac{\mathbb{P}(Y = y_j, X = x_i)}{\mathbb{P}(X = x_i)} \\ &= \frac{\mathbb{P}(Y = y_j) \mathbb{P}(X = x_i)}{\mathbb{P}(X = x_i)} \\ &= \mathbb{P}(Y = y_j) \end{aligned}$$

Dunque per definizione ho che

$$\begin{aligned} H_X(Y) &= - \sum_{i=1}^n \mathbb{P}(X = x_i) \sum_{j=1}^m \mathbb{P}(Y = y_j | X = x_i) \log_2(\mathbb{P}(Y = y_j | X = x_i)) \\ &= - \sum_{i=1}^n \mathbb{P}(X = x_i) \sum_{j=1}^m \mathbb{P}(Y = y_j) \log_2(\mathbb{P}(Y = y_j)) \\ &= \sum_{i=1}^n \mathbb{P}(X = x_i) H(Y) \\ &= H(Y) \sum_{i=1}^n \mathbb{P}(X = x_i) \\ &= H(Y) \end{aligned}$$

Punto 2

\Leftarrow

In questo caso ho che $Y = aX + b$ dunque posso affermare che

$$\mathbb{P}(Y = y_j | X = x_i) = \begin{cases} 0 & j \neq i \\ 1 & j = i \end{cases}$$

Ma questo per un risultato precedente significa che $H_i(Y) = 0$.

\Rightarrow Per un risultato precedente ho che $\forall i, \exists j$

$$\mathbb{P}(Y = y_j | X = x_i) = 1$$

Inoltre

$$\forall k \neq j, \mathbb{P}(Y = y_k | X = x_i) = 0$$

Dunque ciò significa che la sua distribuzione è completamente determinata da X e quindi $\rho(X, Y) = 1$.

Punto 3

Per definizione e sfruttando la disuguaglianza di Gibbs

$$\begin{aligned} H_X(Y) &= - \sum_{i=1}^n \mathbb{P}(X = x_i) \sum_{j=1}^m \mathbb{P}(Y = y_j | X = x_i) \log_2(\mathbb{P}(Y = y_j | X = x_i)) \\ &= \sum_{i=1}^n \mathbb{P}(X = x_i) \left(- \sum_{j=1}^m \mathbb{P}(Y = y_j | X = x_i) \log_2(\mathbb{P}(Y = y_j | X = x_i)) \right) \\ &\leq \sum_{i=1}^n \mathbb{P}(X = x_i) \left(- \sum_{j=1}^m \mathbb{P}(Y = y_j | X = x_i) \log_2(\mathbb{P}(Y = y_j)) \right) \\ &= - \sum_{i=1}^n \sum_{j=1}^m \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j | X = x_i) \log_2(\mathbb{P}(Y = y_j)) \\ &= - \sum_{i=1}^n \sum_{j=1}^m \mathbb{P}(Y = y_j, X = x_i) \log_2(\mathbb{P}(Y = y_j)) \\ &= - \sum_{j=1}^m \log_2(\mathbb{P}(Y = y_j)) \left(\sum_{i=1}^n \mathbb{P}(Y = y_j, X = x_i) \right) \\ &= - \sum_{j=1}^m \mathbb{P}(Y = y_j) \log_2(\mathbb{P}(Y = y_j)) \\ &= H(Y) \end{aligned}$$

Quindi

$$H_X(Y) \leq H(Y)$$

□

Osservazione 6.2.1. *Da quanto visto fino ad ora è dunque ovvio che*

$$H(X, Y) \leq H(X) + H(Y)$$

Esempio 6.2.1. *Siano X, Y due variabili aleatorie tali per cui $X, Y \sim Be(\frac{1}{2})$ e $\rho = 0.2$, devo calcolare l'entropia congiunta.*

Dato che $\rho \neq 0$ allora ho che X, Y non sono indipendenti, inoltre so che

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

Ma per definizione di coefficiente di correlazione ho

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \rho\sqrt{\text{Var}[X]\text{Var}[Y]}$$

Dato che $\mathbb{E}[XY] = \sum_x \sum_y xy \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = 1, Y = 1)$ allora

$$\mathbb{P}(X = 1, Y = 1) = 0.2 \frac{1}{4} + \frac{1}{4} = \frac{3}{10}$$

Quindi $\mathbb{P}(X = 1, Y = 0) = \frac{1}{5}$ e per simmetria, dato che le due variabili hanno la stessa distribuzione, ho che $\mathbb{P}(Y = 1, X = 0) = \frac{1}{5}$ quindi $\mathbb{P}(X = 0, Y = 0) = \frac{3}{10}$ perciò per definizione

$$H(X, Y) = \frac{3}{5} \log_2 \frac{3}{10} + \frac{2}{5} \log_2 \frac{1}{5} = 1.97095$$

Osservazione 6.2.2. *Se x, Y sono indipendenti allora $H_j(Y) = H(Y)$ e $H_X(Y) = H(Y)$*

Osservazione 6.2.3. Date $\{p_k\}_{k=1}^N, \{q_k\}_{k=1}^N$ tali per cui $\sum p_k = \sum q_k = 1$ tutti positivi allora vale la disuguaglianza di Gibbs

$$\sum_k p_k \log_2 \left(\frac{1}{p_k} \right) \leq \sum_k p_k \log_2 \left(\frac{1}{q_k} \right)$$

Dimostrazione. In questa dimostrazione posso usare \ln in quanto si ottiene moltiplicando tutto per una costante, dunque

$$\sum_k p_k \ln \left(\frac{1}{p_k} \right) - p_k \ln \left(\frac{1}{q_k} \right) = \sum_k p_k \ln \left(\frac{q_k}{p_k} \right) \leq \sum_k p_k \left(\frac{q_k}{p_k} - 1 \right) = \sum_k (q_k - p_k) = 0$$

□

6.2.1 Entropia della somma di due variabili aleatorie indipendenti

Teorema 6.2.2. Se X è una v.a. e data una funzione $g: \mathbb{R} \rightarrow \mathbb{R}$ iniettiva allora $G(X) \sim \{(g(x_k), p_k)\}$ dunque

$$H(g(X)) = H(X)$$

Se invece non è iniettiva, quindi per definizione $\exists x_1, x_2 | g(x_1) = g(x_2) = z_1$ allora $G(X) \sim \{(z_1, p_1 + p_2), (g(x_k), p_k)\}$ allora

$$H(g(X)) = (p_1 + p_2) \log_2 \frac{1}{p_1 + p_2} + \sum_{k \geq 1} p_k \log_2 \frac{1}{p_k}$$

Dunque

$$H(g(X)) \leq H(X)$$

Dimostrazione. ?

□

Teorema 6.2.3. Sia X, Y v.a. indipendenti e sia $S = X + Y$ allora vale che

$$\max\{H(X), H(Y)\} \leq H(S) \leq H(X) + H(Y)$$

Dimostrazione. Devo dimostrare le due disuguaglianze perciò procedo con ordine.

$\max\{H(X), H(Y)\} \leq H(S)$
Prima di tutto dimostro che $H_X(S) = H_X(Y)$ infatti per definizione so che

$$\begin{aligned} H_X(S) &= - \sum_x \mathbb{P}(X=x) \sum_s \underbrace{\mathbb{P}(S=s|X=x)}_{\mathbb{P}(Y=s-x|X=x)} \log_2 \mathbb{P}(S=s|X=x) \\ &= - \sum_x \mathbb{P}(X=x) \sum_s \underbrace{\mathbb{P}(Y=s-x|X=x)}_{\mathbb{P}(Y=y|X=x)} \log_2 \mathbb{P}(Y=s-x|X=x) \\ &= - \sum_x \mathbb{P}(X=x) \sum_s \mathbb{P}(Y=y|X=x) \log_2 \mathbb{P}(Y=y|X=x) \\ &= H_X(Y) \end{aligned}$$

Siccome sappiamo che X, Y sono indipendenti allora sappiamo che $H(Y) = H_X(Y)$, dunque sappiamo che

$$H_X(S) = H(Y)$$

Perciò per quanto dimostrato precedentemente ho che

$$H(S) \geq H(Y)$$

Per simmetria

$$H(S) \geq H_Y(S) = H_Y(X) = H(X)$$

Dunque la prima disuguaglianza è dimostrata.

$\frac{H(S) \leq H(X) + H(Y)}{\text{Dal teorema precedente ho che}}$

$$H(S) \leq H(X, Y) = H(X) + H(Y)$$

□

Lemma 6.2.2. *Se (X, Y) si può scrivere come una funzione, quindi $(X, Y) = f(S)$ allora*

$$H(S) = H(X) + H(Y)$$

Dimostrazione. Per quanto dimostrato precedentemente so che

$$H(S) \leq H(X, Y)$$

Ma per il teorema precedente ho che

$$H(X, Y) = H(f(S)) \leq H(S)$$

Dunque la tesi è dimostrata

□

Esempio 6.2.2. *Supponiamo che X, Y siano indipendenti e $Be(\frac{1}{2})$ e che*

$$H(X) = H(Y) = 1$$

Inoltre ho che

$$S = X + Y \sim \begin{cases} 0 & \frac{1}{4} \\ 1 & \frac{1}{2} \\ 2 & \frac{1}{4} \end{cases} \implies H(S) = \left(\frac{1}{4} \log_2 4\right) 2 + \frac{1}{2} \log_2 2 = \frac{3}{2}$$

Dunque

$$\max(H(X), H(Y)) = 1 \leq \frac{3}{2} = H(S) \leq 2 = H(X) + H(Y)$$

Esempio 6.2.3. *Sia $X \sim Be(p)$ con $Y = X$ dunque $S = 2X$ con*

$$H(X) = H(Y) = H(S) = 1$$

In quanto $(X, Y) = f(S)$ perché se $S = 0$ allora $X = Y = 0$ se $S = 2$ allora $X = Y = 1$