Statistica

Mattia Marini

6 dicembre 2024

Indice

1 Tribù e algebre

E' spesso necessario scegliere i possibili eventi all'interno di una classe che presenti caratteristiche particolari, ossia una tribù

Definizione 1: *Tribù*
Data una classe $\mathcal{A}$ su uno spazio campionario ω si dice tribù se soddisfa le seguenti proprietà:

1. $\omega \in \mathcal{A}$
2. se $A \in \mathcal{A}$ allora $A^c \in \mathcal{A}$
3. Se $A_1, A_2, \ldots, A_i, \ldots \in \mathcal{A}$ allora $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{A}$

Definizione 2: *Algebra*
Data una classe $\mathcal{A}$ su uno spazio campionario ω si dice tribù se soddisfa le seguenti proprietà:

1. $\omega \in \mathcal{A}$
2. se $A \in \mathcal{A}$ allora $A^c \in \mathcal{A}$
3. Se $A_1, A_2, \ldots, A_i, \ldots \in \mathcal{A}$ allora $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$

In particolare nota che

- *Proprietà 2:* garantisce che se un evento E si può verificare, allora si può verificare anche E^c
- *Proprietà 3:* garanisce che se si possono verificare gli eventi $E_1, \ldots, E_i$ allora si può verificare anche l'evento $E_1 \cup E_2 \cup \ldots \cup E_i$.

Teorema 1: *Tribù e algeebra*
Un'algebra è anche una tribù, tuttavia non vale il contrario

Dimostrzione:

- Per definizione una tribù contine ω e di ogni elemento contiene anche il complementare. Quindi ogni tribù contiene $\emptyset$.

- Data una tribù, l'unione di un numero finito di suoi elementi sta in questa. Ma anche l'unione di un numero infinito di suoi elementi, in quanto posso aggiungere infiniti insiemi vuoti

Definizione 3: *Evento*
Un elemento di una tribù $\mathcal{A}$ si dice evento.

Definizione 4: *Spazio probabilizzabile*
Dato uno spazio campionario ω e una tribù $\mathcal{A}$ la coppia

$$(\omega, \mathcal{A})$$

è detta spazio probabilizzabile

Definizione 5: *Spazio probabilizzato*
Dato ω e una tribù A di ω e una probabilità P sullo spazio probabilizzabile (ossia la coppia ω, A), chiameremo la terna

$$(\omega, A, P)$$

uno spazio probabilizzato

Definizione 6: *Probabilità (versione assiomatica Kolmogorov)*
Dato uno spazio proabilizzabile $(\omega, \mathcal{A})$ la probabilità è un'applicazione $Pr : \mathcal{A} \to \mathbb{R}^+$ tale che:

- *Non negatività:* Se $A \in \mathcal{A}$ allora $Pr\left(A\right) \geq 0$
- *Normalizzazione:*$\Pr\left(\omega\right) = 1$
- σ *addittività:* Data una serie di eventi incompatibili, la probabilità dell'unione di questi è la somma delle probabilità di ogni evento, formalmente:

$$\forall A_j, A_j \text{ t.c. } A_i \cap A_j = \emptyset i \neq j \quad \Rightarrow \quad Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} Pr\left(A_i\right)$$

Esempio 1

Pendo un insieme e una tribù:

$$\omega = \{1,2,3,4,5,6\} \quad \mathcal{A} = P\left(\omega\right)$$

$$(\omega, \mathcal{A}) \text{ è uno spazio probabilizzabile} \quad Pr\left(\{1\}\right) = Pr\left(\{2\}\right) = \ldots = Pr\left(\{6\}\right) = p$$

Per procedere formalmente devo applicare gli assiomi:

$$Pr\left(\omega\right) = 1 = Pr\left(\{1\} \cup \{2\} \cup \ldots \cup \{6\}\right) = \sum_{k=1}^{6} Pr\left(\{i\}\right) = 6p$$

il tutto è uguale a $6p$ visto che ho supposto che gli eventi elementari avessero la stessa probabilità. Quindi

$$p = \frac{1}{6}$$

Se ora dovessi ricavare la probabilità di un insieme generico posso utilizzare p per farlo:

$$P(\{1,3,5\}) = P(\{1\},\{3\},\{5\}) = P(\{1\}) + P(\{3\}) + P(\{5\}) = \frac{3}{6} = \frac{1}{2}$$

Esempio 2

Immagino di prendere un bersaglio di raggio r e di voler calcolare la probabilità di colpire ognuna delle parti. Assumo che non sia possibile colpire fuori dal pannello e che la probabilità di colpire ciascuna porzione sia direttamente proporzionale alla sua area.

$$\begin{aligned}
P(\{3\}) &= \frac{\pi\left(\frac{1}{3}r\right)^2}{\pi r^2} = \frac{1}{9} \\
P(\{2\}) &= \frac{\pi\left(\frac{2}{3}r\right)^2 - \pi\left(\frac{1}{3}r\right)^2}{\pi r^2} = \frac{1}{3} \\
P(\{1\}) &= \frac{\pi r^2 - \pi\left(\frac{2}{3}r\right)^2}{\pi r^2} = \frac{5}{9}
\end{aligned}$$

1.1 Regole importanti

- **Regola 1** $(\omega, \mathcal{A}, P)$ Allora $Pr(A^c) = 1 - Pr(A)$
- **Regola 2** $(\omega, \mathcal{A}, P)$ se $A, B \in A$ allora $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - Se A, B sono disgiunti, allora posso applicare l'assioma di unione di insiemi disgiunti e sono a posto
  - Se A, B <u>NON</u> sono disgiunti, allora applico il seguente trucco:

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

  - Gli insiemi A e $(A \cap A^c)$ sono disgiunti e posso dunque applicare l'assioma
- **Regola 3** Se $A \subset B$ allora $P(A) \leq P(B)$
  - $B = (A \cap B) \cup (A^c \cap B)$
  - Siccome $A \subset B$ allora $B = A \cup (A^c \cap B)$
- **Regola 4** *(Disuguaglianza di Bonferroni).* Se $A_1, A_2, \cdots, A_n$ sono eventi allora

$$\sum_{i=1}^{n} \Pr(A_i) - \sum_{1 \leq i \leq j \leq n} \Pr(A_i \cap A_j) \leq \Pr(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \Pr(A_i), \quad n \geq 1$$

Definizione 7:
Data una famiglia di spazi probabilizzabili $\{(\omega_i, A_i)\}_{i \in I}$, possiamo costruire uno spazio probabilizzabile prodotto ottenuto come pr odotto cartesiano

Definizione 8: *Disposizioni con ripetizioni*
Sia $S = \{a_1, a_2, \ldots, a_n\} \quad \#s = n$ un insieme costituito da oggetti distinti. Il numero di allineamenti che si possono formare con r oggetti scelti tra gli n si indica con $D^*_{n,r}$. Due allineamenti sono diversi se una delle seguenti condizioni è verificata:

- Contengono oggetti divisersi
- Contengono stessi oggetti in numero diverso
- lo stesso oggetto si ripete un numero diverso di volte

$$D^*_{n,r} = n^r$$

2 Calcolo combinatorio

Definizione 9: *Disposizione senza ripetizione*
Sia $S = \{a_1, a_2, \ldots, a_n\} \quad \#s = n$ un insieme costituito da oggetti distinti. Il numero di allineamenti che si possono formare con r oggetti scelti tra gli n si indica con $D_{n,r}$. Due allineamenti sono diversi se una delle seguenti condizioni è verificata:

- Contengono oggetti divisersi
- Contengono stessi oggetti in numero diverso

$$D_{n,r} = n \cdot (n-1) \cdot \ldots \cdot (n-r+1) = \frac{n!}{(n-r)!}$$

Definizione 10: *Permutazioni*
Le permutazioni di n oggetti non sono altro che le disposizioni senza ripetizione di n oggetti in n posizioni:

$$P_n = D_{n,n} = n!$$

Definizione 11: *Combinazioni*
Sono date da tutti gli isiemi contenenti oggetti diversi (ossia non conta l'ordine con il quale questi vengono disposti).

$$C_{n,r} = \frac{D_{n,r}}{P_r} = \frac{n!}{(n-r)! \cdot r!} = \binom{n}{r}$$

Ossia quanti gruppi di r persone posso estrarne da n

2.1 Esercizi

Esercizio 1

Immaginiamo di avere un comitato di 5 persone su 10.

- Una persona può comparire nel comitato al più una volta

- L'ordine con il quale le persone appaiono i nomi nel comitato non è importante

Ciò significa che devo utilizzare le combinazioni:

$$C_{10,5} = \frac{10!}{5! \cdot 5!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{5!} = 252$$

Esercizio 2

Supponiamo di avere $\omega_i = \{1, \ldots, 100\} \quad i = 1, 2, 3$

$$\omega = \bigtimes_{i=1}^{3} \omega_i$$

supponendo di avere 6 volcali, il numero di parole che abbiano una consonante come primo , mentre una vocale come secondo e terzo è:

$$\frac{20 \cdot 6 \cdot 6}{100^3} = 0,00072$$

Esercizio 3

Supponiamo di avere un insieme così definito:

$$\{A_1, A_2, A_2, A_4, A_5, B_1, B_2, C_1, D_1, R_1, R_2\}$$

immagina di estrarre da un vaso una lettera alla volta, senza riimmetterla dopo ogni estrazione. Calcola la probabilità di aver estratto le lettere in questo ordine:

$$\{A, B, R, A, C, A, D, A, B, R, A\}$$

innanzitutto, $\omega = P_1 1 = 11!$. Inoltre mi immagino che ogni lettera ripetuta può essere permutata (in quanto non cambierebbe avere A_1 o A_2 in posizione 1). Dunque prendendo la stringa *ABRACADABRA* devo considerare ogni possibile ordine delle a, delle b e così via:

$$P\left(\text{"}ABRACADABRA\text{"}\right) = \frac{5! \cdot 2! \cdot 1! \cdot 1! \cdot 2!}{11!} = \frac{1}{83160}$$

Esercizio 4

Viene somministrato un trattamento (T) e un placebo (C) a maschi e femmine Nota che

T	M	S	8		T	M	I	5
C	M	S	4		C	M	I	3
T	F	S	12		T	F	I	15
C	F	S	2		C	F	I	3

la differenza di efficacia fra placebo e trattamento dipende dal sesso in quanto:

$$P(S|T) = \frac{8+12}{8+12+5+15} = \frac{1}{2}$$
$$P(S|c) = \frac{4+12}{4+2+3+3} = \frac{1}{2}$$

tuttavia, nota che:

$$P(S|T,M) = \frac{8}{8+5} = \frac{8}{13}$$
$$P(S|C,M) = \frac{4}{4+3} = \frac{4}{7}$$
$$P(S|T,F) = \frac{12}{12+15} = \frac{12}{27}$$

2.2 Teorema di Bayes

Teorema 2: *Teorema di Bayes*
Abbiamo una successione di eventi $\{A_i\}_{i=1}^{+\infty}$ che formi una partizione di ω. Assumiamo che $P(A_i) > 0 \forall i$. Prendo un $b \in \mathcal{A}$ t.c. $P(B) > 0$

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j) \cdot P(A_j)}$$

Definizione 12: *Indipendenza stocastica*
Dato $(\omega, \mathcal{A}, P)$ e due eventi A, B diremo che sono tra loro stocasticamente indipendenti se e solo se:

$$P(A \cap B) = P(A) \cdot P(B)$$

Ciò significa che i due eventi non si influenzano a vicenda:

$$P(A|B) = \frac{P(A \cap B)}{P(B)(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A)$$

3 Probabilità sui reali

Costruiamo uno spazio probabilizzato che veda come spazio campionario $\mathbb{R}$:

- Per questioni di semplicità, noi considereremo $\Omega = \mathbb{R}$
- Dobbiamo ora definire una *tribù*. La tribù $\mathcal{P}(\mathbb{R})$ tuttavia contiene degli insiemi che non sono di nostro interesse (vedi *insiemi di vitali*)

3.1 Insiemi di Bourelliani

L'idea di base è quela di considerare solo insiemi per i quali ha senso parlare della loro *lunghezza* (o area e volume rispettivamente in $\mathbb{R}^2$ e $\mathbb{R}^3$) Per costruire questo insieme procediamo così:

- Prendiamo l'insieme dato da ogni sottointervallo del tipo $(a, b]$ di $\mathbb{R}$ così definito:

$$\mathcal{F} = \{(a, b] : a \leq b; a, b \in \mathbb{R}\}$$

- Dobbiamo ora creare una tribù a partire da questo insieme. In particolare prendiamo la più piccola tribù che contiene ogni sotto intervallo di $\mathbb{R}$:

$$\mathcal{A}_{\mathcal{F}} = \cap \{\mathcal{A} : \mathcal{A} : \text{ è una tribù e } \mathcal{F} \subseteq \mathcal{A}\}$$

La tribù $\mathcal{A}_{\mathcal{F}}$ è detta tribù dei Boreliani. La tribù contiene i seguenti elementi:

- (a, b)
- $[a, b]$
- $[a, b)$
- $(-\infty, b]$
- (a, ∞)
- i singoletti di $\mathbb{R}$;
- gli insiemi finiti di $\mathbb{R}$;
- gli insiemi numerabili di $\mathbb{R}$.

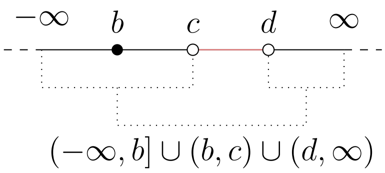
Dimostrazione: Partiamo dimostrando che $(a, b) \in \mathcal{A}_{\mathcal{F}}$.

- (a, b). Definiamo la seguente successione:

$$\bigcup_{k=1}^{\infty} \left(a, b - \frac{1}{k}\right]$$

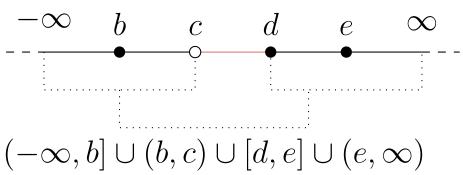
  Chiaramente per $k \to \infty$ la successione tende a (a, b). Visto che l'unione numerabile di elementi di una tribù sta nella tribù stessa, $(a, b) \subset \mathcal{A}_{\mathcal{F}}$

- $(-\infty, b]$. Per lo stesso motivo di prima, $\bigcup_{k=1}^{-\infty} (k, b] \to (-\infty, b)$
- (a, ∞). Visto che $(-\infty, b] \in \mathcal{A}_{\mathcal{F}}$ allora anche il suo negato appartiane
- $[a, b]$. Considero l'insieme seguente: $(-\infty, b] \cup (b, c) \cup (d, \infty)$:



  negando questo intervallo otteniamo chiaramente $[a, b]$

- $[a, b)$. Considero l'insieme seguente: $(-\infty, b] \cup (b, c) \cup [d, e] \cup (e, \infty)$:



Dato uno spazio campionario e una tribù (dei *bourelliani*), manca solo definire la funzione di probabilità. Tale funzione è detta funzione di distribuzione:

Definizione 13: *Funzione di distribuzione o ripartizione*
Una funzione di distribuzione si $(\mathbb{R}, B(\mathbb{R}))$ $F(x)$ è una funzione di distribuzione se:

- Non decrescente: $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$
- Continua a destra: $\lim_{x \to x_0^+} F(x) = F(x_0)$
- $\lim_{x \to -\infty} F(x) = 0$ e $\lim_{x \to +\infty} F(x) = 1$

Data una funzione che rispecchi queste proprietà, la probabilità di un intervallo $P((a,b])$ è data da:

$$F(b) - F(a)$$

mentre la probabilità del singoletto $P(\{x\})$ è data da:

$$F(x) - \lim_{z \to x^-} F(z)$$

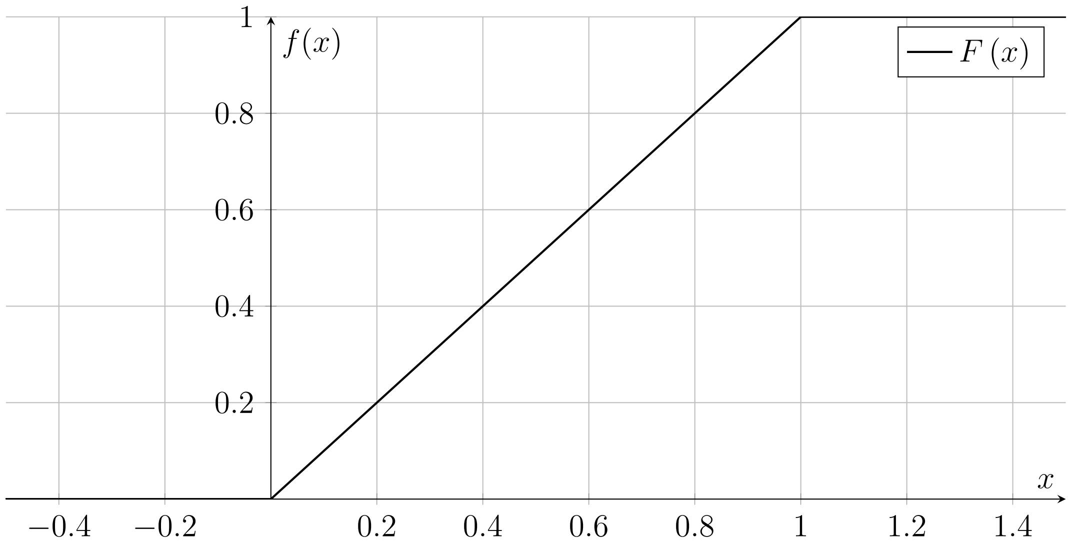
3.2 Esempi

Consideriamo il seguente esperimento: *calcola la probabilità di estrarre un punto nell'intervallo* $[0,1]$ *nei numeri reali* $\mathbb{R}$. Avremo dunque

$$\omega = [0,1] \in \mathbb{R} \quad \mathcal{A}_{\mathcal{F}} = \mathcal{B}[0,1] \quad P((a,b]) = a - b$$

Quindi la probabilità di estrarre un punto appartenente ad un intervallo $(a,b]$ corrisponde alla sua lunghezza. La funzione di probabilità sarà la seguente:

Esempio funzione di distribuzione



Calcoliamo ora le probabilità dei seguenti eventi.

- $A_1 = \left\{\left(0, \frac{1}{4}\right)\right\}$
- $A_2 = \{a\}$

- $A_3 = \{a_1, a_2, \cdots\}$
- $A_4 = \left\{\left[\frac{1}{2}, \frac{3}{4}\right] \cup \left[\frac{4}{5}, 1\right]\right\}$
- $A_5 = [0, 1]/A_3$;
- $A_6 = \cup_{n=1}^{\infty} \left[\frac{k}{k+1}, \frac{k}{k+1} + \frac{1}{10^k}\right]$

dove $a \in [0, 1]$ e $a_i \in [0, 1], i = 1, 2, \cdots$.

Soluzioni

- A_1 è un intervallo e pertanto $\Pr(A_1) = F\left(\frac{1}{4}\right) - F(0) = \frac{1}{4}$; -
- $A_2 = \{a\} = \cap_{n=1}^{\infty} \left(a - \frac{1}{n}, a\right] = \lim_{n\to+\infty} \left(a - \frac{1}{n}, a\right]$ e pertanto (ricordando che la successione degli intervalli è non crescente) per la proprietà di continuità,

$$\begin{aligned}\Pr(A_2) = \Pr(\{a\}) = \Pr\left(\lim_{n\to+\infty}\left(a - \frac{1}{n}, a\right]\right) &= \lim_{n\to+\infty} \Pr\left(\left(a - \frac{1}{n}, a\right]\right) \\ &= \lim_{n\to+\infty}\left(F(a) - F\left(a - \frac{1}{n}\right)\right) \\ &= \lim_{n\to+\infty} \frac{1}{n} = 0\end{aligned}$$

-

- $A_3 = \{a_1, a_2, \cdots\} = \cup_{i=1}^{\infty} \{a_i\}$ e quindi

$$\Pr(A_3) = \Pr(\{a_1, a_2, \cdots\}) = \Pr(\cup_{i=1}^{\infty} \{a_i\}) = \sum_{i=1}^{\infty} \Pr(\{a_i\}) = 0$$

-

- Per A_4 si consideri che i due intervalli sono disgiunti allora

$$\Pr(A_4) = \Pr\left(\left[\frac{1}{2}, \frac{3}{4}\right] \cup \left[\frac{4}{5}, 1\right]\right) = \Pr\left(\left[\frac{1}{2}, \frac{3}{4}\right]\right) + \Pr\left(\left[\frac{4}{5}, 1\right]\right)$$

$$= F\left(\frac{3}{4}\right) - F\left(\frac{1}{2}\right) + F(1) - F\left(\frac{4}{5}\right) = \frac{1}{4} + \frac{1}{5} = \frac{9}{20}$$

- A_5:

$$\Pr(A_5) = \Pr([0, 1]/A_3) = \Pr([0, 1]) - \Pr(A_3) = \Pr([0, 1]) = 1$$

- A_6 è un'unione numerabile. Questo si può dimostrare per induzione (ossia dobbiamo dimostrare che gli intervalli sono tutti disgiunti)
  - Cerchiamo di dimostrare che l'intervallo *n-esimo* non si sovrappone con l'*n+1-esimo*

$$\frac{k+1}{k+2} > \frac{k}{k+1} + \frac{1}{10^k} \quad \forall k \geq 1$$

    semplificando devo dimostrare che

$$10^k > k^2 + 3k + 2$$

- *Passo base:* per $k = 1$ l'ipotesi è verificata
- *Passo induttivo:* assumiamo che $10^k > k^2 + 3k + 2$ sia vera. Devo dimostrare che la proprietà vale anche per $k+1$:

$$10^{k+1} > (k+1)^2 + 3(k+1) + 2$$

$$10^k \cdot 10 > k^2 + 5k + 6$$

$$10^k > \frac{k^2 + 5k + 6}{10}$$

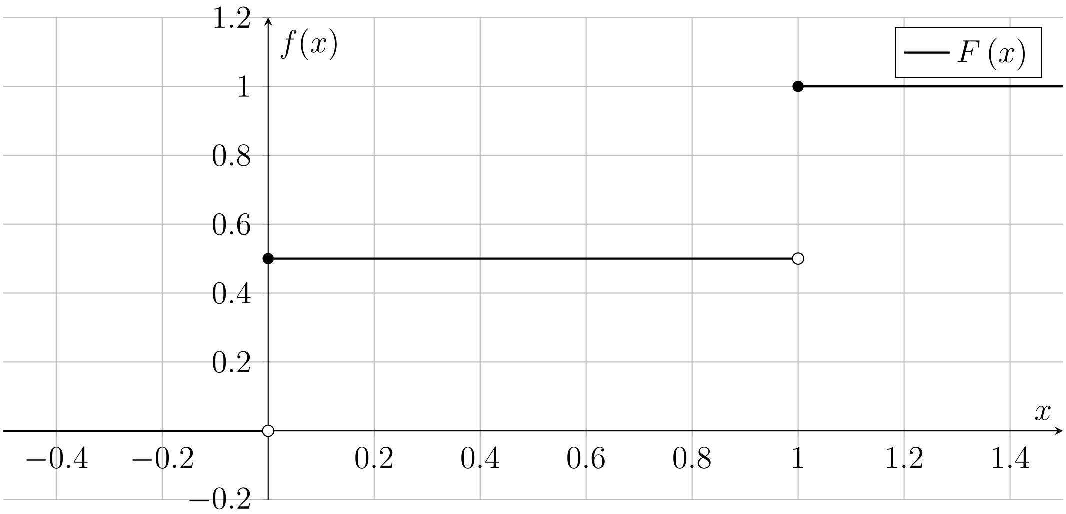
per ipotesi induttiva $10^k > k^2 + 3k + 2$, quindi se $k^2 + 3k + 2 > \frac{k^2+5k+6}{10}$ ho concluso

$$10 \cdot \left(k^2 + 3k + 2\right) > k^2 + 5k + 6 \quad \rightarrow \quad 9k^2 + 25k + 14 > 0 \quad \forall k \geq 1$$

Dunque posso procedere con il calcolo, considerando che tutti gli intervalli sono disgiunti

$$\begin{aligned}
\Pr(A_6) &= \sum_{k=1}^{\infty} \Pr\left(\left[\frac{k}{k+1}, \frac{k}{k+1} + \frac{1}{10^k}\right]\right) \\
&= \sum_{k=1}^{\infty} \left(F\left(\frac{k}{k+1} + \frac{1}{10^k}\right) - F\left(\frac{k}{k+1}\right)\right) \\
&= \sum_{k=1}^{\infty} \frac{1}{10^k} = \frac{1}{1-\frac{1}{10}} - 1 = \frac{\frac{1}{10}}{1-\frac{1}{10}} = \frac{1}{9}
\end{aligned}$$

Altra funzione di probabilità



Nota come in questa funzione ogni intervallo $(a, b)\, a, b \in [0, 1)$ abbia probabilità 0! L'intero "peso" si probabilità è contenuto nei singoletti $\{0\}$ e $\{1\}$

4 Variabili aleatorie

Quando effettuiamo un esperimento aleatorio spesso i risultati sono di natura <u>non numerica</u>. Spesso tuttavia può essere utile averne di natura numerica. Per soddisfare questa esigenza esistono le variabili aleatorie, che permettono di collegare un esito di un esperimento aleatorio di natura non numerica ad un numero reale. Consideriamo il lancio di due dadi a 6 facce. In questo caso avremmo:

$$\Omega = \{(x, y)\, x, y \in 1, \ldots, 6\}$$

tuttavia potrei essere interessato, ad esempio, a trovare la probabilità che *la somma delle due facce sia* x. In questo caso posso usare una variabile aleatoria, che colleghi ogni possibile esito dell'esperimento ad un numero reale (in questo caso ad $x + y$)

Definizione 14: *Variabile aleatoria*
Una variabile aleatoria reale X è una funzione che collega ad ogni elemento di uno spazio campionario Ω ad un numero reale:

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

inoltre una v.a. può essere discreta o continua

Definizione 15: *V.a. discreta e continua*
Una variabile aleatoria si dice:

- *Ddiscreta* se il numero di valori che può assumere è finito o numerabile
- *Continua* se assume un numero di valori più che numerabile

Esempi

- *Discreta* Ad esempio, la variabile che associa al lancio di due dadi la somma dei numero usciti
- *Continua* Ad esempio, la variabile che associa al lancio ripetuto 2 volte di una freccetta la somma delle altezze dei due punti colpiti

4.1 Calcolo probabilità tramite v.a.

Tornando al caso di prima, consideriamo:

$$\Omega = \{(x, y)\, x, y \in 1, \ldots, 6\} \quad \mathcal{A} = P(\Omega)$$

Definiamo una variabile aleatoria che manda ogni possibile combinazione di dadi nella loro somma:

$$X : \Omega \to \tilde{\Omega} = \{2, \ldots, 12\}$$

Se ora volessi calcolare la probabilità che la somma dei lanci sia x, dovrei calcolare la probabilità dell'insieme costituito da tutti gli elementi di omega tali per cui la somma dei due lanci sia x, ossia tutti quelli per cui vale $X(\omega) = x \quad \omega \in \Omega$. Questo insieme chiaramente sta nella tribù scelta ($\mathcal{A}$)

$$Pr(\text{somma} = x) = Pr(\{\omega \in \Omega \text{ t.c. } X(\omega) = x\})$$

Il medesimo discorso può essere fatto se si volesse calcolare la probabilità di un intervallo:

$$Pr(\text{somma} \in (a, b]) = Pr(\{\omega \in \Omega \text{ t.c. } a < X(\omega) \leq b\})$$

La per descrivere la distribuzione di una variabile aleatoria può tornare comodo utilizzare un grafico, detto funzione di probabilità (o densità discreta)

Definizione 16: *Funzione di probabilità discreta*
Sia X una v.a. e sia $R_x = \{z_1, z_2, \ldots\}$ l'insieme di tutti i valori che questa può assumere, allora la funzione seguente definita in $\mathbb{R}$ è detta funzione di probabilità della v.a. X.

$$\mathrm{p}(x) = \begin{cases} \Pr(X = x_i) > 0 & x = x_i \in R_X \\ 0 & x \notin R_X \end{cases}$$

R_X è detto supporto della v.a. X

5 Variabili aleatorie discrete

5.1 Variabile aleatoria di Bernoulli

Capita spesso di dover rappresentare un esperimento aleatorio che possa avere due esiti: successo o insuccesso. Per modellare questo tipo di scenario è possibile adoperare la v.a. di Bernoulli, che associa ad ogni successo il valore 1 e ad ogni insuccesso il valore 0

Definizione 17: *Variabile aleatoria di bernoulli*
Una variabile aleatoria di Bernoulli è definita nel seguente modo:

$$x(\omega) \begin{cases} 0 & \text{se insuccesso} \\ 1 & \text{se successo} \end{cases}$$

Nota che $Pr(1) = p$, mentre $Pr(0) = 1 - p$

5.2 Distribuzione binomiale

Una distribuzione binomiale esprime il numero di successi ottenuti in N tentativi

Questa distribuzione di probabilità regola il numero dei successi (o risultati favorevoli) conseguito in una successione (finita) di prove indipendenti. Si supponga che un certo esperimento venga replicato $N \geq 1$ volte e l'esito di ognuno di essi possa essere favorevole (evento A) oppure non favorevole (evento A^c). Ad ogni prova dell'esperimento associamo una v.a. $X_i, i = 1, 2, \cdots, N$ che ne rappresenta l'esito: $X_i = 1$ se si verifica A e $X_i = 0$ se si verifica A^c. Si supponga che le v.a. $X_i, i = 1, 2, \cdots, N$ siano indipendenti e che $\Pr(X_i = 1) = p, 0 \leq p \leq 1$. Qual è la distribuzione di probabilità del numero totale di successi nelle N prove? Cioè qual è la distribuzione di probabilità della v.a.

$$X = X_1 + X_2 + X_3 + \cdots + X_N?$$

Definizione 18: *Distribuzione Binomiale*
Si dice che una v.a. X si distribuisce secondo la distribuzione di probabilità (o legge) binomiale di parametri $N \geq 1$ (intero) e $0 \leq p \leq 1$, se

$$\Pr(X = x) = \begin{cases} \binom{N}{x} p^x (1-p)^{N-x} & x = 0, 1, \cdots, N \\ 0 & \text{altrove} \end{cases}$$

e scriveremo $X \sim \mathrm{Bi}(N, p)$.

Nel caso speciale in cui $N = 1$ la v.a. è chiamata anche v.a. di Bernoulli.

La funzione di probabilità del numero totale di successi ottenuti in N prove indipendenti con probabilità di successo pari a p ad ogni prova è dato dalla distribuzione di probabilità di una v.a. binomiale

Si consideri una particolare realizzazione per cui il numero di successi è x, ad esempio l'allineamento

$$B = \{\underbrace{A, A, \cdots, A}_{x \text{ volte}} \underbrace{A^c, A^c, \cdots, A^c}_{N-x \text{ volte}}\}$$

Poiché le prove sono indipendenti la probabilità di questo allineamento è

$$\Pr(B) = p^x (1-p)^{N-x}$$

e questa è la probabilità di un qualsiasi allineamento con x successi che sono

$$\binom{N}{x}$$

e quindi

$$\Pr(X = x) = \binom{N}{x} p^x (1-p)^{N-x}$$

Teorema 3: *Somma variabili Bernoulliane*
Siano $X_1, X_2, \cdots, X_N, N$ v.a. Bernoulliane Bi $(1, p)$ stocasticamente indipendenti allora la v.a. X cosi definita

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

è tale che $X \sim Bi(N, p)$

Definizione 19: *Funzione di ripartizione*
Sia X una v.a. Si dice funzione di ripartizione della v.a. X la funzione $y = F(x)$ definita per ogni x reale data da:

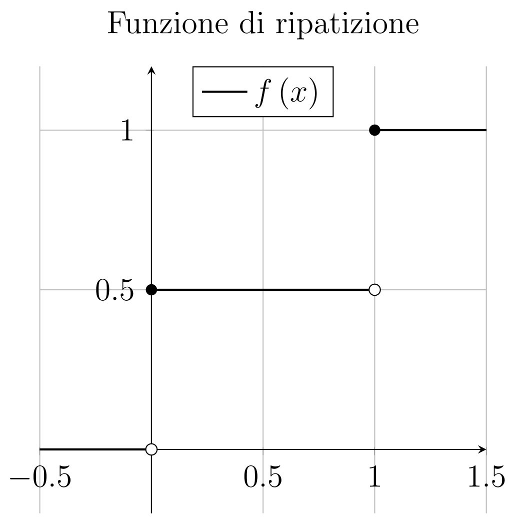
$$F(x) = \Pr(X \leq x) \quad x \in \mathbb{R}$$

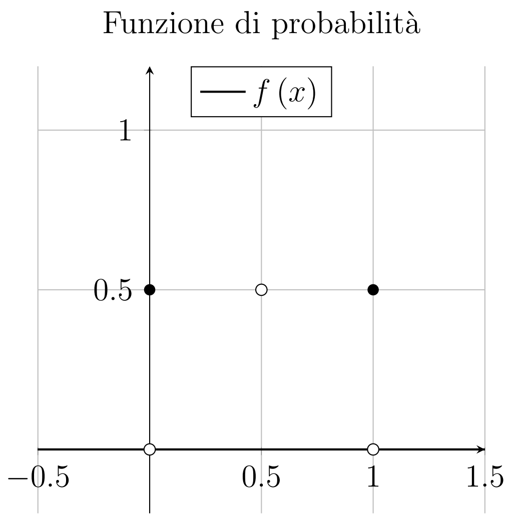
Ad esempio, la funzione di ripartizione di una variabile aleatoria che che esprima la somma del lancio di due dadi sarebbe la funzione che in corrispondenza di ogni x andrebbe esprime la probabilità di ottenere una somma minore di x. Nota che se una v.a. è discreta, posso convertire facilmente da funzione di ripartizione $F(x)$ a funzione di probabilità $p(X = x)$:

$$p(X = x) = F(x) - F\left(x^-\right)$$

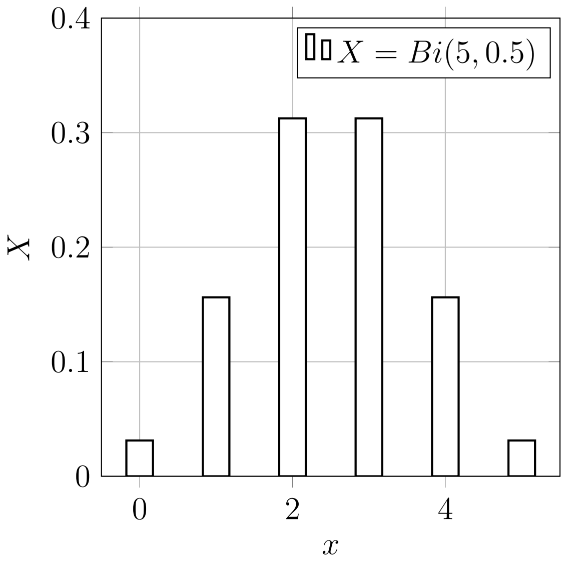
dove $F(x^-)$ è il limite sinistro di F

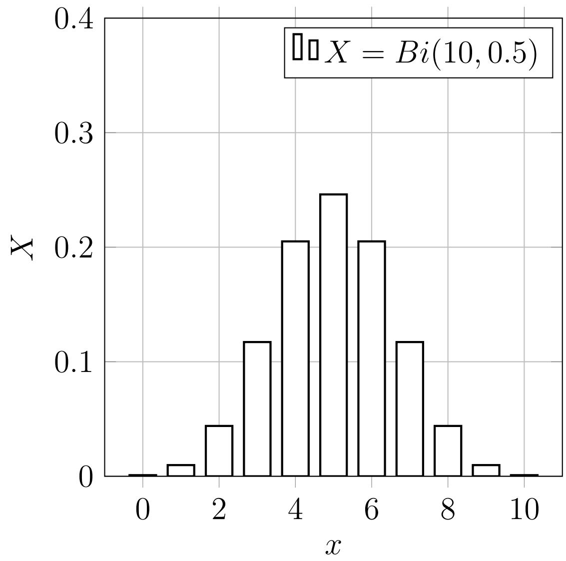
Ad esempio, calcolando la funzione $f(x)$ avro che la sua corrispondende funzione di probabilità sarà:

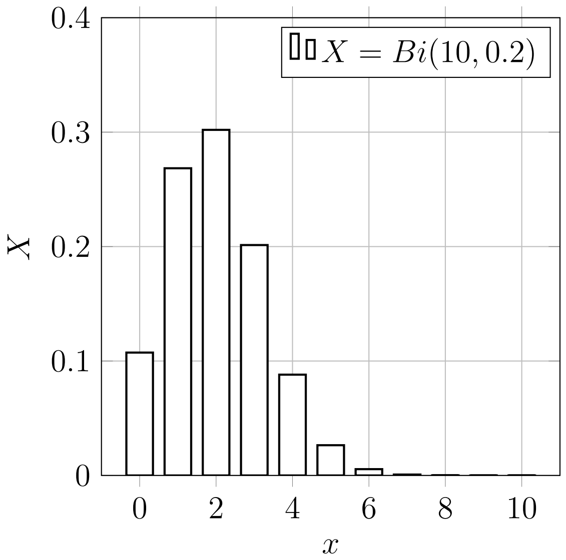


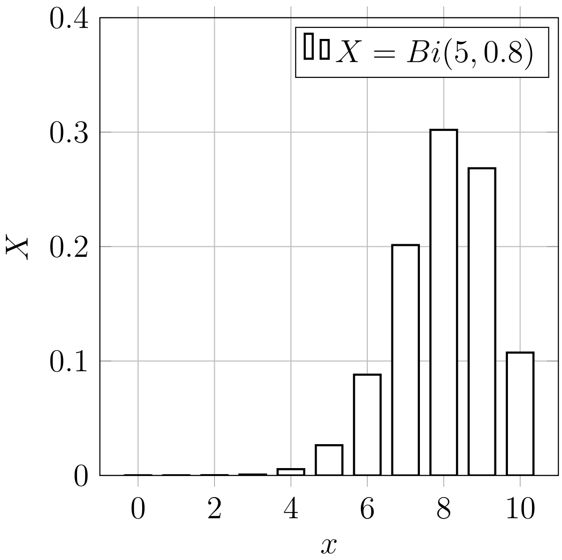


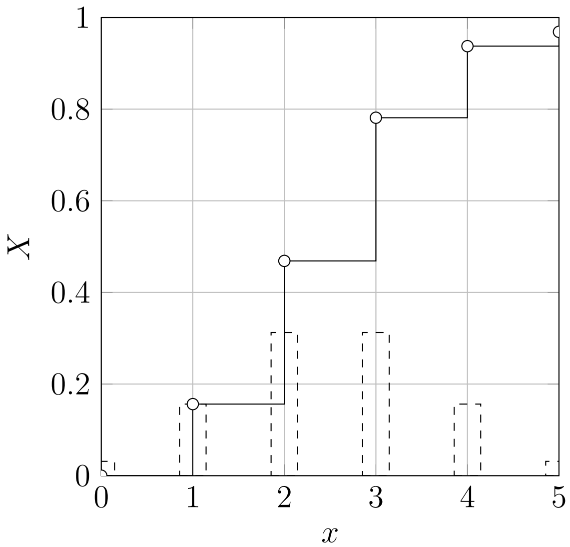
Esempi distribuzioni binomiali

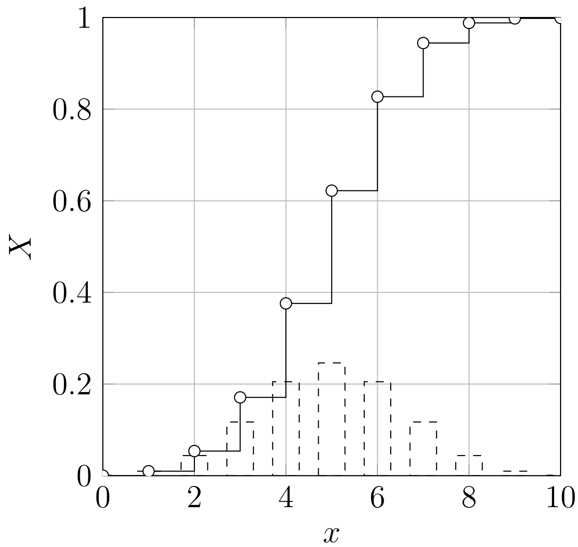


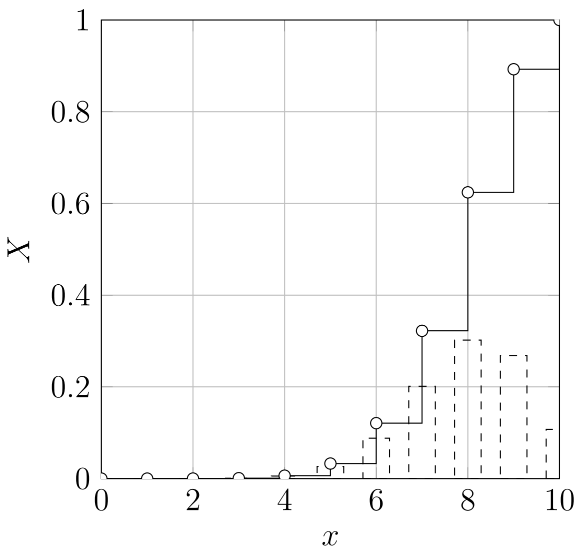


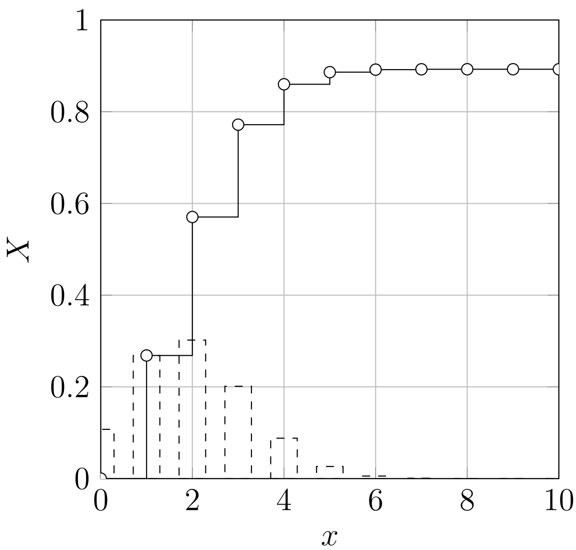

$X = Bi(10, 0.2)$
X
x


$X = Bi(5, 0.8)$
X
x


X
x


X
x


X
x


X
x

5.3 Distribuzione geometrica

Una distribuzione geometrica esprime il numero di lanci necessari al fine di ottenere esattamente almeno un successo

Questa distribuzione è simile alla distribuzione binomiale. Anziché contare il numero di successi in N tentativi, si conta il numero di tentativi prima di ottenere un successo Chiaramente, tutte le prove effettuate devono essere indipendenti

Variabile binomiale	Variabile geometrica
Lanci la moneta n volte e conti il numero di successi ottenuti	Lanci la moneta finchè non ottieni un risultato positivo, e conti quanti lanci hai fatto

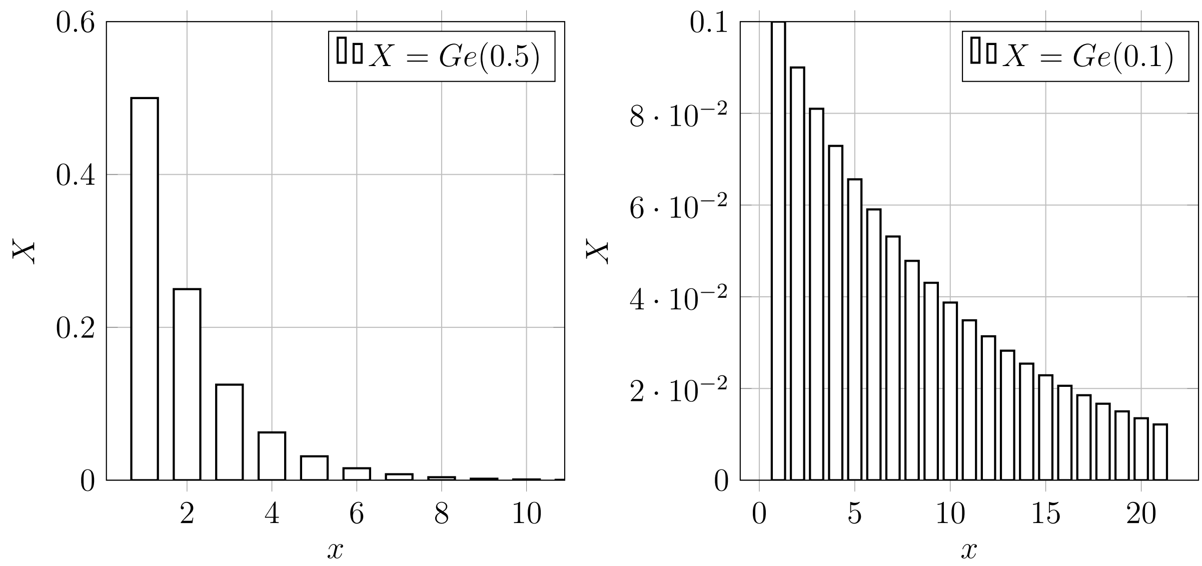
Definizione 20: *Distribuzione Geometrica*
Si dice che una v.a. X si distribuisce secondo una distribuzione geometrica di parametro $0 \leq p \leq 1$ se la sua funzione di probabilità è :

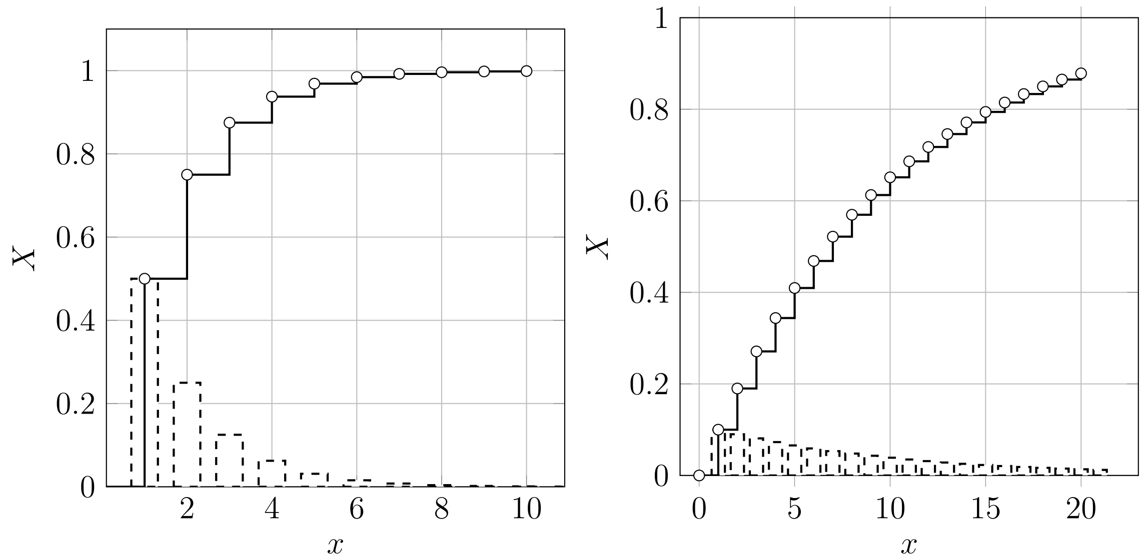
$$Pr\left(X = x\right) = \begin{cases} p\left(1-p\right)^{x-1} & x = 1, 2, 3, \ldots \\ 0 & \text{altrove} \end{cases}$$

scriveremo che $X \sim Ge\left(P\right)$

NB: Talvolta la variabile collega il numero di lanci necessari per ottenere un successo (n). Altre volte la variabile collega il numero di insuccessi prima di ottenere un successo ($n-1$)

Esempi distribuzioni geometriche





5.4 Distribuzione binomiale negativa

Una distribuzione binomiale negativa esprime il numero di lanci necessari al fine di ottenere esattamente r successi

Definizione 21: *Distribuzione binomiale negativa*
Si dice che una v.a. X si distribuisce secondo la distribuzione binomiale negativa di parametri $0 < p \leq 1$ e $r \geq 1$ (intero) se la sua funzione di probabilità è data da:

$$Pr\left(X = x\right) = \begin{cases} \binom{x-1}{r-1} p^r \left(1-p\right)^{x-1} & x = r, r+1, r+2\dots \\ 0 & \text{altrove} \end{cases}$$

e indichiamo con $X \sim BinNe\left(r, p\right)$

Nota che con $r = 1$ la distribuzione coincide con una variabile geometrica.

Teorema 4: *Rapporto distribuzione binomiale e binomiale negativa*
Sia $X \sim BiNe\left(r, p\right)$ e $Z \sim Bi\left(N, p\right)$, allora

$$Pr\left(Z \geq r\right) = \Pr\left(X \leq N\right)$$

Teorema 5: *Rapporto distribuzione geometrica e binomiale negativa*
Siano $X_1, X_2, \dots, X_r$ delle v.a. geometriche stocasticamente indipendenti. Allora la variabile così definita:

$$X = \sum_{i=1}^{r} X_i$$

è una variabile di tipo binomiale negativo: $X \sim BiNe\left(r, p\right)$

5.5 Distribuzione di Poisson

Una distribuzione di Poisson esprime il numero di eventi successi in un determinato lasso di tempo/spazio

Ad esempio, una v.a. di Poisson potrebbe eprimere il numero di camion che passano in un'ora su una data strada

Definizione 22: *Distribuzione di Poisson*
Si dice che una v.a. X si distribuisce secondo la distribuzione di Poisson di parametri $\lambda \geq 0$ se la sua funzione di probabilità è data da:

$$\Pr(X = x) = \begin{cases} \frac{\lambda^x}{x!}e^{-\lambda} & x \geq 0 \\ 0 & \text{altrove} \end{cases}$$

6 Variabili aleatorie continue

Definizione 23: *Densità di probabilità*
se F è assolutamente continua definziamo la densità nel seguente modo:

$$f(x) = F'(x)$$

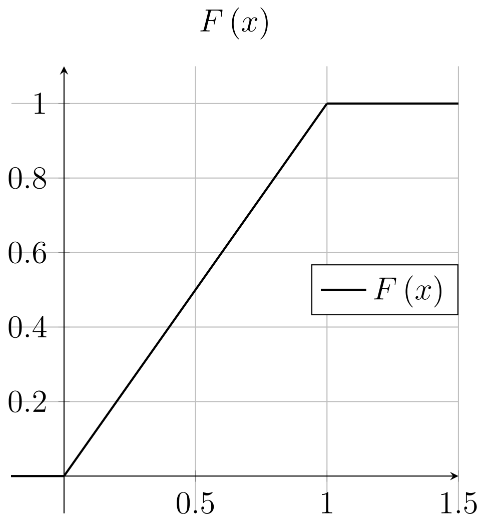
Dove $F(x)$ è la funzione di ripartizione. Per questa ragione la funzione di ripartizione è una primitiva della funzione di densità. La probabilità di un intervallo $(a, b]$ è data da:

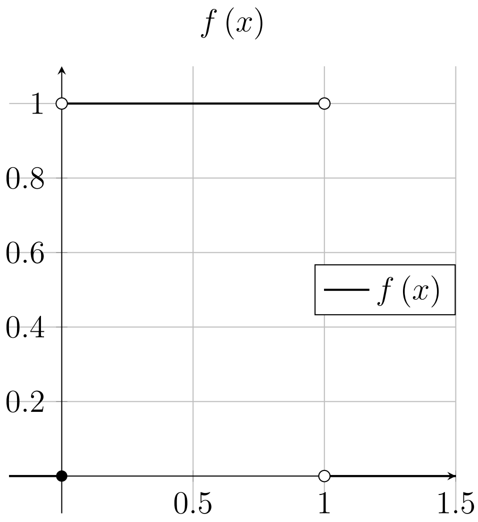
$$\int_a^b f(x)\, dx$$

Nota che la funzione di ripartizione è una primitivà della funzione di densità! L'area del trapezoide della funzione di densità esprime la probabilità del corrispettivo intervallo. La probabilità dell'intervallo $(-\infty, a)$ è data da $\int_{-\infty}^{a} f(x)\ dx$. Tuttavia la probabilità del medesimo intervallo è dato da $F(a)$ dove F è la funzione di ripartizione. Ciò vuol dire che $F'(x) = f(x)$ ossia F è una primitiva f.

Esempio funzione di ripartizione e densità

A sinistra la funzione di ripartizione, a destra la funzione di densità





NB: useremo la teoria dell'integrazione di Riemann, ma di solito si userebbe quella di Lebasgue

Esempio

Sia $\Omega = [0,1]$ e $\mathcal{A} = \mathcal{B}([0,1])$ la Tribù associata. Sia Pr la funzione di probabilità su $\mathcal{A}$ definita dalla funzione $F(x)$

$$F(x) = \begin{cases} 0 & x < 0 \\ x & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

Consideriamo la v.a. $X(\omega) = \omega$ e quindi

$$\begin{aligned} \Pr(a < X \le b) &= \Pr(\{\omega \in [0,1] : a < X(\omega) \le b\}) \\ &= \Pr(\{\omega \in [0,1] : a < \omega \le b\}) \\ &= \Pr((a,b]) \\ &= \begin{cases} b - a = \int_a^b 1dx & 0 \le a < b \le 1 \\ b & a < 0 < b \\ 0 & a < b < 0 \end{cases} \end{aligned}$$

La v.a. X è dotata di densità

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{altrove} \end{cases}$$

Esempio 1.2

Consideriamo sullo stesso spazio la v.a. $Y(\omega) = \omega^2$ allora abbiamo

$$\begin{aligned}
\Pr(a < Y \leq b) &= \Pr(\{\omega \in [0,1] : a < Y(\omega) \leq b\}) \\
&= \Pr\left(\left\{\omega \in [0,1] : a < \omega^2 \leq b\right\}\right) \\
&= \Pr(\{\omega \in [0,1] : \sqrt{a} < \omega \leq \sqrt{b}\}) \\
&= \Pr((\sqrt{a}, \sqrt{b}]) \quad = \begin{cases} \sqrt{a} - \sqrt{b} = \int_a^b \frac{1}{2\sqrt{x}} dx & 0 \leq a < b \leq 1 \\ \sqrt{b} & a < 0 < b \\ 0 & a < b < 0 \end{cases}
\end{aligned}$$

La v.a. Y è dotata di densità

$$f(x) = \begin{cases} \frac{1}{2\sqrt{x}} & 0 \leq x \leq 1 \\ 0 & \text{altrove} \end{cases}$$

6.1 Variabile aleatoria normale o gausiana

Definizione 24: *V.A. Gaussiana*
Si diche che una v.a. X si disttribuisce con legge di probabilità normale (o Gaussiana) di parametri:

- $\mu \in (-\infty, +\infty)$
- $\sigma \in (0, +\infty)$

se possiede la seguente densità

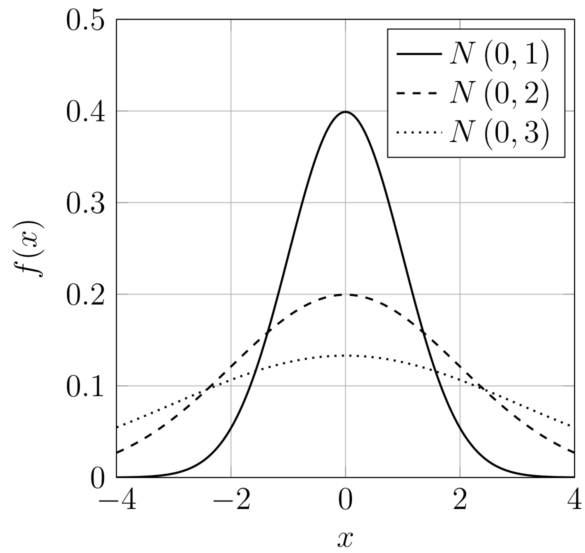
$$f(x, \mu, \omega) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

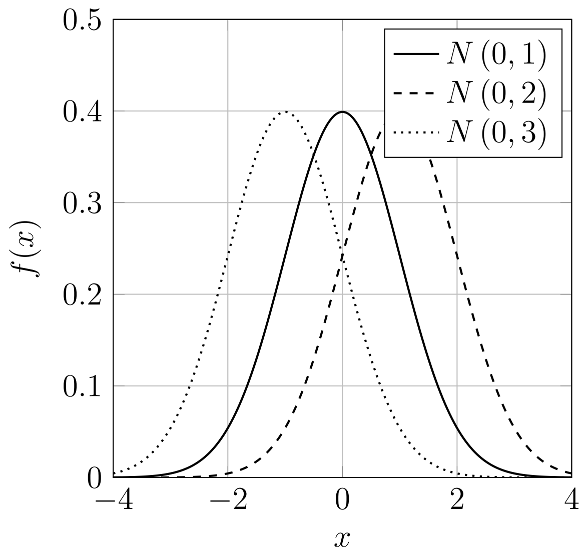
La indicheremo con $X \sim N(\mu, \sigma^2)$

Nota che a differenza delle v.a. discrete, quì stiamo parlando del grafico della densità di probabilità, dunque questultima sarà espressa come area del trapezoide sotteso al suo grafico

La variabile $X \sim N(0,1)$ è chiamata *Normale standard*

Grafici





Come è possibile osservare μ determina il valore al quale è situato il picco della densità, mentre σ regola quanto i risultati sono concentrati intorno a questo picco

Normalizzazione

Visto che la funzione di densità della v.a. *Gaussiana* non è integrabile in forma chiusa, sono disponibili delle tabelle che approssimano l'integrale della normale standard in determinati suoi punti. Se la nostra v.a. non è standard, dobbiamo applicare il processo di normalizzazione. Si può dimostrare che

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

ossia Z è una normale standard Per questa ragione possiamo trasformare la nostra v.a. in una normale standard e calcolare l'area di nostro gradimento:

$$\Pr(\{X \leq c\}) = \Pr\left(\left\{\frac{X-\mu}{\sigma} \leq \frac{c-\mu}{\sigma}\right\}\right)$$

- Voglio calcolare $\Pr(X \leq c)$
- Calcolo $\frac{c-\mu}{\sigma}$
- Vedo sulla tabella l'area della normale standard nell'intervallo $\left(-\infty\frac{c-\mu}{\sigma}\right]$

6.2 Distribuzione Esponenziale

Definizione 25: *Distribuzione esponenziale*
Si diche che una v.a. X ha legge esponenziale con parametro $\lambda > 0$ se la sua funzione di densità ha la forma:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x > \ 0 \\ 0 & \text{altrove} \end{cases}$$

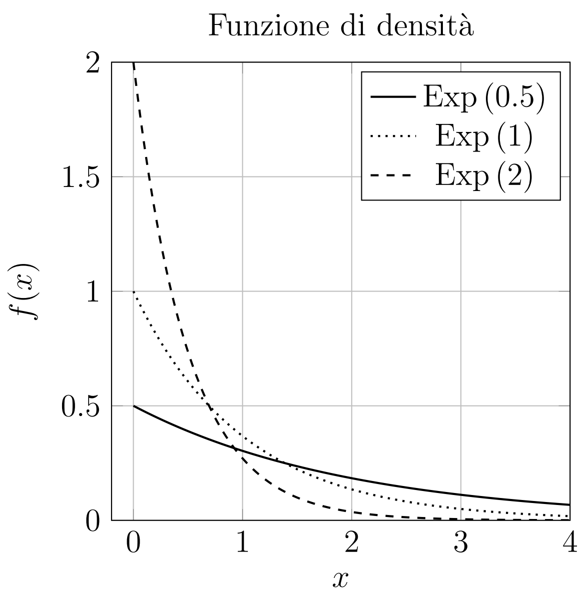
e la indichiamo nel seguente modo $X \sim \text{Exp}(\lambda)$

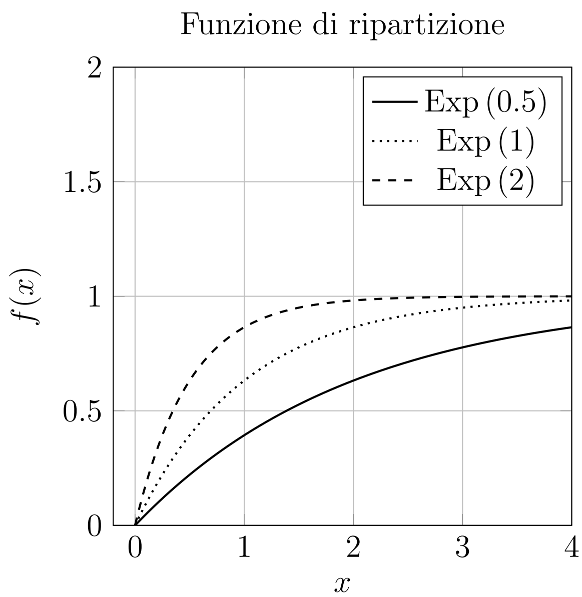
A differenza della v.a. Gaussiana, in questo caso è possibile ricavare l'espressione esplicita della funzione di ripartizione:

$$F(x) = \begin{cases} \int_0^x \lambda e^{-\lambda t}\, dt = 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$\boxed{\text{Come la } \textit{distribuzione geometrica}\text{, anche quella esponenziale } \underline{\text{non ha memoria}}}$

$$\begin{aligned} \Pr(y < X \leq y + x \mid X > y) &= \frac{\Pr(\{y < X \leq y + x\} \cap \{X > y\})}{\Pr(\{X > y\})} \\ &= \frac{\Pr(\{y < X \leq y + x\})}{\Pr(\{X > y\})} \\ &= \frac{e^{-\lambda y} - e^{-\lambda(y+x)}}{e^{-\lambda y}} \\ &= 1 - e^{-\lambda x} = \Pr(X \leq x) \end{aligned}$$





6.3 Cheatsheet e cose che non ho voglia di dimostrare

Variabile	Media	Varianza
Bernoulli $\mathcal{B}(p)$	p	$p(1-p)$
Binomiale $Bi(N, p)$	Np	$Np(1-p)$
Geometrica $\text{Ge}(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson $\mathcal{P}(\lambda)$	λ	λ
Normale $\mathcal{N}(\mu, \sigma^2)$	μ	σ^2
Esponenziale $\text{Exp}(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

7 Scale di misura, moda, quantili, momenti

7.1 Trasformazioni variabili aleatorie

Per alterare come una variabile aleatoria si distribuisce è possibile utilizzare le trasformazioni. Una trasformazione non è altro che una funzione che viene applicata al valore assunto dalla variabile aleatoria. Ad esempio, se condideriamo la funzione $g(x) = x + 2$ e consideriamo:

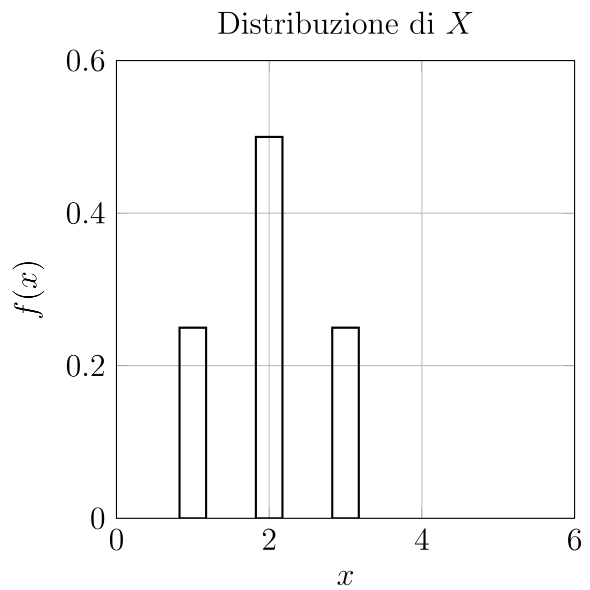
$$\omega = \{ \text{ rosso },\ \text{verde },\ \text{azzurro } \}$$

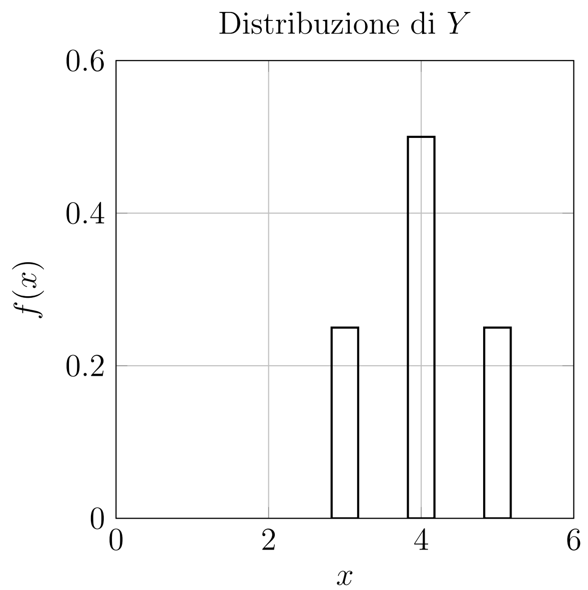
$$X(\text{rosso}) = 1 \quad X(\text{verde}) = 2 \quad X(\text{azzurro}) = 3$$

possiamo applicare la funzione $g(x)$ sulla variabile aleatoria X per ottenere un'altra variabile aleatoria Y:

$$\begin{aligned} Y(\text{rosso}) &= g(X(\text{rosso})) = g(1) = 1 + 2 = 3 \\ Y(\text{verde}) &= g(X(\text{verde})) = g(2) = 2 + 2 = 4 \\ Y(\text{azzurro}) &= g(X(\text{azzurro})) = g(3) = 3 + 2 = 5 \end{aligned}$$

Supponendo che la variabile X si distribuisca come qui a sinistra, la variabile Y si distribuirebbe come qui a destra:





Non tutte le funzioni possono essere applicate come trasformazioni, in quanto deve essere possibile calcolare la probabilità su ogni valore assunto dalla variabile Y trasformata. Per fare ciò è necessario dover poter calcolare la probabilità sul suo insieme controimmagine sei valori assunti da X in corrispondenza del valore assunto da Y. Formalmente:

> **Definizione 26:** *Traformazione di una v.a.*
> Consideriamo una funzione $Y(\omega) = g(X(\omega))$ dove $g : \mathbb{R} \to \mathbb{R}$. Se g è tale che
>
> $$\{x \in \mathbb{R} : g(x) \leq z\} \in \mathcal{B}(\mathbb{R}) \text{ per ogni } z \in \mathbb{R}$$
>
> allora $Y(\omega)$ è una variabile aleatoria

visto che è difficile verificare se questa condizione è verificata, noi useremo il seguente teorema:

Teorema 6: *Funzioni utilizzabili per trasformazioni v.a.*
Una funzione $g(x)$ può essere utilizzata per effettuare una trasformazione di una v.a. se soddifa almeno una delle seguenti conidizoni:

- Continuità
- Monotonia (*crescente o decrescente*)

Quindi formalmente, se poniamo:

$$A(y) = \{x \in \mathbb{R} : g(x) \leq y\}$$

allora abbiamo

$$\Pr(Y(\omega) \leq y) = \Pr(X(\omega) \in A(y))$$

Nel caso discreto abbiamo:

$$\Pr(Y \leq y) = \sum_{\{x:g(x)\leq y\}} p_X(x)$$

mentre nel caso di una v.a. dotata di densità

$$\Pr(Y \leq y) = \int_{\{x:g(x)\leq y\}} f_X(x)dx$$

Teorema 7: *Densità di v.a. trasformata*
Sia X una v.a. con densità $f_X(x)$ con supporto l'intervallo (a, b), eventualmente non limitato. Sia $g(x)$ una funzione strettamente monotona con derivata in (a, b). Allora la v.a. $Y = g(X)$ è dotata di densità

$$f_Y(y) = \begin{cases} f_X\left(g^{-1}(y)\right)\left|D\left[g^{-1}(y)\right]\right| & \alpha < y < \beta \\ 0 & \text{altrove} \end{cases}$$

dove $\alpha = \min(g(a), g(b))$ e $\beta = \max(g(a), g(b))$.

Esempio

Sia X una v.a. esponenziale di parametro λ. Sia $g(x) = e^x$. Allora applicando il teorema precedente abbiamo

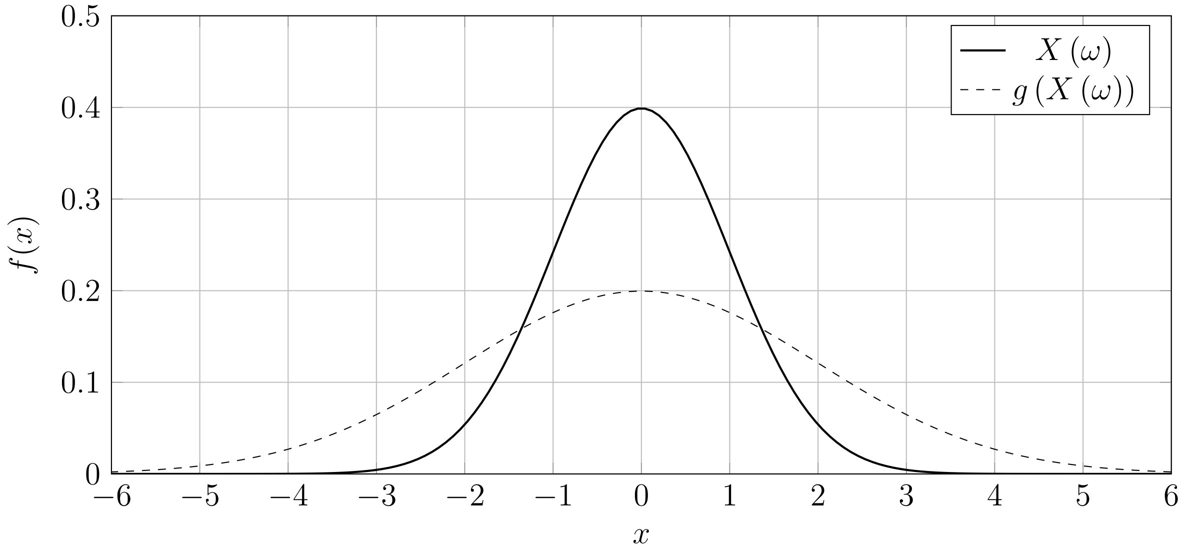
- $\alpha = g(a) = g(0) = 1$
- $\beta = g(b) = g(+\infty) = +\infty$
- $g^{-1}(y) = \log(y)$ per $1 < y < +\infty$
- $D\left[g^{-1}(y)\right] = \frac{1}{y}$

allora

$$f_Y(y) = \lambda \exp(-\lambda \log(y))\frac{1}{y} = \frac{\lambda}{y^{\lambda+1}} \quad 1 < y < +\infty$$

e 0 altrove.

Esempio di una trasformazione di una variabile $X \sim N(0,1)$ tramite la funzione $g(x) = 2x$



Quando il teorema sulle trasformazioni non può essere applicato

Prendiamo come esempio la v.a.

$$X \sim N(0,1)$$

ossia una *normale standard*. Quando la funzione per la quale non è monotona o non ha derivata in un determinato intervallo è possibile comunque calcolarne la probabilità in altri modi. Vediamo un esempio interessante. In seguito affronteremo i momenti. Il momento di ordine α di una v.a. è così definito:

$$u_1 = \mathbb{E}\left(X^2\right)$$

Tuttavia non si può appicare il teorema 7 per trasformare una v.a. tramite la funzione $g(x) = x^2$, in quanto la funzione $g(x)$ non è monotona. Tuttavia, possiamo tenere in mente che:

- La probabilità di un intervallo $(a, b]$ di una variabile trasformata X' è uguale alla probabilità della controimmagine dell'intervallo $(a, b]$ della variabile X
- E' possibile ricavare l'intervallo controimmagine tramite una semplice disequazione, quindi:

$$Pr(X' < a) = \Pr\left(X^2 < a\right) = \Pr\left(X \in \left(-\sqrt{a}, \sqrt{a}\right)\right) = \int_{-\sqrt{a}}^{\sqrt{a}} f(x)\, dx$$

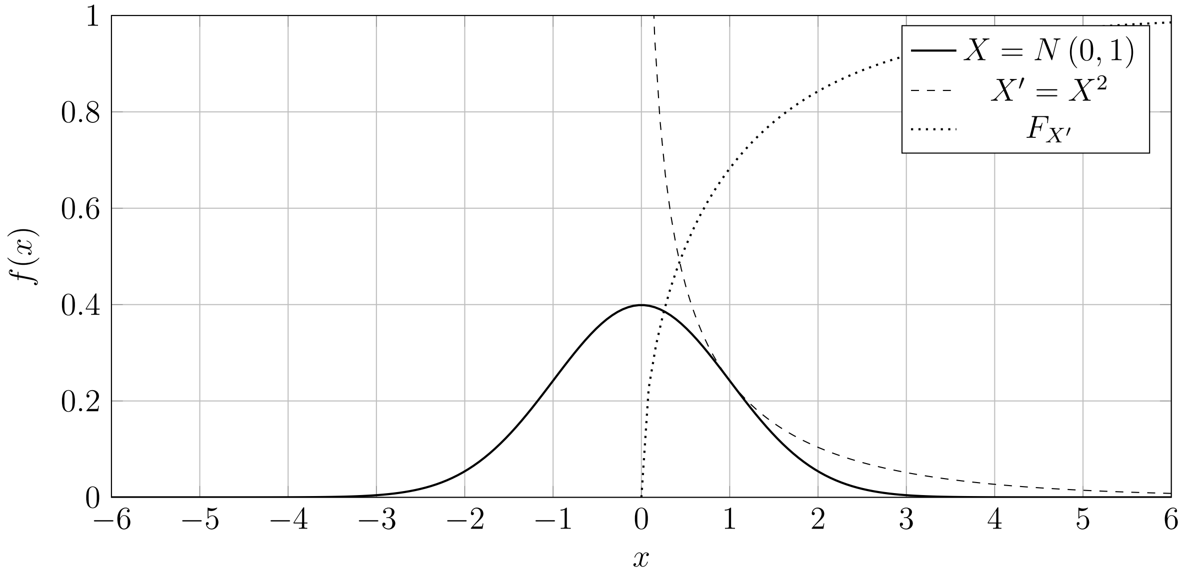
- Per calcolare la forma esplicita della distribuzione della variabile X' posso procedere come segue:

$$\begin{aligned} F_{X'}(a) &= \Pr(X' < a) = \Pr\left(X^2 < a\right) = \Pr\left(X \in \left(-\sqrt{a}, \sqrt{a}\right)\right) \\ &= 2\int_0^{\sqrt{a}} f(x)\, dx \end{aligned}$$

  dove $f(x)$ è la distribuzione di X. Nota che l'ultima eguaglianza è data dal fatto che la funzione di distribuzione di una normale standard è pari

- Sapendo che $f_{X'}(x)$, ossia la distribuzione di X' è la derivata della sua funzione di ripartizione:

$$
\begin{aligned}
f_{X'}(a) &= D\left[f_{X'}(a)\right] = D\left[2\int_0^{\sqrt{a}} f(x)\,dx\right] \\
&= \frac{f\left(\sqrt{a}\right)}{\sqrt{a}}
\end{aligned}
$$



7.2 Scale di misura

Le scale di misura costituiscono i criteri secondo i quali gli esiti di un ipotetico *esperimento aleatorio* vengono categorizzati. Queste sono dei seguenti tipi:

- Qualitativi
  - nominali (categoriali)
  - ordinali
- Quantiti
  - ordinali
  - intervalli
  - rapporti

Scala numinale o categoriale

Esempi:

- colore degli occhi
- i modelli di un determinato prodotto
- il sesso di un individuo

- la religione da questo praticata

in questo caso, un esempio di possibile spazio campionario potrebbe essere:

$$\omega = \{ \text{ verdi },\ \text{azzurri },\ \text{neri },\ \text{marroni } \}$$

Scala ordinale qualitativa

Si parla di scala ordinale nel momento in cui ha senso fare confronti fra i possibili esiti, dicendo che ad esempio esito 1 $<$ esito 2

Esempi:

- Livello di istruzione (*alto, medio, basso*)
- Quanto una torta è dolce (*poco zuccherata, abbastanza zuccherata, molto zuccherata*)

Scala ordinale quantitativa

Simile alla scala ordinale qualitativa, con la differenza che in questo caso le fasce entro cui categorizziamo il dato sono ben definite:

Esempi:

- Classe di reddito annuo ($0 \leq R \leq 5000$, $5000 < R < 10000 \dots$)
- Fascia aliquota irpef

Scala intervallare

Si parla di scala intervallare nel momento in cui ha senso effettuare un confronto fra differenze di esiti ottenuti, ma non esiste uno "zero assoluto".

Esempio:

- Variabile che segna il momento in cui è avvenuto un determinato evento

Ad esempio, se considerassimo calendario gregoriano e mussulmano, un determinato intervallo di tempo rimarrebbe sempre inalterato, tuttavia non avrebbe senso parlare di rapporto, in quanto ad esempio il rapporto fra gli stessi due anni cambierebbe in base al calendario utilizzato

Scala rapporto

Si parla di scala rapporto nel momento in cui esiste uno zero naturale. In questo caso, a differenza che la scala intervallare, ha senso parlare di rapporti fra misure

7.3 Valori di sintesi

Spesso ha senso descrivere il modo in cui si distribuisce una variabile aleatoria indicandone caratteristiche specifiche. Tali caratteristiche vengono definite valori di sintesi. Generalmente, vengono usati i seguenti valori di sintesi:

- moda

- quantili
- momenti (centrari e non centrati)

scala	moda	quantili	momenti
nominale	$\times$		
ordinale	$\times$	$\times$	
intervallo	$\times$	$\times$	$\times$
rapporto	$\times$	$\times$	$\times$

Moda

Definizione 27: *Moda*
Si dice moda, indicata con $Mo(X)$ di una v.a. X l'esito che si presenta più frequentemente

Nel caso di una v.a. discreta, sarà l'esito uscito più volte, mentre nel caso di una v.a. continua sarà il valore di x in corrispondenza del massimo della sua funzione di densità

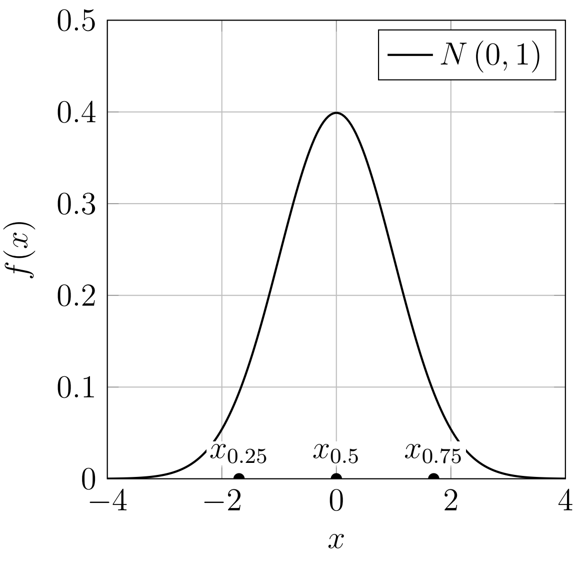
Quantili

I quantili indicana il primo esito x_α tale per cui la probabilità che esca un esito $\leq x_\alpha$ è α. Formalmente:

Definizione 28: *Quantile*
Si dice qualtile di ordine $a \leq \alpha \leq 1$ di una v.a. X l'esito dell'esperimento $Q_\alpha = x_\alpha$ tale che:

$$Pr(X \leq x_\alpha) \geq \alpha \text{ e } \Pr(X \geq x_\alpha) \geq 1 - \alpha$$

Ad esempio considera la seguente distribuzione gaussiana standard:



In particolare, abbiamo:

- centili: $x_{0.001}, x_{0.002}, \ldots, x_1$
- decili: $x_{0.1}, x_{0.2}, \ldots, x_1$
- quartili:
  - *primo quartile* $= q_1 = x_{0.25}$
  - *secondo quartile* $= q_2 = x_{0.5}$
  - *terzo quartile* $= q_3 = x_{0.75}$
  - *quarto quartile* $= q_4 = x_1$

Il *secondo quartile* ha un'importanza fundamentale: questo viene chiamato mediana, in quanto divide in due il grafico della densità: gli eventi a destra della mediana hanno la stessa probabilità di accadere rispetto a quelli a sinistra

Definizione 29: *Mediana*
La mediana di una v.a. corrisponde al suo secondo quartile, ossia l'esito x di un esperimento tale che:

$$Pr(X \leq x) = \Pr(X \geq x)$$

Momenti

L'ultimo valore di sintesi è quello dei momenti. Questo tipo di valore di sintesi permette di ricavarne altri, estremamemnte significativi. Alla base del concetto di momento vi è quello di speranza matematica:

Definizione 30: *Speranza matematica*
Sia Xuna v.a. dotata di densità (*diescreto o continua*). Si chiama speranza matematica di X la quantità (*finita*):

$$\mathbb{E}(X) = \sum_{x \in R_X} x p_X(x)$$

Se X è *discreta*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\, dxx$$

Se X è *continua*

Questo tipo di calcolo equivale ad una sorta di *media ponderata*. Il valore dunque indica il valore attorno al quale sono distribuiti la maggior parte dei risultati

Teorema 8: *Calcolo speranza matematica*
Sia X una v.a. e sia $g(x)$ una funzione che soddisfi i requisiti (teo 6) affinchè questa possa essere usata per effettuare una trasformazione di X. Allora

$$\mathbb{E}(g(X)) = \begin{cases} \sum_{Ry} x f_Y(x)\ dx = \sum_{Rx} g(x) f_X(x)\ dx & \text{se } X \text{ è continua} \\ \int_{Ry} x f_Y(x)\ dx = \int_{Rx} g(x) f_X(x)\ dx & \text{se } X \text{ è continua} \end{cases}$$

Convincitene come segue

Il peso della modalità $g(a)$, nel calcolo della speranza di $g(X)$, è la somma dei pesi delle controimmagini di $g(a)$ tramite la funzione g

E' facile convincersene informalmente. Ragiona così:

- La speranza matematica è una media ponderata fra l'x corrispondente ad un esito e la sua probabilità
- Dunque anzichè trasformare la v.a. per poi calcolarne la speranza matematica, posso "ponderare" con la funzione di distribuzione della v.a. X
- Il valore $g(x)$ avrà un "peso" uguale alla somma degli eventi che trasformati danno, appunto $g(x)$, ad esempio:

  $$\omega = \{-1, 0, 1, 2\} \quad g(-1) = 2 \quad g(0) = 2 \quad g(1) = 2 \quad g(2) = 3$$

  supponendo che tutti gli eventi siano equiprobabili:

  $$\begin{aligned}\mathbb{E} &= g(-1)\Pr(-1) + g(0)\Pr(0) + \ldots + g(2) + \Pr(2) \\ &= 2[\Pr(-1) + \Pr(0) + \Pr(1)] + 3\Pr(2)\end{aligned}$$

  dunque nota come alla fine, il peso che avrà 2 nella media, è la somma dei pesi che avrebbero le controimmagini di 2, il che ha senso

Definizione 31: *Momento non centrato*
Sia X un v.a., si dice momento *non centrato* di ordine r (r intero positivo) il valore:

$$\mu_r = \mathbb{E}(X^r)$$

Definizione 32: *Momento centrato*
Sia X un v.a., si dice momento *centrato* di ordine r (r intero positivo) il valore:

$$\mu_r = \mathbb{E}((X - \mu_1)^r)$$

Vedremo che μ_1 è la media di una v.a. Ciò è significativo in quanto la applicando la trasformazione $X - \mu_1$ *trasliamo* la densità della v.a. facendo si che la sua media cada in $x = 0$

Valori di sinetesi basati sui momenti

I seguenti valori di sintesi di basano sui momenti:

- Media $\mu = \mu_1 = \mathbb{E}(\mathbb{X})$
- Varianza: $\mathrm{X} = \sigma^2 = \overline{\mu}_2 = \mathbb{E}\left((X - \mu_1)^2\right)$
- Deviazione standard: $\sigma = \sqrt{\sigma^2}$
- Coefficiente di variazione: $CV = \frac{\sigma}{\mu}$
- Indice di Asimmetria: $p_3 = \dfrac{\overline{\mu_3}}{\sigma^3}$
- Indice di curtosi: $p_4 = \dfrac{\overline{\mu_4}}{\sigma^4} - 3$

7.4 Proprietà media e varianza

Proprietà operatore media

- *Caso base*: sia X la v.a. che assume valore c con probabilità 1, allora:

$$\mathbb{E}(X) = c$$

- *"Linearità"*:

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

- *Prodotto fra variabili indipendenti*: siano X e Y indipendenti:

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

Proprietà operatore varianza

- *Scomposizione varianza*

$$\operatorname{var}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 = \mu_2 - \mu_1^2$$

- *"Fattore neutro"*: aggiungere una costante ad una v.a. non ne cambia la varianza:

$$\operatorname{var}(X + b) = \operatorname{var}(X)$$

- *"Linearità "* attenzione al fatto che una costante che moltiplica va portata fuori facendone il quadrato:

$$\operatorname{var}(aX + bY) = a^2 \operatorname{var}(X) + b^2 \operatorname{var}(Y)$$

8 Variabili aleatorie doppie

Definizione 33: *Variabile aleatoria doppia*
Sia $(\omega, \mathcal{A}, \Pr)$ uno spazio *probabilizzato*. Siano $X(\omega)$ e $Y(\omega)$ due v.a. definite su ω in modo che:

$$Z(\omega) = (X(\omega), Y(\omega)) : \omega \to \mathbb{R}^2$$

ossia all'esito di un esperimento viene associata la coppia ordinata $(X(\omega), Y(\omega))$. Una v.a. doppia Z ha funzione di ripartizione:

$$F_Z(z) = F_{X,Y}(x, y) = \Pr(\omega : \{X \le x\} \cap \{Y \le y\})$$

Nel caso di una v.a. discreta, la probabilità dell'evento (x, y) è data dalla probabilità che avvengano sia x che y

8.1 V.a. doppie discrete

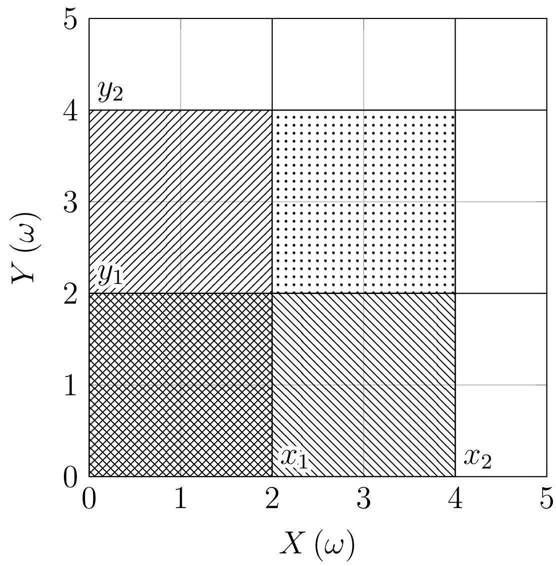
Definizione 34: *Probabilità di v.a. doppia discreta*
per due v.a. discrete X e Y, la v.a. doppai $Z = (X, Y)$ (*che è discreta*) ha funzione di probabilità:

$$p_Z(z) = \begin{cases} P_{X,Y}(x, y) = \Pr(\omega : \{X(\omega) = x\} \cap \{Y(\omega) = y\}) & (x, y) \in R_{X,Y} \\ 0 & \text{altrove} \end{cases}$$

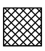
Nota che vale la seguente:

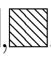
$$\Pr(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1)$$

ciò si può visualizzare graficamente, mettendo sull'asse delle ascisse gli esiti della v.a. X, mentre sull'asse y gli esiti della v.a. Y:



Nota come le aree costituiscano le diverse probabilità:

-  costituisce $\Pr(x_1 < X \leq x_2, y_1 < Y \leq y_2)$
-  costituisce $\Pr(X \leq x_1, Y \leq y_2)$
-  costituisce $\Pr(X \leq x_2, Y \leq y_1)$
-  costituisce $\Pr(X \leq x_1, Y \leq y_1)$

quindi dobbiamo togliere dall'area totale i due rettangoli verticali ed orizzontali ,, e riaggiungere il quadrato piccolo in fondo , dato che è stato tolto due volte

Infine, se è nota la funzione di probabilità congiunta $p_{X,Y}(x, y)$ allora è facile ottenere le funzioni di probabilità delle v.a. X e Y

$$p_X(X = x) = \sum_{y \in R_Y} p_{X,Y}(x, y) \quad p_Y(Y = y) = \sum_{x \in R_X} p_{X,Y}(x, y)$$

e quindi

$$F_X(x) = \sum_{u \leq x} \sum_{y \in R_Y} p_{X,Y}(u, y) \quad F_Y(y) = \sum_{v \leq y} \sum_{x \in R_X} p_{X,Y}(x, v)$$

8.2 V.a. doppie continue

Definizione 35: *Probabilità di v.a. doppia discreta*
La v.a. doppia $Z = (X, Y)$ si dirà dotata di densità se esiste una funzione $f_{X,Y}(x, y)$ tale che - $f_{X,Y}(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx dy = 1$$

A questo punto, la probabilità sarà così definita:

$$\Pr(a < x \leq b, c < y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y) dx dy$$

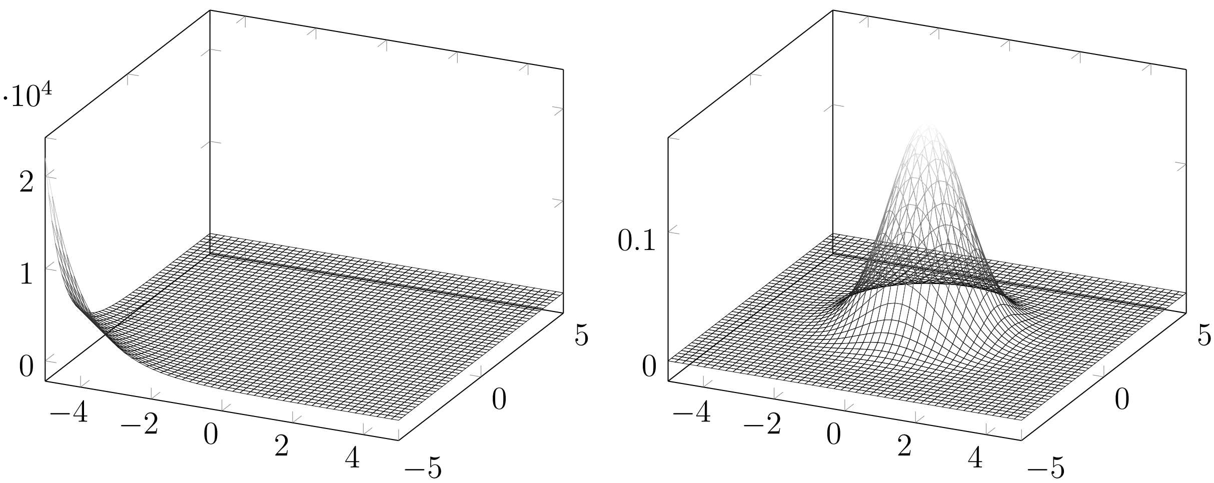
Tale funzione è chiamata densità congiunta $f_Z(z) = f_{X,Y}(x, y)$. Da cui abbiamo

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) du dv$$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, v) dv \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(u, y) du$$

Intuitivamente, la probabilità che $X \in (a, b)$ e $Y \in (c, d)$ è uguale al volume del "trapezoide" con base *abcd* sotteso al grafico della densità di probabilità

Esempi funzioni di probabilità doppie



8.3 V.a. condizionata

Definizione 36: *Distribuzione condizionata di una v.a.*
Sia $(\omega, \mathcal{A}, Pr)$ uno spazio probabilizzato. Sia $X(\omega)$ e $Y(\omega)$ due v.a. discrete. Allora la funzione di probabilità sullo spazio di arrivo è così definita:

$$P_{X|Y}(x|y) = \frac{P(\{X = x\} \cap \{Y = y\})}{P(\{Y = y\})} = \frac{P_{x,y}(x, y)}{P_y(y)}$$

Nota che possiamo definire una v.a. nel modo seguente:

Sia $(X, Y) \sim P_{x,y}(x, y)$ discreta. Allora

$$\forall y \in R_y \quad Z_y = X|y \sim P_{x|y}(x|y)$$

ossia $P_{z_y}(x) = P(X = x|Y = y)$. L'idea è quella di fissare un valore di y, e costruire una funzione di probabilità che associa ad ogni valore di x la probabilità di $x|y$. Chiaramente questa è una v.a. in quanto la somma delle probabilità è 1:

$$\sum \frac{P(x) \cap P(y)}{P(y)} = 1$$

8.4 Esempio

Consideriamo le seguenti v.a. che riguardano il lancio di un dado:

$$X(\omega) = \omega = \text{ numero del dado uscito}$$

$$Y(\omega) = \begin{cases} 0 & \text{esce numero dispari} \\ 1 & \text{esce numero pari minore di 4} \\ 2 & \text{esce numero pari maggiore o uguale a 4} \end{cases}$$

La funzione di probabilità $p_{x,y}(x, y)$ associa ad ogni numero coppia ordinata la probabilità dell'insieme intersezione fra x e y:

	$Y(\omega)$		
$X(\omega)$	0	1	2
1	$\frac{1}{6}$	0	0
2	0	$\frac{1}{6}$	0
3	$\frac{1}{6}$	0	0
4	0	0	$\frac{1}{6}$
5	$\frac{1}{6}$	0	0
6	0	0	$\frac{1}{6}$

Tabella 1: Funzione di probabilità congiunta

$X(\omega)$	$Y(\omega)$ 0	1	2	$F_X(x)$
1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{2}{6}$
3	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{3}{6}$	$\frac{3}{6}$
4	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{4}{6}$
5	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	$\frac{5}{6}$
6	$\frac{3}{6}$	$\frac{4}{6}$	1	1
$F_Y(y)$	$\frac{3}{6}$	$\frac{4}{6}$	1	/

Tabella 2: Ripartizione probabilità congiunta

$X(\omega)$	$Y(\omega)$ 0	1	2
1	$\frac{1}{3}$	0	0
2	0	1	0
3	$\frac{1}{3}$	0	0
4	0	0	$\frac{1}{2}$
5	$\frac{1}{3}$	0	0
6	0	0	$\frac{1}{2}$

Tabella 3: Probabilità condizionata

In quest'ultimo caso, pensa di fissare un valore di Y, considerando dunque $Z_y \sim P(x|y)$ dove y è fissato. Questa è una v.a. (se sommi in colonna le probabilità otterrai chiaramente 1 per ogni colonna). Ora posso calcolare la varianza e media delle v.a. Z_y:

$$
\begin{aligned}
\mathbb{E}(Z_0) &= 1 \cdot \frac{1}{3} + 2 \cdot 0 + 3 \cdot \frac{1}{3} + 4 \cdot 0 + 5 \cdot \frac{1}{3} + 6 \cdot 0 && = 3 \\
\mathbb{E}(Z_1) &= 2 \cdot 1 && = 2 \\
\mathbb{E}(Z_2) &= 4 \cdot \frac{1}{2} + 5 + 6 \cdot \frac{1}{2} && = 5 \\
\mathrm{Var}(Z_0) &= 1^2 \cdot \frac{1}{3} + 2^2 \cdot 0 + 3^2 \cdot \frac{1}{3} + 4^2 \cdot 0 + 5^2 \cdot \frac{1}{3} + 6^2 \cdot 0 - 3^2 && = \frac{8}{3} \\
\mathrm{Var}(Z_1) &= 2^2 \cdot 1 - 2^2 && = 0 \\
\mathrm{Var}(Z_1) &= 4^2 \cdot \frac{1}{2} + 6^2 \cdot \frac{1}{2} - 5^2 && = 1
\end{aligned}
$$

8.5 Variabile aleatoria definita dalla speranza matematica

Data una v.a. condizionale, possiamo definire una seconda variabile aleatoria data da:

$$Z = \mathbb{E}(X|Y = y)$$

	$Y(\omega)$		
$X(\omega)$	0	1	2
1	$\frac{1}{3}$	0	0
2	0	1	0
3	$\frac{1}{3}$	0	0
4	0	0	$\frac{1}{2}$
5	$\frac{1}{3}$	0	0
6	0	0	$\frac{1}{2}$

Ad esempio, supponendo che l'esperimento aleatorio abbia dato come esito la coppia $(3, 0)$, allora la variabile aleatoria Z collegherebbe questo esito alla media della v.a. $X|Y = 0$, ossia $\mathbb{E}(X|Y = 0) = 3$. Tale variabile viene anche chiamata $\mathbb{E}(X|Y)$

La media di $\mathbb{E}(X|Y)$ gode della seguente proprietà:

Teorema 9: *Tower property*
Date due v.a. X e Y e la loro funzione di probabilità condizionata, allora:

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$$

dove $\mathbb{E}(X|Y)$ è la v.a. $\mathbb{E}(X|Y = y)$

Dimostrazione

$$\begin{aligned}
\mathbb{E}(\mathbb{E}(X \mid Y)) &= \sum_{y\in R_Y}\left[\sum_{x\in R_x} x p_{X|Y}(x \mid y)\right] p_Y(y) \\
&= \sum_{y\in R_Y}\sum_{x\in R_x} x p_{X|Y}(x \mid y) p_Y(y) \\
&= \sum_{y\in R_Y}\sum_{x\in R_x} x p_{X,Y}(x, y) \\
&= \sum_{x\in R_x} x \sum_{y\in R_Y} p_{X,Y}(x, y) \\
&= \sum_{x\in R_x} x p_X(x) = \mathbb{E}(X)
\end{aligned}$$

Teorema 10: *Teorema di composizione della varianza*
Date due v.a. X e Y e la loro funzione di probabilità condizionata, allora:

$$\mathrm{Var}(X) = E(\mathrm{Var}(X|Y)) + \mathrm{Var}(E(X|Y))$$

dove $\mathbb{E}(X|Y)$ è la v.a. $\mathbb{E}(X|Y = y)$

Dimosrazione

Partiamo dalla definizione di varianza:

$$Var(X) = \sum_{x\in\mathbb{R}x} (x - \mathbb{E}(x))^2 p_x(x)$$

$$
\begin{aligned}
\mathrm{Var}(X) &= \sum_{x\in R_X} (x-\mathbb{E}(X))^2 p_X(x) \\
&= \sum_{x\in R_X} (x-\mathbb{E}(X))^2 \sum_{x\in A_Y} p_{XY}(x,y) \\
&= \sum_{x\in R_X}\sum_{y\in R_Y} (x-\mathbb{E}(X|Y=y)+\mathbb{E}(X|Y=y)-\mathrm{E}(X))^2 p_{X,Y}(x,y) \\
&= \sum_{x\in R_X}\sum_{y\in R_Y} \big(x-\mathbb{E}(X|Y=y)^2 p_{X,Y}(x,y) \\
&+ \sum_{x\in R_X}\sum_{y\in R_Y} \left(\mathbb{E}(X)Y=y)-\mathbb{E}(X)\right)^2 p_{X,Y}(x,y) \\
&+ 2\sum_{x\in R_X}\sum_{y\in R_Y} (x-\mathbb{E}(X|Y=y))\,(\mathbb{E}(X|Y=y)-E(X))p_{X,Y}\,(x,y) \\
&= \sum_{y\in R_Y}\left[\sum_{x\in R_X} \big(x-\mathbb{E}\,(X|Y=y)^2\, p_{X|Y}(x|y)\big)\right] p_Y(y) \\
&+ \sum_{y\in R_Y}\left[\sum_{x\in R_X} (\mathbb{E}(X|Y=y)-\mathbb{E}\,(\mathbb{E}\,(X|Y)))^2\, p_{X|Y}(x|y)\right] p_Y(y) \\
&+ 0 \\
&= \mathbb{E}(\mathrm{Var}(X|Y)) + \mathrm{Var}(\mathbb{E}(X|Y))
\end{aligned}
$$

Rimane da mostrare

$$
2\sum_{x\in R_X}\sum_{y\in R_Y} (x-\mathbb{E}(X|Y=y))(\mathbb{E}\{X|Y=y)-\mathbb{E}\{X)\}p_{X,Y}(x,y) = 0
$$

infatti

$$
\begin{aligned}
&= \sum_{x\in R_X}\sum_{y\in R_Y} (x-\mathbb{E}(X|Y=y)\rangle(\mathbb{E}(X|Y=v)-\mathbb{E}(X))p_{X,Y}\,(x,y) \\
&= \sum_{y\in R_Y} (\mathbb{E}(X|Y=y)-\mathbb{E}(X))\left[\sum_{x\in R_X} (x-\mathbb{E}(X|Y=y))p_{X|Y}(x|y)\right] p_Y(y) \\
&= \sum_{y\in R_Y} (\mathbb{E}(X|Y=y)-\mathbb{E}(X))\left[\sum_{x\in R_X} xp_{X|Y}\,(x|y)-\mathbb{E}(X|Y=y))\right] p_Y(y) \\
&= \sum_{y\in R_y} (\mathbb{E}(X|Y=y)-\mathbb{E}(X))\,[\mathbb{E}(X|Y=y)-\mathbb{E}(X|Y=y)]\,p_Y(y)
\end{aligned}
$$

Applicazione teorema

Ritornando all'esericio in sub 8.4

$$\begin{aligned}
\mathbb{E}(\mathrm{Var}(X \mid Y)) &= \left[\frac{8}{3}\cdot\frac{1}{2} + 0\cdot\frac{1}{6} + 1\cdot\frac{1}{3}\right] = \frac{8}{6} + \frac{2}{6} = \frac{10}{6} \\
\mathrm{Var}(\mathrm{E}(X \mid Y)) &= \left[3^2\cdot\frac{1}{2} + 2^2\cdot\frac{1}{6} + 5^2\cdot\frac{1}{3}\right] - \mathbb{E}\left(\mathbb{E}\left(X|Y\right)\right)^2 \\
&= \left[\frac{81}{6}\right] - \mathbb{E}\left(X\right)^2 = \frac{27}{2} - \frac{49}{4} = \frac{5}{4} \\
\mathrm{Var}(X) &= \left[(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2\right]\frac{1}{6} = \frac{35}{12}
\end{aligned}$$

Se non avessimo conosciuto la distribuzione di X, avremmo potuto utilizzare il teorema 10:

$$\mathrm{Var}\left(X\right) = \frac{11}{6} + \frac{5}{4} = \frac{35}{12}$$

8.6 Misure di dipendenza fra v.a.

Si possono definire degli indici per quantificare quanto e come una variabile aleatoria X influenzi il risultato di una variabile aleatoria Y

Definizione 37: *Indipendenza stocastica fra v.a.*
(X, Y) sono stocasticamente indipendenti se

$$P_{X,Y}\left(x,y\right) = P_X\left(x\right)P_Y\left(y\right) \quad \forall\left(x,y\right) \in R_X \times R_Y$$

Nota che se due variabili sono stocasticamente indipendenti vale che:

$$P\left(X|Y\right) = \frac{P_{X,Y}\left(x,y\right)}{P\left(y\right)} = \frac{P_X\left(x\right)\cdot P_Y\left(y\right)}{P_Y\left(y\right)} = P_X\left(x\right)$$

Definizione 38: *Indipendenza in media*
Date (X, Y) si dice che X è indipendente in media da Y se :

$$\mathbb{E}\left(X|Y=y\right) = \mathbb{E}\left(X\right)\forall y \in R_Y$$

Mentre l'indipendenza stocastica indica che un evento y non influenza in alcun modo un evento x, l'indipendenza in media indica che un evento y non influenza la media nel verificarsi di x dato y

Nota bene che:

- A differenza dell'indipendenza stocastica, l'indipendenza in media non è simmetrica: il fatto che X sia indipendente in media da Y non implica che Y sia indipendente in media da X
- Se due v.a. sono stocasticamente indipendenti, allora lo sono anche in media

Inoltre è possibile quantificare quanto la media di una variabile X sia influenzata da quella di un'altra Y tramite l'*indice di dipendenza in media*

Definizione 39: *Rapporto di correlazione*
Sia (X, Y) una v.a. doppia discreta, si chiama rapporto di correlazione di X dato Y

$$n^2_{X|Y} = \frac{\text{Var}\left(\mathbb{E}\left(X|Y\right)\right)}{\text{Var}\left(X\right)} = 1 - \frac{\mathbb{E}\left(\text{Var}\left(\mathbb{E}\left(X|Y\right)\right)\right)}{\text{Var}\left(X\right)} \qquad \text{Var}\left(X\right) > 0$$

in modo analogo si definisce $n^2_{Y|X}$

Dalla formula della scomposizione della varianza è facile vedere che

$$0 \leq \eta^2_{X|Y} \leq 1$$

inoltre

- se $\eta^2_{X|Y} = 0$ allora X è indipendente in media da Y;
- se $\eta^2_{X|Y} > 0$ allora X è dipendente in media da Y;
- $\eta^2_{X|Y} = 1$ se e solo se $\Pr(X = \mathbb{E}(X \mid Y)) = 1$.

Definizione 40: *Covarianza e correlazione*
La covarianza e la correlazione sono altri due indici di dipendenza (lineare) tra due v.a..

$$\mathbb{C}ov(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

mentre

$$\rho(X, Y) = \frac{\mathbb{C}ov(X, Y)}{\sqrt{\mathbb{V}ar(X) \cdot \mathbb{V}ar(Y)}}$$

Teorema 11: *Varianza di una combinazione lineare di v.a.*
Sia (X, Y) una v.a. doppia e a e b due costanti. Allora

$$\mathbb{V}ar(aX + bY) = a^2\mathbb{V}ar(X) + b^2\mathbb{V}ar(Y) + 2ab\mathbb{C}ov(X, Y)$$

9 Teoremi limite della probabilità

9.1 La legge dei grandi numeri

Immaginiamo di lanciare un dado n volte. Intuitivamente avremo che il numero di volte che è uscito il numero 3, fratto il numero di lanci, per n "molto grande", convergerà alla probabilità che esca 3. Questo è proprio il ragionamento intuitivo alla base della legge dei grandi numeri

Definizione 41: *Convergenza in probabilità (o debole)*
Sia S una v.a. binomiale (che conta il numero di successi in n tentativi). Sia p la probabilità del successo. Allora si dice che per $n \to \infty$ la v.a. converge in probabilità, ossia accade che per ε piccolo a piacere:

$$\lim_{n \to \infty} \Pr \left\{ \left| \frac{S_n}{n} - p \right| \geq \epsilon \right\} = 0$$

In simboli si scrive che $\frac{S_n}{n} \xrightarrow{p} p$

Si può dire equivalentemente che $Z_i \xrightarrow{p} Z$ se:

$$\lim_{i \to \infty} Pr\left(|Z_i - Z| > \epsilon\right) = 0$$

per ogni ϵ piccolo a piacere

Allo stesso modo, se fossimo interessati agli scotamenti al quadrato, potremmo ricorrere al concetto di convergenza in media quadratica

Definizione 42: *Convergenza in media quadratica*
Sia S una v.a. binomiale (che conta il numero di successi in n tentativi). Sia p la probabilità del successo. Allora si dice che per $n \to \infty$ la v.a. converge in media, ossia accade che per ε piccolo a piacere:

$$\lim_{n \to \infty} \mathbb{E}\left[\left(\frac{S_n}{n} - p \right)^2 \right] = 0$$

In simboli si scrive che $\frac{S_n}{n} \xrightarrow{m.q.} p$

Anche in questo caso si può scrivere che $Z_i \xrightarrow{m.q} Z$ se

$$\lim_{n \to \infty} \mathbb{E}\left[(Z_n - Z)^2 \right] = 0$$

Teorema 12: *Convergenza in media e in probabilità*
La convergenza in media quadratica implica la convergenza in probabilità, ma non viceversa:

$$Y_n \xrightarrow{m.q.} Y \Rightarrow Y_n \xrightarrow{P} Y$$

Somma variabili aleatorie e teoremi utili in statistica

Teorema 13: *Media e varianza della maedia campionaria*
Siano $Y_1, \dots, Y_n$ v.c. indipendenti, tutte con valore atteso μ e varianza σ^2 e sia $\bar{Y}_n = \sum_{i=1}^n Y_i/n$. Allora

$$\mathbb{E}\left(\bar{Y}_n\right) = \sum_{i=1}^n \frac{\mathbb{E}\left(Y_i\right)}{n} = n\frac{\mu}{n} = \mu,$$

$$\operatorname{Var}\left(\bar{Y}_n\right) = \sum_{i=1}^n \frac{\mathbb{V}ar\left(Y_i\right)}{n^2} = n\frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Teorema 14: *Legge debole dei grandi numeri*
Sia $Y_1, Y_2, \ldots$ una successione di v.c. indipendenti, ciascuna con valore atteso μ e varianza σ^2. Sia $\bar{Y}_n = \sum_{i=1}^{n} Y_i/n$, allora, per ogni $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr\left\{\left|\bar{Y}_n - \mu\right| \geq \varepsilon\right\} = 0,$$

ovvero $\bar{Y}_n \xrightarrow{P} \mu$.

Questo esprime il seguente concetto intuitivo: se ripetiamo n volte un esperimento che ha probabilità p di dare successo, allora la media campionaria tenderà a p per n molto grande, ossia:

$$\frac{\text{numero successi}}{\text{numero tentativi}} \to p \qquad \text{per } n \text{ molto grande}$$

Dimostrazione

$$\mathbb{E}\left[(Y-\mu)^2\right] = \text{Var}\left(\overline{Y}_n\right) = \frac{\sigma^2}{n} \to 0$$

Teorema del limite centrale

Teorema 15: *Teorema del limite centrale*
Sia $Y_1, Y_2, \ldots$ una successione di v.c. ciascuna con $\mathbb{E}\left(Y_i\right) = \mu$ e $\mathbb{V}ar\left(Y_i\right) = \sigma^2$. Allora, posto $Z_n = \sqrt{n}\left(\bar{Y}_n - \mu\right)/\sigma$, per ogni z

$$\lim_{n\to\infty} \Pr\left(Z_n \leq z\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt.$$

In simboli il teorema del limite centrale si denota scrivendo

$$Z_n = \sqrt{n}\frac{\left(\bar{Y}_n - \mu\right)}{\sigma} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

dove $\xrightarrow{\mathcal{D}}$ si legge "converge in distribuzione". Una lettura grezza (ma pratica) del teorema del limite centrale è la seguente:

$$\bar{Y}_n \overset{a}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

dove $\overset{a}{\sim}$ significa "distribuita approssimatamente".

In poche parole, si intende che per valori di n molto grandi, una v.a. che rappresenta la media di una distribuzione binomiale può essere approssimata con una distribuzione normale di parametri μ e $\frac{\sigma^2}{n}$

Approssimazione binomiale con v.a. normale

Se sono verificare le seguenti condizioni posso approssimare una v.a. binomiale con una normale:

- n è abbastanza grande

- p e $1-p$ sono piuttosto lontani da 0
- L'intervallo $\left[np \pm \sqrt{np\,(1-p)}\right]$ è contenuto nell'intervallo $[0, n]$

Teorema 16: *Approssimazione della binomiale con una normale*
Una v.a. $X \sim \mathcal{B}(n, p)$ si può approssimare con una v.a. normale con parametri di valore :

$$\mu = np \quad \sigma^2 = np\,(1-p)$$

Nel calolo del valore standardizzato devo aggiungere la cosiddetta correzione per continuità $= 0.5$, che migliora la precisione dell'approssimazione

Ad esempio data $X \sim \mathcal{B}(n, p)$, se dovessi calcolare $\Pr(X < 25)$ allora il valore standardizzato di 25 è

$$\frac{(25 + \mathbf{0.5}) - \mu}{\sigma} = \frac{(25 + \mathbf{0.5}) - np}{\sqrt{np\,(1-p)}}$$

Diseguaglianze utili

Teorema 17: *Disuguaglianza di Markov*
Sia Y una v.a. che assume valori non negativi, allora per ogni numero reale $a > 0$ si ha che :

$$\Pr(Y \geq a) \leq \frac{\mathbb{E}(Y)}{a}$$

Teorema 18: *Disuguaglianza di Chebychev*
Sia Y una v.a. con valore atteso $\mathbb{E}(Y) = \mu$ e varianza $Var(Y) = \sigma^2$, allora:

$$\Pr(|Y - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\mu^2}$$

Tali disuguaglianze sono molto utili in quanto ci permettono di conoscere "i limiti" di una v.a., senza tuttavia conoscerne la sua distribuzione.

Esempio utilizzo diseguaglianze

Supponiamo che una v.a. X conti il numero di auto prodotte in una settimana da una fabbrica. Sappiamo che in media ne vengono prodotte 50, ossia $\mathbb{E}(X) = 50$. Allora possiamo affermare che

- $\Pr(X > 75) \leq \frac{50}{75} = \frac{2}{3}$ per disuguaglianza di Markov
- Supponendo che la varianza sia 25, cosa posso dire circa la probabilità che vengano prodotte fra le 40 e le 60 auto? Procedo con la disuguaglianza di Chebychev

$$\begin{aligned}\Pr(40 < Y < 60) &= \Pr(-10, Y - 50 < 10) \\ &= \Pr(|Y - 50| < 10) \\ &= 1 - \Pr(|Y - 50| \geq 10)\end{aligned}$$

sapendo che per la disuguaglianza di Chebychev

$$\Pr\left(|Y-50|\geq 10\right)\leq \frac{Var\left(Y\right)}{\mathbb{E}\left(X\right)^2}=\frac{25}{100}=\frac{1}{4}$$

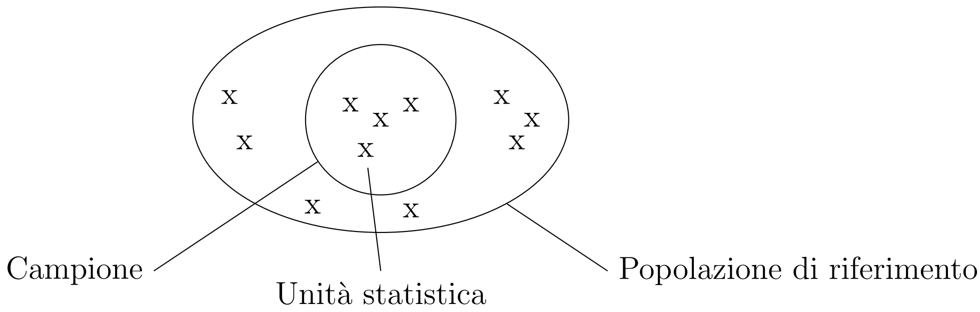
avrò che

$$\Pr\left(40<Y<\ 60\right)=1-\Pr\left(|Y-50|\geq 10\right)\geq \frac{3}{4}$$

10 Inferenza statistica

Tramite l'inferenza statistica, vogliamo utilizzare dei dati raccolti per poter creare modelli probabilistici e dare affermazioni su di una data popolazione

Popolazione, campione, unità statistiica



10.1 Stima e stimatore e metodo dei momenti

Definizione 43: *Stima*
Definiamo stima di un parametro un valore numerico che può essere ottenuto dai dati raccolti

Definizione 44: *Stimatore*
Uno stimatore è una qualsiasi funzione che prenda in input i dati di un campione e stimi il valore di un determinato parametro.

Uno stimatore può essere visto anche come una v.a. funzione delle variabili aleatorie che generano i valori osservati nel campione.

Uno stimatore di un parametro θ viene indicato con $\hat{\theta}$. Talvola $\hat{\theta}$ indica anche la stima del parametro

Ad esempio, suppongo di aver una fabbrica che produce lastre di metallo e ottengo i seguenti valori:

14.33	14.19	14.39	14.43	14.17

in questo caso posso creare uno stimatore della media con il metodo che viene sempre comunemente utilizzato:

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{5}x_i$$

dove $x_1,\dots x_5$ sono le v.a. che indicano lo spessore dell'i-esima lastra

Proprietà stimatori

Idealmente, uno stimatore dovrebbe essere uguale al valore del parametro teorico (ad esempio $\hat{\mu} = \mu$). Tuttavia questo non ha senso in quanto implicherebbe di conoscere con certezza il parametro da stimare. Un requisito più ragionevole è la non distorsione

Definizione 45: *Non distorsione di uno stimatore*
Uno stimatore $\hat{\theta}$ si dice non distorto se

$$\mathbb{E}\left[\left(\hat{\theta} - \theta\right)\right] = \mathbb{E}\left(\hat{\theta}\right) - \theta = 0$$

ossia gli errori in media di compensano, o meglio, le imprecisioni di distribuiscono in manira simmetrica attorno al valore teorico

se invece siamo interessati a capire di quanto erra in media uno stimatore possiamo ricorrere al concetto di errore quadratico medio

Definizione 46: *Errore quadratico medio di uno stimatore*
L'errore quadratico medio di uno stimatore $\hat{\theta}$ è così definito:

$$EQM\left(\hat{\theta}\right) = \mathbb{E}\left[\left(\hat{\theta} - \theta\right)^2\right]$$

Dati due simatori è da preferire in genere quello con l'errore quadratico medio minore

Nota che:

$$EQM\left(\hat{\theta}\right) = \underbrace{\mathbb{V}ar\left(\hat{\theta}\right)}_{\text{varianza}} + \underbrace{\left(\mathbb{E}\left(\hat{\theta}\right) - \theta\right)^2}_{\text{distorsione}}$$

Consistenza stimatori

Definizione 47: *Consistenza stimatorii*
Uno stimatore si dice consistente media quadratica se

$$EQM\left(\hat{\theta}\right) \to 0$$

mentre si dice consistente in senso debole se

$$\hat{\theta} \xrightarrow{P} \theta$$

Nota che ha senso parlare di convergenza in quanto ogni osservazione (ad esempio ogni misura di una lastra di metallo), corrisponde ad una v.a. Dunque uno stimatore sarà dato da una funzione di n v.a.

Metodo dei momenti

Un modo per "creare" degli stimatori è secondo il metodo dei momenti. Innanzitutto diamo la definizione di momento campionario:

Definizione 48: *Momento campionario*
Il momento campionario di ordine r è così definito:

$$\mu_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$$

se dovessi calcolare il momento capionario di una trasformazione $g(X)$ di v.a. (ad esempio il momento centrato), allora vale che

$$\mu_r = \frac{1}{n}\sum_{i=1}^{n} g(X_i)^r$$

Il metodo dei momenti consiste nel eguagliare l'espressione di un momento teorico con il suo corrispondente momento campionario, ad esempio, supponendo di avere $X \sim \mathcal{N}(\mu, \sigma^2)$, allora

$$\mu = \mathbb{E}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\sigma^2 = \mathbb{E}\left[(X-\mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n} (X_i - \mu)^2$$

10.2 Metodo della massima verosimiglianza

Supponiamo di avere un set di campioni derivati dal lancio di una moneta:

$$\boxed{T \quad T \quad C \quad C \quad T \quad T \quad T \quad T \quad C \quad C}$$

Voglio stimare il parametro p, ossia la probabilità che esca T. Allora

Il metodo della massima verosimiglianza ricerca il valore di p tale per cui la probabilità che escano i dati del campione sia massima

La verosimiglianza è quindi così definita:

$$L(\theta) = \Pr(\{T, T, C, C, T, C, T, T, C, C\}) = p^6 (1-p)^4$$

- La verosimiglianza è dunque una funzione di p ossia del parametro che vogliamo stimare. Per ottenere una legge di distribuzione quanto più accurata possibile devo cercare di massimizzare la probabilità che si verifichi una formazione come quella del campione ($\{T, T, C, C, T, T, T, T, C, C\}$)

- La massima verosimiglianza (o MLE) è il valore di p che rende massima la verosimiglianza

- Visto che per trovare il massimo bisogna derivare un prodotto e questo può essere molto tedioso, spesso si utilizza la log-likelihood, ossia si applica il logaritmo alla massima verosimiglianza (il log è crescente e non altera i massimi), in modo da trasformare i prodotti in somme

Valori noti di MLE

- Normale

$$\hat{\mu} = \overline{X}_n \qquad \hat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

I valori sono distribuiti uniformemente intorno alla media μ, per questo la media campionaria approssima bene la media teoria. La varianza è data dalla media di quanto u valore si discosta dalla media

- Esponenziale

$$\hat{\lambda} = \frac{1}{\overline{X}_n}$$

λ è inversamente proporzionale alla media delle osservazioni, più la media è vicina a zero, più lambda sarà alto

- Bernoulli

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

Immagina il lancio ripetuto di una moneta per convincertene. E' chiaro che la miglior approssimazione di μ è data dal rapporto fra successi e insuccessi

- Poisson

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

- Uniforme, che ha quindi densità

$$X_i \sim f = \begin{cases} \frac{1}{\sigma} & \text{se } x \in [0, \sigma] \\ 0 & \text{altrove} \end{cases}$$

allora

$$\hat{\sigma} = \max\{X_1, X_2, \ldots, X_n\}$$

Se ci pensi ciò significa che i dati si distribuiscono in maniera uniforme nel range $[0, \sigma]$ e non ce ne possono essere maggiori di σ stesso. La stima migliore è il dato più alto registrato

10.3 Errore di stima ed intervalli di confidenza

Definizione 49: *Errore di stima*
Dato un parametro θ, si definisce errore di stima la seguente quantità:

$$\hat{\theta} - \theta$$

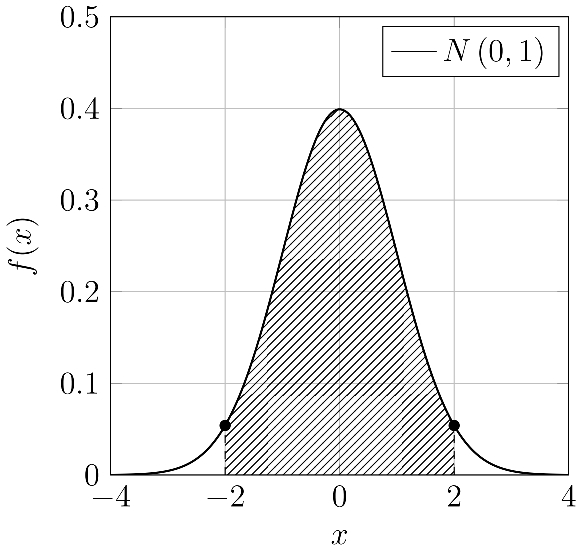
Si può calcolare la probabilità che un errore di stima sia "piccolo", ossia minore della varianza. Ciò significa prendere la v.a. $\hat{\theta} - \theta \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$ e calcolare la probabilità dell'intervallo $\left[\overline{X}_n - \mathbb{V}ar, \overline{X}_n + \mathbb{V}ar\right]$

Definizione 50: *Intervallo di confidenza*
Un intervallo è detto <u>intervallo di confidenza</u> di livello $\alpha \in (0, 1)$ se la probabilità che la stima di un parametro ignoto rientri all'interno di esso è α

Per ragioni di comodità di calcolo, spesso l'intervallo di confidenza si indica con $1 - \alpha$

Praticamente questo servirà quando si andrà a calcolare intorno alla media di una normale l'intervallo di livello α, ossia l'intervallo tale che

$$\Pr\left(x \in [\mu + x, \mu - x]\right) = \alpha$$



L'area del trapezoide  è uguale ad α, supponendo che l'intervallo $-2, 2$ sia di livello α