Dispense di Probabilità e Statistica

Leonardo De Faveri

Indice

Capitolo Nr.1

Introduzione

1.1 Definizione di Probabilità

La Probabilità permette di misurare l'incertezza; l'incertezza può essere dovuta del tempo, da una misurazione, dallo spazio e dai casi e/o soggetti.

1.2 Statistica descrittiva e inferenziale

La *Statistica descrittiva* studia tutta la popolazione, mentre la *Statistica inferenziale* ne studia solo un campione e successivamente usa quel risultato per stimare la quantità interessata nell'intera popolazione.

1.3 Cenni di teoria degli insiemi

Un insieme è definito come una collezione di oggetti detti elementi.
La nozione di appartenenza può essere messa a fondamento di tutta la teoria degli insiemi:

Definizione 1.3.1. Scriveremo $x \in A$ intendendo che x è un elemento dell'insieme A, mentre diremo che $y \notin A$ intendendo che y non è un elemento di A.

Proprietà. Dato un insieme A, $\forall x$ si deve essere in grado di stabilire se x appartiene o meno all'insieme A.

Valgono i seguenti assiomi:

(i) *Estensionalità*: Dati 2 insiemi A, B, vale $A = B \Leftrightarrow \forall x : x \in A \Leftrightarrow x \in B$

(ii) *Insieme vuoto*: Il simbolo $\emptyset$ identifica un insieme privo di elementi

(iii) *Separazione*: Sia X un insieme. Supponiamo che a ogni elemento $x \in X$ sia associata un'affermazione $P(x)$ dipendente da x. Allora:

$$\{x|x \in X,\ P(x) \text{ è vera}\} = \{x \in X|P(x)\}$$

è un insieme.

Definizione 1.3.2. Siano X e Y due insiemi. Scriveremo $X \subset Y$ se è vero che $\forall x \in X \Rightarrow x \in Y$, ma non il contrario. E diremo che X è un *sottoinsieme proprio* di Y.

Definizione 1.3.3. Siano X e Y due insiemi. Se $X \subset Y$ e $X = Y$ allora X viene detto *sottoinsieme* di Y e in simboli si scrive $x \subseteq Y$.

1.3.1 Operazioni sugli insiemi

Siano X e Y due insiemi, definiamo:

- *Unione*: $X \cup Y = \{x|x \in X \text{ o } x \in Y\}$
- *Intersezione*: $X \cap Y = \{x|x \in X,\ x \in Y\}$
- *Differenza*: $X\backslash Y = \{x|x \in X, x \notin Y\}$
- *Differenza simmetrica*: $X \triangle Y = (X\backslash Y) \cup (Y\backslash X)$
- *Prodotto cartesiano*: $X \times Y = \{(x, y)|x \in X,\ y \in Y\}$

Se $Y \subset X$ definiamo l'insieme complementare come segue:

$$C_X(Y) = Y^c = \{x|x \in X,\ x \notin Y\} = X\backslash Y$$

NB. La notazione Y^c non specifica rispetto a quale insieme si sta calcolando il complementare.

Sia I un insieme non vuoto. $\forall i \in I$ sia A_i un insieme. Indichiamo il dato I (A_i con $i \in I$) scrivendo $\{A_i\}_{i\in I}$. Su questa famiglia d'insiemi definiamo:

- *Unione arbitraria*: $\bigcup_{i\in I} A_i = \{x|\exists i \in I,\ x \in A_i\}$
- *Intersezione arbitraria*: $\bigcap_{i\in I} A_i = \{x|\forall i \in I,\ x \in A_i\}$

Definizione 1.3.4. Dato un insieme A e sua una famiglia di sottoinsiemi S. S è detto *partizione* se $\forall x \in A\ \exists! B \in S|x \in B$.

Proprietà. Dato un insieme A e una sua partizione S valgono le seguenti:

(i) $A = \bigcup_{B\in S} B$

(ii) $\forall B, C \in S$ con $B \neq C$ vale $B \cap C = \emptyset$

1.3.2 Cardinalità di un insieme

Definizione 1.3.5. Dato un insieme A se ne definisce cardinalità il numero di elementi in esso contenuti e scriveremo in simboli $\#A$.

Proprietà. Valgono le seguenti:

1. Siano A un insieme e $S = \{E_i\}_{i=1}^n$ una sua partizione. Allora $\#A = \sum_{i=1}^n \#E_i$
2. Siano A e B due insiemi. $\#(A \times B) = \#A \cdot \#B$

   NB. $(A \times B) \neq (B \times A)$, ma $\#(A \times B) = \#(B \times A)$

   Oss. $\# \otimes_{i=1}^n A_i = \prod_{i=1}^n A_i$

3. $\#(A_1 \cup A_2) = \#A_1 + \#A_2 - \#(A_1 \cap A_2)$

   Oss. $\#(A_1 \cup A_2 \cup A_3) = \#A_1 + \#A_2 + \#A_3 - \#(A_1 \cap A_2) - \#(A_1 \cap A_3) - \#(A_2 \cap A_3) + \#(A_1 \cap A_2 \cap A_3)$
   Vale un ragionamento analogo per famiglie più numerose di insiemi.

Capitolo Nr.2

Principi di calcolo combinatorio

Nella probabilità classica la probabilità di evento è calcolata come il rapporto tra il numero di casi favorevoli e il numero di casi totali:

$$P(E) = \frac{\#F}{\#T}$$

Ma come si contano i casi favorevoli e totali? Usando il calcolo combinatorio.

2.1 Permutazioni

Le permutazioni permettono di rispondere alla domanda: *In quanti modi posso mettere in fila n elementi di un insieme A?*
Risposta: $n!$

Esempio Quanti sono gli anagrammi della parola "Prendiamo"?
Provo a procedere una lettera alla volta. La prima volta posso scegliere una tra 9 lettere, la seconda 8, le terza 7 e cosi via. Alla fine arrivo ad avere: $9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 9!$

2.1.1 Permutazioni con ripetizioni

Cosa succede se un elemento si ripete m volte?
Risposta: $\frac{n!}{m!}$

Esempio Quanti sono gli anagrammi della parola "Anagramma"?
Noto che le lettere a ed m si ripetono rispettivamente 4 e 2 volte. Gli anagrammi possibili sono quindi:

$$\frac{n!}{a! \cdot m!} = \frac{9!}{4! \cdot 2!} = \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{4 \cdot 3 \cdot 2 \cdot 1 \cdot 2 \cdot 1} = \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5}{2 \cdot 1} = 9 \cdot 8 \cdot 7 \cdot 5 \cdot 3$$

2.2 Disposizioni

Le disposizioni rispondono alla domanda: *In quanti modi posso mettere in fila m elementi tra gli n di un insieme A?*
Risposta: $\frac{n!}{(n-m)!}$

Esempio Quante sono le parole di 4 lettere, tutte diverse, se l'alfabeto utilizzato ha 26 lettere?

Provo a scegliere una lettera alla volta. La Prima volta ho 26 possibilità, la seconda 25, le terza 24 e la quarta 23. Posso quindi ottenere il risultato facendo: $26 \cdot 25 \cdot 24 \cdot 23$.

E usando la formula?
Devo innanzitutto scegliere i valori di m e n. Siccome l'alfabeto ha 26 lettere pongo $n = 26$ e, di conseguenza, $m = 4$. Ora applico la formula:

$$\frac{n!}{(n-m)!} = \frac{26!}{(26-4)!} = \frac{26!}{22!} = \frac{26 \cdot 25 \cdot 24 \cdot 23 \cdot 22!}{22!} = 26 \cdot 25 \cdot 24 \cdot 23$$

2.2.1 Disposizioni con ripetizioni

Cosa succede se alcuni elementi si ripetono?
Risposta: n^k

Esempio Utilizzando le cifre $1, 2, 3$ quanti numeri di 4 cifre si possono formare?

Essendo le cifre tra cui scegliere meno rispetto a quelle da dover utilizzare ci saranno sicuramente delle ripetizioni. In particolare, ogni cifra può ripetersi fino a 4 volte. Se ora provo a immaginare di avere tanti cassetti quante sono le cifre del numero da creare, quindi 4, e provo a creare il numero prendendo una cifra alla volta, la prima volta potrò scegliere una cifra su 3. Siccome le cifre si possono ripetere la seconda sarà ancora tra 3 cifre e così anche le altre. Quindi avrò $3 \cdot 3 \cdot 3 \cdot 3$ modi di formare un numero di 4 cifre.

E usando la formula?
È sufficiente prendere $k = 4$ e $n = 3$ e applicare la formula:

$$n^k = 3^4$$

In particolare ho posto k pari al numero di cassetti e n pari al numero di scelte che ho per ogni cassetto.

2.3 Combinazioni

Le combinazioni rispondono alla domanda: *In quanti modi posso scegliere m elementi (in qualsiasi ordine) tra gli n di un insieme A?*
Risposta: $\frac{n!}{(n-m)! \cdot m!} = \binom{n}{m}$

Esempio Quante sono le sestine formate da 6 numeri scelti tra 90?

Posso procedere in due passi

1. Conto in quanti modi possono essere ordinati 6 numeri scelti tra 90
2. Conto in quanti modi posso ordinare 6 numeri

Procedo:

1. Il problema di contare i possibili ordinamenti di 6 numeri scelti tra 90 è un tipico problema di *disposizioni semplici*, quindi: $n =$ "*#sestine ordinate*" $= \frac{90!}{(90-6)!}$
2. I possibili ordinamenti di 6 numeri sono invece calcolabili sfruttando le *permutazioni semplici*, quindi: $m =$ "*#numero ordinamenti*" $= 6!$

A questo punto mettendo insieme i risultati si ottiene:

$$\frac{n}{m} = \frac{90!}{(90-6)!} \cdot \frac{1}{6!} = \frac{90!}{(90-6)! \cdot 6!} = \binom{90}{6}$$

2.3.1 Combinazioni con ripetizioni

Cosa succede se alcuni elementi si ripetono?
Risposta: $\frac{(n+m-1)!}{m!(n-1)!}$

Esempio Sono assegnati due contagocce: il primo contenente 5 gocce di colore bianco e il secondo 5 gocce di colore nero. Mischiando 5 gocce scelte tra i due colori, quanti colori diversi si possono formare?

In questo esempio è evidente che l'ordine non ha importanza, inoltre per formare il colore servono 5 gocce, ed essendoci soltanto 2 colori, ci saranno sicuramente delle ripetizioni. Riprendendo l'esempio dei cassetti di prima, possiamo immaginare di avere 5 cassetti da riempire con una goccia di colore scelta tra i due possibili. Pongo poi $m = 5$ e $n = 2$ e vado ad applicare la formula:

$$\frac{(n+m-1)!}{m! \cdot (n-1)!} = \frac{(2+5-1)!}{5! \cdot (2-1)!} = \frac{6!}{5! \cdot 1!} = 6$$

2.4 Coefficiente binomiale

Durante la discussione sulle combinazioni è stato definito il coefficiente binomiale come segue:

$$\binom{n}{m} = \frac{n!}{(n-m)! \cdot m!} \text{ con } 0 \leq m \leq n$$

Proprietà. Valgono le seguenti proprietà:

1. $\binom{n}{m} = \binom{n}{n-m}$
2. $\binom{n}{n} = \binom{n}{0}$
3. $\binom{n}{m} + \binom{n}{m+1} = \binom{n+1}{m+1}$
4. $\sum_{m=0}^{n} \binom{n}{m} = 2^n$

Capitolo Nr.3

Eventi aleatori

3.1 Insieme degli esiti

Definizione 3.1.1. Un esperimento si dice aleatorio o casuale se con i dati a disposizione il risultato è incerto.

Definizione 3.1.2. I risultati a 2 a 2 incompatibili di un esperimento aleatorio si chiamano esiti. Li vediamo come elementi di un insieme dello spazio degli esiti, detto anche spazio campionario, popolazione o universo, indicato con Ω o $\mho$.

Esempio Un'urna contiene biglie bianche, nere e rosse. Se estraggo una biglia qual è lo spazio degli esiti?

$$\Omega = \{B, N, R\}$$

E se estraggo due biglie, reinserendo la prima biglia?

$$\Omega = \{B, N, R\} \times \{B, N, R\}$$

E se non reinserisco la prima biglia?
Non ho abbastanza informazioni per rispondere

Definizione 3.1.3. Due insiemi A e B sono equipotenti, cioè hanno la stessa cardinalità, se e solo se esiste una biezione $f : A \to B$

Definizione 3.1.4. Dato un insieme A si definisce l'insieme $\mathcal{P}(A)$, detto *insieme potenza* o *insieme delle parti*, come l'insieme dei sottoinsiemi di A.

Proprietà. Dati un insieme A e il suo insieme potenza $\mathcal{P}(A)$, vale $\#\mathcal{P}(A) = 2^{\#A}$.

Teorema 3.1.1 (Teorema di Cantor)**.** Non esiste alcuna funzione suriettiva tra un insieme A e $\mathcal{P}(A)$. In particolare $\#A < \#\mathcal{P}(A)$.

Prop. Valgono le seguenti:

- $\#\mathcal{P}(\mathbb{N}) = \#\mathbb{R}$
- $\#\mathcal{P}(\mathbb{R}) = 2^{\#\mathbb{R}} = 2^{\mathcal{P}(\mathbb{N})}$

3.2 Algebra e σ-algebra

Noi vorremmo definire la probabilità sugli eventi, ma se $\Omega = \mathbb{R}$ e gli eventi fossero tutti i sottoinsiemi di Ω dovremmo definirla su troppi punti e sarebbe difficile.
Idea: invece di avere come eventi tutti i sottoinsiemi di Ω, se ne può considerare solo una parte. Una famiglia $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ e ipotizziamo che $\mathcal{F}$ sia stabile rispetto a unione, intersezione e complementare e che contenga Ω.

Definizione 3.2.1. $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ è un algebra se:

(i) $\Omega \in \mathcal{F}$

(ii) Se $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$

(iii) $\forall A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$

Oss. Potremmo anche scrivere $\{A_i\}_{i=1}^n \subseteq \mathcal{F} \Rightarrow \bigcup_{i=1}^n A_i \in \mathcal{F}$

Prop. Se $\mathcal{F}$ è un algebra ($\mathcal{F} \subseteq \mathcal{P}(\Omega)$) allora valgono le seguenti:

(i) $\emptyset \in \mathcal{F}$

(ii) Se $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$ ($\mathcal{F}$ è chiuso rispetto all'intersezione)

(iii) Se $\{A_i\}_{i=1}^n \subseteq \mathcal{F} \;\Rightarrow \bigcap_{i=1}^n A_i \in \mathcal{F}$

(iv) Se $A, B \in \mathcal{F} \Rightarrow A \backslash B \in \mathcal{F} \wedge B \backslash A \in \mathcal{F}$

(v) Se $A, B \in \mathcal{F} \Rightarrow A \triangle B \in \mathcal{F}$

Esempio Sia $\Omega = \{0, 1, 2\}$. L'insieme $\mathcal{F} = \{\emptyset, \{0\}, \{1, 2\}, \Omega\}$ è un'algebra?
Provo a vedere se rispetta le 3 condizioni necessarie:

(i) $\Omega \in \mathcal{F}$ ✓

(ii) $\forall A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ ✓

(iii) $\forall A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$ ✓

Segue che $\mathcal{F}$ è un'algebra.

Oss. Gli insiemi $\mathcal{F} = \{\emptyset, \Omega\}$ e $\mathcal{F} = \mathcal{P}(\Omega)$ sono sempre valide algebre.

Esempio Sia $\Omega = \{0, 1, 2\}$. L'insieme $\mathcal{F} = \{\emptyset, \{0\}, \{1\}, \{1, 2\}, \Omega\}$ è un'algebra?
$\mathcal{F}$ non è un'algebra in quanto $\{0\} \cup \{1\} \notin \mathcal{F}$

Definizione 3.2.2. $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ è una *tribù* o *σ-algebra* (sigma-algebra) se sono verificate le seguenti condizioni:

(i) $\Omega \in \mathcal{F}$

(ii) $\forall A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

(iii) $\forall$ famiglia numerabile $\{A_i\}_{i=1}^{+\infty} \subseteq \mathcal{P}(\Omega) \Rightarrow \bigcup_{i=1}^{+\infty} A_i \in \mathcal{F}$

Oss. La differenza tra *algebra* e *σ-algebra* sta nel fatto che, nel caso di un'*algebra*, la terza proprietà è vera solo per unioni finite di insiemi, mentre è vera per un'infinità numerabile di insiemi nel caso di una *σ-algebra*.

Oss. Se un insieme è una *tribù* allora è anche un'*algebra*.

3.3 Definizione di probabilità sugli eventi

Definizione 3.3.1. Sia $\mathcal{F}$ una tribù su Ω. Ogni $E \in \mathcal{F}$ (cioè ogni $E \subseteq \Omega$) si dice evento. I singoletti si chiamano *eventi elementari.* Un evento E si verifica se il risultato osservato dell'esperimento aleatorio appartiene ad E.

Definizione 3.3.2. Dato Ω e una sua tribù $\mathcal{F}$, si definisce *spazio probabilizzante* la coppia $(\Omega, \mathcal{F})$.

Vogliamo definire una funzione di probabilità su $\mathcal{F}$. Gli *assiomi di Kolmogorov* ci dicono quali proprietà una funzione deve soddisfare per poter essere definita *funzione di probabilità.*

Definizione 3.3.3. Sia $(\Omega, \mathcal{F})$ uno spazio probabilizzante. Una funzione $P : \mathcal{F} \to \mathbb{R}$ si dice *funzione* o *misura di probabilità* se soddisfa i seguenti assiomi:

(i) *Non negatività*: $\forall E \in \mathcal{F}\ P(E) \geq 0$

(ii) *Normalizzazione*: $P(\Omega) = 1$

(iii) *σ-additività*: Data una famiglia numerabile $\{E_i\}_{i=1}^{+\infty}$ di eventi a 2 a 2 disgiunti vale $P(\bigcup_{i=1}^{+\infty} E_i) = \sum_{i=1}^{+\infty} P(E_i)$

Oss. Se vale *(iii)* vale anche l'*additività finita.*

NB. $P(E)$ di dice *probabilità dell'evento E.*

Esempio Siano $\Omega = \{0,1\}$ e $\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, \{0\}, \{1\}, \{0,1\}\}$. Sia definita una funzione $P : \mathcal{F} \to \mathbb{R}$ tale che:

$$P : \begin{cases} P(\emptyset) & = 0 \\ P(\{0\}) = P(\{1\}) & = \frac{1}{2} \\ P(\Omega) & = 1 \end{cases}$$

P così definita una valida funzione di probabilità su $(\Omega, \mathcal{F})$?
Bisogna vedere se rispetta gli *Assiomi di Kolmogorov*:

(i) $P(E) \geq 0\ \forall E \in \mathcal{F}$ ✓

(ii) $P(\Omega) = 1$ ✓

(iii) $P(\{0,1\}) = P(\{0\}) + P(\{1\})$ ✓

Esempio Siano $\Omega = \{C, Q, F, P\}$ $\mathcal{F} \subseteq \mathcal{P}(\Omega)$. Di P si sa che:

$$P(\emptyset) = 0,\ P(F) = P(Q) = \frac{1}{3},\ P(\{F, Q, P\}) = \frac{7}{9},\ P(\{P\}) = q,\ P(\{C\}) = p$$

Per quali valori di p, q P è una valida funzione di probabilità?

$$P(\{F, Q, P\}) = P(\{F\}) + P(\{Q\}) + P(\{P\}) = \frac{1}{3} + \frac{1}{3} + q = \frac{2}{3} + q = \frac{7}{9}$$

$$\Rightarrow q = \frac{7}{9} - \frac{2}{3} \Rightarrow q = P(\{P\}) = \frac{1}{9}$$

$$p = P(\{C\}) = 1 - P(\{F, Q, P\}) = 1 - \frac{7}{9} = \frac{2}{9}$$

3.3.1 Proprietà della probabilità

1. $P(\emptyset) = 0$
2. Se $E \in \mathcal{F}$ allora $P(E^c) = 1 - P(E)$
3. Siano $E, F \in \mathcal{F}$. $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

3.1. $P(E \cup F) \leq P(E) + P(F)$

3.2. Sia $\{E_i\}_{i=1}^n \subset \mathcal{F}$.
$P(\bigcup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i) - \sum_{i<j} P(E_i \cap E_j) + \sum_{i<j<k} P(E_i \cap E_j \cap E_k)... =$
$\sum_{k \in \mathcal{P}(\{1,..,n\})} (-1)^{\#k} P(\bigcap_{i \in k+1} E_i)$

3.3. Sia $\{E_i\}_{i=1}^{+\infty} \subset \mathcal{F}$. $P(\bigcup_{i=1}^{+\infty} E_i) \leq \sum_{i=1}^{+\infty} P(E_i)$

4. *Disuguaglianza di Bo*: $\sum_{i=1}^{+\infty} P(E_i) - \sum_{i<j} P(E_i \cap E_j) \leq P(\bigcup_{i=1}^{+\infty} E_i) \leq \sum_{i=1}^{+\infty} P(E_i)$

Dim. Dimostriamo ora alcune delle proprietà:

1. Valgano le seguenti uguaglianze: $\Omega \cap \emptyset = \emptyset$ e $\Omega \cup \emptyset = \Omega \Rightarrow \Omega$ e $\emptyset$ formano una partizione.
$P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) \Leftrightarrow P(\emptyset) = 0$
2. Valgano le seguenti uguaglianze: $E \cap E^c = \emptyset$ e $E \cup E^c = \Omega$.
$P(\Omega) = P(E \cup E^c) = P(E) + P(E^c) = 1 \Leftrightarrow P(E) = 1 - P(E^c)$
3. Valgano le seguenti uguaglianze: $E \cup F = (E \backslash F) \cup F$, $(E \backslash F) \cap F = \emptyset$, $E = (E \backslash F) \cup (E \backslash F)$ e $E \backslash F \cap (E \cap F) = \emptyset$.
Allora $P(E \cup F) = P(E \backslash F) + P(F)$ e $P(E) = P(E \backslash F) + P(F) \Rightarrow P(E \backslash F) = P(E) - P(E \cap F) \Rightarrow P(E \cup F) = P(E) - P(E \cap F) + P(F)$

3.4 Probabilità condizionata

Idea: dato che la probabilità è una misura dell'informazione/incertezza, se avessimo nuove informazioni, queste potrebbero influire sulla probabilità.

Definizione 3.4.1. Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità. Siano $E, F \in \mathcal{F}$ con $P(F) = 0$. Allora, la probabilità di E *condizionata* F è:

$$P(E|F) \coloneqq \frac{P(E \cap F)}{P(F)}$$

Oss. Se $F \in \mathcal{F}$ e $P(F) \neq 0$ allora $P(\bullet|F)$ è una probabilità. Infatti $\forall E \in \mathcal{F}$ valgono le seguenti:

(i) $P(E|F) = \frac{P(E \cap F)}{P(F)} \geq 0$

(ii) $P(\Omega|F) = \frac{P(\Omega \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1$

(iii) Data una famiglia $\{E_i\}_{i=1}^{+\infty}$ di eventi a 2 a 2 disgiunti vale $P(\bigcup_{i=1}^{+\infty} E_i|F) = \frac{P(\bigcup_{i=1}^{+\infty}(E \cap F))}{P(F)} = \frac{\sum_{i=1}^{+\infty} P(E \cap F)}{P(F)}$

Teorema 3.4.1 (Teorema di fattorizzazione o delle probabilità totali)**.** Dati, uno spazio di probabilità $(\Omega, \mathcal{F}, P)$, una famiglia di insiemi $\{E_i\}_{i\in I}$ con $E_i \in \mathcal{F}, E_i \cap E_j = \emptyset$ se $i \neq j$. Se valgono le seguenti:

(i) $\bigcup_{i\in I} E_i = \Omega$

(ii) $\forall i \in I \ P(E_i) \neq 0$

Preso un $E \in \mathcal{F}$ allora $P(E) = \sum_{i\in I} P(E \cap E_i) = \sum_{i\in I} P(E|E_i) \cdot P(E_i)$

Dim. Se vale:

$$E = E \cap \Omega = E \cap \left(\bigcup_{i\in I} E_i\right) = \bigcup_{i\in I}(E \cap E_i)$$

allora vale anche:

$$P(E) = P\left(E \cap \Omega\right) = P\left(E \cap \left(\bigcup_{i\in I} E_i\right)\right) = P\left(\bigcup_{i\in I}(E \cap E_i)\right)$$

$$= \sum_{i\in I} P(E \cap E_i) = \sum_{i\in I} P(E|E_I) \cdot P(E_i)$$

Teorema 3.4.2. (Teorema di Bayes)

(i) Se $P(E) \neq 0 \neq P(F)$ allora $P(E|F) = \frac{P(F|E)\cdot P(E)}{P(F)}$

(ii) Se ho $\{E_i\}_{i\in I}$ $t.c.$ $E_i \in \mathcal{F}, E_i \cap E_j = \emptyset$ con $i \neq j, \bigcup_{i\in I} E_i = \Omega, P(E_i) \neq 0$ allora $P(E_j|F) = \frac{P(F|E_j)\cdot P(E_j)}{\sum_{i\in I}(P(F|E_i)\cdot P(E_i))}$

Esempio Studenti del primo anno del corso di informatica:

Scuola di provenienza	% sul corso	Rapporto $\frac{♀}{totale}$
Liceo Scientifico	13%	$\frac{3}{8}$
Liceo Scienze applicate	18%	$\frac{2}{11}$
Liceo Classico	3%	$\frac{1}{2}$
ITT	57%	$\frac{1}{34}$
Altro	9%	$\frac{4}{5}$

Prendendo uno studente a caso, qual è la probabilità che sia femmina?

Sia $E = ♀ =$ "*É una femmina*" l'evento di cui studiare la probabilità. Dalla tabella posso ricavare direttamente la probabilità che, preso uno studente, questo venga da una certa scuola:

- $P(\text{"}Viene\ dallo\ scientifico\text{"}) = P(E_1) = \frac{13}{100}$
- $P(\text{"}Viene\ dalle\ scienze\ applicata\text{"}) = P(E_2) = \frac{18}{100}$
- $P(\text{"}Viene\ dal\ classico\text{"}) = P(E_3) = \frac{3}{100}$
- $P(\text{"}Viene\ dall'ITT\text{"}) = P(E_4) = \frac{57}{100}$
- $P(\text{"}Viene\ da\ altre\ scuole\text{"}) = P(E_5) = \frac{9}{100}$

Posso ricavare anche la probabilità che preso uno studente proveniente da una certa scuola, questo sia una femmina:

- $P(♀|E_1) = \frac{3}{8}$
- $P(♀|E_2) = \frac{2}{11}$
- $P(♀|E_3) = \frac{1}{2}$
- $P(♀|E_4) = \frac{1}{34}$
- $P(♀|E_5) = \frac{4}{5}$

Noto che gli eventi E_i con $i \in \{1, 2, 3, 4, 5\}$ formano una partizione. Infatti, sia $\{E_i\}_{i=1}^5$ una famiglia di insiemi. Valgono le seguenti:

(i) $\bigcup_{i=1}^5 E_i = \Omega$

(ii) $\forall i, j \in \{1, .., 5\}$ con $i \neq j$ $E_i \cap E_j = \emptyset$

A questo punto, noto che sono verificate tutte le ipotesi del *Teorema di fattorizzazione*, posso quindi applicarlo per calcolare $P(E)$.

$$P(E) = \sum_{i=1}^{5} P(E \cap E_i) = \sum_{i=1}^{5} P(E|E_i) \cdot P(E_i) =$$

$$= P(♀|E_1) \cdot P(E_1) + P(♀|E_2) \cdot P(E_2) + P(♀|E_3) \cdot P(E_3) + P(♀|E_4) \cdot P(E_4) + P(♀|E_5) \cdot P(E_5) =$$

$$= \frac{3}{8} \cdot \frac{13}{100} + \frac{2}{11} \cdot \frac{18}{100} + \frac{1}{2} \cdot \frac{3}{100} + \frac{1}{34} \cdot \frac{57}{100} + \frac{4}{5} \cdot \frac{9}{100} = \frac{18.5}{100} = 18.5\%$$

Qual è la probabilità che presa una studentessa questa venga dal classico?

Mi si chiede di calcolare $P(E_3|♀)$, ma siccome conosco già $P(♀|E_3)$ e $P(♀)$ dal quesito precedente, posso applicare il *Teorema di Bayes*:

$$P(E_3|♀) = \frac{P(♀|E_3) \cdot P(E_3)}{P(♀)} = \frac{\frac{1}{2} \cdot \frac{3}{100}}{\frac{18.5}{100}} = \frac{3}{200} \cdot \frac{100}{18.5} = \frac{3}{37} = 0.0811 = 8.11\%$$

3.5 Indipendenza degli eventi

Definizione 3.5.1. Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità. Dati due eventi $E, F \in \mathcal{F}$, questi sono *indipendenti* o *stocasticamente indipendenti* se e solo se:

$$P(E \cap F) = P(E) \cdot P(F)$$

Oss. Se $E, F \in \mathcal{F}$ sono eventi indipendenti, valgono le seguenti uguaglianze:

- $P(E|F) \cdot P(F) = P(E \cap F) = P(E) \cdot P(F) \Rightarrow P(E|F) = P(E)$
- $P(F|E) \cdot P(E) = P(F \cap E) = P(F) \cdot P(E) \Rightarrow P(F|E) = P(F)$

Oss. Se due eventi sono indipendenti, la probabilità di uno di questi eventi non influisce sulla probabilità dell'altro e viceversa.

NB. L'indipendenza è una proprietà dello *spazio di probabilità* e non degli *eventi*.

Definizione 3.5.2. Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità. Dati n eventi $E_1, .., E_n$ essi sono *indipendenti tra loro* se per ogni sottoinsieme finito $I = \{1, .., m\}$ con $m \leq n$ vale:

$$P\left(\bigcup_{i=1}^{m} E_i\right) = \prod_{i=1}^{m} P(E_i)$$

Definizione 3.5.3. Dati uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ e due sottoinsiemi $\mathcal{F}_1, \mathcal{F}_2 \in \mathcal{F}$ (cioè due tribù contenute in $\mathcal{F}$), allora $\mathcal{F}_1$ e $\mathcal{F}_2$ sono *indipendenti tra loro* se ogni elemento di $\mathcal{F}_1$ è indipendente a ogni elementi di $\mathcal{F}_2$.

Oss. Siano $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità e $F \in \mathcal{F}$ un evento con $P(F) \neq 0$. $P(\bullet|F)$ è una *funzione di probabilità*, per cui ha senso parlare di indipendenza condizionata a F.

Definizione 3.5.4. Dati due eventi $E_1, E_2 \in \mathcal{F}$ in uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ e fissato un terzo evento $F \in \mathcal{F}$, E_1 ed E_2 sono indipendenti condizionatamente a F se:

$$P(E_1 \cap E_2|F) = P(E_1|F) \cdot P(E_2|F)$$

Capitolo Nr.4

Funzioni di probabilità

Dato uno spazio di probabilità $(\Omega, \mathcal{F}, P)$, come di definisce in esso una funzione di probabilità?

4.1 Ω è finito o numerabile

In questo caso, Ω è dato e come tribù prendiamo $\mathcal{P}(\Omega)$, che è un insieme grande, ma ancora "maneggiabile". Per quanto riguarda la *funzione di probabilità* assegniamo a ogni singoletto $\{\omega\} \in \mathcal{F}$ con $\omega \in \Omega$, una probabilità P tale che:

(i) $P(\{\omega\}) \geq 0$

(ii) $\sum_{\omega \in E} P(\{\omega\}) = 1$

A questo punto $\forall E \in \mathcal{F}$ $P(E) \coloneqq \sum_{\omega \in \Omega} P(\{\omega\})$

Oss. Questo può funzionare perché le somme sono al più numerabili.

4.2 Spazi prodotto

Cosa succede se consideriamo più ripetizioni di un esperimento o se mettiamo assieme più esperimenti distinti?

Esempio Lancio di una moneta infinite volte.
$\Omega = \{0,1\}^{\mathbb{N}} = \{\{\omega\}_{i=1}^{+\infty} | \omega_i \in \{0,1\} \ \forall i \in \mathbb{N}\}$ quindi $\#\Omega = \#\mathbb{R}$

L'insieme così definito sarebbe più che numerabile, ma voglio sfruttare il fatto che sia una partizione dell'esperimento $(\Omega_E, \mathcal{F}_E, P_E)$ (singolo lancio di moneta).

Per scegliere $\mathcal{F}$ comincio considerando solo il lancio di due monete. In particolare voglio che $\mathcal{F}$ sia una tribù e che contenga tutti gli eventi $E_1 \times E_2$ con $E_1, E_2 \in \mathcal{F}_E$.

NB. Siccome voglio che $\mathcal{F}$ sia una tribù non basta prendere tutti e soli i prodotti $E_1 \times E_2$.

Definizione 4.2.1. Data una famiglia di sottoinsiemi di Ω detta $\mathcal{A}$, si definisce la tribù $\mathcal{F}_{\mathcal{A}}$ come la più piccola tribù contenente $\mathcal{A}$:

$$\mathcal{F}_{\mathcal{A}} = \sigma(\mathcal{A}) = \cap\{\mathcal{G} : \mathcal{G} \text{ è tribù di } \Omega \text{ e } \mathcal{A} \subseteq \mathcal{G}\}$$

Tornando ora, allo spazio prodotto $(\Omega_E \times \Omega_E)$, la tribù sarà quella generata dalle coppie $E_1 \times E_2$:

$$\mathcal{F} = \mathcal{F}_E \otimes \mathcal{F}_E = \sigma(\{E_1 \times E_2 : E_1, E_2 \in \mathcal{F}_E\})$$

Oss. Non cambia molto se invece di ripetere lo stesso esperimento ne 2 diversi.

Oss. Il procedimento è facilmente applicabile a un numero finito di esperimenti, ma per un numero infinito? *Idea*: fissiamo $\forall n \in \mathbb{N}$ le prime n componenti:

$$E_1 \times E_2 \times .. \times E_n \times \Omega_E \times \Omega_E \times ..$$

e poi, genero la tribù a partire da queste sequenze finite, dette *cilindri*.

Riassumendo, se abbiamo una famiglia finita o numerabile di esperimenti $\{(\Omega, \mathcal{F}, P)\}_{i\in I}$ (ad esempio $I \subset \mathbb{N}, I = \mathbb{N}$), lo spazio prodotto prende la seguente forma:

- $\Omega = \bigotimes_{i\in I} \Omega_i$
- $\mathcal{F} = \bigotimes_{i\in I} \mathcal{F}_i = \sigma(\prod_{i\in I} E_i : E_i \in \mathcal{F}_i$ e $\exists n | \forall j \geq n \ E_j = \Omega_i)$
- $P = \bigotimes_{i\in I} P_i$, cioè $P(\prod_{i\in I} E_i) = \prod_{i\in I} P(E_i)$

4.3 Ω è più che numerabile

Come si definisce uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ quando $\#\Omega = 2^{\aleph_0}$ (ad esempio quando $\Omega = [0,1]$ o $\Omega = \mathbb{R}$).

Oss. Se prendessimo $\mathcal{F} = \mathcal{P}(\Omega)$, $\#\mathcal{F} = 2^{2^{\aleph_0}}$ che è troppo grande.

Se, ad esempio, fissiamo $\Omega = [0,1]$, quali sono i sottoinsiemi che ci interessano?

- Sicuramente ci saranno $\emptyset, [0,1]$
- $\{0\}, \{1\}$ e in generale tutti i singoletti per $x \in [0,1]$
- $[0, \frac{1}{2}]$ e in generale tutti gli intervalli (di qualsiasi tipo)
- Combinazioni ragionevoli (unione, intersezione, complementare, ...) dei precedenti

NB. Se fossimo su $\mathbb{R}$ ci interesserebbero anche le semirette.

Idea: prendiamo le semirette generate da questi sottoinsiemi. Iniziamo prendendo la tribù generata dagli intervalli chiusi $[a,b]$ con $0 \leq a \leq b \leq 1$. Che cosa c'è dentro? Quali sottoinsiemi di $[0,1]$ si possono scrivere come unione, intersezione, (complementare, differenza, ...) di intervalli chiusi?

- I singoletti ci sono? Si, perché $[c,a] \cap [a,b] = [a,a] = \{a\}$
- Gli intervalli aperti? Si, perché $([0,a] \cup [b,1])^c = (a,b)$
- Gli intervalli semiaperti e della forma $[a,a)$ o $(b,b]$?...

Quindi in $\mathcal{F} = \sigma([a,b]$ con $0 \leq a \leq b \leq 1)$ troviamo tutti i sottoinsiemi che ci interessano.

Definizione 4.3.1. La tribù generata dagli intervalli chiusi in $[0,1]$ si chiama *tribù dei Boreliani su* $[0,1]$ e si indica con $\mathcal{B}([0,1])$.

Oss. Possiamo ottenere $\mathcal{B}([0,1])$ come tribù generata da (a,b), $(a,b]$ e $[a,b)$, inoltre $\#\mathcal{B}([0,1]) = \#[0,1] = 2^{\aleph_0}$.

Passiamo ora a definire una funzione di probabilità. Una possibilità è prendere P che misuri la lunghezza dei segmenti (interpretando gli intervalli come segmenti):

$$P([0,1]) = P(\Omega) = length([0,1]) = 1$$

Possiamo quindi prendere $P([a,b]) = b - a$. In questo modo abbiamo definito P sui generatori della tribù dei *Boreliani* $\mathcal{B}([0,1])$ e possiamo estenderla in modo unico a tutto $\mathcal{B}([0,1])$.

- $P(\{\frac{1}{3}\}) = P([\frac{1}{3}, \frac{1}{3}]) = \frac{1}{3} - \frac{1}{3} = 0$
- $P([a,b)) = P([a,b]) - P([b,b]) = P([a,b]) - P(\{b\}) = b - a - 0 = b - a$

NB. Questa funzione di probabilità, detta *uniforme*, non è la sola possibile! Ad esempio, possiamo generalizzare e chiedere che $P((a,b]) = F(b) - F(a)$ per quale opportuna funzione F.

Esempio Sia $F : [0,1] \to \mathbb{R}$ così definita:

$$F : \begin{cases} x & \text{se } 0 \leq x < \frac{1}{2} \\ \frac{1}{2} \cdot (x+1) & \text{se } \frac{1}{2} \leq x \leq 1 \end{cases}$$

Quanto vale $P((\frac{3}{4}, \frac{7}{8}))$?

Innanzitutto, noto che la funzione non può analizzare direttamente quell'intervallo in quanto è definita soltanto su intervalli semiaperti a sinistra ($(a,b]$).

$$P\left(\left(\frac{3}{4}, \frac{7}{8}\right)\right) = P\left(\bigcup_{n \geq n_0} \left(\frac{3}{4}, \frac{7}{8} - \frac{1}{n}\right]\right) = \lim_{n \to +\infty} P\left(\frac{3}{4}, \frac{7}{8} - \frac{1}{n}\right] =$$

$$\lim_{n \to +\infty} F\left(\frac{7}{8} - \frac{1}{n}\right) - F\left(\frac{3}{4}\right) = F\left(\frac{7}{8}\right) - F\left(\frac{3}{4}\right) =$$

$$\frac{1}{2} \cdot \left(\frac{7}{8} + 1\right) - \frac{1}{2} \cdot \left(\frac{3}{4} + 1\right) = \frac{1}{2} \cdot \left(\frac{7}{8} - \frac{6}{8}\right) = \frac{1}{16}$$

Quanto vale $P([\frac{1}{8}, \frac{1}{4}])$?

Posso seguire lo stesso ragionamento di prima:

$$P\left(\left[\frac{1}{8}, \frac{1}{4}\right]\right) = P\left(\left\{\frac{1}{8}\right\}\right) + P\left(\left(\frac{1}{8}, \frac{1}{4}\right]\right) = P\left(\left\{\frac{1}{8}\right\}\right) + F\left(\frac{1}{4}\right) + F\left(\frac{1}{8}\right) =$$

$$0 + F\left(\frac{1}{4}\right) + F\left(\frac{1}{8}\right) = F\left(\frac{1}{4}\right) + F\left(\frac{1}{8}\right) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$$

Quanto vale $P((\frac{1}{8}, \frac{1}{2}))$?

$$P\left(\left(\frac{1}{8}, \frac{1}{2}\right)\right) = P\left(\bigcup_{n \geq n_0} \left(\frac{1}{8}, \frac{1}{2} - \frac{1}{n}\right]\right) = \lim_{x \to \frac{1}{2}^-} \left(F(x) - F\left(\frac{1}{8}\right)\right) = \lim_{x \to \frac{1}{2}^-} x - \frac{1}{8} = \frac{3}{8}$$

Quanto vale $P((\frac{1}{8}, \frac{1}{2}])$?

$$P\left(\frac{1}{8}, \frac{1}{2}\right) = F\left(\frac{1}{2}\right) - F\left(\frac{1}{8}\right) = \frac{1}{2} \cdot \left(\frac{1}{2} + 1\right) - F\left(\frac{1}{8}\right) = \frac{1}{2} \cdot \frac{3}{2} - \frac{1}{8} = \frac{3}{4} - \frac{1}{8} = \frac{5}{8}$$

Oss. Possiamo osservare che $P((\frac{1}{8},\frac{1}{2}]) \neq P((\frac{1}{8},\frac{1}{2}))$, infatti se ne calcolo la differenza ottengo $P((\frac{1}{8},\frac{1}{2}]) - P((\frac{1}{8},\frac{1}{2})) = \frac{5}{8} - \frac{3}{8} = \frac{2}{8}$. Questo è il valore di probabilità del singoletto $\{\frac{1}{2}\}$, infatti $(\frac{1}{8},\frac{1}{2}]\backslash(\frac{1}{8},\frac{1}{2}) = [\frac{1}{2}] = \{\frac{1}{2}\}$. Se ne deduce che i singoletti possono avere probabilità diversa da 0.

Oss. Sappiamo che $F(1) = 1$ e $F(0) = 0$. Inoltre F è monotona crescente, ma non è necessariamente continua.

Oss. $P(\{b\}) = P((A,b]) - P((a,b)) = F(b) - F(a) - (\lim_{x\to 0^-} F(x) - F(a)) = F(b) - \lim_{x\to 0^-} F(x)$. Da questa catena di uguaglianze segue che se F è continua a sinistra in b (cioè se $\lim_{x\to 0^-} F(x) = F(x)$) allora $P(\{b\}) = 0$.

Cosa succede se $\Omega = \mathbb{R}$?
Idea: possiamo usare l'idea vista per $\Omega = [0,1]$ e definire $\mathcal{B}(\mathbb{R})$ come la tribù generata dagli intervalli chiusi, ma anche dalle semirette $(-\infty, b]$. A questo punto però non possiamo più usare la lunghezza dei segmenti come misura di probabilità, ma possiamo comunque riciclare l'idea che $P((a,b]) = F(b) - F(a)$ che ora decliniamo come:

$$F(b) = P((-\infty, b])$$

Oss. Definire P su $(-\infty, b]$ è equivalente a definire P su $(a,b]$.

Ma quali sono le funzioni $F : \mathbb{R} \to \mathbb{R}$ che vanno bene?

(i) F monotona non decrescente

(ii) $\lim_{x\to -\infty} F(x) = 0$

(iii) $\lim_{x\to +\infty} F(x) = 1$ poiché $P(\mathbb{R}) = 1 = P((-\infty, +\infty))$

(iv) $\forall x \in \mathbb{R}$ F è continua a destra, cioè $\exists \lim_{x\to x_0^+} F(x) = F(x_0)$ e limitata a sinistra, cioè $\exists \lim_{x\to x_0^-} F(x) \leq F(x_0)$

Oss. Valgono le seguenti:

- Se F è continua in $\mathbb{R}$ allora $P(\{x\}) = 0$ $\forall x \in \mathbb{R}$ e più in generale se F è continua in x_0 $P(\{x_0\}) = 0$
- Se F è discontinua in x_0 $P(\{x_0\}) = P((-\infty, x_0]) - P(\bigcup_{n\geq n_0}(-\infty, x_0 - \frac{1}{n}]) = F(x_0) - \lim_{x\to x_0^-} F(x) > 0$
- P è definita su tutta $\mathcal{B}(\mathbb{R})$ e si calcola facilmente per eventi che sono intervalli, semirette o punti, ma non necessariamente

Esempio Un'azienda produttrice di calcolatori, spende 1000€ per produrre ogni calcolare e lì vende poi a 2000€ l'uno. Durante la produzione un calcolatore si guasta con probabilità $P = 10\%$.
Con che probabilità l'azienda guadagna dalla vendita di un calcolatore?
L'azienda continua a produrre calcolatori finché non ne ha uno funzionante:

- Al 1^o tentativo con $1 - P = 90\%$
- Al 2^o tentativo con $P \cdot (1 - P) = 9\%$
- Al 3^o tentativo con $P^2 \cdot (1 - P)$
- Al $n - esimo$ tentativo con $P^{n-1} \cdot (1 - P)$

Per guadagnare l'azienda deve produrre al più n calcolatori con n tale che $2000 - 1000 \cdot n > 0$, quindi se $n = 1$ guadagna, se $n = 2$ pareggia, altrimenti va in perdita. Ne deriva che la probabilità di guadagnare è 90%, la probabilità di guadagnare o pareggiare è $90\% + 9\% = 99\%$, mentre la probabilità di perdere è $100\% - 99\% = 1\%$.

Non abbiamo ancora definito Ω, ma prendendo come esiti il numero di calcolatori prodotti fino al primo funzionante, possiamo definite $\Omega = \mathbb{N} \backslash \{0\}$.

Con che probabilità l'azienda guadagna con un ordine di 3 calcolatori?

In questo caso $\Omega = \mathbb{N} \backslash \{0, 1, 2\}$ perché deve produrre almeno 3 calcolatori:

- Ordine evaso con 3 calcolatori prodotti con probabilità $(1 - P)^3 = 0.729 \approx 73\%$
- Ordine evaso con 4 calcolatori prodotti con probabilità $\binom{3}{1} \cdot P \cdot (1 - P)^3 = 0.2187 \approx 22\%$
- Ordine evaso con 4 calcolatori prodotti con probabilità $\binom{4}{2} \cdot P^2 \cdot (1 - P)^3 = 0.04374 \approx 4\%$
- Ordine evaso con n calcolatori prodotti con probabilità $\binom{n-1}{n-3} \cdot P^{n-3} \cdot (1 - P)^3$

NB. È stato usato il *coefficiente binomiale* perché non ha importanza l'ordine in cui vengono prodotti i calcolatori funzionanti e difettosi.

Quindi la probabilità che ci guadagni è: $P_3 + P_4 + P_5 = 99\%$

Oss. In questo esempio, tuttavia, abbiamo usato poco la forma $(\Omega, \mathcal{F}, P)$.

Capitolo Nr.5

Variabili aleatorie

5.1 Definizione di una variabile aleatoria

Nell'ultimo esempio del precedente capitolo non è stata sfruttata la forma $(\Omega, \mathcal{F}, P)$, ma vorremmo sfruttare le probabilità già definite prima, usando il fatto che ci interessa una funzione dell'esperimento aleatorio.

Definizione 5.1.1. Dato uno spazio probabilizzante $(\Omega, \mathcal{F})$, si dice *variabile aleatoria (casuale)* (v.a.) ogni funzione $X : \Omega \to \mathbb{R}$ tale che $\forall x \in \mathbb{R}$ l'insieme $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$

Oss. X si dice *variabile aleatoria* di A perché il valore della funzione dipende dall'esito di un *esperimento aleatorio* $X(\omega) \in \mathbb{R}$.

Partiamo da uno spazio probabilizzante $(\Omega, \mathcal{F})$. Chiediamo che $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ così da potergli assegnare una probabilità. Ma perché abbiamo scelto insiemi del tipo: $X(\omega) \leq x \Rightarrow X(\omega) \in (-\infty, x]$? L'abbiamo fatto perché su $\mathbb{R}$ conosciamo l'insieme $\mathcal{B}$ dei Boreliani generato da semirette di quel tipo.

Teorema 5.1.1. Sia $(\tilde{\Omega}, \tilde{\mathcal{F}})$ uno spazio probabilizzante. Siano inoltre Ω un insieme e $X : \Omega \to \tilde{\Omega}$ una funzione. Allora $\mathcal{F} = \{X^{-1}(E) : E \in \tilde{\mathcal{F}}\}$.

Dim.

1. $\Omega = X^{-1}(\tilde{\Omega}) \Rightarrow \Omega \in \mathcal{F}$
2. $X^{-1}(\tilde{E}^c) = X^{-1}(\tilde{\Omega} \backslash \tilde{E}) = \Omega \backslash X^{-1}(\tilde{E}) = (X^{-1}(\tilde{E}))^c \Rightarrow \mathcal{F}$ è chiuso rispetto al complementare
3. $X^{-1}(\bigcup_{i=1}^{+\infty} \tilde{E}_i) = \bigcup_{i=1}^{+\infty} X^{-1}(\tilde{E}_i)$

Teorema 5.1.2. Sia $(\Omega, \mathcal{F})$ uno spazio probabilizzante. Siano $\tilde{\Omega}$ un insieme e $X : \Omega \to \tilde{\Omega}$ una funzione. Allora $\tilde{\mathcal{F}} = \{\tilde{E} \subseteq \tilde{\Omega} : X^{-1}(\tilde{E}) \in \mathcal{F}\}$ è una tribù.

Dim.

1. $\Omega = X^{-1}(\tilde{\Omega}) \Rightarrow \tilde{\Omega} \in \tilde{\mathcal{F}}$
2. Sia $\tilde{E} \in \tilde{\mathcal{F}}$, allora $X^{-1}(\tilde{E}) \in \mathcal{F}$, ma $\mathcal{F}$ è una tribù quindi $(X^{-1}(\tilde{E}))^c \in \mathcal{F} \Rightarrow \Omega \backslash X^{-1}(\tilde{E}) \in \mathcal{F} \Rightarrow X^{-1}(\tilde{E}^c) \in \mathcal{F} \Rightarrow \tilde{E}^c \in \tilde{\mathcal{F}}$
3. Se abbiamo $\{\tilde{E}_i\} \subseteq \tilde{\mathcal{F}}$, allora $\{X^{-1}(\tilde{E}_i\}_i \subseteq \mathcal{F} \Rightarrow \mathcal{F} \supseteq \bigcup_{i=1}^{+\infty} X^{-1}(\tilde{E}_i) = X^{-1}(\bigcup_{i=1}^{+\infty} \tilde{E}_i)$ e quindi $\bigcup_{i=1}^{+\infty} \tilde{E}_i \in \tilde{\mathcal{F}}$

Oss. A noi interessa un caso speciale, il caso, cioè, in cui $\tilde{\Omega} = \mathbb{R}$, $\tilde{\mathcal{F}} = \mathcal{B} \Rightarrow X : \Omega \to \mathbb{R}$.

In questo conteso:

$$\sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}\}$$

è una tribù detta *tribù generata da* X e inoltre $\sigma(X) \subseteq \mathcal{F}$. Per tutti gli eventi in $\mathcal{F}$ abbiamo un probabilità nel momento in cui aggiungiamo un P a $(\Omega, \mathcal{F})$, quindi ogni elemento $\sigma(X)$ ha una probabilità. Allo stesso tempo, $\sigma(X)$ contiene gli eventi che hanno a che fare con X.

NB. Possiamo portare la probabilità da $(\Omega, \mathcal{F}, P)$ a $(\mathbb{R}, \mathcal{B}, \bullet)$.

Teorema 5.1.3. Se $X : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ è una funzione *continua* e *monotona* (crescente o decrescente), allora è una *variabile aleatoria.*

Esempio Alberto e Barbara lanciano una moneta a turno finché qualcuno non ottiene testa e vince. Con che probabilità vince Alberto?

Lo spazio di probabilità è $(\Omega, \mathcal{F}, P)$ dove:

- $\Omega = \{T, C\}^{\mathbb{N}}$ in quanto ho potenzialmente infiniti lanci da fare
- $\mathcal{F}$ è la tribù generata dai cilindri
- P è la probabilità prodotto

Considero come X la funzione che indica qual è il primo lancio in cui esce testa in modo che si possa poi calcolare la probabilità che X appartenga all'insieme dei numeri dispari (ipotizzo che prima lancia Alberto e poi Barbara), ovvero $P(X \in \{2 \cdot k + 1,\ k \in \mathbb{N}\})$:

$$X(\omega) = inf\{i \geq 1 : \omega_i = T\}$$

Com'è fatto $\sigma(X)$? Provo a fissare il numero di tentativi, ad esempio suppongo che esca testa al 4^o lancio:

$$H_4 := X^{-1}(\{4\}) = \{\omega \in \Omega : \omega_1 = \omega_2 = \omega_3 = C,\ \omega_4 = T\} \text{ } (H_4 \text{ è un cilindro})$$

In generale $\sigma(X)$ è costituita da unioni numerabili o complementari di cilindri nella forma:

$$H_k := \{\omega \in \Omega : \omega_i = C\ i < k,\ \omega_k = T\}$$

Quindi, la probabilità che X sia dispari è:

$$P(\text{X è dispari}) = \sum_{i=0}^{+\infty} P(\omega \in H_{2\cdot i+1}) = \sum_{i=1}^{+\infty} \frac{1}{2^{2\cdot i+1}} = \sum_{i=0}^{+\infty} \frac{1}{4^i \cdot 2} = \frac{1}{2} \sum_{i=0}^{+\infty} 4^{-i} = \frac{1}{2} \cdot \frac{1}{1 - \frac{1}{4}} = \frac{2}{3}$$

Oss. Potevamo notare che $P(\text{X è pari}) + P(\text{X è dispari}) = 1 \Rightarrow P(\text{X è pari}) = \frac{1}{2} \cdot P(\text{X è dispari})$

5.2 Tipi di variabili aleatorie

Le *variabili aleatorie* sono funzioni da uno spazio probabilizzabile, ad $\mathbb{R}$ con la tribù dei *Boreliani*: $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$.

5.2.1 Variabili aleatorie degeneri

Sia $c \in \mathbb{R}$. La funzione $X(\omega) = c\ \forall \omega \in \Omega$ è una variabile aleatoria detta essere degenere.

Esempio Possiamo chiedere:

$$P(X = a) = \begin{cases} 1 & \text{se } a = c \\ 0 & \text{se } a \neq c \end{cases}$$

Se $A \in \mathcal{B}$, $P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\}) = P(X^{-1}(A)) = \begin{cases} 1 & \text{se } c \in A \\ 0 & \text{se } c \notin A \end{cases}$

Qual è la tribù $\sigma(X)$ generata da X?

$$\sigma(X) = \{\emptyset, \Omega\} \Rightarrow X^{-1}(A) = \begin{cases} \Omega & \text{se } c \in A \\ \emptyset & \text{se } c \notin A \end{cases}$$

5.2.2 Variabili aleatorie indicatrici

Siano $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità ed $E \in \mathcal{F}$ un evento. La *variabile aleatoria indicatrice* di E che si indica con I_E o $\mathbb{1}_E$ è definita così:

$$I_E(\omega) = \mathbb{1}_E(\omega) = \begin{cases} 1 & \text{se } \omega \in E \\ 0 & \text{se } \omega \notin E \end{cases}$$

Com'è fatta $\sigma(I_E)$?

$$\sigma(I_E) = \{\emptyset, \Omega, E = I_E^{-1}(\{1\}), E^C = I_E^{-1}(\{0\})\}$$

Ogni $B \in \mathcal{B}$ ha:

$$I_E^{-1} = \begin{cases} \emptyset & \text{se } 0 \notin B \land 1 \notin B \\ \Omega & \text{se } 0 \in B \land 1 \in B \\ E & \text{se } 0 \notin B \land 1 \in B \\ E^c & \text{se } 0 \in B \land 1 \notin B \end{cases} \text{ e } P(I_E^{-1} \in B) = \begin{cases} P(\emptyset) = 0 & \text{se } 0 \notin B \land 1 \notin B \\ P(\Omega) = 1 & \text{se } 0 \in B \land 1 \in B \\ P(E) & \text{se } 0 \notin B \land 1 \in B \\ P(E^c) & \text{se } 0 \in B \land 1 \notin B \end{cases}$$

5.2.3 Variabili aleatorie semplici: combinazioni lineari di v.a. indicatrici

Siano $E, F \in \mathcal{F}$ e $X = I_E - 3I_F$

$$X(\omega) = \begin{cases} 0 & \text{se } \omega \notin E \cup F \\ 1 & \text{se } \omega \in E \cap F^c \\ -3 & \text{se } \omega \in F \cap E^c \\ -2 & \text{se } \omega \in E \cap F \end{cases}$$

Chi è $\sigma(X)$?

$$\sigma(X) = \sigma(E, F) =$$

$$\{\emptyset, E, F, E^c, F^c, E \cap F, (E \cap F)^c, E \cup F, (E \cup F)^c, E \cap F^c, E^c \cup F, E^c \cap F, E \cup F^c, E \triangle F, (E \triangle F)^c\}$$

Per $A \in \mathcal{B}$, quanto vale $P(X \in A)$?

Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità. $X : \Omega \to \mathbb{R}$, $A \in \mathcal{B}$, allora:

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

A è un Boreliano, quindi lo possiamo generare a partire dalle semirette nella forma $(-\infty, a]$ con $a \in \mathbb{R}$, quindi grazie alla richiesta che $X \leq a$ ($X \in (-\infty, a]$) sia in $\mathcal{F}$ (nella definizione di v.a.) anche $X^{-1}(A) \in \mathcal{F}$.

NB. Avremmo potuto definire $P(A \in X)$ per casi come nell'esempio precedente, ma ce ne sarebbero stati molti.

Definizione 5.2.1. Dati uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ e una variabile aleatoria $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$. Si dice *legge* o *distribuzione* di X, la funzione di probabilità $P(X)$ definita su $(\mathbb{R}, \mathcal{B})$ da $P_X(A) \coloneqq P(X \in A) = P(X^{-1}(A))$.

Esempio In un esperimento aleatorio vengono lanciati due dadi a 6 facce e poi viene calcolata la somma tra i valori usciti. Qual è la probabilità dei vari risultati?

Possiamo definire una variabile aleatoria S che tracci i risultati dell'esperimento aleatorio. Se per esempio volessi calcolare la probabilità che la somma dei risultati sia 7, ci basterebbe calcolare $P(S = 7)$. Inoltre, per la definizione precedente ci permette di scrivere $P_S(\{7\})$.

Definizione 5.2.2. Siano $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ e $Y : (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P}) \to (\mathbb{R}, \mathcal{B})$ due variabili aleatorie. Siano P_X e P_Y le loro leggi:

$$P_X(\bullet) = P(X^{-1}(\bullet)),\ \tilde{P}_Y(\bullet) = \tilde{P}_Y(Y^{-1}(\bullet))$$

Se P_X e P_Y sono uguali (cioè $\forall A \in \mathcal{B}\ P_X(A) = P_Y(A)$) diciamo che X e Y sono identicamente stabilite e scriveremo: $X \sim Y$

Esempio C'è un'urna con 50 biglie bianche e 50 biglie nere. X è la variabile aleatoria indicatrice di *"è uscita una biglia bianca"*, allora $\forall A \in \mathcal{B}$:

$$P_X(A) = \begin{cases} 1 & \text{se } \{0,1\} \subseteq A \\ \frac{1}{2} & \text{se } \#(\{0,1\} \cap A) = 1 \\ 0 & \text{se } \{0,1\} \cap A = \emptyset \end{cases}$$

Oss. Quindi, con le *variabili aleatorie* ci portiamo sempre su $(\mathbb{R}, \mathcal{B})$. Dobbiamo definire le leggi su $(\mathbb{R}, \mathcal{B})$ e per farlo, basta sfruttare i generatori di $\mathcal{B}$ e, in particolare, le semirette $(-\infty, a]$ al variare di $a \in \mathbb{R}$. Di conseguenza, possiamo definire le leggi tramite funzioni su $\mathbb{R}$.

5.3 Funzione di ripartizione

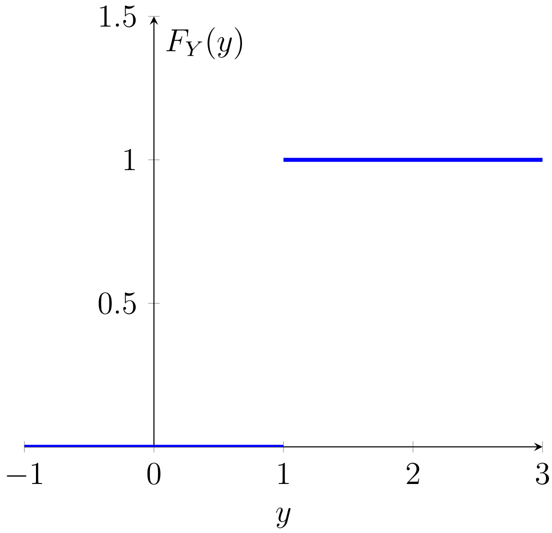
Definizione 5.3.1. Data una *variabile aleatoria* X su $(\Omega, \mathcal{F}, P)$ la *funzione di ripartizione* o *funzione cumulativa* (cdf) di X è la funzione $F_X : \mathbb{R} \to \mathbb{R}$ definita $\forall y \in \mathbb{R}$, come:

$$F_X(y) \coloneqq P_X((-\infty, y]) = P(X \in (-\infty, y]) = P(\{\omega \in \Omega | X(\omega) \leq y\}) = P(X \leq y)$$

Scriviamo $X \sim F_X$ per dire che X ha *cdf* F_X.

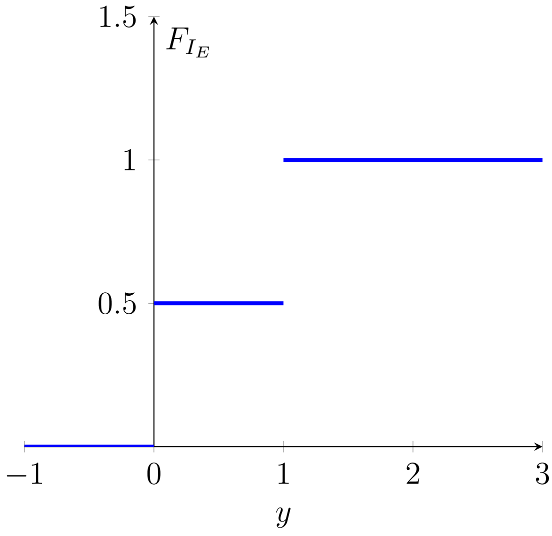
Esempio Sia $X = c$ una *variabile aleatoria degenere.* La sua funzione di ripartizione è:

$$F_X(y) = P(X \leq y) = \begin{cases} 0 & \text{se } y < c \\ 1 & \text{se } y \geq c \end{cases}$$



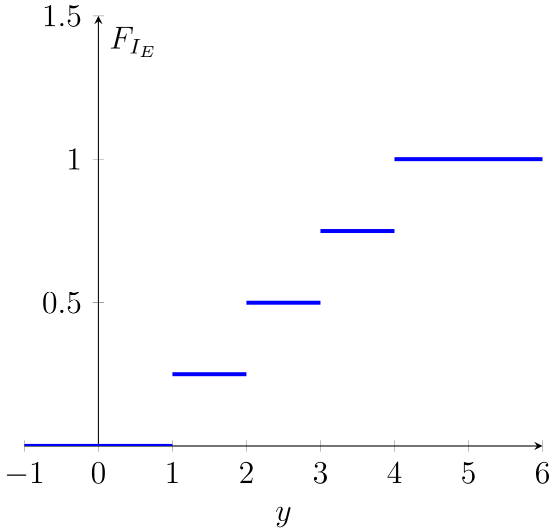
Esempio Fissato un evento $E \in \mathcal{F}$ consideriamo la sua *variabile aleatoria indicatrice* I_E:

$$I_E(\omega) = \begin{cases} 0 & \text{se } \omega \in E^c \\ 1 & \text{se } \omega \in E \end{cases} \Rightarrow F_{I_E}(y) = P(I_E \leq y) = \begin{cases} 0 & \text{se } y < 0 \\ P(E^c) & \text{se } 0 \leq y < 1 \\ 1 & \text{se } y \geq 1 \end{cases}$$



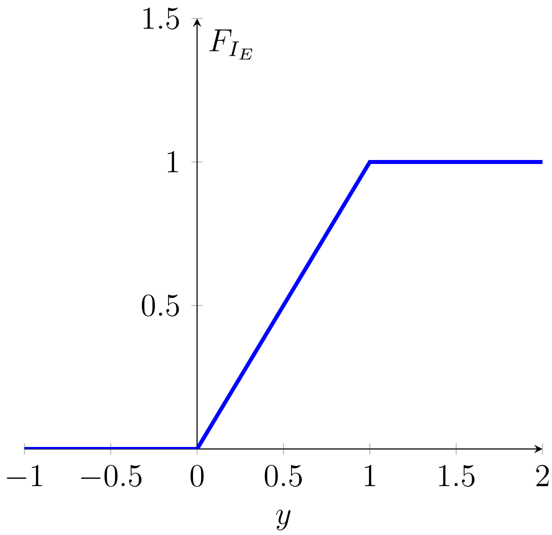
Esempio Sia $D4$ un dado a 4 facce, e sia F_{D4} la sua *funzione di ripartizione*:

$$F_{D4}(y) = P(D4 \leq y) = \begin{cases} 0 & \text{se } y < 1 \\ \frac{1}{4} & \text{se } 1 \leq y < 2 \\ \frac{2}{4} & \text{se } 2 \leq y < 3 \\ \frac{3}{4} & \text{se } 3 \leq y < 4 \\ 1 & \text{se } y \geq 4 \end{cases}$$



Esempio Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità con $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$ e $P([a, b]) = b - a$. Prendiamo $X = Id : [0, 1] \to \mathbb{R}$.

$$F_X(y) = P(X \leq y) = P(\{\omega \in \Omega : X(\omega) \leq y\}) = \begin{cases} P(\emptyset) & \text{se } y < 0 \\ P([0, y]) = y & \text{se } 0 \leq y < 1 \\ P(\Omega) & \text{se } y \geq 1 \end{cases}$$



5.3.1 Proprietà della funzione di ripartizione

Sia X una *variabile aleatoria* e sia F_X la sua *funzione di ripartizione*. Per F_X valgono le seguenti proprietà:

(i) Non è decrescente

(ii) Esistono i limiti per $x \to +\infty$ e $x \to -\infty$:

  (a) $\lim_{x\to-\infty} F_X(x) = 0$

  (b) $\lim_{x\to+\infty} F_X(x) = 1$

(iii) È *cadlag*, cioè continua a destra ($\lim_{x\to x_0^+} F_X(x) = F_X(x_0)$) e limitata a sinistra ($\lim_{x\to x_0^-} F_X(x) = F_X(x_0) - P(X = x_0)$)

Dim.

(i) Siano $s, t \in \mathbb{R}$ con $s \leq t$. È vero che $F_X(s) \leq F_X(t)$?
$F_X(s) = P(X \leq s) = P(\{\omega \in \Omega : X(\omega) \leq \Omega\}) \leq P(\{\omega \in \Omega : X(\omega) \leq t\}) = P(X \leq t) = F_X(t)$

(ii) La dimostrazione è immediata dalla definizione

(iii) $\lim_{x\to x_0^+} F(x) = \lim_{x\to x_0^+} P(X \leq x) = P(\bigcup_{n=1}^{+\infty}\{X \leq x_0 + \frac{1}{n}\}) = P(X \leq x_0) = F_X(x_0)$
$\lim_{x\to x_0^+} F_X(x) \lim_{x\to x_0^-} P(X \leq x) = P(\bigcup_{n=1}^{+\infty}\{X \leq x_0 - \frac{1}{n}\}) = P(X < x_0) =$
$P(X \leq x_0) - P(X = x_0) = F_X(x_0) - P(X = x_0)$

Oss. Chiedere che F_X sia continua a destra in x_0 (cioè $\lim_{x\to x_0^+} F_X(x) = F(x_0)$) equivale a chiedere che $P(X = x_0) = 0$.

Oss. Data F_X posso calcolare la probabilità di un qualunque *Boreliano*:

- $P_X((a, b]) = P(X \in (a, b]) = P(X \in (-\infty, b]) - P(X \in (-\infty, a]) = F_X(b) - F_X(a)$
- $P_X((a, +\infty)) = 1 - F_X(a)$
- $P((a, b)) = \lim_{x\to b^-} F_X(x) - F_X(a) = F_X(b) - P(X = b) - F_X(a)$
- $P_X([a, +\infty)) = 1 - F_X(a) + P(X = a) = 1 - \lim_{x\to a^-} F_X(a)$

5.4 Classi di variabili aleatorie

5.4.1 Variabili aleatorie discrete

Definizione 5.4.1 (Variabili aleatorie discrete)**.** Una *variabile aleatoria* X si dice *discreta* se può assumere al più un numero finito e numerabile di valori.

Oss. Una *variabile aleatoria* X è discreta se e solo se la sua funzione di *ripartizione* F_X è discontinua e costante a tratti con un numero al più numerabile di discontinuità (punti di discontinuità).

Definizione 5.4.2. Sia X una *variabile aleatoria discreta*. Chiamiamo *densità discreta* o *funzione di massa di probabilità* (pmf) di X la funzione: $\varphi_X : \mathbb{R} \to [0, 1]$ *t.c.* $\varphi_X(x) = P(X = x)$. Inoltre, l'insieme $\mathcal{R}_X$ dei punti in cui $\varphi_X(x) \neq 0$ è detto *supporto di* φ_X ed ha cardinalità finita o numerabile:

$$\mathcal{R}_X = \{x_i\}_{i\in I}$$

Oss. Sia X una *variabile aleatoria discreta* e sia φ_X la sua *funzione di densità.* Valgono le seguenti proprietà:

(i) $\forall x \; \varphi_X(x) \geq 0$

(ii) $\forall x \in \mathcal{R}_X^c \; \varphi_X(x) = 0$

(iii) $\sum_{x \in \mathcal{R}_X} \varphi_X(x) = 1$

(iv) Se $A \in \mathcal{B}$ allora $P_X(A) = \sum_{x \in \mathcal{R}_X \cap A} \varphi_X(x) = \sum_{x \in \mathcal{R}_X} \mathbb{1}_A(x) \cdot \varphi_X(x)$

Dim. (i), (ii) seguono direttamente dalle definizioni, mentre (iii) segue da (iv).

(iv) $P_X(A) = P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\} \cap \Omega) = P(\{\omega \in \Omega : X(\omega) \in A\} \cap \bigcup_{x \in \mathcal{R}_X} \{\omega \in \Omega : X(\omega) = x\}) = P(\bigcup_{x \in \mathcal{R}_X} (\{\omega \in \Omega : X(\omega) \in A)\} \cap \{\omega \in \Omega : X(\omega) = x\})) = \sum_{x \in \mathcal{R}_X} P((X \in A) \cap (X = x)) = \sum_{x \in \mathcal{R}_X} \mathbb{1}_A(x) \cdot \varphi_X(x) = \sum_{x \in \mathcal{R}_X \cap A} \varphi_X(x)$

Se $A = \mathbb{R}$ segue che $\mathcal{R}_X \cap A = \mathcal{R}_X \cap \mathbb{R} = \mathcal{R}_X \Rightarrow \sum_{x \in \mathcal{R}_X \cap \mathbb{R}} \varphi_X(x) = \sum_{x \in \mathcal{R}_X} \varphi_X(x) = 1$.
Da questa catena di uguaglianze risulta dimostrata la proprietà (iii).

Oss. Data X *variabile aleatoria discreta* con *pmf* φ_X, la sua *funzione di ripartizione* è:

$$F_X(y) = \sum_{x \in \mathcal{R}_X} \mathbb{1}_{(-\infty, y)} \cdot \varphi_X(x)$$

5.4.2 Variabili aleatorie continue

Definizione 5.4.3 (Variabili aleatorie continue)**.** Una *variabile aleatoria* X si dice *continua* se la sua *funzione di ripartizione* F_X è continua. Se inoltre esiste una funzione non negativa $f_x : \mathbb{R} \to \mathbb{R}$ tale che:

$$F_X(x) = \int_{-\infty}^{x} f_X(y) dy \; \forall x \in \mathbb{R}$$

X si dice *assolutamente continua.*

Oss. Se X è una *variabile aleatoria continua*, $\varphi_X \equiv 0$, quindi non ci dà informazioni.

Definizione 5.4.4. Sia X una *variabile aleatoria assolutamente continua.* La funzione non negativa $f_X : \mathbb{R} \to \mathbb{R}$ tale che:

$$F_X(x) = \int_{-\infty}^{x} f_X(y) dy$$

prende il nome di *densità di probabilità* di X (pdf).
Possiamo definire il *supporto di* f_X come l'insieme:

$$\mathcal{R}_X = \{x : f_X(x) \neq 0\}$$

Oss. Sia X una *variabile aleatoria assolutamente continua* e sia f_X la sua *funzione di densità.* Valgono le seguenti proprietà:

(i) $\int_{-\infty}^{+\infty} f_X(x) dx = 1$

(ii) $\int_a^b f_X(x) dx = F_X(b) - F_X(a)$

Dim.

(i) $\int_{-\infty}^{+\infty} f_X(x) dx = \lim_{x \to +\infty} F_X(x) = 1$

(ii) $\int_a^b f_X(x) dx = \int_{-\infty}^{b} f_X(x) dx - \int_{-\infty}^{a} f_X(x) dx = F_X(b) - F_X(a)$

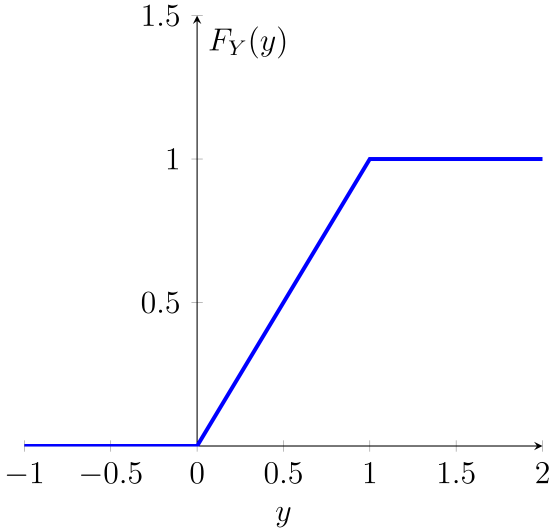
Oss. Nei punti in cui f_X è differenziabile, $F'_X(x) = f_X(x)$, e quindi:

$$P(x \in [x-\epsilon, x+\epsilon]) = \int_{x-\epsilon}^{x+\epsilon} f_X(y)dy \approx 2 \cdot \epsilon \cdot f_X(x)$$

Ci possono essere dei punti in cui f_X non è differenziabile. In questi punti non possiamo ricavare f_X univocamente da F_X. Questo non è un problema per il fatto che, quando parliamo di probabilità, abbiamo l'integrale f_X che ignora i singoli valori.

Esempio Sia X una *variabile aleatoria* uniforme su $[0, 1]$.

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ x & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases} \Rightarrow \text{ tralasciando 0 e 1 } \Rightarrow f_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 & \text{se } 0 < x < 1 \\ 0 & \text{se } x > 1 \end{cases}$$

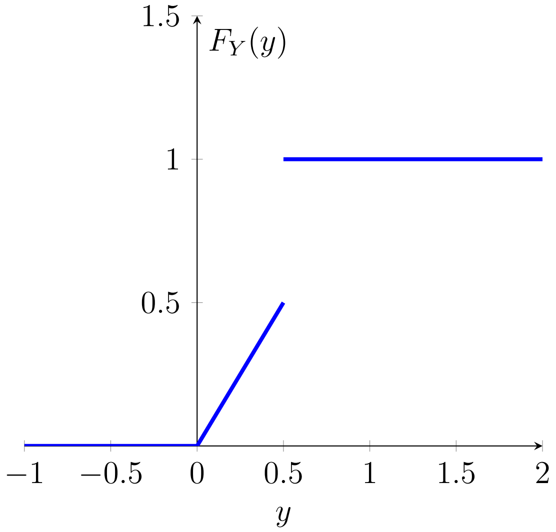


5.4.3 Variabili aleatorie miste

Definizione 5.4.5. Una *variabile aleatoria* che non è né *discreta* né *continua* si dice *mista*.

Esempio Si prenda la seguente *funzione di ripartizione*:

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ x & \text{se } 0 \leq x < \frac{1}{2} \\ 1 & \text{se } x \geq \frac{1}{2} \end{cases}$$



Dal grafico si vede che la funzione non è continua quindi la *variabile aleatoria* non è *continua*, inoltre la funzione non è nemmeno costante a tratti quindi non si tratta nemmeno di una *variabile aleatoria discreta*. Se ne deduce che la *variabile aleatoria* è *mista*.

5.5 Costante di rinormalizzazione

Si supponga di avere una *variabile aleatoria assolutamente continua*, con *densità* f. Perché f sia una *funzione di densità* deve essere non negativa e avere $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Se abbiamo una funzione f non negativa, ma di integrale finito diverso da 1, come possiamo ricavare una densità? Ovvero, esiste un $c \in \mathbb{R}$ tale che $c \cdot f$ sia una densità?

Innanzitutto, per la non negatività di f, possiamo fissare $c \geq 0$.

$$1 = \int_{-\infty}^{+\infty} c \cdot f(x)dx = c \cdot \int_{-\infty}^{+\infty} f(x)dx = c \cdot A \text{ con } A \in \mathbb{R}\backslash\{1\} \wedge A > 0$$

$$\Rightarrow c = \left(\int_{-\infty}^{+\infty} f(x)dx\right)^{-1}$$

Una tale costante prende il nome di *costante di rinormalizzazione*.

Esempio Sia $f(x) = e^{-x}$ per $x \in (0, 1)$. Come si può rendere f una *densità*?

(i) Devo controllare che sia non negativa: $e^{-x} > 0 \; \forall x \in \mathbb{R}$

(ii) Devo controllare che l'integrale sia finito: lo è (la dimostrazione è banale)

A questo punto ricavo la *costante di rinormalizzazione*:

$$c = \left(\int_{-\infty}^{+\infty} f(x)dx\right)^{-1} = \left(\int_{0}^{1} f(x)dx\right)^{-1} = ([-e^{-x}]_0^1)^{-1} = (1 - e^{-1})^{-1} = \frac{1}{1 - e^{-1}}$$

Ora posso definire le funzione di *densità* e *ripartizione*:

$$f_X(x) = \begin{cases} \frac{1}{1-e^{-1}} \cdot e^{-x} & \text{se } x \in (0,1) \\ 0 & \text{se } x \notin (0,1) \end{cases} \Rightarrow F_X(X) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{1-\frac{1}{e}} \cdot e^{-x} & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases}$$

Capitolo Nr.6

Trasformazioni di variabili aleatorie

6.1 Trasformazioni lineari

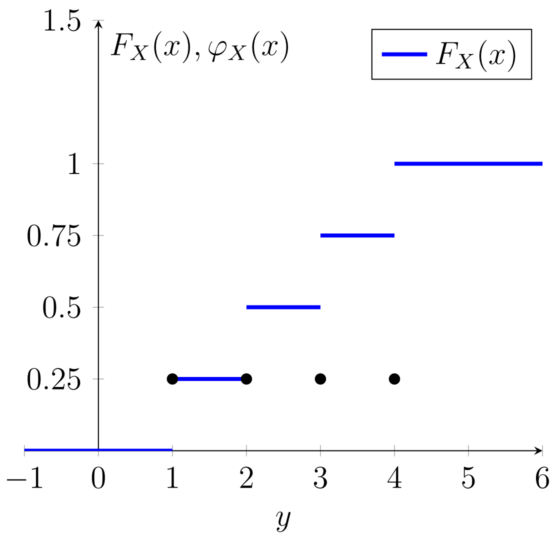
Esempio Siano, X una *variabile aleatoria* e Y una funzione lineare:

$$Y = 2 \cdot X + 3$$

Consideriamo ora il lancio di un dado a 4 facce, e poniamo che X ne tracci il risultato (X indica la faccia uscita). In particolare X in questo caso è una *variabile aleatoria discreta*.

$$\varphi_X(x) = P(X = x) = \begin{cases} \frac{1}{4} & \text{se } x \in \{1,2,3,4\} \\ 0 & \text{se } x \notin \{1,2,3,4\} \end{cases}, \quad F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{se } x < 1 \\ \frac{1}{4} & \text{se } 1 \leq x < 2 \\ \frac{2}{4} & \text{se } 2 \leq x < 3 \\ \frac{3}{4} & \text{se } 3 \leq x < 4 \\ 1 & \text{se } x \geq 4 \end{cases}$$

Il grafico delle funzioni di *massa* e *ripartizione* di X è il seguente:



Ma come sono definite le funzioni di *massa* e *ripartizione* di $Y = 2 \cdot X + 3$?

$$\varphi_Y(y) = P(Y = y) = P(2 \cdot X + 3 = y) = P\left(X = \frac{y-3}{2}\right) = \varphi_X\left(\frac{y-3}{2}\right)$$

$$F_Y(y) = P(Y \leq y) = P(2 \cdot X + 3 \leq y) = P\left(X \leq \frac{y-3}{2}\right) = F_X\left(\frac{y-3}{2}\right)$$

Esempio Sia X una *variabile aleatoria* uniforme su $[0,1]$ e valga $Y = 2 \cdot X + 3$. Come cambia f_Y?

Sappiamo che $\int_{-\infty}^{+\infty} f_X(x)dx = \int_{\mathcal{R}_X} f_X(x)dx = 1$, possiamo quindi aspettarci che anche per Y valga qualcosa del tipo $\int_{-\infty}^{+\infty} f_Y(y)dy = \int_{\mathcal{R}_Y} f_Y(y)dy = c$ per qualche $c \in \mathbb{R}$. Ma qual è il *supporto di* Y?

Il *supporto di* X è $\mathcal{R}_X = (0,1)$, quindi è ragionevole ipotizzare che il *supporto di* Y sia $\mathcal{R}_Y = (3,5)$.

Ora rimane da determinare il valore di c. È possibile che $c = 1$?

$$c = 1 = \int_{-\infty}^{+\infty} f_Y(y)dy = \int_3^5 f_Y(y)dy = \int_3^5 1dy = [y]_3^5 = 5 - 3 = 2$$

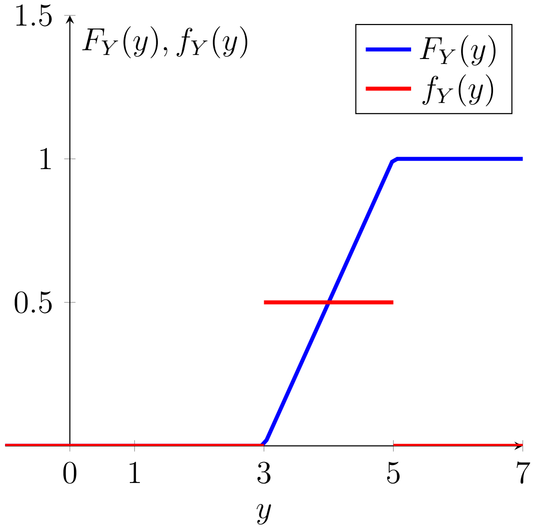
Questa catena di uguaglianze smentisce l'ipotesi che $c = 1$ e anzi dimostra che per questa particolare trasformazione $c = \frac{1}{2}$.

Rimangono ora da definire le *funzioni di massa* e di *ripartizione*:

$$F_Y(y) = P(Y \leq y) = P(2 \cdot X + 3 \leq y) = P\left(X \leq \frac{y-3}{2}\right) = F_X\left(\frac{y-3}{2}\right) =$$

$$\begin{cases} 0 & \text{se } \frac{y-3}{2} < 0 \\ \frac{y-3}{2} & \text{se } 0 \leq \frac{y-3}{2} < 1 \\ 1 & \text{se } \frac{y-3}{2} \geq 1 \end{cases} = \begin{cases} 0 & \text{se } y < 0 \\ \frac{y-3}{2} & \text{se } 3 \leq y < 5 \\ 1 & \text{se } y \geq 5 \end{cases}$$

$$f_Y(y) = F'_Y(y) = \begin{cases} 0 & \text{se } y \in [3,5]^c \\ \frac{1}{2} & \text{se } y \in [3,5] \end{cases}$$



Prop. Sia X una *variabile aleatoria* e Y una sua trasformazione lineare tale che:

$$Y = a \cdot X + b, \ a, b \in \mathbb{R} \ a \neq 0$$

La funzione è così definita:

$$F_Y(y) = \begin{cases} F_X\left(\frac{y-b}{a}\right) & \text{se } a > 0 \\ 1 - F_X\left(\frac{y-b}{a}\right) & \text{se } a < 0 \wedge X \text{ è continua} \\ 1 - F_X\left(\frac{y-b}{a}\right) + \varphi_X\left(\frac{y-b}{a}\right) & \text{se } a < 0 \wedge X \text{ è discreta} \end{cases}$$

Inoltre, se X è *assolutamente continua* la *funzione di densità* è:

$$f_Y(y) = \frac{1}{|a|} \cdot f_X\left(\frac{y-b}{a}\right)$$

Mentre, se X è *discreta* la *funzione di massa* è:

$$\varphi_Y(y) = \varphi_X\left(\frac{y-b}{a}\right)$$

Dim.
Funzione di ripartizione

$$F_Y(y) = P(Y \leq y) = P(a \cdot X + b \leq y) = P(A \cdot X \leq y - b) = \cdot$$

$$\text{Se } a > 0 \cdot = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

$$\text{Se } a < 0 \cdot = P\left(X \geq \frac{y-b}{a}\right) = 1 - P\left(X < \frac{y-b}{a}\right) = 1 - \lim_{x \to \left(\frac{y-b}{a}\right)^-} F_X(x) =$$

$$\begin{cases} 1 - F_X\left(\frac{y-b}{a}\right) & \text{se } X \text{ è continua} \\ 1 - F_X\left(\frac{y-b}{a}\right) + P\left(X = \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right) + \varphi_X\left(\frac{y-b}{a}\right) & \text{se } X \text{ è discreta} \end{cases}$$

Funzione di densità Se X è *continua* $f_X = F'_X$, quindi f_Y vale:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \begin{cases} \frac{d}{dy}F_X\left(\frac{y-b}{a}\right) & \text{se } a > 0 \\ \frac{d}{dy}\left(1 - F_X\left(\frac{y-b}{a}\right)\right) & \text{se } a > 0 \end{cases} = \begin{cases} F'_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a} & \text{se } a > 0 \\ -F'_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a} & \text{se } a < 0 \end{cases}$$

A questo punto si può notare che se $a < 0$ è possibile semplificare i due $-$ e ottenere:

$$f_Y(y) = \frac{1}{|a|} \cdot F'_X\left(\frac{y-b}{a}\right)$$

Funzione di massa Se X è *discreta* abbiamo:

$$\varphi_Y(y) = P(Y = y) = P(a \cdot X + b = y) = P\left(X = \frac{y-b}{a}\right) = \varphi_Y\left(\frac{y-b}{a}\right)$$

6.2 Trasformazioni non lineari

Sia X una *variabile aleatoria* con legge F_X e, con *densità discreta* φ_X se X *v.a. discreta*, o *denstià* f_X se *v.a. continua*. Sia poi $g : \mathbb{R} \to \mathbb{R}$ una funzione generica. Come si definisce la legge di $Y = g(X)$?

NB. La funzione g non è necessariamente iniettiva o suriettiva.

6.2.1 Variabili aleatorie discrete

Se X è una *variabile aleatoria discreta* possiamo definire la sua funzione di massa come segue:

$$\varphi_Y(y) = \sum_{x \in g^{-1}(\{y\})} \varphi_X(x)$$

Il *supporto* di Y è $\mathcal{R}_Y = g(\mathcal{R}_X)$ e la *funzione di ripartizione* si ottiene sommando le φ_Y:

$$F_Y(y) = \sum_{y \in \mathcal{R}_Y} \mathbb{1}_{(-\infty, y)} \cdot \varphi_Y(y)$$

6.2.2 Variabili aleatorie assolutamente continue

Se X è una *variabile aleatoria assolutamente continua* un possibile modo per ricavare la legge di $Y = g(X)$ è quello di usare la forma di X, cioè di F_X, e della funzione g.

Esempio Sia X una *variabile aleatoria assolutamente continua* di *densità*:

$$f_X(x) = \begin{cases} e^{-x} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

Se $g : \mathbb{R} \to \mathbb{R}$ è definita come $g = e^{-x}$, qual è la legge di $Y = e^{-X}$?

Inizio studiando X per ricavarne la legge.

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{0}^{x} f_X(t)dt = \int_{0}^{x} e^{-t}dt = [-e^{-t}]_0^x = 1 - e^{-x} \Rightarrow \begin{cases} 0 & \text{se } x \leq 0 \\ 1 - e^{-x} & \text{se } x > 0 \end{cases}$$

A questo punto posso passare allo studio di Y:

$$F_Y(y) = P(Y \leq y) = P(e^{-X} \leq y) = P(-X \leq \ln(y)) \Rightarrow \begin{cases} 0 & \text{se } y \leq 0 \\ P(-X \leq \ln(y)) & \text{se } y > 0 \end{cases} \Rightarrow$$

$$\begin{cases} 0 & \text{se } x \leq 0 \\ P(X \geq \ln(y)) \text{ se} & 0 < x < 1 \\ 1 & \text{se } x \geq 1 \end{cases} = \begin{cases} 0 & \text{se } x \leq 0 \\ 1 - F_X(\ln(y)) & \text{se } 0 < x < 1 \\ 1 & \text{se } x \geq 1 \end{cases} =$$

$$\begin{cases} 0 & \text{se } x \leq 0 \\ 1 - (1 - e^{-(\ln(y))}) & \text{se } 0 < x < 1 \\ 1 & \text{se } x \geq 1 \end{cases} = \begin{cases} 0 & \text{se } x \leq 0 \\ y & \text{se } 0 < x < 1 \\ 1 & \text{se } x \geq 1 \end{cases}$$

Oss. In questo caso Y è una *variabile aleatoria* uniforme su $[0, 1]$.

Ora che abbiamo la legge di Y possiamo anche ricavarne la *densità* f_Y:

$$f_Y(y) = F_Y'(y) = \frac{d}{dy}(1 - f_X(-\ln(y))) = -F_X'(-\ln(y)) \cdot \frac{d}{dy}(-\ln(y)) = -f_X(-\ln(y)) =$$

$$\frac{1}{y} \cdot e^{-(-\ln(y))} \text{ se } \ln(y) < 0 \Rightarrow \begin{cases} 0 & \text{se } y \notin (0, 1) \\ \frac{1}{y} \cdot y & \text{se } y \in (0, 1) \end{cases}$$

Da questa forma è facile passare alla forma della *densità* unforme su $[0, 1]$:

$$f_Y(y) = \begin{cases} 0 & \text{se } y < 0 \\ 1 & \text{se } 0 < y < 1 \\ 0 & \text{se } y > 1 \end{cases}$$

Esempio Sia X una *varibile aleatoria assolutamente continua* con *funzione di densità*:

$$f_X(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ e^{-x} & \text{se } x > 0 \end{cases}$$

Sia poi, $Z = g(x) = (1-x)^2$. Qual è la legge di Z?

Inizio definendo la *funzione di ripartizione* di X:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_0^x e^{-x} = [-e^{-t}]_o^x = 1 - e^{-x} \Rightarrow \begin{cases} 0 & \text{se } x \leq 0 \\ 1 - e^{-x} & \text{se } x > 0 \end{cases}$$

Passo, ora, allo studio di Z:

$$F_Z(z) = P(Z \leq z) = P((1-x)^2 \leq z) = \begin{cases} 0 & \text{se } x < 0 \\ P((1-x)^2 \leq z) & \text{se } x \geq 0 \end{cases} \Rightarrow$$

$$P(|1-x| \leq \sqrt{z}) = P(-\sqrt{z} \leq 1 - x \leq \sqrt{z}) = P(-1-\sqrt{z} \leq -x \leq -1+\sqrt{z}) =$$
$$P(1+\sqrt{z} \geq x \geq 1-\sqrt{z}) = F_X(1+\sqrt{z}) - F_X(1-\sqrt{z}) =$$
$$(1 - e^{-(1+\sqrt{z})}) \cdot \mathbb{1}_{1+\sqrt{z}>0} - (1 - e^{(1-\sqrt{z})}) \cdot \mathbb{1}_{1-\sqrt{z}>0} \Rightarrow$$

$$F_Z(z) = \begin{cases} 0 & \text{se } x < 0 \\ e^{-(1-\sqrt{z})} - e^{-(1+\sqrt{z})} & \text{se } 0 \leq z < 1 \\ 1 - e^{-(1+\sqrt{z})} & \text{se } z \geq 1 \end{cases}$$

Adesso, posso ricavare la *densità* f_Z:

$$f_Z(z) = F_Z'(z) = \frac{d}{dz}(e^{-(1-\sqrt{z})} - e^{-(1+\sqrt{z})}) = f_X(1+\sqrt{z}) \cdot \frac{d}{dz}(1+\sqrt{z}) - f_X(1-\sqrt{z}) \cdot \frac{d}{dz}(1-\sqrt{z}) =$$
$$\frac{1}{2 \cdot \sqrt{z}} \cdot (f_X(1+\sqrt{z}) + f_X(1-\sqrt{z})) \Rightarrow$$

$$f_Z(z) = \begin{cases} 0 & \text{se } z < 0 \\ \frac{1}{2 \cdot \sqrt{z}} \cdot (e^{-(1+\sqrt{z})} + e^{-(1-\sqrt{z})}) & \text{se } 0 < z < 1 \\ \frac{1}{2 \cdot \sqrt{z}} \cdot e^{-(1+\sqrt{z})} & \text{se } z < 1 \end{cases}$$

Oss. Questo metodo è particolarmente vantaggioso nel caso in cui la *variabile* X e la funzione g abbiano buone proprietà, ma la costruzione va "inventata" ogni volta, per cui è facile sbagliare.

Teorema del cambio di variabile

Un altro modo per ricavare la legge di $Y = g(X)$ sfrutta il *Teorema del cambio di variabile.*

Teorema 6.2.1 (Teorema del cambio di variabile)**.** Sia X una *variabile aleatoria assolutamente continua* di *densità* f_X. Valga poi, $Y = g(x)$ con $g : \mathbb{R} \to \mathbb{R}$, continua a tratti e tale per cui $P(g(x) = 0) = 0$. Allora

$$f_Y(y) = \sum_{x \in g^{-1}(\{y\})} \frac{f_X(x)}{|g'(x)|}$$

Oss. Cosa significa $P(g(x) = 0) = 0$?

$$\{g'(x) = 0\} = \{\omega \in \Omega : g'(X(\omega) = 0)\} = \bigcup_{x : g(x) = 0} \{\omega \in \Omega : \omega \in X^{-1}(\{x\})\}$$

Oss. Perchè si calcola una somma invece di un integrale?

Poiché $x \in g^{-1}(\{y\}) = \{x : g(x) = y\}$, siccome $g'(x) = 0$ in un numero al più numerabile di punti, l'insieme $\{x \in g^{-1}(\{y\})\}$ ha al più un numero numerabile di elementi, quindi è possibile fare una somma.

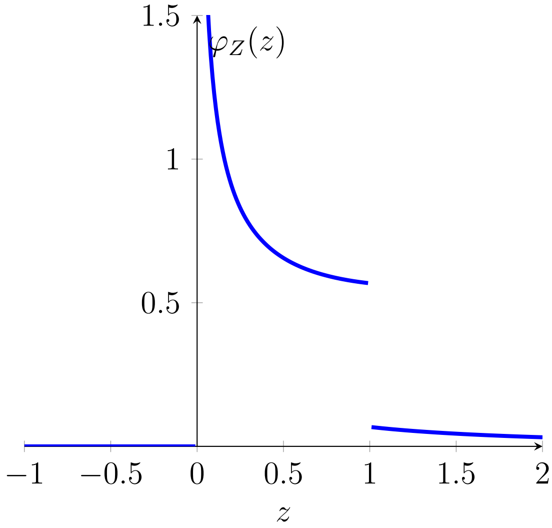
Esempio Siano X come negli esempi precedenti, $g(x) = (1-x)^2$ e $Z = g(X)$.

La derivata di $g(x)$ $g'(x) = -2 \cdot (1-x) = 2 \cdot (x-1)$ si annulla solo in $x = 1$. Vediamo $g^{-1}(\{z\})$ al variare di $z \in \mathbb{R}$:

$$g^{-1}(\{z\}) = \begin{cases} \emptyset & \text{se } x < 0 \\ \{1\} & \text{se } z = 0 \\ \{1-\sqrt{z}, 1+\sqrt{z}\} & \text{se } z > 0 \end{cases}$$

A questo punto, possiamo definire f_Z:

$$f_Z(z) = \begin{cases} 0 & \text{se } z < 0 \\ \frac{f_X(1)}{|g'1|} & \text{se } z = 0 \\ \frac{f_X(1+\sqrt{z})}{g'(1+\sqrt{z})} + \frac{f_X(1-\sqrt{z})}{|g'(1-\sqrt{z})|} & \text{se } 0 < z < 1 \\ \frac{f_X(1+\sqrt{z})}{g'(1+\sqrt{z})} & \text{se } z > 1 \end{cases} = \begin{cases} 0 & \text{se } z < 0 \\ \frac{e^{-(1+\sqrt{z})}}{2\cdot\sqrt{z}} + \frac{e^{-(1+\sqrt{z})}}{2\cdot\sqrt{z}} & \text{se } 0 < x < 1 \\ \frac{e^{-(1+\sqrt{z})}}{2\cdot\sqrt{z}} & \text{se } z > 1 \end{cases}$$



Capitolo Nr.7

Vettori aleatori

7.1 Coppie di variabili aleatorie

Dato uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ si considerino due *variabili aleatorie* $X, Y : \Omega \to \mathbb{R}$. Come si calcola la probabilità di eventi che coinvolgono entrambe le *variabili aleatorie*? Ad esempio $P(X \leq Y)$? Dobbiamo considerare le coppie $(X, Y) : \Omega \to \mathbb{R}^2$.

Definizione 7.1.1. Dati, uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ e X, Y *variabili aleatorie* su di esso si chiama *coppia di variabili aleatorie*, *variabile aleatoria doppia* o *2-vettore aleatorio*, la funzione:

$$V : \Omega \to \mathbb{R}^2 \ t.c. \ V(\omega) = (X(\omega), Y(\omega))$$

Il *supporto* del *vettore aleatorio* V è definito come:

$$\mathcal{R}_V = \mathcal{R}_{X,Y} = \mathcal{R}_X \times \mathcal{R}_Y = \{(X, Y) \in \mathbb{R}^2 : x \in \mathcal{R}_X, y \in \mathcal{R}_Y\}$$

Definizione 7.1.2. Sia (X, Y) una *coppia di variabili aleatorie* definite sullo stesso spazio di probabilità, la sua *funzione di probabilità* è:

$$F_{X,Y}((X, Y)) = F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

$F_{X,Y}$ si chiama anche *funzione di ripartizione congiunta* di X, Y.

Oss. Di solito, note F_X e F_Y non è possibile ricavare $F_{X,Y}$.

Viceversa, nota $F_{X,Y}$ possiamo ricavare F_X e F_Y, che in questo caso prendono il nome di *marginalità*:

$$F_X(x) = P(X \leq x) = P(X \leq x, \forall Y) = P(X \leq x, Y < +\infty) = \lim_{y \to +\infty} F_{X,Y}(x, y)$$

$$F_Y(y) = P(Y \leq y) = P(\forall X, Y \leq y) = P(X < +\infty, Y \leq y) = \lim_{x \to +\infty} F_{X,Y}(x, y)$$

Definizione 7.1.3. Data (X, Y) *coppia di variabile aleatorie*, chiamiamo *funzione di ripartizione* di X condizionata Y la funzione:

$$F_{X|Y}(x|y) = \frac{F_{X,Y}(x, y)}{F_Y(y)}$$

Oss.

$$F_{X|Y}(x, y) = \frac{P(X \leq x, Y \leq y)}{P(Y \leq y)} = P(X \leq x | Y \leq y)$$

Definizione 7.1.4.

(i) Dati uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ e due sue tribù $\mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{F}$, $\mathcal{F}_1$ e $\mathcal{F}_2$ sono indipendenti se ogni evento in $\mathcal{F}_1$ è indipendente da ogni evento in $\mathcal{F}_2$

(ii) Date due *variabili aleatorie* X, Y su $(\Omega, \mathcal{F}, P)$, diciamo che X e Y sono indipendenti se lo sono le tribù $\sigma(X)$ e $\sigma(Y)$ da esse generate

Prop. X e Y sono indipendenti se e solo se $\forall (x,y) \in \mathbb{R}^2 \; F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$.

Oss. Nel caso di *n-vettori aleatori* per verificare l'indipendenza delle *n variabili aleatorie* è necessario controllare tutti i possibili raggruppamenti.

7.2 Vettori aleatori discreti

Nei vettori (in realtà stiamo parlando di *2-vettori*) tutte le *variabili aleatorie* sono *discrete.*

Definizione 7.2.1. Siano X, Y *variabili aleatorie discrete* su $(\Omega, \mathcal{F}, P)$, chiamiamo *densità discreta congiunta* la funzione $\varphi_{X,Y} : \mathbb{R}^2 \to [0,1]$ definita come:

$$\varphi_{X,Y}(x,y) = P(X = x, Y = y)$$

La densità discreta di X condizionata Y è:

$$\varphi_{X|Y}(x|y) = \begin{cases} 0 & \text{se } y \notin \mathcal{R}_Y \Leftrightarrow P(Y = y) = 0 \\ P(X = x | Y = y) & \text{se } y \in \mathcal{R}_Y \Leftrightarrow P(Y = y) \neq 0 \end{cases}$$

Oss. Valgono le seguenti proprietà:

(i) $\forall (x,y) \in \mathbb{R}^2 \; 0 \leq \varphi_{X,Y}(x,y) \leq 1$

(ii) $\varphi_{X,Y}(x,y) = 0 \; \forall x \in \mathcal{R}_X^c \vee \forall y \in \mathcal{R}_Y^c$

(iii) $\sum_{x,y \in \mathbb{R}^2} \varphi_{X,Y}(x,y) = \sum_{(x,y) \in \mathcal{R}_X \times \mathcal{R}_Y} \varphi_{X,Y}(x,y) = 1$

Prop. Sia (X, Y) una *coppia di variabili aleatorie discrete*:

(i) $\forall (x,y) \in \mathbb{R}^2 \; F_{X,Y}(x,y) = \sum_{(x,y) \in \mathcal{R}_X \times \mathcal{R}_Y} \varphi_{X,Y}(\xi, \eta \in \mathcal{R}_X \times \mathcal{R}_Y) \varphi_{X,Y}(\xi, \eta) \cdot \mathbb{1}_{\xi \leq \eta} \cdot \mathbb{1}_{\eta \leq \xi}$

(ii) $\varphi_{X,Y}(x,y) = \varphi_{X,Y}(x|y) \cdot \varphi_Y(y)$

(iii) $\forall x \in \mathbb{R} \; \varphi_X(x) = \sum_{y \in \mathcal{R}_Y} \varphi_{X,Y}(x,y)$

(iv) X, Y sono indipendenti se e solo se $\varphi_{X,Y}(x,y) = \varphi_X(x) \cdot \varphi_Y(y)$

(v) X, Y sono indipendenti se e solo se $\varphi_X(x) = \varphi_{X|Y}(x|y)$ e $\varphi_Y(y) = \varphi_{Y|X}(y|x)$

Dim.

(i) Dalle definizioni

(ii) Se $y \notin \mathcal{R}_Y$ l'identità è verificata da $0 = 0$
Se $y \in \mathcal{R}_Y \; \varphi_{X|Y}(x|y) \cdot \varphi_Y(y) = P(X = x | Y = y) \cdot P(Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)} \cdot P(Y = y) = P(X = x, Y = y) = \varphi_{X,Y}(x,y)$

(iii) $\sum_{y \in \mathcal{R}_Y} \varphi_{X,Y}(x,y) = \sum_{y \in \mathcal{R}_Y} \varphi_{Y|X}(y|x) \cdot \varphi_X(x) = \varphi_X(x) \cdot \sum_{y \in \mathcal{R}_Y} \varphi_{Y|X}(y|x) = \varphi_X(x) \cdot \sum_{y \in \mathcal{R}_Y} P(Y = y | X = x) = \varphi_X(x) \cdot 1 = \varphi_X(x)$

(iv) [$\Rightarrow$] dalle definizioni
[$\Leftarrow$] uso la i e l'analogo risultato visto per le funzioni di ripartizione

(v) Segue da *ii* e *iv*

Esempio Siano X, Y *variabili aleatorie discrete*. X è il lancio di una moneta bilanciata:

$$Y = \begin{cases} \text{"Lancio di un d6"} & \text{se } X = 0 \\ \text{"Lancio di un d8"} & \text{se } X = 1 \end{cases}$$

Qual è la legge di Y?

$$\varphi_{Y|X}(y|x=0) = \begin{cases} 0 & \text{se } y \notin \{1,..,6\} \\ \frac{1}{6} & \text{se } y \in \{1,..,6\} \end{cases}, \ \varphi_{Y|X}(y|x=1) = \begin{cases} 0 & \text{se } y \notin \{1,..,8\} \\ \frac{1}{8} & \text{se } y \in \{1,..,8\} \end{cases}$$

$$\Rightarrow \varphi_{Y|X}(y|x) = \begin{cases} \frac{1}{6} & \text{se } x = 0 \wedge y \in \{1,..,6\} \\ \frac{1}{8} & \text{se } x = 1 \wedge y \in \{1,..,8\} \\ 0 & \text{altrimenti} \end{cases}$$

Possiamo ricavare $\varphi_{X,Y}(x,y) = \varphi_{Y|X}(y|x) \cdot \varphi_X(x)$

$$\varphi_{X,Y}(x,y) = \begin{cases} \frac{1}{6} \cdot \frac{1}{2} & \text{se } x = 0 \wedge y \in \{1,..,6\} \\ \frac{1}{8} \cdot \frac{1}{2} & \text{se } x = 1 \wedge y \in \{1,..,8\} \\ 0 & \text{altrimenti} \end{cases} = \begin{cases} \frac{1}{12} & \text{se } x = 0 \wedge y \in \{1,..,6\} \\ \frac{1}{16} & \text{se } x = 1 \wedge y \in \{1,..,8\} \\ 0 & \text{altrimenti} \end{cases}$$

Ora possiamo definire $\varphi_Y(y)$:

$$\varphi_Y(y) = \sum_{x \in \mathcal{R}_X} \varphi_{X,Y}(x,y) = \begin{cases} \frac{1}{12} + \frac{1}{16} & \text{se } y \in \{1,..,6\} \\ \frac{1}{16} & \text{se } y \in \{1,..,8\} \\ 0 & \text{altrimenti} \end{cases} = \begin{cases} \frac{7}{48} & \text{se } y \in \{1,..,6\} \\ \frac{1}{16} & \text{se } y \in \{7,8\} \\ 0 & \text{altrimenti} \end{cases}$$

Controlliamo se sono indipendenti:

$$\varphi_X(x) \cdot \varphi_Y(y) = \begin{cases} \frac{7}{48} \cdot \frac{1}{2} & \text{se } x = 0 \wedge y \in \{1,..,6\} \\ \frac{7}{48} \cdot \frac{1}{2} & \text{se } x = 1 \wedge y \in \{1,..,6\} \\ \frac{1}{16} \cdot \frac{1}{2} & \text{se } x = 1 \wedge y \in \{7,8\} \\ 0 & \text{altrimenti} \end{cases} = \begin{cases} \frac{7}{96} & \text{se } x = 0 \wedge y \in \{1,..,6\} \\ \frac{7}{96} & \text{se } x = 1 \wedge y \in \{1,..,6\} \\ \frac{1}{32} & \text{se } x = 1 \wedge y \in \{7,8\} \\ 0 & \text{altrimenti} \end{cases}$$

Dunque X, Y non sono tra loro indipendenti, in quanto $\varphi_{X,Y}(x,y) \neq \varphi_X(x) \cdot \varphi_Y(y)$

Prop. Siano X, Y due *variabili aleatorie discrete* definite sullo stesso spazio di probabilità con *densità discreta congiunta* $\varphi_{X,Y}$. La loro somma $X + Y$ ha *densità discreta*:

$$\varphi_{X+Y}(z) = \sum_{x \in \mathcal{R}_X} \varphi_{X,Y}(x, z - x)$$

Dim. Vale la seguente catena di uguaglianze:

$$\varphi_{X+Y}(z) = P(X+Y=z) = P\left(\bigcup_{x \in \mathcal{R}_X} \{X = x, X + Y = z\}\right) = P\left(\bigcup_{x \in \mathcal{R}_X} \{X = x, Y = z - x\}\right)$$

$$= \sum_{x \in \mathcal{R}_x} P(X = x, Y = z - x) = \sum_{x \in \mathcal{R}_X} \varphi_{X,Y}(x, z - x)$$

Oss. Se X e Y sono indipendenti, vale:

$$\varphi_{X+Y}(z) = \sum_{x \in \mathcal{R}_X} \varphi_X(x) \cdot \varphi_Y(z - x)$$

Esempio Siano X, Y *variabili aleatorie discrete* associate ciascuna ai risultati del lancio di un dado a 10 facce. Sia $S = X + Y$, qual è la sua legge? Quali sono $\varphi_{S,X}$ e $\varphi_{S|X}$?
Inizio definendo $\varphi_X(x)$ e $\varphi_Y(y)$:

$$\varphi_X(x) = \varphi_Y(y) = \begin{cases} \frac{1}{10} & \text{se } x \in \{1, .., 10\} \\ 0 & \text{altrimenti} \end{cases}$$

A questo punto $\varphi_S(z)$ è immediato dalla proposizione:

$$\varphi_S(z) = \varphi_{X+Y}(z) = \sum_{x \in \mathcal{R}_X} \varphi_X(x) \cdot \varphi_Y(z - x) = \sum_{i=1}^{10} \varphi_X(x) \cdot \varphi_Y(z - x) \Rightarrow$$

$$\varphi_S(z) = \begin{cases} \frac{1}{100} & \text{se } z = 2 \lor z = 20 \\ \frac{2}{100} & \text{se } z = 3 \lor z = 19 \\ \frac{3}{100} & \text{se } z = 4 \lor z = 20 \\ \dots & \\ \frac{11}{100} & \text{se } z = 11 \\ 0 & \text{altrimenti} \end{cases}$$

Ora, passo a ricavare $\varphi_{S,X}(z, x)$:

$$\varphi_{S,X}(z, x) = P(S = z, X = x) = (X + Y = z, X = x) = P(Y = z - X, X = x) =$$

$$\varphi_{X,Y}(x, z - x) = \varphi_X(x) \cdot \varphi_Y(z - x) = \begin{cases} \frac{1}{100} & \text{se } z \in \{2, .., 20\} \land x \in \{1, .., 10\} \\ 0 & \text{altrimenti} \end{cases}$$

Infine:

$$\varphi_{S|X}(z|x) = \frac{\varphi_{S,X}(z, x)}{\varphi_X(x)} = \frac{\varphi_X(x) \cdot \varphi_Y(z - x)}{\varphi_X(x)} = \varphi_Y(z - x) \Rightarrow$$

$$\varphi_{S|X}(z|x) = \begin{cases} \frac{1}{10} & \text{se } x \in \{1, .., 10\} \land z \in \{x + 1, .., x + 10\} \\ 0 & \text{altrimenti} \end{cases}$$

7.3 Vettori aleatori assolutamente continui

Definizione 7.3.1. Siano X, Y *variabili aleatorie assolutamente continue* definite su $(\Omega, \mathcal{F}, P)$. Chiamiamo densità congiunta di X e Y la funzione $F_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ tale che:

$$\forall E \in \mathcal{B} \times \mathcal{B} \; P((X, Y) \in E) = \int\int_E f_{X,Y}(x, y) dx \; dy$$

Oss. $P((X, Y) \in E) = P(\{\omega \in \Omega : (X(\omega), Y(\omega) \in E)\})$

Prop. Siano X e Y due *variabili aleatorie assolutamente continue.* Allora:

(i) $\forall (x, y) \in \mathbb{R}^2 \; F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s, t) dt \; ds$

(ii) $\forall (x, y) \in \mathbb{R}^2$ le densità marginali sono:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, t) dt, \; f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(s, y) ds$$

(iii) X, Y sono indipendenti se e solo se $\forall (x, y) \in \mathbb{R}^2 \; f_{X,Y}(x, y) = f_X(y) \cdot f_Y(y)$

Dim.

(i) Segue dalla definizione di $F_{X,Y}$ con $E \in (-\infty, x] \times (-\infty, x]$.
Possiamo scrivere la (i) anche come: $\frac{\partial^2}{\partial x \cdot \partial y} \cdot F_{X,Y}(x,y) = f_{X,Y}(x,y)$

(ii) Segue dai teoremi di integrazione

(iii) [$\Rightarrow$] dalle definizioni
[$\Leftarrow$] segue da (i) più un'analoga proprietà per $F_{X,Y}$

Oss.

- Per ogni $(x,y) \in \mathbb{R}^2$ $f_{X,Y}(x,y) \geq 0$ ma non necessariamente ≤ 1
- $\int\int_{\mathbb{R}^2} f_{X,Y}(x,y)dx\ dy = 1$ percui anche in questo caso possiamo parlare di *costanti di rinormalizzazione*

Esempio Siano X, Y due *variabili aleatorie assolutamente continue* e sia così definita la *funzione di densità congiunta*:

$$f_{X,Y(x,y)} = \begin{cases} e^{-x} & \text{se } 0 \leq y \leq x \\ 0 & \text{altrimenti} \end{cases}$$

Qual è $F_X(x)$?

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \begin{cases} \int_0^x e^{-x}dy & \text{se } x > 0 \\ 0 & \text{se } x < 0 \end{cases} = \begin{cases} x \cdot e^{-x} & \text{se } x > 0 \\ 0 & \text{se } x < 0 \end{cases}$$

e se volessimo $F_X(x)$:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} \int_0^x t \cdot e^{-t}dt & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases} = \begin{cases} 1 - e^{-x} \cdot (x+1) & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$

oppure

$$F_X(x) = \lim_{y \to +\infty} F_{X,Y}(x,y) = \lim_{y \to +\infty} \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s,t)dt\ ds = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f_{X,Y}(s,t)dt\ ds$$

Definizione 7.3.2. Siano X e Y due *variabili aleatorie assolutamente continue* sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, P)$. Chiamiamo *densità di X condizionata Y* la funzione definita come:

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{se } y \in \mathcal{R}_Y \\ 0 & \text{se } y \notin \mathcal{R}_Y \end{cases}$$

Oss. $f_{X,Y}(x,y) = f_{X|Y}(x|y) \cdot f_Y(y) = fY|X(y|x) \cdot f_X(x)$

Esempio Siano X, Y *variabili aleatorie assolutamente continue* con la seguente *funzione di densità congiunta*:

$$f_{X,Y}(x,y) = \begin{cases} 6e^{-2x}e^{-3y} & \text{se } x, y > 0 \\ 0 & \text{altrimenti} \end{cases}$$

X e Y sono indipendenti?

Per essere indipendenti deve valere:

$$f_X(x) \cdot f_Y(y) = f_{X,Y}(x,y)$$

quindi, procedo ricercando $f_X(x)$ e $f_Y(y)$:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \begin{cases} \int_0^{+\infty} 6e^{-2x}e^{-3y}dy & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

$$= \begin{cases} 6e^{-2x} \cdot \int_0^{+\infty} e^{-3y}dy & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases} = \begin{cases} 2e^{-2x} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx \begin{cases} \int_0^{+\infty} 6e^{-2x}e^{-3y}dx & \text{se } y > 0 \\ 0 & \text{se } y \leq 0 \end{cases}$$

$$= \begin{cases} 6e^{-3y} \cdot \int_0^{+\infty} e^{-2x}dx & \text{se } y > 0 \\ 0 & \text{se } y \leq 0 \end{cases} = \begin{cases} 3e^{-3y} & \text{se } y > 0 \\ 0 & \text{se } y \leq 0 \end{cases}$$

Ora, verifico la condizione di indipendenza:

$$f_X(x) \cdot f_Y(y) = \begin{cases} 2e^{-2x} \cdot 3e^{-3y} & \text{se } x, y > 0 \\ 0 & \text{altrimenti} \end{cases} = \begin{cases} 6e^{-2x}e^{-3y} & \text{se } x, y > 0 \\ 0 & \text{altrimenti} \end{cases} = f_{X,Y}(x,y)$$

Quindi, X e Y sono indipendenti.

Prop (Somma di *v.a. assolutamente continue*)**.** Siano X e Y *variabili aleatorie assolutamente continue*, definite nello stesso spazio di probabilità $(\Omega, \mathcal{F}, P)$, con *densità congiunta* $f_{X,Y}$. La *variabile aleatoria* $X + Y$ ha *densità*:

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

Dim. Consideriamo la *funzione di ripartizione* $F_{X+Y}(z) = P(X + Y \leq z)$. Questa probabilità può essere vista come $P((X,Y) \in E)$ per qualche $E \in \mathcal{B} \times \mathcal{B}$, infatti:

$$X + Y \leq z \Leftrightarrow Y \leq z - X$$

Se lo si rappresenta nel piano $\mathbb{R}^2$, questa condizione è verificata da tutti i punti al di sotto della retta $y = -x + z$, quindi:

$$F_{X,Y}(z) = P((X,Y) \in E) = \int \int_E f_{X,Y}(x,y)dy\ dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f_{X,Y}(x,y)dy\ dx$$

Per ricavare la densità è sufficiente derivare in z.

7.4 Vettori aleatori misti

Per quanto riguarda i *vettori aleatori misti* esamineremo soltanto il caso dei *vettori* composti da una *variabile aleatoria discreta* e una *assolutamente continua.*

Esempio All'Università di Trento, gli studenti si dividono tra un 52% iscritto a materie umanistiche e un 48% iscritto a materie scientifiche. Tra chi studia materie scientifiche il tempo di studio giornaliero è uniformemente distribuito tra 155 e 180 minuti, mentre per chi studia materie umanistiche il tempo varia tra 143 e 166 minuti.

Siano X la *variabile aleatoria* che traccia l'ambito di studio degli studenti, e Y la *variabile aleatoria* che ne indica il tempo di studio giornaliero.

1. Qual è, se esiste, la legge congiunta di X e Y?
2. Qual è la probabilità che un qualunque studente studi al più 160 minuti?
3. Come sono suddivisi tra i due indirizzi gli studenti che studiano meno di 160 minuti?
4. E come sono distribuiti quelli che studiano esattamente 160 minuti?

Osservo che X è una *variabile aleatoria discreta*, mentre Y è *assolutamente continua.* Inizio quindi defininendo le *funzioni di densità* di X e Y condizionata X:

$$\varphi_X(x) = \begin{cases} 0 & \text{se studia materie scientifiche} \\ 1 & \text{se studia materie umanistiche} \end{cases}$$

$$f_{Y|X}(y|0) = \begin{cases} cs & \text{se } y \in [155, 180] \\ 0 & \text{altrimenti} \end{cases} ; \; f_{Y|X}(y|1) = \begin{cases} cu & \text{se } y \in [143, 166] \\ 0 & \text{altrimenti} \end{cases}$$

Devo trovare i valori delle costanti cs e cu:

$$\int_{-\infty}^{+\infty} f_{Y|X}(y|0)dy = [cs \cdot y]_{155}^{180} = cs \cdot (180 - 155) = cs \cdot 25 \Rightarrow cs = \frac{1}{25}$$

$$\int_{-\infty}^{+\infty} f_{Y|X}(y|1)dy = [cu \cdot y]_{143}^{166} = cu \cdot (166 - 143) = cu \cdot 23 \Rightarrow cu = \frac{1}{23}$$

Da qui è facile definire $f_{Y|X}(y|x)$:

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{25} & \text{se } x = 0 \wedge y \in [155, 180] \\ \frac{1}{23} & \text{se } x = 1 \wedge y \in [143, 166] \\ 0 & \text{altrimenti} \end{cases}$$

A questo punto possiamo ricavare la legge congiunta di X e Y:

$$f_{X,Y}(x, y) = f_{Y|X}(y|x) \cdot f_X(x) = \begin{cases} \frac{1}{25} \cdot 0.48 = 0.0192 & \text{se } x = 0 \wedge y \in [155, 180] \\ \frac{1}{23} \cdot 0.52 = 0.0226087 & \text{se } x = 1 \wedge y \in [143, 166] \\ 0 & \text{altrimenti} \end{cases}$$

Per rispondere al secondo quesito è necessario ricavare $F_Y(y)$. Prima però serve la *funione di densità* di Y. Per farlo marginalizziamo $f_{X,Y}$ sommando tutti i 2 valori di X:

$$f_Y(y) = f_{X,Y}(0, y) + f_{X,Y}(1, y) = \begin{cases} 0.0226087 & \text{se } y \in [143, 155) \\ 0.0226087 + 0.0192 = 0.0418087 & \text{se } y \in [155, 166) \\ 0.0192 & \text{se } y \in [166, 180] \\ 0 & \text{altrimenti} \end{cases}$$

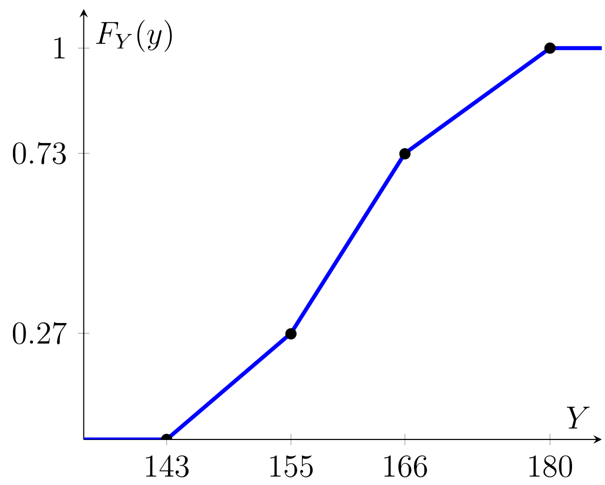
Ora, ricaviamo $F_Y(y)$:

$$F_Y(y) = \begin{cases} 0 & \text{se } y < 143 \\ 0.0226087 \cdot (y - 143) & \text{se } 143 \leq y < 155 \\ 0.0418087 \cdot (y - 155) + 0.2713044 & \text{se } 155 \leq y < 166 \\ 0.0192 \cdot (y - 166) + 0.7312001 & \text{se } 166 \leq y < 180 \\ 1 & \text{se } x \geq 180 \end{cases}$$

NB. Le costanti 0.2713044 e 0.7312001 sono state calcolate come segue:

- $0.2713044 = 0.0226087 \cdot (155 - 143)$
- $0.7312001 = 0.0418087 \cdot (166 - 155) + 0.2713044$

Anche l'uno può essere ottenuto allo stesso modo, infatti: $1 = 0.0192 \cdot (180 - 166) + 0.7312001$.



A questo punto, per rispondere al secondo quesito, basta calcolare $F_Y(160)$:

$$F_Y(160) = P(Y \leq 160) = 0.0418087 \cdot (160 - 155) + 0.2713044 \approx 0.48$$

Al terzo quesito si risponde calcolando la probabilità di X condizionata Y:

$$P(X = x | Y < 160) = \frac{P(X = x, Y < 160)}{P(Y < 160)} = \frac{\int_{143}^{160} f_{X,Y}(x,y)dy}{F_Y(160)} =$$

$$\begin{cases} \frac{[0.0192 \cdot y]_{155}^{160}}{0.48} \approx 0.20 & \text{se } X = 0 \\ \frac{[0.0226087 \cdot y]_{143}^{160}}{0.48} \approx 0.80 & \text{se } X = 1 \end{cases} = \begin{cases} 20\% & \text{studia materie scientifiche} \\ 80\% & \text{studia materie umanistiche} \end{cases}$$

Per il quarto quesito non si può calcolare $P(X = x, y = 160)$ perchè la probabilità di un punto è 0, quindi calcoliamo:

$$f_{X|Y}(x|160) = \frac{f_{X,Y}(x, 160)}{f_Y(160)} = \begin{cases} \frac{0.0192}{0.0418087} & \text{se } x = 0 \\ \frac{0.0226087}{0.0418087} & \text{se } x = 1 \end{cases} \approx \begin{cases} 0.46 & \text{se } x = 0 \\ 0.54 & \text{se } x = 1 \end{cases}$$

Capitolo Nr.8

Modelli di variabili aleatorie discrete

8.1 Bernoulliane

Definizione 8.1.1. Una *variabile aleatoria* si dice *Bernoulliana* di paramentro $p \in [0,1]$, e indicheremo in simboli $X \sim bin(1,p)$, se ha *densità discreta*:

$$\varphi_X(x) = \begin{cases} p & \text{se } x = 1 \\ 1-p & \text{se } x = 0 \\ 0 & \text{altrimenti} \end{cases}$$

La *funzione di ripartizione* di una *Bernoulliana* di parametro p è:

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1-p & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases}$$

Oss. È una *variabile aleatoria indicatrice* in cui $E =$ "*successo*".

8.2 Binomiali

Ipotiziamo di avere n *variabili aleatorie Bernoulliane* di parametro p indipendenti (ad esempio n lancio di una moneta bilanciata). Se ne prendiamo la somma S, S indica il numero di successi.

Definizione 8.2.1. Sia X una *variabile aleatoria discreta binomiale* di n parametri $p \in [0,1]$ con $n \in \mathbb{N} \backslash \{0\}$. Se X è definita come la somma di n *Bernoulliane* di parametro p indipendenti scriveremo $X \sim bin(n,p)$.

Prop. Se $X \sim bin(n,p)$ allora $\varphi_X(k) = \begin{cases} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} & \text{se } k \in \{0,..,n\} \\ 0 & \text{altrimenti} \end{cases}$

Dim. Sia $\mathcal{R}_X = \{0,..,n\}$ il *supporto* di X. Preso $k \in \mathcal{R}_X$ $\varphi_X(k) = P(X=k) = P(\sum_{i=1}^n X_i = k)$ con $X_i \sim bin(1,p)$ $\forall i$ indipendenti tra loro. Chiamiamo $N = \{1,..,n\}$ l'insieme degli indici delle *Bernoulliane* e indichiamo con $\mathcal{I}_k \subseteq \mathcal{P}(N)$ la famiglia dei sottoinsiemi di N con esattamente k elementi:

$$\mathcal{I}_k = \{I_k \in \mathcal{P}(N) : \#I_k = k\}$$

Per ciascun insieme I_k definiamo:

$$E_{I_k} = \bigcap_{i \in I_k} \{X_i = 1\} \cap \bigcap_{i \in I_k^c} \{X_i = 0\}$$

Preso atto che $\{X = k\} = \bigcup_{I_k \in \mathcal{I}_k} E_{I_k}$ vale:

$$P(E_{I_K}) = P\left(\bigcap_{i \in I_k} \{X_i = 1\} \cap \bigcap_{i \in I_k^c} \{X_i = 0\}\right)$$

$$= \prod_{i \in I_k} \cdot P(X_i = 1) \cdot \prod_{i \in I_k^c} \cdot P(X_i = 0) = p^k \cdot (1-p)^{n-k}$$

A questo punto la *funzione di massa* è:

$$\varphi_X(k) = \sum_{I_k \in \mathcal{I}_k} p^k \cdot (1-p)^{n-k} = p^k \cdot (1-p)^{n-k} \cdot \sum_{I_k \in \mathcal{I}_k} 1 = p^k \cdot (1-p)^{n-k} \cdot \binom{n}{k}$$

Oss. La *funzione di ripartizione* è definita come:

$$F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} \varphi_X(k) = \sum_{k=0}^{min\{\lfloor x \rfloor, n\}} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

e in particolar,e se $X \geq 1$ $F_X(x) = 1$, infatti:

$$\sum_{k=0}^{n} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = (p + (1-p))^n = 1$$

Esempio Un'azienda produttrice di chiavette USB, produce chiavette difettose con probabilità del 2%. Vende le pennette in confezioni da 15 e se in una confezione ce ne sono almeno due difettose rimborsa il cliente.

1. Quale percentuale di confezioni viene rimborsata?
2. Comprando 4 confezioni con che probabilità esattamente una confezione è rimborsabile?

   Definisco due *variabili aleatorie*:

   (i) $O \sim bin(1, 0.02)$: indica se una pennetta è difettosa o meno

   (ii) $N \sim bin(15, 0.02)$: conta le pennette difettose in una confezione

Per rispondere al primo quesito, calcolo la probabilità che in una confezione ci siano più di una pennetta USB difettose:

$$P(N > 1) = \sum_{k=2}^{15} P(N = k) = 1 - P(N = 0) - P(N = 1) = 1 - \varphi_N(0) - \varphi_N(1) =$$

$$1 - \binom{15}{0} \cdot 0.02^0 \cdot (1 - 0.02)^{15} - \binom{15}{1} \cdot 0.02^1 \cdot (1 - 0.02)^{14} = 1 - 0.98^{15} - 15 \cdot 0.02 \cdot 0.98^{14} \approx 3.5\%$$

Per rispondere invece, al secondo quesito definisco una nuova *variabile aleatoria* S che conta il numero di scatole rimborsabili: $S \sim bin(4, P(N > 1)) = bin(4, 0.035)$.

$$\varphi_S(1) = \binom{4}{1} \cdot 0.035^1 \cdot (1 - 0.035)^3 \approx 12.7\%$$

8.2.1 Bernoulliane e binomiali in R

Presa una *binomiale* la *funzione di densità discreta* è la funzione:

`dbinom(x, size, prob)`

dove `x` è il punto in cui vogliamo calcolare la *densità* φ, `size` è il numero di tentativi (nella definizione 8.2.1 è stato chiamato n) e `prob` è la probabilità di successo di ogni tentativo (nella def. 8.2.1 era p).

Esempio Sia $X \sim bin(44, 0.2)$ una *binomiale*. Qual è la probabilità di $\varphi_X(11)$?

$\varphi_X(11) =$ `dbinom(x = 11, size = 44, prob = 0.2) =`
`dbinom(size = 44, x = 11, prob = 0.2) = dbinom(11, 44, 0.2)`

La *funzione di ripartizione* è invece la funzione:

`pbinom(q, size, prob, lower.tail = TRUE)`

i cui parametri `size`, `prob` sono gli stesso di `dbinom`, mentre `q` è il punto in vogliamo calcolare la *funzione di ripartizione* F_X e `lower.tail` è un parametro logico che che determina se stiamo calcolando la *funzione di ripartizione* F_X nel punto `q` (ossia $P(X \leq q)$, la "coda inferiore") in corrispondenza del valore `TRUE` o il suo complementare $1 - F_X(q)$ (cioè $P(X > q)$, la "coda superiore") in corrispondeza del valore `FALSE`.

Esempio Sia $X \sim bin(23, 0.5)$ una *binomiale*. Qual è la probabilità di $\mathcal{F}_X(12.5)$ e $P(X > 10)$?

$F_X(12.5) =$ `pbinom(q = 12.5, size = 23, prob = 0.5, lower.tail = TRUE) =`
`dbinom(size = 23, q = 12.5, prob = 0.5)= dbinom(12.5, 23, 0.5)`
$P(X > 10) =$ `pbinom(q = 10, size = 23, prob = 0.5, lower.tail = FALSE) =`
`dbinom(12.5, 23, 0.5, FALSE)`

Ci sono anche altre funzioni per le *binomiali* in R: `rbinom` e `qbinom`.
`rbinom(n, size, prob)` è un generatore casuale di risultati distribuiti come una binomiale dei parametri assegnati, cioè genera dei valori $X(\omega) \in \mathbb{R}$ con $X \sim bin(n, p)$. I parametri `size` e `prob` sono gli stessi delle funzioni precedenti, mentre `n` indica il numero di valori da generare. `qbinom` è la funzione *quantile*.

Esempio Se vogliamo un campione di 100 esiti di lanci di una moneta bilanciata, che è una *Bernoulliana* di parametro $p = 0.5$, possiamo usare:

`rbinom(n = 100, size = 1, prob = 0.5)`

8.3 Schema o processo di Bernoulli

Definizione 8.3.1. Si definisce *schema* o *processo di Bernoulli* una successione infinita di *Bernoulliane* indipendenti e identicamente distribuite:

$$\{X_i\}_{i \in \mathbb{N}} \; X_i \sim bin(1, p)$$

Dato uno *schema di Bernoulli* definiamo lo spazio di probabilità $(\Omega, \mathcal{F}, P)$ prendendo:

- $\Omega = \{0, 1\}^{\mathbb{N}}$, cioè successioni a valori in $\{0, 1\}$

- Siccome l'universo è uno *spazio prodotto infinito*, come *tribù* prendiamo la *tribù* $\mathcal{F}$ generata dai *cilindri*. I *cilindri* in questo caso sono i sottoinsiemi $C \subseteq \Omega$ tali che esistono un naturale $n \in \mathbb{N}$ e un vettore $v = \{0,1\}^n$:

$$C = \{\omega \in \Omega : \omega_i = v_i \text{ per } 1 \leq i \leq n\}$$

- P è il prodotto delle probabilità delle componenti

Esempio Sia $\{X_i\}_{i\in\mathbb{N}}$ $X_i \sim bin(1,p)$ uno *schema di Bernoulli*.

1. Qual è la probabilità di avere un successo seguito da 2 insuccessi?
2. E che il primo successo sia al k-esimo tentativo?
3. E la probabilità che il terzo esito sia un successo?

1. Il cilindro corrispondente è:
$$(1,0,0,*,*,*,\dots)$$
cioè $n = 3$ e $v = 1,0,0$, mentre la probabilità è:
$$P = p \cdot (1-p)^2 \cdot \prod_{i=1}^{+\infty} 1 = p \cdot (1-p)^2$$

2. Il cilindro stavolta è:
$$(0,0,\dots,0,1,*,*,\dots)$$
con $k-1$ insuccessi e un successo come k-esimo esito. La probabilità diventa:
$$P = (1-p)^{k-1} \cdot p \cdot \prod_{i=1}^{+\infty} 1 = (1-p)^{k-1} \cdot p$$

3. Infine, nel terzo quesito il cilindro è:
$$(*,*,1,*,*,\dots)$$
Questo di per sè non sarebbe un cilindro in quanto non inizia con un vettore, ma posso posso comunque immaginarlo come unione di cilindri:
$$(*,*,1,*,*,\dots) = (0,0,1,*,*,\dots) \cup (1,0,1,*,*,\dots) \cup (0,1,1,*,*,\dots) \cup (1,1,1,*,*,\dots)$$
percui, la probabilità si calcola come:
$$P = P(0,0,1,*) + P(1,0,1,*) + P(0,1,1,*) + P(1,1,1,*) =$$
$$(1-p)^2 \cdot p + p \cdot (1-p) \cdot p + (1-p) \cdot p2 + p^3 = p \cdot ((1-p)^2 + 2p \cdot (1-p) + p^2) =$$
$$p \cdot ((1-p) + p)^2 = p$$

8.4 Geometriche

Preso uno *schema di Bernoulli* di parametro p, siamo interessati al numero di insuccessi che si verificano prima del primo successo. Definiamo quindi una *variabile aleatoria* che tracci l'indice del primo successo:

$$T_1 = inf\{i \geq 1 : \omega_i = 1\}$$

Questa particolare *variabile aleatoria* si chiama *tempo di arresto*, ma è anche una *variabile aleatoria indicatrice* dell'istante di primo successo. Quindi, per ricavare il numero di insuccessi basta considerare $T_1 - 1$.

Definizione 8.4.1. Una *variabile aletoria* è una *geometrica* di parametro p, e scriveremo $X \sim geom(p)$, se è l'istante precedente al primo successo in un *processo di Bernoulli* di parametro p.

Oss. Qual è la *funzione di massa* φ_X di una *geometrica*?

$\varphi_X(k) = P(X = k)$ equivale a chiedere quale sia la probabilità che si verifichino k insuccessi prima del primo successo:

$$\varphi_X(k) = P(X = k) = \begin{cases} 0 & \text{se } k \notin \mathbb{N} \\ (1-p)^k \cdot p & \text{se } k \in \mathbb{N} \end{cases}$$

NB. In alcuni testi la *geometrica* indica l'istante di primo successo (T_1) invece del numero di insuccessi ($T_1 - 1$), ma la probabilità risultante è la stessa.

Oss.

$$\sum_{k=0}^{+\infty} \varphi_X(k) = \sum_{k=0}^{+\infty} (1-p)^k \cdot p = p \cdot \sum_{k=0}^{+\infty} (1-p)^k = p \cdot \frac{1}{1-(1-p)} = p \cdot \frac{1}{p} = 1$$

NB. $\sum_{k=1}^{+\infty} (1-p)^k$ è una *serie geometrica*. Da qui il nome *variabile aleatoria geometrica*.

Oss. Qual è la *funzione di ripartizione* F_X di una *geometrica*?

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ \sum_{k=0}^{\lfloor x \rfloor} \varphi_X(x) & \text{se } x \geq 0 \end{cases} = \begin{cases} 0 & \text{se } x < 0 \\ p \cdot \sum_{k=0}^{\lfloor x \rfloor} (1-p)^k & \text{se } x \geq 0 \end{cases} =$$

$$\begin{cases} 0 & \text{se } x < 0 \\ p \cdot \frac{1-(1-p)^{\lfloor x \rfloor + 1}}{1-(1-p)} & \text{se } x \geq 0 \end{cases} = \begin{cases} 0 & \text{se } x < 0 \\ 1 - (1-p)^{\lfloor x \rfloor + 1} & \text{se } x \geq 0 \end{cases}$$

NB. Avremmo potuto raggiungere in modo più diretto lo stesso risultato, considerando il complementare

Prop (Assenza di memoria)**.** $\forall n \in \mathbb{N}\ P(X \geq n + k | X \geq n) = P(X \geq k)$

Dim.

$$P(X \geq n + k | X \geq n) = \frac{P((X \geq n + k) \cap (X \geq n))}{P(X \geq n)} = \frac{P(X \geq n + k)}{P(X > n - 1)} =$$

$$\frac{PXn + k - 1}{P(X > n - 1)} = \frac{(1-p)^{n+k-1+1}}{(1-p)^{n-1+1}} = \frac{(1-p)^{n+k}}{(1-p)^n} = (1-p)^k = P(X \geq k)$$

Oss. Cosa significa *assenza di memoria*?

La proposizione può essere letta come: "Se dopo n tentantivi non si è verificato alcun successo, possiamo iniziare un nuovo *schema di Bernoulli* con uguale parametro e calcolare la probabilità che si verifichino almeno k insuccessi prima di un successo." Cioè, sapere quanti insuccessi ci sono stati non ci dà alcuna informazione aggiuntiva sull'istante in cui si verificherà il prossimo successo.

Esempio Nel Superenalotto il 55 non esce da 60 estrazioni.

1. Con che probabilità uscirà alla prossima estrazone?
2. E che esca la prossima volta tra almeno 30 estrazioni?

1. Calcolare la probabilità che esca 55 equivale a calcolare la probabilità che non esca e considerarne il complementare:

$$P(\text{"esce 55"}) = 1 - P(\text{"non esce 55"}) = 1 - \frac{\binom{89}{6}}{\binom{90}{6}} = \frac{\binom{89}{5}}{\binom{90}{6}} = \frac{6}{90} = \frac{1}{15} \approx 6.6\%$$

   Avrei potuto rispondere al quesito, anche sfruttando gli *schemi di Bernoulli*.
   Sia X la *variabile indicatrice* di "esce 55", $X \sim bin(1, \frac{1}{15})$ uno *schema di Bernoulli* di parametro $\frac{1}{15}$.

$$P(\text{"55 esce alla } 61^a \text{ estrazione"}) = P(X = 60 | X \geq 60) = \frac{P(X = 60)}{P(X > 59)} =$$

$$\frac{(1 - \frac{1}{15})^{60} \cdot \frac{1}{15}}{(1 - \frac{1}{15})^{59+1}} = \frac{1}{15}$$

   Si può notare che il risultato è lo stesso del primo quesito. Questo avviene per effetto dell'*assenza di memoria*. Infatti, sapere che ci sono stati 60 insuccessi non influenza la probabilità che se ne verifichi uno.

2. Utilizzando lo stesso *schema di Bernoulli* si ha:

$$P(X \geq 60 + 30 | X \geq 60) = P(X \geq 30) = (1 - \frac{1}{15})^{30} \approx 12.6\%$$

8.4.1 Geometriche in R

Come per le *Binomiali*, anche per le *geometriche* c'è una famiglia di funzioni che ne permettono la manipolazione:

- `dgeom(x, prob)`
  Questa funzione permette di calcolare la *funzione di massa* nel punto `x`, sapendo che la probabilità di successo, cioè che il parametro della *geometrica*, è `prob`.

- `pgeom(q, prob, lower.tail = TRUE)`
  Calcola il valore della *cdf* nel punto `q`, considerando di nuovo come parametro `prob`. Inoltre, se `lower.tail = TRUE` calcola direttamente la *cdf*, altrimenti ne considera il complementare.

- `rgeom(n, size, prob)`
  Genera casualmente `n` risultati distribuiti come una *geometrica*.

- `qgeom`
  È la funzione *quantile*.

8.5 Binomiali negative

Si consideri l'istante dell'ennesimo successo di uno *schema di Bernoulli* ($n \in \mathbb{N}\backslash\{0\}$) e lo si chiami I_n. I_n sarà definito come:

$$T_n = inf\{i \geq 1 : \sum_{k=1}^{i} \omega_k = n\}$$

NB. T_n è anche definibile in modo ricorsivo, come segue:

$$\begin{cases} T_1 = inf\{i \geq 1 : \omega_i = 1\} \\ T_{n+1} = inf\{1 > T_n : \omega_i = 1\} \text{ se } n \geq 1 \end{cases}$$

Definizione 8.5.1. X è una *variabile aleatoria binomiale negativa* (o di *Pascal*) di parametri n e p se conta il numero di insuccessi precedenti all'n-esimo successo di uno *schema di Bernoulli* di parametro p. Scriveremo in simboli $X \sim NB(n, p)$.

Oss. Qual è la *funzione di massa* φ_X di una *binomiale negativa*?

Sia $k \in \mathbb{N}$ e $X \sim NB(n, p)$:

$$\varphi_X(k) = P(X = k) = P(T_n = k + n) = P\left(\omega_{n+k} = 1, \sum_{j=1}^{n+k-1} \omega_j = n - 1\right) =$$

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

Esempio Un personaggio è intrappolato nel fondo di una buca. Se ottiene 3 risultati maggiori di 15 lanciando un dado a 20 facce riesce a uscire dalla buca. Ogni lancio corrisponde a 5 minuti nel gioco. Con che probabilità impiega al più mezz'ora di gioco per uscire?

Ogni lancio è una *Bernoulliana* con $p = \frac{20-15}{20} = \frac{1}{4}$. Cerchiamo la probabilità che $5 \cdot$ numero tentativi $\leq 30 \Rightarrow$ numero tentativi ≤ 6.

Vogliamo scriverlo come *binomiale negativa* in cui gli insuccessi devono essere al più 3:

$$X \sim NB\left(3, \frac{1}{4}\right) \Rightarrow P(X = k) = \sum_{k=0}^{3} \varphi_X(k) = \sum_{k=0}^{3} \left(\binom{k+n-1}{n-1}\right) \cdot \left(\frac{1}{4}\right)^3 \cdot \left(\frac{3}{4}\right)^k$$

8.5.1 Binomiali negative in R

Anche per le *binomiali negative* esiste una famiglia di funzioni in R:

- `dnbinom(x, size, prob)`
  Calcola il valore della *funzione di massa* in `x`. size e `prob` rappresentano i parametri n e p indicati nella definizione.

- `pnbinom(q, size, prob, lower.tail = TRUE)`
  Calcola il valore della *funzione di ripartizione* in `q`.

- `rnbinom(n, size, prob)`
  Genera `n` risultati casuali distribuiti come una *binomiale negativa*.

- `qnbinom`
  È la funzione *quantile*.

8.6 Riproducibilità

Definizione 8.6.1. Una famiglia di *variabili aleatorie* si dice *riproducibile* se sommando 2 *variabili aleatorie* indipendenti appartenenti a quella famiglia abbiamo ancora una *variabile aleatoria* della medesima famiglia.

Esempio Siano X e Y indipendenti e identicamente distribuite (iid) *geometriche* di parametro p. Cosa possiamo dire della loro somma?

$$\varphi_S(k) = \sum_{j \in \mathcal{R}_x} \varphi_X(j) \cdot \varphi_Y(k-j) = \sum_{j=0}^{+\infty} (1-p)^j \cdot p \cdot \mathbb{1}_{(k-j)\geq 0} \cdot (1-p)^{k-j} =$$

$$\sum_{j=0}^{k} p^2 \cdot (1-p)^k = (k+1) \cdot p^2 \cdot (1-p)^k$$

NB. Non è la *densità* di una *geometrica*, quindi S non è una *geometrica*, ma una *binomiale negativa*: $NB(2,p)$. Questo perché:

$$\binom{k+n-1}{n-1} = \binom{k+2-1}{2-1} = \binom{k+1}{1} = k+1$$

Da qui deriva che:

$$X \sim geom(p) \Leftrightarrow X \sim NB(1,p)$$

Prop. La famiglia delle *binomiali negative* a parametro p fissato è riproducibile. In particolare se $X \sim NB(n,p)$ e $Y \sim NB(m,p)$ allora $X+Y \sim NB(n+m,p)$.

Prop. La famiglia delle *binomiali* a parametro p fissato è riproducibile. $X \sim bin(n,p), Y \sim bin(m,p), X, Y$ sono indipendenti $\Rightarrow X+Y \sim bin(n+m,p)$.

Dim.

$$\varphi_{X+Y}(l) = \sum_{k=0}^{n} \varphi_X(k) \cdot \varphi_Y(l-k) =$$

$$\sum_{k=0}^{n} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \cdot \mathbb{1}_{0 \leq l-k \leq m} \cdot \binom{m}{1-k} \cdot p^{l-k} \cdot (1-p)^{m-(l-k)} =$$

$$p^l \cdot (1-p)^{n+m-l} \cdot \sum_{k=0}^{n} \cdot \mathbb{1}_{0 \leq l-k \leq m} \binom{n}{k} \cdot \binom{n}{l-k}$$

Basta dimostrare che:

$$\sum_{k=0}^{n} \mathbb{1}_{0 \leq l-k \leq m} \binom{n}{k} \cdot \binom{m}{l-k} = \binom{n+m}{l}$$

Procediamo per induzione su n:

- $n = 0$: base dell'induzione

$$\binom{0}{0} \cdot \binom{m+0}{l} = \binom{m}{l} = \binom{m+0}{l}$$

- $n \Rightarrow n+1$: passo induttivo

$$\sum_{k=0}^{n+1}\binom{n+1}{k}\cdot\binom{m}{l-k}=\sum_{k=0}^{n+1}\left(\binom{n}{k}+\binom{n}{k+1}\right)\cdot\binom{m}{l-k}=$$

$$\sum_{k=0}^{n}\binom{n}{k}\cdot\binom{m}{l-k}+\sum_{k=1}^{n+1}\binom{n}{k-1}\cdot\binom{m}{l-k}$$

A questo punto applico il *passo induttivo* e pongo $h=k-1$:

$$\binom{n+m}{l}+\sum_{h=0}^{n}\binom{n}{h}\cdot\binom{m}{l-h-1}=\binom{n+m}{l}+\binom{n+m}{l-1}=\binom{m+(n+1)}{l}$$

8.7 Ipergeometriche

Presa un'urna con m biglie bianche e n biglie nere, se ne estraggono k.

- Se lo facciamo con reimmissione abbiamo una *binomiale* $bin\left(k,\frac{n}{m+n}\right)$
- E se lo facciamo senza reimmissione?

Definizione 8.7.1. Si chiama *ipergeometrica* di parametri k,m,n la *variabile aleatoria discreta* che conta il numero di biglie bianche tra le k estratte senza reimmissione da un urna con m biglie bianche e n nere. Se X è *ipergeometrica* di parametri k,m,n scriviamo $X\sim hyp(k,m,n)$.

Oss. Qual è la *funzione di massa* φ_X?
Posto che devono valere le seguenti:

(i) $0\le k\le m+n$

(ii) $\varphi_X(b)\neq 0$ se $0\le b\le k$
In realtà devo prendere $max\{0,k-n\}\le b\le min\{k,m\}$, in quanto $max\{0,k-n\}$ considera il caso in cui finisco le biglie nere, mentre $min\{k,m\}$ considera l'esaurimento delle biglie bianche.

A questo punto, $\varphi_X(b)$ è definita come:

$$\varphi_X(b)=\begin{cases}\frac{\binom{n}{b}\cdot\binom{n}{k-b}}{\binom{m+n}{k}} & \text{se } max\{0,k-n\}\le b\le min\{k,m\}\\ 0 & \text{altrimenti}\end{cases}$$

Oss. Se $X\sim hyp(k,m,n)$ allora

$$\sum_{b=0}^{k}\varphi_X(b)=\sum_{b=max\{0,k-n\}}^{min\{k,m\}}\varphi_X(b)=\frac{\sum_b\binom{m}{b}\cdot\binom{n}{k-b}}{\binom{n+m}{k}}=1$$

Esempio Un'azienda produce 400 tastiere al giorno di cui 10 difettose. Inoltre, ogni giorno, l'azienda controlla 5 tastiere. Com'è distribuito tra queste il numero di tastiere difettose?
Per rispondere basta definire una *variabile aleatoria ipergeometrica* $D\sim hyp(5,10,390)$.

8.7.1 Ipergeometriche in R

- `dhyper(x, m, n, k)`
  dove `x` è il punto in cui si calcola φ_X $X \sim hyp(k, m, n)$.

- `phyper(q, m, n, k, lower.tail = TRUE)`
  la funzione si comporta come le corrispondenti nelle altre famiglie.

- `rhyper(nn, m, n, k)`
  genera `nn` risultati casuali distribuiti come `ipergeometriche`.

- `qhyper`
  funzione *quantile*.

Esempio Nel gioco del Blackjack si gioca con un mazzo da 52 carte. Le carte hanno un punteggio: 10 le figure, 1 o 11 l'asso, tutte le altre valgono tanto quanto il proprio valore nominale. A ogni giocatore vengono poi date due carte.

Qual è la probabilità di fare Blackjack, cioè 21, con 2 carte?

Procediamo per passi:

(i) Qual è la probabilità che le due carte in mano valgano 10 o siano assi?

Considero un'*ipergeometrica* con $k = 2, m = 4 + 4 + 4 + 4 + 4 = 20, n = 52 - 20 = 32 \Rightarrow$ $X \sim hyp(2, 20, 32)$. Per calcolare la probabilità basta usare R:
$\varphi_X(2) =$`dhyper(2, 2, 20, 32)` $=$ A

(ii) Qual è la probabilità di avere due assi in mano?

$Y \sim hyp(2, 4, 48) \Rightarrow \varphi_Y(2) =$`dhyper(2, 2, 4, 48)` $=$ B

(iii) Qual è la probabilità di avere due carte di valore 10 in mano?

$Z \sim hyp(2, 16, 36) \Rightarrow \varphi_Z(2) =$`dhyper(2, 2, 16, 36)` $=$ C

A questo punto la probabilità di fare Blackjack è:

$$P(\text{"Blackjack"}) = A - B - C$$

Esempio Ogni giocatore ha 2 carte in mano da combinare con 5 carte sul tavolo. Un giocatore ha in mano il 3$\heartsuit$ e il 7$\spadesuit$. Le prime 2 carte sul tavolo sono il 3$\spadesuit$, $Q\spadesuit$. Con che probabilità le prossime 3 carte gli faranno fare un poker o un colore?

Per quanto riguarda il colore, dobbiamo calcolare la probabilità che escano 3 carte di cuori o 2 carte di picche:

(i) Consideriamo il seme cuori:
Per fare colore devono esserci 5 carte dello stesso seme (non in scala). In questo caso, una carta è già in gioco (3$\heartsuit$ in mano). Ne rimangono quindi 12 nel mazzo. I parametri della *ipergeometrica* sono:

- $k = 3$ perché devo pescare 3 carte
- $m = 12$ perché le carte di cuori rimaste sono 12
- $n = 36$ perché dalle 52 carte del mazzo tolgo le 12 di cuori e le 4 giù in gioco

Definiamo adesso la *variabile aleatoria* come: $X_\heartsuit \sim hyp(3, 12, 36)$.

(ii) Consideriamo il seme picche:
Facendo lo stesso ragionamento possiamo definire $X_\spadesuit$ come: $X_\spadesuit \sim hyp(3, 10, 38)$.

A questo punto possiamo calcolare la probabilità di fare colore come:

$$\varphi_{X_\heartsuit}(3) + \varphi_{X_\spadesuit}(2)$$

NB. Stiamo calcolando la *massa* in 3 e 2 perché per fare colore abbiamo bisogno rispettivamente di 3 carte di cuori o due 2 picche.

NB. In realtà bisognerebbe sottrarre dal risultato la probabilità di fare scala colore.

Ora calcoliamo la probabilità di fare poker. Ho 3 casistiche:

(i) Poker di 3: $Y_3 \sim hyp(3, 2, 46)$ $\varphi_{Y_3}(2)$

(ii) Poker di 7: $Y_7 \sim hyp(3, 3, 45)$ $\varphi_{Y_7}(3)$

(iii) Poker di Q: $Y_Q \sim hyp(3, 3, 45)$ $\varphi_{Y_Q}(3)$

Quindi la probabilità di fare poker è:

$$\varphi_{Y_3}(2) + \varphi_{Y_7}(3) + \varphi_{Y_Q}(2)$$

Prop. Supponiamo di avere due successioni di interi non negativi $\{a_i\}_{i\in\mathbb{N}}$ e $\{b_i\}_{i\in\mathbb{N}}$ che tendono in modo monotono a $+\infty$. In particolare varrà $\lim_{i\to+\infty}\{a_i\}_{i\in\mathbb{N}} = \lim_{i\to+\infty} = \{b_i\}_{i\in\mathbb{N}} = +\infty$ e le due successioni saranno crescenti. Inoltre, si supponga che le successioni siano tali che:

$$\frac{a_i}{a_i + b_i} = \alpha \text{ con } \alpha \in [0, 1]$$

Allora:

$$\lim_{i\to+\infty} \frac{\binom{a_i}{k} \cdot \binom{b_i}{n-k}}{\binom{a_i+b_i}{n}} = \binom{n}{k} \cdot \alpha^k \cdot (1-\alpha)^{n-k}$$

Oss. La proposizione precedente equivale a dire che, ad esempio, se si ha una successione di urne contenenti un certo numero di biglie bianche e nere, tale che il rapporto tra il numero di bianche e il totale sia un valore finito. Allora sappiamo che le *ipergeometriche* di parametri a_i e b_i tendono a una *binomiale* che ha come probabilità di successo α.

Dim.

(i) $\frac{b_i}{a_i+b_i} = 1 - \frac{a_i}{a_i+b_i} \xrightarrow[i\to+\infty]{} 1 - \alpha$

(ii) Siano c, d costanti:

$$\frac{a_i - c}{a_i + b_i - d} = \frac{a_i}{a_i + b_i} \cdot \frac{1 - \frac{c}{a_i}}{1 - \frac{d}{a_i+b_i}} \xrightarrow[i\to+\infty]{} \alpha$$

(iii) Siano c, d costanti:

$$\frac{b_i - c}{a_i + b_i - d} = \frac{b_i}{a_i + b_i} \cdot \frac{1 - \frac{c}{b_i}}{1 - \frac{d}{a_i+b_i}} \xrightarrow[i\to+\infty]{} 1 - \alpha$$

(iv)

$$\frac{\binom{a_i}{k}\cdot\binom{b_i}{n-k}}{\binom{a_i+b_i}{k}} = \frac{a_i!\cdot b_i!\cdot n!\cdot(a_i+b_i-n)!}{k!\cdot(a_i-k)!\cdot(n-k)!\cdot(b_i-n+k)!\cdot(a_i+b_i)!}$$

$$= \binom{n}{k}\cdot\frac{a_i!}{(a_i-k)!}\cdot\frac{b_i!}{(b_i-(n-k))!}\cdot\frac{(a_i+b_i-n)!}{(a_i+b_i)!}$$

$$= \binom{n}{k}\frac{a_i\cdot(a_i-1)\cdot\dots\cdot(a_i-(k-1))\cdot b_i\cdot(b_i-1)\cdot\dots\cdot(b_i-(n-k-1))}{(a_i+b_i)\cdot(a_i+b_i-1)\cdot\dots\cdot(a_i+b_i-(n-1))}$$

$$= \binom{n}{k}\cdot\frac{a_i}{a_i+b_i}\cdot\dots\cdot\frac{a_i-(k-1)}{a_i+b_i-(k-1)}\cdot\frac{b_i}{a_i+b_i-k}\cdot\dots\cdot\frac{b_i-(n-k-1)}{a_i+b_i-(n-1)}$$

$$= \binom{n}{k}\cdot\alpha^k\cdot(1-\alpha)^{n-k}$$

NB. Nel penultimo passaggio abbiamo usato i punti (ii) e (iii) della dimostrazione.

8.8 Poissoniane

Esempio In una partita di Premier League vengono segnati, in media, 2.5 goal a partita. Come possiamo descrivere con una *variabile aleatoria* il numero di goal a partita?

Come prima idea, possiamo dividere i 90 minuti della partita in 5 slot da 18. A questo punto pensiamo ogni intervallo come una *Bernoulliana* di parametro $p = \frac{1}{2} \Rightarrow bin(1, \frac{1}{2})$ e quindi nella partita abbiamo $X_1 \sim bin(5, \frac{1}{2})$.

Oss. Con questo approccio risultano evidenti alcuni problemi:

(i) Stiamo limitando a 1 il numero di goal per intervallo

(ii) Stiamo limitando a 5 il numero di goal per partita

Proviamo quindi a ridurre il tempo per intervallo. Avremo ora 10 intervalli da 9 minuti: $X_2 \sim bin(10, \frac{1}{4})$.

Abbiamo gli stessi problemi di prima, ma va già meglio. Continuiamo a dividere: $X_3 \sim bin(20, \frac{1}{8}), X_4 \sim bin(40, \frac{1}{16}), X_5 \sim bin(45, \frac{1}{18}), X_6 \sim bin(90, \frac{1}{36})$.

Otteniamo in questo modo una successione che converge a una *variabile aleatoria di Poisson.*

Definizione 8.8.1. X è una *variabile aleatoria di Poisson* o *Poissoniana* di parametro λ con $\lambda \in \mathbb{R}^+$ se:

$$\varphi_X(k) = \begin{cases} \frac{\lambda^k}{k!}\cdot e^{-\lambda} & \text{se } k \in \mathbb{N} \\ 0 & \text{altrimenti} \end{cases}$$

Oss. Alcuni casi pratici in cui usiamo una *Poissoniana* sono quali in cui vogliamo descrivere una *binomiale* con parametri n e p grandi, piccoli o non noti con precisione. Ad esempio:

- Numero di email ricevute in un giorno da un certo utente
- Numero di morti sul lavoro in un anno in Italia
- Numero di domande di ammissione al 1^o di informatica

Prop. Sia $\{P_n\}_n$ una successione di numeri in $[0,1]$ tale che $\lim_{n\to+\infty} n\cdot p_n = \lambda$ con $\lambda \in \mathbb{R}^+$. Allora:

$$\forall k \in \mathbb{N} \lim_{n\to+\infty}\binom{n}{k}\cdot P_n^k\cdot(1-P_n)^{n-k} = \frac{\lambda^k}{k!}\cdot e^{-\lambda}$$

Dim.

$$\binom{n}{k} \cdot P_n^k \cdot (1-P_n)^{n-k} = \frac{n!}{k! \cdot (n-k)!} \cdot P_n^k \cdot (1-P_n)^{n-k}$$
$$= \frac{n \cdot (n-1) \cdot \dots \cdot (n-(k-1))}{k!} \cdot P_n^k \cdot (1-P_n)^{n-k}$$
$$= \frac{n \cdot (n-1) \cdot \dots \cdot (n-(k-1))}{k! \cdot n^k} \cdot n^k \cdot P_n^k \cdot (1-P_n)^{n-k}$$
$$= \frac{1}{k!} \cdot (n \cdot P_n)^k \cdot \left[\frac{n}{n} \cdot \frac{n+1}{n} \cdot \dots \cdot \frac{n-(k-1)}{n}\right] \cdot \left(1 - \frac{n \cdot P_n}{n}\right)^{n-k}$$
$$= \frac{(n \cdot P_n)^k}{k!} \cdot \left(1 - \frac{n \cdot p_n}{n}\right)^{-k} \cdot \left(1 - \frac{n \cdot P_n}{n}\right)^{n}$$

A questo punto, mandando $n \to +\infty$ otteniamo:

$$\frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

NB.

$$\left(1 - \frac{n \cdot P_n}{n}\right)^n = e^{-n \cdot P_n} = e^{-\lambda} \text{ per } n \to +\infty$$

Questa uguaglianza deriva dal prodotto notevole:

$$\lim_{n \to +\infty} \left(1 - \frac{a}{n}\right)^n = e^{-a}$$

Prop. Le *variabili aleatorie Poissoniane* sono *riproducibili*, cioè:

$$X \sim pois(\lambda_1) \; Y \sim pois(\lambda_2) \Rightarrow X + Y \sim pois(\lambda_1 + \lambda_2)$$

8.8.1 Poissoniane in R

La famiglia di funzioni si chiama `pois`:

- `dpois(x, lambda)`
- `dpois(q, lambda, lower.tail = TRUE)`
- `rpois(n, lambda)`
- `qpois`

Esempio Siano $X \sim pois(\lambda_1)$ e $Y \sim pois(\lambda_2)$ indipendenti e sia $S = X + Y$. Qual è la legge di $X|S$?

Partiamo dalla definizione:

$$\varphi_{X|S}(k|n) = P(X = k|S = n) = \frac{P(X = k, S = n)}{P(S = n)} = \frac{P(X = k, Y = n-k)}{P(S = n)}$$
$$= \frac{P(X = k) \cdot P(Y = n-k)}{P(S = n)} = \frac{\lambda_1^k}{k!} \cdot e^{-\lambda_1} \cdot \frac{\lambda_2^{n-k}}{(n-k)!} \cdot e^{-\lambda_2} \cdot \frac{n!}{(\lambda_1 + \lambda_2)^n} \cdot e^{\lambda_1 + \lambda_2}$$
$$= \frac{\lambda_1^k}{k!} \cdot \frac{\lambda_2^{n-k}}{(n-k)!} \cdot \frac{n!}{(\lambda_1 + \lambda_2)^n} = \binom{n}{k} \cdot \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \cdot \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}$$

Notiamo che $\varphi_{X|S}(x|s) \sim bin(s, \frac{\lambda_1}{\lambda_1 + \lambda_2})$.

8.9 Riepilogo

Riproponiamo ora tutte le *variabili aleatorie* viste in questo capitolo:

- *Bernoulliane* [$X \sim bin(1,p)$]: è una *variabile aleatoria indicatrice* di un evento $E =$ "*successo*" che si verifica con probabilità p.
- *Binomiali* [$X \sim bin(n,p)$]: conta il numero di successi che si verificano in n tentativi.

  $\varphi_X(k)$: indica con che probabilità si verificano k successi
- *Geometriche* [$X \sim bin(p)$]: indica l'istante precedente al primo successo.

  $\varphi_X(k)$: indica con che probabilità si verificano k insuccessi prima del primo successo
- *Binomiali negative* [$X \sim bin(n,p)$]: conta il numero di insuccessi precedenti all'n-esimo successo.

  $\varphi_X(k)$: indica con che probabilità si verificano k insuccessi prima dell'n-esimo successo
- *Ipergeometriche* [$X \sim hyp(k,m,n)$]: conta il numeri di oggetti di tipo a estratti, senza reimmissione, da un insieme di m oggetti di tipo a ed n di tipo b.

  $\varphi_X(k)$: indica con che probabilità vengono estratti k oggetti di tipo a tra gli m di tipo a e gli n di tipo b
- *Poissoniane* [$X \sim pois(\lambda)$]: si usa quando si ha una successione di *ipergeometriche* di parametri a_i e b_i che tende a una *binomiale* con probabilità di successo α.

  $\varphi_X(k)$: se, ad esempio, si stanno considerando i morti giornalieri sul lavoro, e in media, ne muoiono $3(= \lambda)$ ogni giorno, indica la probabilità che un giorno ne muoiano k

Capitolo Nr.9

Indicatori di una variabile aleatoria

9.1 Valore atteso

9.1.1 Variabili aleatorie discrete

Definizione 9.1.1. Il *valore atteso* o *media* o *speranza* di una *variabile aleatoria discreta* X è il baricentro o centro di massa della sua distribuzione:

$$\mathbf{E}[X] = E[X] = \sum_{x \in \mathcal{R}_X} x \cdot \varphi_X(x)$$

Esempio Esiste un gioco in cui se lanciando una moneta bilanciata la prima testa esce al lancio n si vincono 2^n monete. Sia X la *variabile aleatoria* associata all'ammontare della vincita, qual è la sua media?

$$\varphi_X(2^n) = P(X = 2^n) = P(T_1 = n) = \left(\frac{1}{2}\right)^{n-1} \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^{n} = \frac{1}{2^n}$$

Oss. Se prendo un evento H nella mia *tribù* quanto vale $P(\dot{|}H)$?

$$\sum_{x \in \mathcal{R}_X} x \cdot P(X = x|H) = E[X|H] = \text{"speranza di X condizionata ad H"}$$

Esempio Sia H l'evento $Y = y$.

$$E[X|Y = y] = \sum_{x \in \mathcal{R}_X} x \cdot P(X = x|Y = y) = \sum_{x \in \mathcal{R}_X} x \cdot \varphi_{X|Y}(x|y)$$

Esempio Siano X una *variabile aleatoria discreta* con *massa* φ_X e $Y = g(X)$. Allora:

$$E[Y] = \sum_{k \in \mathcal{R}_X} g(k) \cdot \varphi_X(k)$$

Dim.

$$E[Y] = \sum_{y \in \mathcal{R}_Y} y \cdot \varphi_X(y) = \sum_{y \in \mathcal{R}_Y} y \cdot \sum_{x \in g^{-1}(\{y\})} \varphi_X(x) = \sum_{y \in \mathcal{R}_Y} \sum_{x \in g^{-1}(\{y\})} g(x) \cdot \varphi_X(x)$$

$$= \sum_{x \in \mathcal{R}_X} g(x) \cdot \varphi_X(x)$$

Teorema 9.1.1. Sia (X, Y) un *vettore aleatorio discreto* con *densità congiunta* $\varphi_{X,Y}$. Sia poi $Z = g(X, Y)$ $g : \mathbb{R}^2 \to \mathbb{R}$. Allora:

$$E[Z] = \sum_{j\in\mathcal{R}_X} \sum_{k\in\mathcal{R}_Y} g(j, k) \cdot \varphi_{X,Y}(j, k)$$

Esempio Siano X, Y due dadi a 20 facce indipendenti e $X = min(X, Y)$. Quanto vale $E[Z]$?

$$E[Z] = \sum_{j\in\mathcal{R}_X} \sum_{k\in\mathcal{R}_Y} \frac{min(j,k)}{20\cdot 20} = \frac{1}{400} \cdot \sum_{j=1}^{20}\sum_{k=1}^{20} min(j,k) = \frac{1}{400} \cdot \sum_{j=1}^{20} \left[\sum_{k=1}^{j} k + \sum_{k=j+1}^{20} j\right]$$

$$= \frac{1}{400} \cdot \sum_{j=1}^{20} \left[\frac{j\cdot(j+1)}{2} + j \cdot (20-j)\right] = \frac{1}{400} \cdot \sum_{j=1}^{20} \left(\frac{-j^2}{2} + \frac{41}{2} \cdot j\right)$$

$$= \frac{1}{800} \cdot \sum_{j=1}^{20} (-j^2 + 41 \cdot j) = \frac{5740}{800} = 7.175$$

Prop. Il *valore atteso* (per *variabili aleatorie discrete*) gode delle seguenti proprietà:

(i) *Linearità*: se X e Y sono *variabili aleatorie discrete* e $a, b \in \mathbb{R}$, allora:

$$E[aX + Y + b] = a \cdot E[X] + E[Y] + b$$

(ii) *Prodotto di v.a. indipendenti*: siano X, Y due *variabili aleatorie indipendenti*:

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

(iii) *Monotonia*: sia X una *variabile aleatoria discreta*, se $X \geq 0$ allora $E[X] \geq 0$. In più $E[X] = 0 \Leftrightarrow X = 0$

Dim.

(i) *Linearità*: siano (X, Y) *variabili aleatorie discrete* e sia $g(x, y) = ax + y + b$, allora:

$$E[aX + Y + b] = E[g(X, Y)] = \sum_{x\in\mathcal{R}_X} \sum_{y\in\mathcal{R}_Y} g(x, y) \cdot \varphi_{X,Y}(x, y)$$

$$= \sum_{x\in\mathcal{R}_X} \sum_{y\in\mathcal{R}_Y} (ax + y + b) \cdot \varphi_{X,Y}(x, y)$$

$$= \sum_{x\in\mathcal{R}_X} ax \cdot \sum_{y\in\mathcal{R}_Y} \varphi_{X,Y}(x, y) + \sum_{x\in\mathcal{R}_X} \sum_{y\in\mathcal{R}_Y} y \cdot \varphi_{X,Y}(x, y) + b \cdot \sum_{x\in\mathcal{R}_x} \sum_{y\in\mathcal{R}_Y} \varphi_{X,Y}(x, y)$$

$$= a \cdot \sum_{x\in\mathcal{R}_X} x \cdot \varphi_X(x) + \sum_{y\in\mathcal{R}_Y} y \cdot \sum_{x\in\mathcal{R}_X} \varphi_{X,Y}(x, y) + b$$

$$= a \cdot \sum_{x\in\mathcal{R}_X} x \cdot \varphi_X(x) + \sum_{y\in\mathcal{R}_Y} y \cdot \varphi_Y(y) + b = a \cdot E[x] + E[Y] + b$$

(ii) *Prodotto di v.a. indipendenti*: siano (X, Y) *variabili aleatorie discrete* e sia $g(x, y) = x \cdot y$, allora:

$$E[X \cdot Y] = E[g(x, y)] = \sum_{x\in\mathcal{R}_X} \sum_{y\in\mathcal{R}_Y} g(x, y) \cdot \varphi_{X,Y}(x, y) = \sum_{x\in\mathcal{R}_X} \sum_{y\in\mathcal{R}_Y} x \cdot y \cdot \varphi_X(x) \cdot \varphi_Y(y)$$

$$= \sum_{x\in\mathcal{R}_x} x \cdot \varphi_X(x) \cdot \sum_{y\in\mathcal{R}_Y} y \cdot \varphi_Y(y) = E[X] \cdot E[Y]$$

(iii) *Monotonia*: $E[X] = \sum_{x\in\mathcal{R}_X} x \cdot \varphi_X(x) \geq 0$ perché tutti gli addendi lo sono, mentre la somma fa zero se e solo se $x = 0 \Leftarrow \varphi_x(0) = 1$

Corollario 9.1.1.1. Se X e Y sono due *variabili aleatorie discrete* tali che $P(X \geq Y) = 1$ allora $E[X] \geq E[Y]$. In più $E[X] = E[Y] \Leftrightarrow X = Y$.

Dim. Sia $Z = X - Y$. Siccome $X \geq Y$ $Z \geq 0 \Rightarrow 0 \leq E[Z] = E[X] - E[Y] \Rightarrow E[X] \geq E[Y]$.

9.1.2 Riepilogo

- *Bernoulliane* $[X \sim bin(1,p)]$: $E[X] = p$
- *Binomiali* $[X \sim bin(n,p)]$:

$$X = \sum_{i=1}^{n} Y_i \text{ con } Y_i \sim bin(1,p) \text{ allora } E[X] = E\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} E[Y_i] = \sum_{i=1}^{n} p = n \cdot p$$

- *Geometriche* $[X \sim geom(p)]$:

$$E[X] = \sum_{k \in \mathcal{R}_X} k \cdot \varphi_X(k) = \sum_{k=0}^{+\infty} k \cdot P(X = k) = \sum_{k=0}^{+\infty} \sum_{j=0}^{k-1} P(X = k) = \sum_{j=0}^{+\infty} \sum_{k=j+1}^{+\infty} P(X = k)$$

$$= \sum_{j=0}^{+\infty} P(X > j) = \sum_{j=1}^{+\infty} (1-p)^{j+1} = \frac{1-p}{1-(1-p)} = \frac{1-p}{p}$$

- *Binomiali negative* $[X \sim NB(n,p)]$: sono come le *geometriche*, quindi se $E[X] = n \cdot E[Y]$ con $Y \sim geom(p)$ segue che:

$$E[X] = \frac{n \cdot (1-p)}{p}$$

- *Ipergeometriche* $[X \sim hyp(k,m,n)]$: definiamo una *variabile aleatoria indicatrice* Y_i come:

$$Y_i = \begin{cases} 1 & \text{se la i-esima biglia è bianca} \\ 0 & \text{se la i-esima biglia è nera} \end{cases} \quad \text{con } i \in \{1, \dots, k\} \Rightarrow X = \sum_{i=1}^{k} Y_i$$

  Qual è φ_{Y_1}?

$$\varphi_{Y_1}(h) = \begin{cases} \frac{m}{m+n} & \text{se} h = 1 \\ \frac{n}{m+n} & \text{se} h = 0 \\ 0 & \text{altrimenti} \end{cases}$$

  Se non sappiamo cosa è uscito prima anche $Y_1 \sim bin(1, \frac{m}{m+n})$, allora:

$$E[X] = \sum_{i=1}^{k} E[Y_i] = k \cdot \frac{m}{m+n}$$

- *Poissoniane* $[X \sim pois(\lambda)]$:

$$E[X] = \sum_{x \in \mathcal{R}_X} k \cdot \varphi_X(k) = \sum_{k=0}^{+\infty} k \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} = e^{-\lambda} \cdot \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} = \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \cdot e^{-\lambda} \cdot \lambda$$

  Poniamo adesso $h = k - 1$

$$\Rightarrow E[X] = \lambda \cdot \sum_{h=0}^{+\infty} \cdot \frac{\lambda^h}{h!} \cdot e^{-\lambda} = \lambda \cdot \sum_{h=0}^{+\infty} \varphi_X(h) = \lambda$$

9.1.3 Variabili aleatorie assolutamente continue

Definizione 9.1.2. Se X è una *variabile aleatoria assolutamente continua*, allora:

$$E[X] = \int_{-\infty}^{+\infty} x \cdot f_X(x)dx$$

Teorema 9.1.2. Se X è una *variabile aleatoria assolutamente continua* e $Y = g(x)$ per qualche g, allora:

$$E[Y] = \int_{\mathbb{R}} g(x) \cdot f_X(x)dx$$

Teorema 9.1.3. Sia (X, Y) è un *vettore aleatorio assolutamente continuo* e $g : \mathbb{R} \to \mathbb{R}$. Se $Z = g(x, y)$, allora:

$$E[Z] = \int\int_{\mathbb{R}^2} g(x, y) \cdot f_{x,y}(x, y)dx\ dy$$

Teorema 9.1.4. Sia (X, Y) un *vettore aleatorio misto* con X *discreta* e Y *assolutamente continua* e con una *densità ibrida (mista)* $f_{X,Y}$. Se $g : \mathbb{R}^2 \to \mathbb{R}$ e $X = g(x, y)$, allora:

$$E[Z] = \sum_{x \in \mathcal{R}_X} \int_{-\infty}^{+\infty} g(x, y) \cdot f_{X,Y}(x, y)dy$$

Prop. Siano X e Y *variabili aleatorie assolutamente continue.* Il *valore atteso* gode delle seguenti proprietà:

(i) *Linearità*:

$$E[aX + Y + b] = a \cdot E[X] + E[Y] + b$$

(ii) *Prodotto di v.a. indipendenti*:

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

(iii) *Monotonia*:

- $X \geq 0 \Rightarrow E[X] \geq 0$
- $X = 0 \Rightarrow E[X] = 0$

Esempio Sia X una *variabile aleatoria* di *densità*:

$$f_X(x) = \begin{cases} e^{-x} & \text{se } x > 0 \\ 0 & \text{altrimenti} \end{cases}$$

Qual è $E[X]$? E $[\sqrt{x}]$?

$$E[X] = \int_{-\infty}^{+\infty} x \cdot f_X(x)dx = \int_0^{+\infty} x \cdot e^{-x}dx = -[x \cdot e^{-x}]_0^{+\infty} + \int_0^{+\infty} e^{-x}dx = 0 + [-e^{-x}]_0^{+\infty}$$

$$= 0 - (-e^{-x}) = 0 - (-1) = 1$$

$$E[\sqrt{x}] = \int_0^{+\infty} \sqrt{x} \cdot f_X(x)dx = \int_0^{+\infty} \sqrt{x} \cdot e^{-x}dx \underset{x=\frac{\xi^2}{2}}{=} \int_0^{+\infty} \frac{\xi}{\sqrt{2}} \cdot e^{-\frac{-\xi^2}{2}} \cdot \xi d\xi$$

$$= \frac{1}{\sqrt{2}} \cdot \int_0^{+\infty} e^{\frac{-\xi^2}{2}} d\xi = \sqrt{\frac{\pi}{2}} \cdot \frac{1}{\sqrt{2}} = \frac{\sqrt{\pi}}{2}$$

Esempio Si consideri la seguente funzione:

$$f_{X,Y}(x,y) = \begin{cases} k \cdot (5x^2 + 10y^2) & \text{se } (x,y) \in [0,1] \times [0,1] \\ 0 & \text{altrimenti} \end{cases}$$

(i) Per quale valore di k $f_{X,Y}(x,y)$ è una *funzione di densità*?

(ii) Siano X, Y *variabili aleatorie* con *densità congiunta* la funzione determinata al punto (i). X e Y sono indipendenti?

(iii) Quanto vale la differenza tra la *media* di X e Y?

(iv) Qual è il *valore atteso* di $X \cdot Y - 0.375$?

Per soddisfare la prima richiesta basta integrare la funzione e imporre k in modo che tale integrale valga 1.

$$\int_0^1 \int_0^1 f_{X,Y}(x,y)dy\ dx = 1 \Rightarrow k = \left(\int_0^1 \int_0^1 (5x^2 + 10y^2)dy\ dx \right)^{-1}$$

$$= \left(\int_0^1 \left(5x^2 + \left[\frac{10}{3}y^3 \right]_0^1 \right) dx \right)^{-1} = \left(\int_0^1 \left(5x^2 + \frac{10}{3} \right) dx \right)^{-1}$$

$$= \left(\left[\frac{5}{3}x^3 + \frac{10}{3}x \right]_0^1 \right)^{-1} = \left(\frac{15}{3} \right)^{-1} \Rightarrow k = \frac{3}{15} = \frac{1}{5}$$

A questo punto la funzione diventa:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{5}(5x^2 + 10y^2) & \text{se } (x,y) \in [0,1] \times [0,1] \\ 0 & \text{altrimenti} \end{cases} \Rightarrow \begin{cases} x^2 + 2y^2 & \text{se } (x,y) \in [0,1] \times [0,1] \\ 0 & \text{altrimenti} \end{cases}$$

Ora, per rispondere al secondo quesito devo ricavare le *marginalizzate* e vedere se vale $f_X(x) \cdot f_Y(y) = f_{X,Y}(x,y)$:

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)dy = \int_0^1 (x^2 + 2y^2)dy = x^2 + \left[\frac{2}{3}y^3 \right]_0^1 = x^2 + \frac{2}{3}$$

$$f_Y(y) = \int_0^1 f_{X,Y}(x,y)dx = \int_0^1 (x^2 + 2y^2)dx = \left[\frac{x^3}{3} \right]_0^1 + 2y^2 = 2y^2 + \frac{1}{3}$$

È evidente, a questo punto, che $f_X(x) \cdot f_Y(y) \neq f_{X,Y}(x,y)$ dunque X e Y non sono indipendenti.

Il terzo quesito chiede di calcolare $E[X] - E[Y]$, quindi calcolo le *medie*:

$$E[X] = \int_0^1 x \cdot f_X(x)dx = \int_0^1 \left(x^3 + \frac{2}{3}x \right) dx = \left[\frac{x^4}{4} + \frac{2}{6}x^2 \right]_0^1 = \frac{1}{4} + \frac{2}{6} = \frac{7}{12}$$

$$E[Y] = \int_0^1 y \cdot f_Y(y)dy = \int_0^1 \left(2y^3 + \frac{y}{3} \right) dy = \left[\frac{y^4}{2} + \frac{y^2}{6} \right]_0^1 = \frac{1}{2} + \frac{1}{6} = \frac{8}{12}$$

Quindi, la differenza tra le *medie* vale:

$$E[X] - E[Y] = \frac{7}{12} - \frac{8}{12} = -\frac{1}{12}$$

Per rispondere all'ultimo quesito sfrutto la *linearità* e quindi riscrivo $E[X \cdot Y - 0.375]$ come $E[X \cdot Y - 0.375] = E[X \cdot Y] - 0.375$. Adesso però, abbiamo un *vettore aleatorio* (X, Y) e una nuova *variabile aleatoria* $Z = g(X, Y) = X \cdot Y$, il cui *valore atteso* è:

$$E[Z] = E[X \cdot Y] = \int_0^1 \int_0^1 g(x, y) \cdot f_{X,Y}(x, y) dy\ dx = \int_0^1 \int_0^1 x \cdot y \cdot (x^2 + 2y^2) dy\ dx$$

$$= \int_0^1 \int_0^1 (x^3 \cdot y + 2x \cdot y^3) dy\ dx = \int_0^1 \left[\frac{x^3}{2} y^2 + \frac{x}{2} y^4\right]_0^1 dx = \int_0^1 \left(\frac{x^3}{2} + \frac{x}{2}\right) dx$$

$$= \left[\frac{x^4}{8} + \frac{x^2}{4}\right]_0^1 = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}$$

Ora, non ci resta che sottrarre 0.375 al risultato appena calcolato:

$$E[X \cdot Y - 0.375] = E[X \cdot Y] - 0.375 = \frac{3}{8} - 0.375 = 0$$

9.2 Momento di una variabile aleatoria

Definizione 9.2.1. $\forall n \in \mathbb{N} \backslash \{0\}$ si dice *momento n-esimo* di X *variabile aleatoria* il numero reale $E[X^n]$, inoltre si dice *momento centrato n-esimo* di X il reale $E[(X - E[X])^n]$. Il *momento secondo centrato* di X è detto *varianza* di X e si indica con $Var[X]$.

Prop.

$$Var[X] = E[X^2] - (E[X])^2$$

Dim. Posto $\mu = E[X]$, vale:

$$Var[X] = E[(X - \mu)^2] = E[X^2 - 2x \cdot \mu + \mu^2] = E[X^2] - 2\mu \cdot E[X] + \mu^2$$

$$= E[X^2] - \mu^2 = E[X^2] - (E[X])^2$$

Oss. La *proposizione* precedente semplifica molto il calcolo effettivo della *varianza*.

Prop. Sia X una *variabile aleatoria* con *varianza* $Var[X]$. Valgono le seguenti:

(i) $Var[X] \geq 0$; $Var[X] = 0 \Leftrightarrow X = c$ con $c \in \mathbb{R}$

(ii) Siano $a, b \in \mathbb{R}$: $Var[aX + b] = a^2 \cdot Var[X]$

Dim.

(i) $Var[X] = E[(X - (E[X])^2)] \geq 0$ grazie alla *monotonia*. In particolare $Var[X] = 0 \Leftrightarrow X - E[X] = 0 \Leftrightarrow X = E[X]$

(ii) $Var[aX + b] = E[(aX + b - a \cdot E[x] - b^2)^2] = E[a^2 \cdot (X - E[X^]))^2] = a^2 \cdot Var[X]$

Prop. Se X, Y sono *variabile aleatorie* indipendenti, allora:

$$Var[X + Y] = Var[X] + Var[Y]$$

Dim.

$$Var[X + Y] = E[(X + Y)^2] - E[X + Y]^2$$

$$= E[X^2] + 2 \cdot E[X \cdot Y] + E[Y^2] - E[X]^2 - 2 \cdot E[X] \cdot E[Y] - E[Y^2]$$

$$= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 - 2 \cdot (E[X \cdot Y] - E[X] \cdot E[Y])$$

$$Var[X] + Var[Y] - 2 \cdot (E[X \cdot Y] - E[X \cdot Y]) = Var[X] + Var[Y]$$

Esempio Sia X una *variabile aleatoria assolutamente continua* con *densità*:

$$f_X(t) = \begin{cases} c \cdot t^{-8} & \text{se } t > 6 \\ 0 & \text{altrimenti} \end{cases}$$

(i) Quanto vale c?

(ii) Qual è la *speranza* di X?

(iii) Qual è la *varianza* di X?

Per rispondere al primo quesito dobbiamo trovare una *costante di rinormalizzazione* c tale che:

$$\int_6^{+\infty} c \cdot t^{-8} dt = 1 \Rightarrow c = \left(\int_6^{+\infty} t^{-8} dt \right)^{-1}$$

Risolvendo l'integrale si ottiene $c = 7 \cdot 6^7 = 1959552$.

Per il calcolo della *speranza* e della *varianza* applico le definizioni:

$$E[X] = \int_{-\infty}^{+\infty} t \cdot f_X(t)dt = \int_6^{+\infty} c \cdot t \cdot t^{-8} dt = c \cdot \int_6^{+\infty} t^{-7} dt = c \cdot \left[\frac{t^{-}-6}{6} \right]_6^{+\infty} = c \cdot \frac{6^{-6}}{6} = \frac{c}{6^7}$$

$$Var[X] = E[X^2] - (E[X])^2 = c \cdot \int_6^{+\infty} t^2 \cdot f_X(t)dt - (E[X])^2 = c \cdot \frac{6^{-5}}{5} - (E[X])^2$$

Oss (Significato della varianza)**.** La *varianza* misura la "*larghezza*" al quadrato della distribuzione.

Definizione 9.2.2. Chiamiamo *deviazione standard* σ_X di una *variabile aleatoria* X il numero non negativo:

$$\sigma_X = \sqrt{Var[X]}$$

Oss. Sia $Y = \alpha X$.

$$\sigma_Y = \sqrt{Var[Y]} = \sqrt{\alpha^2 Var[X]} = \alpha \sqrt{Var[X]} = \alpha \cdot \sigma_X$$

Cioè, la *deviazione standard* è lineare.

9.2.1 Varianza di modelli discreti

- *Bernoulliane* $[X \sim bin(1, p)]$:

$$Var[X] = E[X^2] - (E[X])^2 = 1^2 \cdot p + 0^2 \cdot (1-p) - p^2 = p \cdot (1-p)$$

- *Binomiali* $[X \sim bin(n, p)]$:

$$X = \sum_{i=1}^{n} Y_i \text{ con } Y_i \sim bin(1, p), Y_i \text{ iid}$$

$$\Rightarrow Var[X] = Var\left[\sum_{i=1}^{n} Y_i \right] = \sum_{i=1}^{n} Var[Y_i] = n \cdot p \cdot (1-p)$$

- *Geometriche* [$X \sim geom(p)$]:
  Inizio definendo una nuova *variabile aleatoria* Y che conta il numero di insuccessi prima del primo successo ignorando il primo lancio:

$$Y = inf\{n \geq 2 : \omega_n = 1\} - 2$$

NB. Il -2 toglie l'istante di primo successo e il primo esito, che infatti voglio ignorare.

Se scrivo la legge di Y:

$$\varphi_Y(k) = P(Y = k) = 1 \cdot (1-p)^k \cdot p$$

noto che $Y \sim X \sim geom(p)$.

$$E[X] = \sum_{k=0}^{+\infty} k \cdot \varphi_X(k) = \sum_{k=0}^{+\infty} k \cdot P(X = k) =$$

$$\sum_{k=0}^{+\infty} k \cdot (P(X = k|\omega_1 = 0)) \cdot P(\omega_1 = 0) + P(X = k|\omega_1 = 1) \cdot P(\omega_1 = 1)$$

$$= \sum_{k=0}^{+\infty} k \cdot P(X = k|\omega_1 = 0) \cdot (1-p) + \sum_{k=0}^{+\infty} k \cdot P(X = k|\omega_1 = 1) \cdot p$$

Se $\omega_1 = 0$ $Y = X - 1 \Rightarrow X = Y + 1$, se invece, $\omega_1 = 1$ non continuo ed ho 0 insuccessi.

$$\Rightarrow \sum_{k=0}^{+\infty} k \cdot P(Y + 1 = k|\omega_1 = 0) \cdot (1-p) + 0 = \left(\sum_{k=0}^{+\infty} k \cdot P(Y + 1 = k)\right) \cdot (1-p)$$

$$= (1-p) \cdot \sum_{k=1}^{+\infty} k \cdot P(Y = k - 1) = \left(\sum_{k=1}^{+\infty} k \cdot P(Y + 1 = k)\right) \cdot (1-p)$$

NB. Facciamo partire k da 1 perché siccome Y traccia istanti, non è ammissibile l'istante negativo che ci sarebbe con $k = 0$. Non ammissibile nel senso che la probabilità sarebbe 0, quindi in una sommatoria non darebbe alcun contributo.

$$\left(\sum_{k=1}^{+\infty} k \cdot P(Y + 1 = k)\right) \cdot (1-p) \Rightarrow E[Y + 1] \cdot (1-p) = (E[Y] + 1) \cdot (1-p)$$

Ma, siccome $Y \sim X$, vale $E[Y] = E[X]$, allora:

$$E[Y + 1] = (E[X] + 1) \cdot (1-p) = E[X] \cdot (1-p) + 1 - p \Leftrightarrow E[X] = \frac{1-p}{p}$$

Siamo interessati alla *varianza*, quindi cerchiamo $E[X^2]$:

$$E[X^2] = E[X^2|\omega_1 = 0] \cdot P(\omega_1 = 0) + E[X^2|\omega_1 = 1] \cdot P(\omega_1 = 1) = E[(Y+1)^2|\omega_1 = 0] \cdot (1-p)$$

$$= E[Y^2 + 2Y + 1|\omega_1 = 0] \cdot (1-p) = (1-p) \cdot (E[Y^2] + 2 \cdot E[Y] + 1)$$

$$= (1-p) \cdot E[Y^2] + (1-p) \cdot \left(2 \cdot \frac{1-p}{p} + 1\right) \Rightarrow E[X^2] = \frac{2 \cdot (1-p) \cdot p}{p^2} \cdot (1-p)$$

Dunque, la *varianza* di X è:

$$Var[X] = E[X^2] - (E[X])^2 = \frac{2 \cdot (1-p)^2}{p^2} + \frac{(1-p) \cdot p}{p^2} - \frac{(1-p)^2}{p^2} = \frac{1-p}{p^2}$$

- *Binomiali negative* $[X \sim NB(n,p)]$:

$$X = \sum_{i=1}^{n} Y_i \text{ con } Y_i \sim geom(p), Y_i \text{ iid}$$

$$Var[X] = Var\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} Var[Y_i] = n \cdot \frac{1-p}{p^2}$$

- *Poissoniane* $[X \sim pois(\lambda)]$:
Sappiamo già che $E[X] = \lambda$, quindi cerchiamo $E[X^2]$. Proviamo a cercare $E[X^2 - X]$ e poi ricaviamo $E[X^2]$ come $E[X^2] = E[X^2 - X] + E[X]$.

$$E[X^2 - X] - E[X \cdot (X-1)] = \sum_{k=0}^{+\infty} k \cdot (k-1) \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

$$= \sum_{k=2}^{+\infty} k \cdot (k-1) \cdot \frac{\lambda^{k-2} \cdot \lambda^2 \cdot e^{-\lambda}}{(k-2) \cdot (k-1) \cdot k!} \underset{h=k-2}{=} \lambda^2 \cdot \sum_{h=0}^{+\infty} \frac{\lambda^h}{h!} \cdot e^{-\lambda} = \lambda^2$$

Quindi, ora ricavo $E[X^2]$:

$$E[X^2] = E[X^2 - X] + E[X] = \lambda^2 + \lambda$$

A questo punto, ricavare la *varianza* è banale:

$$Var[X] = E[X^2] - (E[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Oss. Se $X \sim pois(\lambda)$ $E[X] = Var[X]$.

9.3 Disuguaglianze

Prop (Disuguaglianza di Markov)**.** sia X una *variabile aleatoria* non negativa. Per ogni $a > 0$ vale:

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Dim.

(i) Se $P(X \geq a) = 0$ è già verificata

(ii) Se $P(X \geq a) \neq 0$ vale:

$$E[X] = E[X|X < a] \cdot P(X < a) + E[X|x \geq a] \cdot P(X \geq a)$$
$$\geq E[X|X \geq a] \cdot P(X \geq a) \geq a \cdot P(X \geq a)$$

L'ultimo passaggio è giustificato dal fatto che $X - a|X \geq a$ è una *variabile aleatoria* non negativa, quindi $E[X - a|X \geq a] \geq 0 \Rightarrow E[X|X \geq a] - a \geq 0 \Leftrightarrow E[X|X \geq a] \geq a$.

Prop (Disuguaglianza di Chebychev)**.** Sia X una *variabile aleatoria*. Per ogni $a > 0$ vale:

$$P(|X - E[X]| \geq a) \leq \frac{Var[X]}{a^2}$$

Dim.

$$P(|X - E[X]| \geq a) = P((X - E[X])^2 \geq a) \underset{\text{Markov}}{\leq} \frac{E[(X - E[X])^2]}{a^2} = \frac{Var[X]}{a^2}$$

Oss. Possiamo prendere $a \cdot \sigma_X$ al posto di a e la *disuguaglianza di Chebychev* diventa:

$$P(|X - E[X]| \geq a \cdot \sigma_X) \leq \frac{\sigma_X^2}{a^2 \cdot \sigma_X^2} = \frac{1}{a^2}$$

9.4 Covarianza e correlazione

Definizione 9.4.1. Date X e Y *variabili aleatorie* si chiama *covarianza* di X e Y il numero:

$$Cov[X, Y] = E[(X - E[X] \cdot (Y - E[Y]))]$$

Oss. Se $X = Y$, vale:

$$Cov[X, Y] = E[(X - E[X])^2] = Var[X]$$

NB. Per la *covarianza* vale anche la seguente:

$$Cov[X, Y] = E[X, Y] - E[X] \cdot E[Y]$$

Valgono le seguenti proprietà:

(i) Se X e Y sono indipendenti allora $Cov[X, Y] = 0$

(ii) $Var[X + Y] = Var[X] + Var[Y] + 2 \cdot Cov[X, Y]$

(iii) *Simmetria*: $Cov[X, Y] = Cov[Y, X]$

(iv) *Linearità*: $Cov[a \cdot x + b \cdot Y, Z] = a \cdot Cov[X, Z] + b \cdot Cov[Y, Z]$

(v) *Bilinearità*: Siano $(a_i)_{i=1}^n$ e $(b_j)_{j=1}^m$ *vettori reali* e $(X_i)_{i=1}^n$ e $(Y_j)_{j=1}^m$ *vettori aleatori*, vale:

$$Cov\left[\sum_{i=1}^{n} a_i \cdot X_i, \sum_{j=1}^{m} b_j \cdot Y_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i \cdot b_j \cdot Cov[X_i, Y_j]$$

Definizione 9.4.2. La matrice $Cov[X_i, Y_i]$ si chiama *matrice di covarianza* e la si indica con Σ:

$$Cov[\overrightarrow{a} \cdot \overrightarrow{X}, \overrightarrow{b} \cdot \overrightarrow{Y}] = \overrightarrow{a} \Sigma \overrightarrow{b}$$

Definizione 9.4.3. Due *variabili aleatorie* X, Y si dicono *scorrelate* se e solo se $Cov[X, Y] = 0$.

Esempio Siano X e Y *variabili aleatorie* con *massa congiunta* $\varphi_{X,Y}(x, y)$ definita come:

$$\varphi_{X,Y}(x, y) = \begin{cases} \frac{1}{4} & \text{se } (x, y) \in \{(-1, -1), (1, -1)\} \\ \frac{1}{2} & \text{se } (x, y) = (0, 1) \\ 0 & \text{altrimenti} \end{cases}$$

Si dimostri che X e Y sono *scorrelate*, ma non indipendenti.

Prima di tutto ricaviamo le *marginalizzate* $\varphi_X(x)$ e $\varphi_Y(y)$:

$$\varphi_X(x) = \varphi_{X,Y}(x, -1) + \varphi_{X,Y}(x, 1) = \begin{cases} \frac{1}{4} & \text{se } x = \pm 1 \\ \frac{1}{2} & \text{se } x = 0 \\ 0 & \text{altrimenti} \end{cases}$$

$$\varphi_Y(y) = \sum_{x \in \mathcal{R}_X} \varphi_{X,Y}(x, y) = \begin{cases} \frac{1}{2} & \text{se } y = \pm 1 \\ 0 & \text{altrimenti} \end{cases}$$

Risulta evidente che $\varphi_X(x) \cdot \varphi_Y(y) \neq \varphi_{X,Y}(x, y)$, quindi X e Y non sono indipendenti. Per vedere se sono *scorrelate* calcolo la *covarianza*:

$$Cov[X, Y] = E[X \cdot Y] - E[X] \cdot E[Y]$$

$$= \sum_{(x,y) \in \mathcal{R}_X \times \mathcal{R}_Y} x \cdot y \cdot \varphi_{X,Y}(x, y) - \sum_{x \in \mathcal{R}_x} x \cdot \varphi_X(x) \cdot \sum_{y \in \mathcal{R}_Y} y \cdot \varphi_Y(y) = ... = 0$$

Quindi, X e Y sono *scorrelate*, ma non indipendenti.

Prop. Vale la seguente disuguaglianza:

$$-\sqrt{Var[X] \cdot Var[Y]} \leq Cov[X,Y] \leq \sqrt{Var[X] \cdot Var[Y]}$$

Definizione 9.4.4. Date due *variabili aleatorie* X e Y si dice *coefficiente di correlazione* il numero:

$$\rho(X,Y) = corr[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X] \cdot Var[Y]}}$$

9.5 Altri indicatori

9.5.1 Mediana e mediana impropria

Definizione 9.5.1. Si dice *mediana* di una *variabile aleatoria* X un numero m_X tale che:

$$P(X \leq m_X) = P(X \geq m_X)$$

NB. Entrambe quelle probabilità varranno $\frac{1}{2}$.

Oss. Ci sono diversi problemi:

(i) Nel caso di *variabili aleatorie discrete* non è detto che un tale valore esista

(ii) Non è detto che quel valore sia unico

Oss. Se X è una *variabile aleatoria* valgono le seguenti uguaglianze:

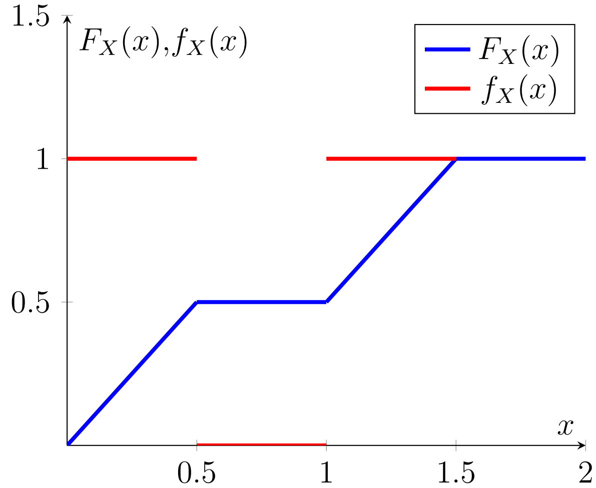
$$P(X \leq m_X) = F_X(m_X)$$

$$P(X \geq m_X) = 1 - F_X(m_X) + P(X = m_X)$$

In particolare, se X è *assolutamente continua*, vale:

$$F_X(m_X) = 1 - F_X(m_X) = \frac{1}{2} \Rightarrow m_X \in F_X^{-1}\left(\left\{\frac{1}{2}\right\}\right)$$

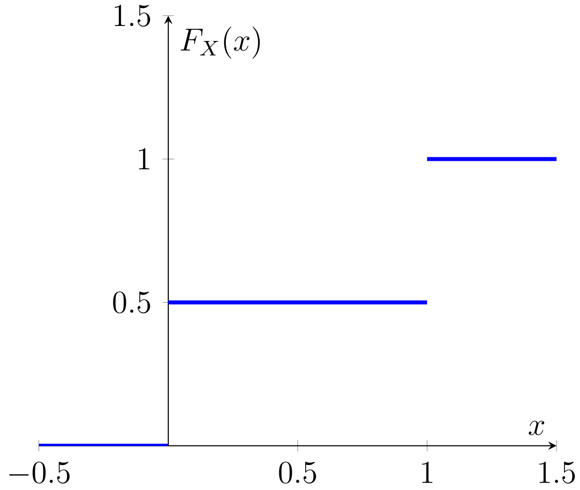
Esempio Sia X una *variabile aleatoria assolutamente continua*, con il seguente grafico delle sue *funzioni di densità* e *ripartizione*:



Ogni punto in $(\frac{1}{2}, 1)$ è *mediana*, inoltre per ogni X *assolutamente continua* esiste una *mediana*, ma non è detto che sia unica. Lo è, solo se F_X è invertibile in m_X.

Oss. Se X è una *variabile aleatoria discreta*, la *mediana* può non essere unica, ma non è detto che esista.

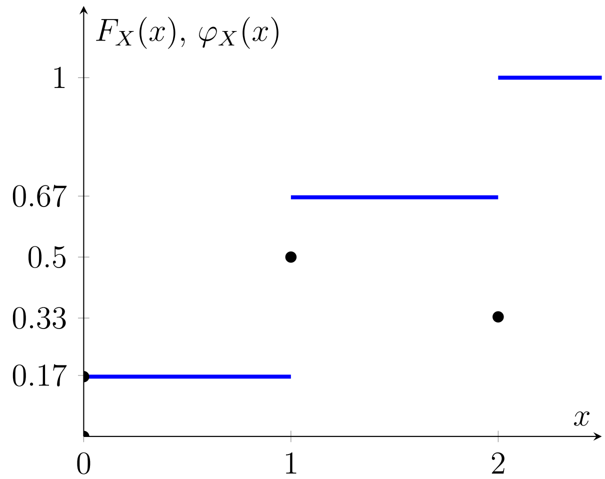
Esempio Sia $X \sim bin(1, \frac{1}{2})$.



Ogni punto in $(0, 1)$ è *mediana.*

Esempio Sia X una *variabile aleatoria discreta* così definita:

$$X = \begin{cases} 0 & \text{con probabilità } \frac{1}{6} \\ 2 & \text{con probabilità } \frac{1}{1} \\ 2 & \text{con probabilità } \frac{1}{3} \end{cases}$$



In questo caso abbiamo:

$$x \in (-\infty, 1) \ P(X \leq x) \leq P(X = 0) = \frac{1}{6} \Rightarrow P(X \geq x) \geq P(X = 1, X = 2) = \frac{5}{6}$$

cioè tutti i valori per $x \in (-\infty, 1)$ sono sbilanciati verso sinistra e quindi non posso essere una *mediana.*

$$x \in (1, +\infty) \ P(X \leq x) \geq P(X \leq 1) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3} \Rightarrow P(X \geq x) \leq P(X = 2) = \frac{1}{3}$$

Qui invece abbiamo il problema opposto in quanto i valori sono troppo sbilanciati verso destra. L'ultima possibilità per avere una *mediana* è che le condizioni necessarie vengano soddisfatte per $x = 1$.

$$x = 1 \ P(X \leq x) = \frac{2}{3} \neq \frac{5}{6} = P(X \geq 1)$$

Dunque, in questo caso, possiamo vedere come non esista alcun punto $m_X \in (-\infty, +\infty)$ tale che $P(X \leq m_X) = P(X \geq m_X)$.

Definizione 9.5.2. Chiamiamo *mediana impropria* un numero reale $\tilde{m}_X$ tale che:

$$P(X \leq \tilde{m}_X) \geq \frac{1}{2} \wedge P(X \geq \tilde{m}_X \geq \frac{1}{2})$$

Oss. Nell'esempio sopra $x = 1$ è una *mediana impropria*.

9.5.2 Quantile

Definizione 9.5.3. Data una *variabile aleatoria* X con legge F_X e $p \in (0,1)$, chiamiamo *p-quantile* (*quantile p-esimo*, *quantile p*) il numero reale:

$$Q_X(p) = inf\{x \in \mathbb{R} : F_X(x) \geq p\}$$

Oss. La funzione *quantile* $Q_X : p \mapsto Q_X(p)$ con $p \in (0,1)$ e $Q_X(p) \in \mathbb{R}$, è qualcosa di simile all'inversa di F_X.

Esempio `qpois(p, lambda, lower.tail=TRUE)`

- Se $p = \frac{k}{4}$ con $k = 1, 2, 3$ parliamo di *quartili*
- Se $p = \frac{k}{10}$ con $k = 1, ..., 9$ parliamo di *decili*
- Se $p = \frac{k}{100}$ con $k = 1, ..., 100$ parliamo di *percentili*

9.5.3 Moda

Definizione 9.5.4. Chiamiamo *moda* di una *variabile aleatoria* X il un numero $x \in \mathcal{R}_X$ tale che:

- Se X è *discreta* φ_X è massima in x, cioè $x \in argmax_y \varphi_X(y)$
- Se X è *assolutamente continua* f_X è massima in x, cioè $x \in argmax_y f_X(y)$

Oss. Intuitivamente possiamo pensare alla moda come ai valori più probabili. Proprio perché non è necessariamente unica distinguiamo tra *variabili aleatorie unimodali* se hanno un'unica *moda*, *bimodali* se ne hanno 2 e *multimodali* se ne hanno molte.

Capitolo Nr.10

Modelli di variabili aleatorie assolutamente continue

10.1 Uniformi

Definizione 10.1.1. Dati due numeri $a, b \in \mathbb{R}$ con $a < b$, chiamiamo una *variabile aleatoria* X *uniforme* su $[a, b]$ se la sua *densità* f_X è costante in $[a, b]$ e nulla altrove. Scriveremo $X \sim unif(a, b)$ oppure $X \sim unif[a, b]$.

$$1 = \int_{-\infty}^{+\infty} f_X dx = \int_a^b c \cdot dx = [c \cdot x]_a^b = c \cdot (b - a) \Rightarrow c = \frac{1}{b - a}$$

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in (a, b) \\ 0 & \text{altrimenti} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{se } x \leq a \\ \int_{-\infty}^{+\infty} \frac{1}{b-a} dt = \int_a^x \frac{1}{b-a} dt & \text{se } a < x \leq b \\ 1 & \text{se } x > b \end{cases} \Rightarrow \begin{cases} 0 & \text{se } x \leq a \\ \frac{x-a}{b-a} & \text{se } a < x \leq b \\ 1 & \text{se } x > b \end{cases}$$

Oss. Data $X \sim unif(a, b)$ gli indicatori visti finora assumono i seguenti valori:

- $E[X] = \int_a^b x \cdot f_X(x) dx = \frac{1}{b-a} \cdot \int_a^b x dx = \frac{1}{b-a} \cdot \left[\frac{x^2}{2}\right]_a^b = \frac{1}{b-a} \cdot \frac{b^2-a^2}{2} = \frac{b+a}{2}$
- $Var[X] = \int_a^b x^2 \cdot f_X(x) dx - \left(\frac{a+b}{2}\right)^2 = \frac{1}{b-a} \cdot \left[\frac{x^3}{3}\right]_a^b - \frac{(a+b)^2}{4} \left[\frac{x^2}{2}\right]_a^b = \frac{b^3-a^3}{3 \cdot (b-a)} - \frac{(a+b)^2}{2} = \frac{(b-a)^2}{12}$
- $m_X = E[X]$ cioè la *mediana* coincide con la *media*
- *Moda* qualunque valore in (a, b)

10.1.1 Uniformi in R

La famiglia di funzioni si chiama `unif`:

- `dunif(x, min = 0, max = 1)`
- `punif(q, min = 0, max = 1, lower.tail = TRUE)`
- `runif(n, min = 0, max = 1)`
- `qunif(p, min = 0, max = 1, lower.tail = TRUE)`

Esempio Un autobus passa ogni 15 minuti. Possiamo descrivere il tempo di attesa come $X \sim unif(0, 15)$. $P(X < 15)$? $P(X > 10|X > 5)$?

$$P(X > 5) = 1 - F_X(5) = 1 - \frac{5-0}{15-0} = 1 - \frac{5}{15} = 1 - \frac{1}{3} = \frac{2}{3}$$

$$P(X > 10|X > 5) = \frac{P(X > 10, X > 5)}{P(X > 5)} = \frac{P(X > 10)}{\frac{2}{3}} = \frac{1 - F_X(10)}{\frac{2}{3}} = \frac{1 - \frac{10}{15}}{\frac{2}{3}} = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

10.2 Esponenziali

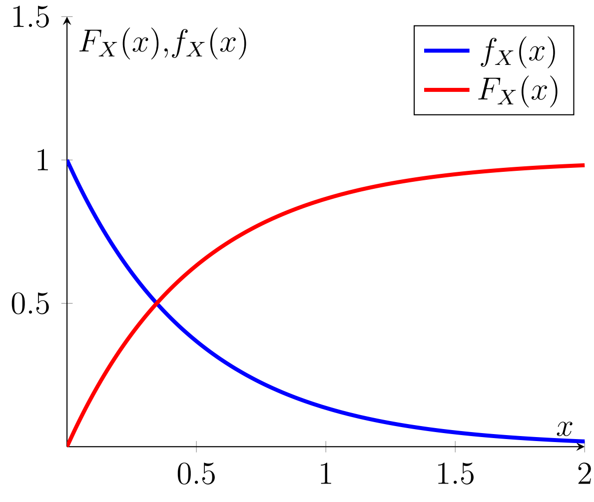
Definizione 10.2.1. Una *variabile aleatoria* X è detta *esponenziale*, di parametro $\lambda \in \mathbb{R}$ con $\lambda > 0$, se:

$$f_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ c \cdot e^{-\lambda x} & \text{se } x \geq 0 \end{cases}$$

In questo caso indicheremo in simboli $X \sim exp(\lambda)$ o $X \sim expo(\lambda)$. λ viene detto anche *rate* o *intensità*.

Non è difficile dimostrare che $c = \lambda$. La *funzione di ripartizione* di una *variabile aleatoria esponenziale* è:

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ \int_0^x \lambda \cdot e^{-\lambda t} dt = \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x} & \text{se } x \geq 0 \end{cases}$$



Oss. Data $X \sim exp(\lambda)$ gli indicatori visti finora assumono i seguenti valori:

- $E[X] = \int_0^{+\infty} x \cdot \lambda \cdot e^{-\lambda x} dx = \lambda \cdot \left[e^{-\lambda x} \cdot \frac{1}{\lambda^2} \cdot (\lambda x - 1)\right]_0^{+\infty} = \frac{1}{\lambda}$
- $Var[X] = \int_0^{+\infty} x^2 \cdot \lambda \cdot e^{-\lambda x} dx - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$
- Vogliamo $F_X(m_X) = \frac{1}{2} \Rightarrow 1 - e^{-\lambda x} = \frac{1}{2} \Rightarrow e^{-\lambda x} = \frac{1}{2} \Rightarrow -\lambda \cdot x = \ln\left(\frac{1}{2}\right) \Rightarrow x = -\frac{\ln\left(\frac{1}{2}\right)}{\lambda} = \frac{\ln(2)}{\lambda} \Rightarrow m_X = \frac{\ln(2)}{\lambda}$
- *Moda* è 0 perché è il punto di massimo di f_X

10.2.1 Esponenziali in R

La famiglia di funzioni si chiama `exp`:

- `dexp(x, rate = 1)`
- `pexp(q, rate = 1, lower.tail = TRUE)`
- `rexp(n, rate = 1)`
- `qexp(p, rate = 1, lower.tail = TRUE)`

Esempio A una fermata dell'autobus il tempo medio di attesa è $\frac{15}{2}$. Supponiamo di poter descrivere il tempo d'attesa come $X \sim exp(\lambda)$. $P(X < 5)$? $P(X > 10|X > 5)$?

Sapendo la *media*, possiamo ricavare facilmente il valore del parametro λ, come $\lambda = \frac{1}{E[X]} = \frac{2}{15}$.

$$P(X > 5) = 1 - F_X(5) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x} = e^{-\frac{15}{2}\cdot 5} = e^{-\frac{2}{3}} \approx 0.51$$

$$P(X > 10|X > 5) = \frac{P(X > 10, X > 5)}{P(X > 5)} = \frac{1 - F_X(10)}{1 - F_X(5)} = e^{-\frac{20}{15}} \cdot e^{\frac{2}{3}} = e^{-\frac{2}{3}} \approx 0.51$$

Oss. Sapere di avere già aspettato 5 minuti, non ci dice nulla riguardo al futuro. Ciò ci suggerisce che probabilmente le *variabili aleatorie esponenziali* non hanno memoria.

Prop. Sia $X \sim exp(\lambda)$. X ha assenza di memoria, cioè presi $t, s > 0$, vale:

$$P(X > s + t|X > s) = P(X > t)$$

Dim.

$$P(X > s + t|X > s) = \frac{P(X > s + t, X > s)}{P(X > s)} = \frac{1 - F_X(s + t)}{1 - F_X(s)}$$

$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda(s)}} = e^{-\lambda t} = 1 - F_X(t) = P(X > t)$$

NB. Le *esponenziali* sono l'equivalente *continuo* delle *geometriche*.

10.3 Gaussiane o normali

Definizione 10.3.1. Una *variabile aleatoria* X è una *normale standard* o se ha *densità*:

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

Scriveremo in simboli: $X \sim \mathcal{N}(0, 1)$.

Oss. Possiamo notare le seguenti:

- $e^{-\frac{x^2}{2}}$ da una forma a campana
- Il fattore $\frac{1}{2}$ a esponente è presente solo per comodità
- $\frac{1}{\sqrt{2\pi}}$ è la *costante di rinormalizzazione*

Proprietà. La *funzione di densità* di una *normale standard* gode delle seguenti:

(i) *Simmetria assiale*: è simmetrica rispetto all'asse $x = 0 \Rightarrow f_X(-x) = f_X(x)$

(ii) Ha il punto di massimo in $x = 0$ e vale $f_X(0) = \frac{1}{\sqrt{2\pi}} \approx 0.4$

(iii) Ha dei punti di flesso in $x = \pm 1$ e in quei punti vale $f_X(1) = f_X(-1) \approx 0.24$

(iv) $f_x(2) = f_X(-2) \approx 0.05$

(v) $f_x(3) = f_X(-3) \approx 0.004$

Oss. La *funzione di ripartizione* di una *normale standard* è la seguente:

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{-t^2}{2}} dt = \Phi(x)$$

Proprietà. La *funzione di ripartizione* gode delle seguenti:

(i) $F_X(0) = \Phi(0) = \frac{1}{2}$

(ii) *Simmetria centrale*: è simmetrica rispetto a $(0, \frac{1}{2})$, quindi $\Phi(-x) = 1 - \Phi(x)$

(iii) $\Phi(-3) \approx 0.0013$, $\Phi(3) = 1 - \Phi(-3) \approx 0.9987$

(iv) $\Phi(-3) \approx 0.0228$, $\Phi(3) = 1 - \Phi(-3) \approx 0.9772$

(v) *Monotonia*: è *monotona strettamente crescente*, quindi è *invertibile*

Oss. Essendo Φ *invertibile*, possiamo definire Φ^{-1} come la *funzione quantile*:

$$\Phi^{-1}(p) = x \Leftrightarrow F_X(x) = p \Leftrightarrow P(X \leq x) = p$$

$\Phi^{-1} : [0, 1] \to \mathbb{R}$ è simmetrica rispetto al punto $(0, \frac{1}{2})$, cioè $\Phi^{-1}(p) = -\Phi^{-1}(1 - p)$.

Oss. Data $X \sim \mathcal{N}(0, 1)$ *media* e *varianza* assumono i seguenti valori:

- $E[X] = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx = \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{0} x \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx + \int_{0}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx \underset{\xi = -x}{=} \int_{-\infty}^{0} -\xi \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx + \int_{0}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} = 0$

  NB. Saremmo potuti giungere allo stesso risultato sfruttando il fatto che $x \cdot f_X(x)$ è una funzione *dispari* e quindi simmetrica rispetto all'origine degli assi.

- $Var[X] = \int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx = \left[-x \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} = 1$

Definizione 10.3.2. Sia $Z \sim \mathcal{N}(0, 1)$. X è una *normale* o *Gaussiana* di parametri $\mu, \sigma \in \mathbb{R}$ con $\sigma > 0$ se:

$$X = \sigma \cdot Z + \mu$$

Scriveremo in simboli $X \sim \mathcal{N}(\mu, \sigma)$.

Oss. Dalle osservazioni fatte nel caso delle *normali standard*, possiamo affermare che:

$$F_X(x) = P(X \leq x) = P(\sigma \cdot Z + \mu \leq x) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

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

Oss. Partendo dalle informazioni sulle *normali standard* possiamo ricavare i valori della *media* e della *varianza* nel caso delle *Gaussiane*.

Siano $Z \sim \mathcal{N}(0,1)$ e $X = \sigma \cdot Z + \mu$, con $\mu \in \mathbb{R}, \sigma > 0$:

$$E[X] = E[\sigma \cdot Z + \mu] = \sigma \cdot E[Z] + \mu = \mu$$

$$Var[X] = Var[\sigma^2 \cdot Z + \mu] = \sigma^2 \cdot Var[Z] = \sigma^2$$

Prop. Siano $X \sim \mathcal{N}(\mu, \sigma)$ e $Z \sim \mathcal{N}(0,1)$. X eredita tutte le proprietà di Z tenendo conto di dilatazione e traslazione.

Prop. La famiglia delle *Gaussiane* è *riproducibile*, cioè prese $X_1 \sim \mathcal{N}(\mu_1, \sigma_1)$ e $X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$, vale:

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$$

10.3.1 Normali in R

La famiglia di funzioni si chiama `norm`:

- `dnorm(x, mean = 0, sd = 1)`
- `pnorm(q, mean = 0, sd = 1, lower.tail = TRUE)`
- `qnorm(p, mean = 0, sd = 1, lower.tail = TRUE)`
- `rnorm(n, mean = 0, sd = 1)`

10.4 Chi quadro

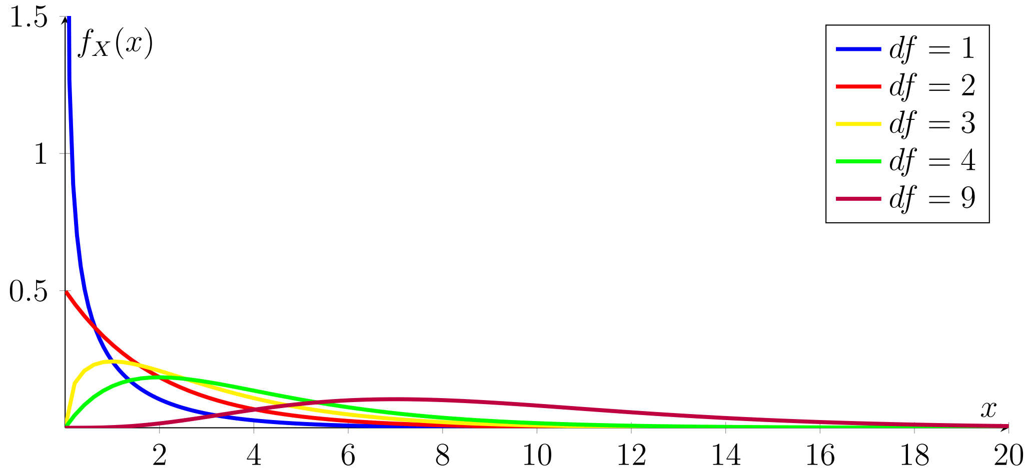
Definizione 10.4.1. Se X è la *variabile aleatoria* definita come somma di quadrati di n *normali standard* indipendenti, diciamo che X è una *Chi quadro* con n *gradi di libertà* (*df*) e scriviamo $X \sim \chi_n^2$ oppure $X \sim \chi^2(n)$.

$$\{Y_i\}_{i=1}^n \; Y_i \sim \mathcal{N}(0,1) \; X = \sum_{i=1}^n Y_i^2$$

Oss. Se sommiamo due quadrati otteniamo un'*esponenziale* di parametro $\lambda = \frac{1}{2}$. La *funzione di densità* è quindi definita in questo modo:

$$f_X(x) = c_n \cdot x^{\frac{x}{2}} \cdot e^{-\frac{x}{2}} \text{ con } c_n \text{ opportuna } \textit{costante di rinormalizzazione}$$

Una volta definita la *funzione di densità* può essere utile esaminare come ne cambi la curva al variare dei *gradi di libertà*.



Oss. Le *Chi quadro* sono riproducibili per definizione. Quindi, se $X \sim \chi_2(n)$ e $Y \sim \chi_2(m)$, vale:

$$X + Y \sim \chi_2(n+m)$$

Oss. Se $X \sim \chi^2(n)$ $X = \sum_{i=1}^{n} Y_i^2$ con $Y_i \sim \mathcal{N}(0,1)$, vale:

$$E[X] = E\left[\sum_{i=1}^{n} Y_i\right] \underset{\text{indipendenza}}{=} \sum_{i=1}^{n} E\left[Y_i^2\right] = n$$

$$Var[X] = \sum_{i=1}^{n} Var[Y_i] = \sum_{i=1}^{n} E\left[Y_i^4\right] - \left(E\left[Y_i^2\right]\right)^2 = \sum_{i=1}^{n} 2 = 2n$$

10.4.1 Chi quadro in R

La famiglia di funzioni si chiama `chisq`:

- `dchisq(x, df)`
- `pchisq(q, df, lower.tail = TRUE)`
- `qchisq(p, df, lower.tail = TRUE)`
- `rchisq(n, df)`

10.5 Distribuzione t di Student

Definizione 10.5.1. Siano $Z \sim \mathcal{N}(0,1)$ e $W \sim \chi^2(n)$ con Z e W indipendenti. X è una t di *Student* a n *gradi di libertà* se:

$$X = \frac{Z}{\sqrt{\frac{W}{n}}}$$

Scriveremo in simboli $X \sim t(n)$ oppure $X \sim t_n$.

NB. Il grafico di una t di *Student* è simile al grafico di una *normale*, con la differenza che la curva tende a stringersi e alzarsi all'origine all'aumentare dei *gradi di libertà.*

Proprietà. Data $X \sim t(n)$, X gode delle seguenti proprietà:

(i) *Densità*: $f_X(-x) = f_X(x)$

(ii) *Ripartizione*: $F_X(-x) = 1 - F_X(x)$ con simmetria in $(0, \frac{1}{2})$

(iii) *Quantile*: $F_X^{-1}(p) = -F_X^{-1}(1-p)$ con simmetria in $(\frac{1}{2}, 0)$

Oss. Se $X \sim t(n)$, vale:

$$E[X] = 0 \ \forall n$$

$$Var[X] = \begin{cases} \text{indefinita} & \text{se } n = 1 \\ +\infty & \text{se } n = 2 \\ \frac{n}{n-2} & \text{se } n > 2 \end{cases}$$

NB. La *speranza* di una t di *Student* è 0 perché la *speranza* di una *normale standard* è costantemente 0.

Oss. Si noti che se $X \sim t(n)$ $\lim_{n\to+\infty} Var[X] = 1$.

Oss. Se $X \sim t(n)$, dalla definizione sappiamo che $X = \frac{Z}{\sqrt{\frac{W}{n}}}$ con $Z \sim \mathcal{N}(0,1)$ e $W \sim \chi^2(n)$. Proviamo a studiare il denominatore $\frac{W}{n}$:

$$E\left[\frac{W}{n}\right] = \frac{E[W]}{n} = \frac{n}{n} = 1$$

$$Var\left[\frac{W}{n}\right] = \frac{Var[W]}{n^2} = \frac{2n}{n^2} = \frac{2}{n}$$

Si noti che $\lim_{n\to+\infty} Var[\frac{W}{n}] = 0$, quindi, in un certo senso, al crescere di n le t di *Student* si distribuiscono come una *normale standard*.

10.5.1 t di Student in R

La famiglia di funzioni si chiama `t`:

- `dt(x, df)`
- `pt(q, df, lower.tail = TRUE)`
- `qt(p, df, lower.tail = TRUE)`
- `rt(n, df)`

Capitolo Nr.11

Convergenza di variabili aleatorie

11.1 Tipi di convergenza

Definizione 11.1.1 (Convergenza quasi certa)**.** Sia $(\Omega, \mathcal{F}, P)$ uno *spazio di probabilità* e siano X una *variabile aleatoria* su di esso e $\{X_n\}_{n\in\mathbb{N}}$ una successione di *variabili aleatorie* sullo stesso spazio. Diciamo che $\{X_n\}_{n\in\mathbb{N}}$ *converge quasi certamente* o *puntualmente* a X se:

$$\exists E \in \mathcal{F} \text{ con } P(E) = 1 \; t.c. \; \forall\omega \in E \; \lim_{n\to+\infty} X_n(\omega) = X(\omega)$$

Scriveremo in simboli:

$$X_n \xrightarrow[n\to+\infty]{q.c} X$$

Definizione 11.1.2 (Convergenza in probabilità)**.** Siano $\{X_n\}_{n\in\mathbb{N}}$ una successione di *variabili aleatorie* e X una *variabile aleatoria* su $(\Omega, \mathcal{F}, P)$. Diciamo che $\{X_n\}_{n\in\mathbb{N}}$ *converge in probabilità* se:

$$\forall\epsilon > 0 \; \lim_{n\to+\infty} P(|X_n - X| > \epsilon) = 0$$

Scriveremo in simboli:

$$X_n \xrightarrow[n\to+\infty]{P} X$$

Definizione 11.1.3 (Convergenza in media quadratica)**.** Siano $\{X_n\}_{n\in\mathbb{N}}$ una successione di *variabili aleatorie* e X una *variabile aleatoria* su $(\Omega, \mathcal{F}, P)$. Diciamo che $\{X_n\}_{n\in\mathbb{N}}$ *converge in media quadratica* se:

$$\lim_{n\to+\infty} E\left[|X_n - X|^2\right] = 0$$

Scriveremo in simboli:

$$X_n \xrightarrow[n\to+\infty]{L^2} X$$

Prop. La *convergenza in media quadratica* implica la *convergenza in probabilità*, cioè:

$$X_n \xrightarrow[n\to+\infty]{L^2} X \Rightarrow X_n \xrightarrow[n\to+\infty]{P} X$$

Dim. Fissiamo $\epsilon > 0$:

$$P(|X_n - X| > \epsilon) = P(|X_n - X|^2 > \epsilon^2) \underset{Markov}{\leq} \frac{E\left[|X_n - X|^2\right]}{\epsilon^2}$$

A questo punto se prendiamo il limite per $n \to +\infty$ otteniamo:

$$\lim_{n\to+\infty} P(|X_n - X| > \epsilon) \leq \frac{\lim_{n\to+\infty} E\left[|X_n - X|^2\right]}{\epsilon^2} = 0$$

L'ultima uguaglianza è garantita dalla definizione di *convergenza in media quadratica.*

Prop. La *convergenza quasi certa* implica la *convergenza in probabilità*, cioè:

$$X_n \xrightarrow[n\to+\infty]{q.c.} X \Rightarrow X_n \xrightarrow[n\to+\infty]{P} X$$

Oss. Dalle precedenti proposizioni otteniamo che:

$$X_n \xrightarrow[n\to+\infty]{L^2} X \Rightarrow X_n \xrightarrow[n\to+\infty]{P} X$$

$$X_n \xrightarrow[n\to+\infty]{q.c.} X \Rightarrow X_n \xrightarrow[n\to+\infty]{P} X$$

Quindi, la *convergenza in media quadratica* e la *convergenza quasi certa* non sono confrontabili.

Definizione 11.1.4. Siano $\{X_n\}_{n\in\mathbb{N}}$ una successione di *variabili aleatorie* su $(\Omega, \mathcal{F}, P)$ e X una *variabile aleatoria* su $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$. Diciamo che $\{X_n\}_{n\in\mathbb{N}}$ *converge in distribuzione*, in *legge* o *debolmente* a X se:

$$\forall x \in \mathbb{R} \lim_{n\to+\infty} P(X_n \leq x) = P(X \leq x), \text{ ossia } \lim_{n\to+\infty} P(X_n \leq x) = F_X(x)$$

Scriveremo in simboli:

$$X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X, X_n \xrightarrow[n\to+\infty]{d} X \text{ o } X_n \xrightarrow[n\to+\infty]{\omega} X$$

Prop. La *convergenza in probabilità* implica la *convergenza debole*, cioè:

$$X_n \xrightarrow[n\to+\infty]{P} X \Rightarrow X_n \xrightarrow[n\to+\infty]{d} X$$

NB. Per *transitività* gli altri due tipi di convergenza implicano *convergenza debole*, cioè:

$$X_n \xrightarrow[n\to+\infty]{L^2} X \Rightarrow X_n \xrightarrow[n\to+\infty]{P} X \Rightarrow X_n \xrightarrow[n\to+\infty]{d} X$$

$$X_n \xrightarrow[n\to+\infty]{q.c.} X \Rightarrow X_n \xrightarrow[n\to+\infty]{P} X \Rightarrow X_n \xrightarrow[n\to+\infty]{d} X$$

11.2 Teoremi limite

Prop. Sia $(X_1, X_2, \ldots, X_n)$ un *vettore aleatorio* con componenti indipendenti tra loro, di *media* e *varianza* comuni μ e σ^2. Sia $S_n = \sum_{i=1}^n X_i$ la *variabile aleatoria* somma. Allora, vale:

$$E\left[\frac{S_n}{n}\right] = \mu$$

$$Var\left[\frac{S_n}{n}\right] = \frac{\sigma^2}{n}$$

Dim. Per *linearità* della *speranza* otteniamo:

$$Var\left[\frac{S_n}{n}\right] = \frac{1}{n^2} \cdot \sum_{i=1}^n Var[X_i] = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

Teorema 11.2.1 (Legge debole dei grandi numeri)**.** Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di *variabili aleatorie* indipendenti, ciascuna di *media* μ e *variazione finita* σ^2. Sia inoltre $S_n = \sum_{i=1}^n X_i$ la *variabile aleatoria* somma. La *variabile aleatoria* $\frac{S_n}{n} \xrightarrow[n\to+\infty]{P} \mu$, cioè:

$$\forall \epsilon > 0 \lim_{n\to+\infty} P(|S_n - \mu| > \epsilon) = 0$$

Dim.

$$P\left(|\frac{S_n}{n}-\mu|>\epsilon\right)=P\left(|\frac{S_n}{n}-E\left[\frac{S_n}{n}\right]|>\epsilon\right)\underset{Chebychev}{\leq}\frac{Var\left[\frac{S_n}{n}\right]}{\epsilon^2}=\frac{\sigma^2}{\epsilon^2}\cdot\frac{1}{n}$$

A questo punto facendo il limite per $n\to+\infty$ si ottiene:

$$\lim_{n\to+\infty}\frac{\sigma^2}{\epsilon^2}\cdot\frac{1}{n}=0$$

Esempio Sia $p=\frac{1}{2}$. Consideriamo il *processo di Bernoulli* di parametro p: $\{X_n\}_{n\in\mathbb{N}}$ con $X_i\sim bin(1,\frac{1}{2})$ iid e quindi $S_n\sim bin(n,\frac{1}{2})$. La *legge debole dei grandi numeri* ci dice che $\frac{S_n}{n}\xrightarrow[n\to+\infty]{P}\frac{1}{2}$.

Vogliamo confrontare S_n con $\frac{1}{2}$

$$\frac{S_n}{n}\xrightarrow[n\to+\infty]{P}\frac{1}{2}\Rightarrow\frac{S_n}{n}-\frac{1}{2}\xrightarrow[n\to+\infty]{P}0\Rightarrow\frac{S_n-\frac{n}{2}}{n}\xrightarrow[n\to+\infty]{P}0$$

Teorema 11.2.2 (Teorema centrale del limite)**.** Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di *variabili aleatorie* indipendenti di *media* μ e *varianza finita* σ^2. Se $S_n=\sum_{i=1}^{n}X_i$ allora, vale:

$$\frac{S_n-n\cdot\mu}{\sigma\cdot\sqrt{n}}\xrightarrow[n\to+\infty]{d}\mathcal{N}(0,1),\text{ cioè }\lim_{n\to+\infty}P\left(\frac{S_n-n\cdot\mu}{\sigma\cdot\sqrt{n}}\leq x\right)=\Phi(x)$$

Oss. Nella pratica, non si usa mai il *teorema centrale dei grandi numeri* in quanto, non si lavora mai su infinite *variabili aleatorie* X_i, ma su un numero finito. Quindi, useremo il teorema per avere delle approssimazioni nel caso n sia sufficientemente grande. In qual caso otteniamo:

$$\frac{S_n-n\cdot\mu}{\sigma\cdot\sqrt{n}}\dot{\sim}\mathcal{N}(0,1)\Rightarrow\frac{S_n}{n}\dot{\sim}\mathcal{N}(\mu,\frac{\sigma}{\sqrt{n}})\Rightarrow S_n\dot{\sim}\mathcal{N}(n\cdot\mu,\sqrt{n}\cdot\sigma)$$

Esempio Un calcolatore somma 10^6 numeri con un errore di arrotondamento in ogni operazione. Gli errori sono distribuiti in modo uniforme sull'intervallo $[-0.5\cdot10^{-10},0.5\cdot10^{-10}]$. Qual è la probabilità che l'errore totale sia minore di $0.5\cdot10^{-7}$?

Sia $X_i\sim unif(-0.5\cdot10^{-10},0.5\cdot10^{-10})$ la *variabile aleatoria* che descrive l'errore di una singola somma. L'errore totale sarà $S_n=\sum_{i=1}^{n}X_i$ con $n=10^6$.

Cerchiamo $P(|S_n|\leq0.5\cdot10^{-7})=P(|S_n|<0.5\cdot10^{-7})$. In questo caso non ci interessa distinguere tra $<$ e $\leq$ in quanto stiamo lavorando con *variabili aleatorie assolutamente continue.*

$$P(|S_n|<0.5\cdot10^{-7})=P(S_n<0.5\cdot10^{-7})-P(S_n<-0.5\cdot10^{-7})$$

Dal *teorema centrale del limite* sappiamo che

$$\frac{S_n-n\cdot E[X_i]}{\sqrt{n}\cdot\sqrt{Var[X_i]}}\dot{\sim}\mathcal{N}(0,1)\Leftrightarrow P\left(\frac{S_n-n\cdot E[X_i]}{\sqrt{n}\cdot\sqrt{Var[X_i]}}\leq x\right)\simeq\Phi(x)$$

Ora, sarebbe comodo riscrivere il risultato ottenuto sopra in modo da arrivare a una forma del tipo $P(S_n\leq y)$:

$$P\left(S_n\leq\sqrt{n\cdot Var[X_i]}\cdot x+n\cdot E[X_i]\right)\simeq\Phi(x)$$

A questo punto ricaviamo x:

$$n\cdot Var[X_i]\cdot x+n\cdot E[X_i]=y\Leftrightarrow x=\frac{y-n\cdot E[X_i]}{\sqrt{n}\cdot\sqrt{Var[X_i]}}$$

Notiamo che saremmo potuti arrivare allo stesso risultato ricordando che $S_n \dot\sim \mathcal{N}(y - n \cdot E[X_i], \sqrt{n} \cdot \sqrt{Var[X_i]})$, e di conseguenza, sfruttando le proprietà delle *Gaussiane*, arrivare ad avere:

$$P(S_n \leq x) = F_{S_n}(y) = \Phi\left(\frac{y - n \cdot E[X_i]}{\sqrt{n} \cdot \sqrt{Var[X_i]}}\right)$$

Ora non ci resta che svolgere i calcoli veri e proprio, inserendo nell'espressione, i seguenti valori:

- $n = 10^6$
- $E[X_i] = 0$
- $Var[X_i] = \frac{10^{-20}}{12}$
- $y = \pm 0.5 \cdot 10^{-7}$

Quindi

$$P(S_n < 0.5 \cdot 10^{-7}) = P(S_n < 0.5 \cdot 10^{-7}) - P(S_n < -0.5 \cdot 10^{-7}) =$$
$$F_{S_n}(0.5 \cdot 10^{-7}) - F_{S_n}(-0.5 \cdot 10^{-7}) = F_{S_n}(0.5 \cdot 10^{-7}) - (1 - F_{S_n}(0.5 \cdot 10^{-7}))$$
$$2 \cdot F_{S_n}(0.5 \cdot 10^{-7}) - 1 \simeq 2 \cdot \Phi\left(\frac{0.5 \cdot 10^{-7}}{10^3 \cdot \frac{10^{-10}}{\sqrt{12}}}\right) - 1 \simeq 2 \cdot \Phi(1.75) - 1 = 1.9108 - 1 \approx 91\%$$

La probabilità che l'errore totale sia minore di $0.5 \cdot 10^{-7}$ è circa 91%.

Oss. Nell'esempio appena visto abbiamo trattato la relazione $<$ come quella $\leq$ e viceversa. Lo abbiamo potuto fare perché stavamo lavorando con *variabili aleatorie assolutamente continue*, ma, nel caso di *variabili aleatorie discrete*, questo tipo di semplificazioni avrebbe potuto provocare errori.

Per risolvere questo inconveniente si può fare affidamento sulla *correzione di continuità.*

Oss. In precedenza ci siamo posti il problema di poter lavorare solo con un numero finito di *variabili aleatorie* e di dover quindi ricorrere ad approssimazioni e che per ottenere un risultato vicino al valore reale c'è bisogno di una parametro n che si sufficientemente grande. Ma quanto grande deve essere?

La risposta dipende dal modello delle *variabili aleatorie* che si stanno considerando:

- *Normali* $[X_i \sim \mathcal{N}]$: grazie alla *riproducibilità* è sufficiente $n \geq 1$
- *Uniformi* $[X_i \sim unif]$: $n \geq 5$
- *Esponenziali* $[X_i \sim exp]$: $n \geq 15$
- *Geometriche* $[X_i \sim geom]$: $n \geq 15$
- *Chi quadro* $[X_i \sim \chi^2]$: $n \geq 25$.
  Possiamo però sfruttare la *riproducibilità* $X_i \sim \chi^2_n \dot\sim \mathcal{N}(n, \sqrt{2n})$, quindi se sommiamo $\chi^2(1)$ ne occorrono almeno 25, ma se sommiamo $\chi^2(9)$ ne bastano circa 3
- *Binomiali* $[X_i \sim bin]$: in questo caso è necessario che p non sia troppo "sbilanciata", cioè che non sia troppo vicina ai valori limite 0 e 1. In tal caso possiamo sfruttare il *teorema centrale del limite* per approssimare la distribuzione stessa:

  $$bin(n, p) \dot\sim \mathcal{N}(n \cdot p, \sqrt{n \cdot p \cdot (1 - p)})$$

  La condizione su p dipende da n e in generale si richiede che $n \cdot p \cdot (1 - p) \gtrsim 3$

- *Poissoniane* $[X_i \sim pois]$: sfruttiamo la *riproducibilità*

$$pois(\lambda) \dot{\sim} \mathcal{N}(\lambda, \sqrt{\lambda})$$

Ricordiamo infatti, che una *Poissoniana* di parametro λ può essere vista come la somma di n *Poissoniane* di parametro $\tilde{\lambda} = \frac{\lambda}{n}$. Per $\lambda \gtrsim 30$ otteniamo una buona approssimazione.

Capitolo Nr.12

Statistica

12.1 Generalità sulla statistica

Definizione 12.1.1. Si definisce *popolazione di riferimento* l'insieme di elementi, distinti tra loro, sui quali viene condotta un'indagine. I singoli elementi vengono detti *individui*, *esemplari* o *unità statistiche.*

Riprendendo le definizioni di *statistica descrittiva* e *inferenziale*, possiamo esaminarne in modo più preciso le differenze. In entrambe vogliamo studiare alcune caratteristiche di una popolazione, ma, mentre nella *statistica descrittiva* abbiamo dati sulla totalità della popolazione, la *statistica inferenziale*, ne considera invece solo un *campione* e utilizza poi metodi probabilistici per ricavare (o inferire) informazioni generalizzate.

Definizione 12.1.2. Un *campione* è un sottoinsieme della popolazione.

Il concetto di *campione* è molto simile a quello di *evento* in probabilità. Sia i *campioni* che gli *eventi* infatti, sono rispettivamente, sottoinsiemi della *popolazione* e dell'*universo.* È quindi logico pensare che ci siano delle regole da rispettare per scegliere un *campione* adatto, proprio come ce n'erano per la definizione di una *tribù.*

Oss. Esistono diversi tipi di campionamento, tra cui:

- *Campionamento casuale semplice*
- *Campionamento stratificato*
- *Campionamento a grappolo*

Definizione 12.1.3. Le caratteristiche che misuriamo prendono il nome di *variabili* e i loro valori prendono il nome di *modalità* o *livelli.* Le *variabili* possono essere di tipo:

- *Qualitative o categoriche*: se sono grandezze non numeriche
  - *Nominali*: se non hanno un ordine naturale (e.g. colore dei capelli)
  - *Ordinali*: se hanno un ordine naturale (e.g. titolo di studio)
- *Quantitative o numeriche*: se sono grandezze numeriche
  - *Discrete*: se sono descritte da numeri interi (e.g. numero di persone)
  - *Continue*: se sono descritte da numeri reali (e.g. una misura di tempo o distanza)

Nel caso di *variabili numeriche*, la scala di misurazione può essere di due tipi:

- *Intervallo*: se lo 0 è scelto in modo arbitrario
- *Rapporto*: se lo 0 è fissato in modo naturale

Definizione 12.1.4. Una *statistica* è una funzione calcolabile a partire dalla misurazione di un *campione.*

Esempio I seguenti sono esempi di *statistiche* calcolabili su un campione $(x_1, \dots, x_n)$:

- *Media campionaria*: $\bar{x} := \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$
- *Varianza campionaria a media (o speranza) μ nota*: $s_*^2 := \frac{1}{n} \cdot \sum_{i=1}^{n}(x_i - \mu)^2$
- *Varianza campionaria a media (o speranza) μ ignota*: $s^2 := \frac{1}{n-1} \cdot \sum_{i=1}^{n}(x_i - \bar{x})^2$
- Il numero di misurazione eccedenti una certa soglia $c := \#\{i \in \{1, \dots, n\} : x_i > c\}$
- Il primo esemplare del campione la cui misurazione è inferiore a una certa soglia: $c := \inf\{i \in \{1, \dots, n\} : x_i < c\}$

Oss. Se gli x_i sono indicati con la minuscola, stiamo parlando di quantità, valori numerici, se invece scriviamo X_i, indichiamo le *variabili aleatorie* rappresentanti i singoli *esemplari.* La stessa notazione è estesa alle *funzioni statistiche*, ad esempio s^2 sarà un numero, mentre S^2 una *variabile aleatoria* o il risultato di un *esperimento aleatorio.*

Esempio Una ditta produce bulloni con un diametro di $7mm$. Un bullone è accettabile se ha un diametro compreso tra $6.5mm$ e $7.5mm$. Preso un bullone ne misuriamo il diametro effettivo. Il nostro obiettivo è usare queste misurazioni per risalire a una funzione f_X che possa essere usata per descrivere il diametro di tutti i bulloni prodotti.

Ci sono molti modi in cui la funzione f_X può essere ignota. Noi ci limiteremo a considerare il caso in cui sia noto il tipo di distribuzione della *variabile aleatoria*, ma non ne conosciamo i parametri. Ad esempio, potremmo sapere che X è una *Poissoniana*, ma non conosciamo λ. Il nostro scopo è quindi usare i dati a nostra disposizione per ricavare λ, o equivalentemente, il *valore atteso.*

12.2 Stimatori e stime

Si dice *stimatore* di un parametro una *variabile aleatoria* che sia una *statistica* e il cui valore sia "spesso vicino" al parametro che ci interessa.

Il valore deterministico assunto dallo *stimatore* usando i dati ricavati prende il nome di *stima.*

NB. Si noti che lo *stimatore* è una funzione definita sul campione che ha come parametri i dati delle osservazioni, mentre la *stima* è un numero.

Esempio Sia $(X_1, \dots, X_n)$ un *vettore di variabili aleatorie* indipendenti e identicamente distribuite. Come parametro di interesse abbiamo $E[X_i] = \mu$.

Uno *stimatore* della *media* può essere la *media campionaria* $\hat{\mu} = \bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$. $\bar{X}$ è uno *stimatore* per μ, cioè $\bar{X} \approx \mu$, in quanto, per la *legge dei grandi numeri*, vale:

$$\frac{\sum_{i=1}^{n} X_i}{n} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i = \{\bar{X}\}_n \xrightarrow[n \to +\infty]{P} \mu$$

NB. Se Θ è uno *stimatore* di ϑ, possiamo scrivere Θ_n per evidenziare la numerosità del campione. In particolare, vale che $\Theta_n = \hat{\vartheta}(\{X_i\}_{i=1}^n)$.

Oss. Essendo Θ una *stima*, è ragionevole aspettarsi un qualche scostamento dal valore effettivo. Questo errore, detto *errore di stima*, è anch'esso una *variabile aleatoria*. Siccome nella definizione di *stimatore* abbiamo chiesto che il valore stimato fosse vicino al parametro interessato, ci aspettiamo che l'errore sia piccolo.

Definizione 12.2.1. Uno *stimatore* Θ di ϑ è:

- *Corretto, non distorto o unbiased* se $E[\Theta] = \vartheta$
- *Scorretto o biased* se $E[\Theta] \neq \vartheta$ e chiamiamo *bias* il valore $E[\Theta] - \vartheta$

In particolare, se $\lim_{n\to+\infty} E[\Theta_n] = \vartheta$, allora Θ è detto *asintoticamente non distorto.*

Definizione 12.2.2. Si dice *errore quadratico medio* di Θ, il valore:

$$MSE[\Theta] = E\left[(\Theta - \vartheta)^2\right]$$

Oss. L'*errore quadratico medio* può essere riscritto in una maniera un po' diversa:

$$MSE[\Theta] = E\left[(\Theta - \vartheta)^2\right] = E\left[(\Theta - E[\Theta] + E[\Theta] - \vartheta)^2\right]$$
$$= E\left[(\Theta - E[\Theta])^2\right] + 2E\left[(\Theta - E[\Theta]) \cdot (E[\Theta] - \vartheta)\right] + E[(E[\Theta] - \vartheta)^2]$$
$$= Var[\Theta] + 2E\left[(E[\Theta] - E[\Theta]) \cdot (\Theta - \vartheta)\right] + (bias)^2 = Var[\Theta] + (bias)^2$$

In particolare se Θ è *corretto*, il *bias* è nullo e quindi $MSE[\Theta] = Var[\Theta]$.

Definizione 12.2.3. Diciamo che uno *stimatore* Θ è *consistente*, se Θ_n *converge in probabilità* a ϑ, cioè se vale:

$$\Theta_n \xrightarrow[n\to+\infty]{P} \vartheta$$

Se, inoltre, Θ_n *converge in media quadratica* a ϑ, ovvero se vale:

$$\Theta_n \xrightarrow[n\to+\infty]{L^2} \vartheta$$

diciamo che Θ è *consistente in media quadratica.*

Prop. Se Θ è *asintoticamente non distorto* e $\lim_{n\to+\infty} Var[\Theta] = 0$, allora Θ è uno *stimatore consistente in media quadratica* e quindi è anche *consistente.*

Dim. Chiedere che Θ sia *consistente in media quadratica* significa chiedere che:

$$\lim_{n\to+\infty} E\left[(\Theta_n - \vartheta)^2\right] = 0$$

ossia che $\lim_{n\to+\infty} MSE[\Theta_n] = 0$, ma per quanto visto nell'ultima osservazione

$$MSE[\Theta_n] = Var[\Theta_n] + (E[\Theta_n] - \vartheta)^2$$

e la convergenza a 0 è assicurata dalle ipotesi applicate ai due addendi del secondo membro. La *consistenza* segue dal fatto che siccome la *convergenza in media quadratica* implica *convergenza in probabilità*, la *consistenza in media quadratica* non può che implicare *consistenza.*

Oss. Uno *stimatore* può essere *corretto*, ma non *consistente.* Ad esempio, se $(X_1, \ldots, X_n)$ è un *vettore di variabili aleatorie* indipendenti e identicamente distribuite, ogni X_i è uno *stimatore* della *media* μ, poiché $E[X_i] = \mu \ \forall i \in \{1, \ldots, n\}$. Se prendiamo $\sigma^2 = Var[X_i] \neq 0$, allora $P(|X_i - \mu| > \epsilon) \neq 0$ e rimane tale per $n \to +\infty$, quindi non vi è *convergenza in probabilità.*

12.2.1 Alcuni stimatori

Supponiamo di avere come *campione* $(X_1, \dots, X_n)$, cioè un vettore di *variabili aleatorie iid* di *media* $E[X_i] = \mu$ e *varianza* $Var[X_i] = \sigma^2$:

- La *media campionaria* $\hat{\mu} = \hat{\mu_n} = \bar{X} = \bar{X_n} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$ è uno *stimatore corretto e consistente* della *speranza*, ovvero $E[X_i] = \mu$. Valgono infatti le seguenti:

$$E[\hat{\mu}] = E\left[\frac{1}{n} \cdot \sum_{i=1}^n X_i\right] = \frac{1}{n} \cdot \sum_{i=1}^n E[X_i] = \frac{1}{n} \cdot \sum_{i=1^n} \mu = \frac{1}{n} \cdot n \cdot \mu = \mu$$

$$Var[\hat{\mu}] = Var\left[\frac{1}{n^2} \cdot \sum_{i=1}^n X_i\right] = \frac{1}{n} \cdot \sum_{i=1}^n Var[X_i] = \frac{1}{n^2} \cdot n \cdot \sigma^2 \xrightarrow[n \to +\infty]{} 0$$

- La *varianza campionaria a media nota* $S_\cdot^2 = S_{\cdot n}^2 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \mu)^2$ è uno *stimatore corretto e consistente*:

$$E[S_\cdot^2] = \frac{1}{n} \cdot \sum_{i=1}^n E[(X_i - \mu)^2] = \frac{1}{n} \cdot \sum_{i=1}^n Var[X_i] = \frac{1}{n} \cdot n \cdot \sigma^2 = \sigma^2 = Var[X_i]$$

$$S_\cdot^2 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \mu)^2 \xrightarrow[n \to +\infty]{P} E[(X_i - \mu)^2] = Var[X_i] = \sigma^2$$

  Ovviamente, se non conoscessimo la *media* μ non potremmo usare questo *stimatore*.

- La *varianza campionaria a media ignota* è uno *stimatore corretto e consistente*, ma a patto di usare qualche accorgimento. Se ci si trovasse a dover stimare la *varianza*, ma non si conoscesse μ, la prima tentazione sarebbe quella di usare $\hat{\mu}$, ovvero:

$$\frac{1}{n} \cdot \sum_{i=1}^n (X_i - \hat{\mu})^2 = \frac{1}{n} \cdot \sum (X_i^2 - 2\hat{\mu}X_i + \hat{\mu}^2) = \frac{1}{n} \cdot \sum_{i=1}^n X_i^2 - 2\frac{1}{n}\hat{\mu}^2 \cdot \sum_{i=1}^n X_i + \frac{1}{n} \cdot \sum_{i=1}^n \hat{\mu}^2$$

$$= \frac{1}{n} \cdot \sum_{i=1}^n X_i^2 - 2\hat{\mu}^2 + \frac{1}{n} n\hat{\mu}^2 = \frac{1}{n} \cdot \sum_{i=1}^n X_i^2 - 2\hat{\mu}^2 + \hat{\mu}^2 = \frac{1}{n} \cdot \sum_{i=1}^n X_i^2 - \hat{\mu}^2$$

  Se però ora ne calcoliamo il *valore attesto* per verificarne la *consistenza*, otteniamo:

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

$$= \frac{1}{n} \cdot \sum_{i=1}^n E[X_i^2 - \mu^2] - E[\hat{\mu}^2 - \mu^2] = \frac{1}{n} \cdot Var[X_i] - Var[\hat{\mu}] = \frac{1}{n} \cdot n \cdot Var[X_i] - \frac{1}{n} \cdot \sigma^2$$

$$= \sigma^2 - \frac{1}{n} \cdot \sigma^2 = \frac{n-1}{n} \cdot \sigma^2$$

  Il risultato ottenuto ci dimostra che lo *stimatore* è *corretto* per la *legge dei grandi numeri*, ma non è *consistente*, in quanto, se lo fosse stato, avremmo ottenuto come risultato 0. Tuttavia, è facile correggere questo *stimatore* e definirne uno *consistente*:

$$S^2 = S_n^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \hat{\mu})^2$$

Lo *stimatore* S^2 così definito è *corretto e consistente*:

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

Utilizzando la *legge dei grandi numeri*, non è difficile dimostrarne la *consistenza*.

Oss. Lo *stimatore* S^2 può essere espresso anche come:

$$S^2 = \frac{1}{n-1} \cdot \left(\sum_{i=1}^{n} X_i^2 - n \cdot \hat{\mu}^2 \right)$$

Tuttavia, questa forma è computazionalmente più onerosa e instabile.

12.2.2 Distribuzione degli stimatori

Supponiamo di avere una *popolazione* distribuita come una *Gaussiana* di parametri ignori μ e σ, cioè ogni esemplare X_i del *campione* ha legge $\mathcal{N}(\mu, \sigma)$. Come possiamo stimarne i parametri?

Abbiamo già dimostrato che $\bar{X}$ e S^2 sono *stimatori corretti e consistenti* della *speranza* e della *varianza a media ignota*.