

Appunti di Calcolo delle Probabilità e Statistica

Lezioni della prima parte di Marco Ghimenti
Lezioni della seconda parte di Maaurizio Pratelli
A cura di Alessandro Cheli

A.A 2019-2020

Indice

I	Probabilità Discreta e Catene di Markov	
	Lezioni di Marco Ghimenti	1
1	Probabilità Discreta e Condizionata	3
1.1	Probabilità Discreta e Formule Combinatorie	3
1.1.1	Principio di induzione	4
1.1.2	Permutazioni di n elementi	4
1.1.3	Coefficiente Binomiale	4
1.1.4	Disposizioni	5
1.2	Probabilità Condizionata	5
1.3	Esercizi	6
2	Spazio Probabilizzato	9
2.1	Lo Spazio Probabilizzato	9
2.2	Formula di fattorizzazione	12
2.3	Formula di Bayes	12
2.4	Esercizi	13
3	Variabili Aleatorie	15
3.1	Variabili Aleatorie Discrete	15
3.2	Variabili Aleatorie notevoli	16
3.3	Valore Atteso	20
3.4	Esercizi	25
3.5	Esercitazione del 29/10/19	27
4	Catene di Markov	31
4.1	Catene di Markov e Processi Stocastici	31
4.2	Calcolo Algebrico su catene di Markov	33
4.3	Esercizi	37
II	Variabili Aleatorie con Densità e Introduzione alla Statistica	
	Lezioni di Maurizio Pratelli	43
5	Variabili Aleatorie con densità (continue)	45
5.1	Svolgere integrali multipli	45
5.2	Introduzione alle Variabili Aleatorie con Densità	47
5.3	Proprietà principali delle variabili aleatorie con densità	48
5.4	Principali tipologie di densità	52
5.5	Teoremi e leggi	62
5.6	Accenni di Inferenza Statistica	64

5.7	Esercizi	64
-----	--------------------	----

Parte I

Probabilità Discreta e Catene di Markov Lezioni di Marco Ghimenti

Capitolo 1

Probabilità Discreta e Condizionata

1.1 Probabilità Discreta e Formule Combinatorie

Definizione 1.1.1. Probabilità: **Attendibilità confortata da motivi ragionevoli**

La probabilità (discreta) di un evento si può definire, in maniera intuitiva,

Definizione 1.1.2.

$$P(\text{evento}) = \frac{\# \text{ casi favorevoli}}{\# \text{ casi possibili}} \quad (1.1)$$

Esempio 1.1.1. Prendiamo ad esempio il lancio di un dado, voglio ottenere un numero ≥ 5 , la probabilità dell'evento è

$$P = 2/6 = 1/3$$

Un altro esempio può essere la probabilità, lanciando 2 dadi, che almeno uno dei due renda un numero ≥ 4 .

$$P = \frac{27}{6^2} = \frac{3}{4}$$

I casi favorevoli sono 27 perché lanciando se lanciando il primo dado ottenendo un numero ≤ 3 significa che ho 3 possibili casi per ognuno dei lanci del primo dado per ottenere un numero ≥ 4 dal lancio del secondo dado ($3 \cdot 3$), a cui si aggiungono ($3 \cdot 6$) casi se ottengo un numero ≥ 4 dal primo lancio (tutti i casi del secondo lancio sono validi.)

Esempio 1.1.2. Qual è la probabilità di ottenere almeno un asso pescando 2 carte da un mazzo di 54?

$$P = \frac{(50 \cdot 4) + (53 \cdot 4)}{54 \cdot 53} = \frac{206}{1431}$$

Per i casi possibili, ho 54 casi per la prima pescata e 53 per la seconda, per i casi favorevoli ho

$$\begin{cases} \text{Se pesco un Asso alla prima e una carta qualsiasi alla seconda} \implies 4 \cdot 53 \\ \text{Se non pesco un Asso alla prima e un Asso alla seconda} \implies 50 \cdot 4 \end{cases}$$

Esercizio 1.1.3. Calcolare la probabilità di pescare esattamente 2 assi pescando 5 carte

Esercizio 1.1.4. Calcolare la probabilità di pescare esattamente 2 due donne pescando 5 carte sapendo che la prima carta uscita è una figura

Svolgere questi due esercizi contando i casi è molto macchinoso. Serve introdurre un po' di calcolo combinatorio e, per il secondo esercizio, il concetto di probabilità condizionata.

1.1.1 Principio di induzione

Vogliamo dimostrare una proposizione che dipende da un indice $n \in \mathbb{N}$. Una possibile strategia dimostrativa è il *principio di induzione*. Se riusciamo a dimostrare che

1. La proposizione è verificata per un certo indice n_0
2. Se assumiamo per verificata la proposizione per un generico indice n , allora riusciamo a dimostrare la proposizione per l'indice $n + 1$

allora la proposizione è verificata per ogni $n \geq n_0$.

1.1.2 Permutazioni di n elementi

Definizione 1.1.3. Una permutazione è uno scambio dell'ordine di una sequenza di elementi che possono essere di qualunque tipo. L'obiettivo è trovare il numero di tutte le permutazioni (cioè tutte le sequenze con ordine) possibili dato un certo numero n di elementi.

Proposizione 1.1.1. *Le permutazioni di un insieme di n elementi sono definite come*

$$\text{Perm}(n) = n! \quad (1.2)$$

Dimostrazione. Dimostriamo per induzione. Come passo base possiamo verificare immediatamente che $\text{Perm}(1) = 1$.

Il passo induttivo sarà

$$\begin{aligned} \text{Perm}(n) = n! &\implies \text{Perm}(n+1) = (n+1)! \\ \text{Perm}(n+1) &= (n+1) \cdot \text{Perm}(n) \\ &= (n+1) \cdot n! = (n+1)! \end{aligned}$$

□

1.1.3 Coefficiente Binomiale

Definizione 1.1.4. Il coefficiente binomiale è un numero intero non negativo definito dalla formula

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \quad (1.3)$$

Proposizione 1.1.2. *Indicato con $S_{n,k}$ il numero di modi possibili di scegliere k oggetti da un insieme di n elementi vale $S_{n,k} = \binom{n}{k}$*

Dimostrazione. Fissato $k \geq 2$ dimostriamo per induzione su $n \geq k$

Il primo passo iniziale è, per $n = k$

$$S_{k,k} = 1 = \binom{k}{k} = \frac{k!}{k!(k-k)!}$$

(c'è un solo modo di prendere k elementi da un insieme di k oggetti: prenderli tutti)

Per questa dimostrazione serve anche un ulteriore passo iniziale: dobbiamo vedere in quanti modi si possono scegliere $k-1$ elementi da un insieme di k oggetti. Abbiamo

$$S_{k,k-1} = k = \binom{k}{k-1} = \frac{k!}{(k-1)!(k-k+1)!}$$

(infatti dobbiamo semplicemente scegliere quale elemento non prendere, e quindi abbiamo k scelte possibili)

Passo induttivo: consideriamo che $S_{n+1,k} = S_{n,k} + S_{n,k-1}$ infatti posso scegliere i k elementi dai primi n , e scartare l'ultimo, oppure sceglierne $k-1$ dai primi n e prendere l'ultimo. Questa formula spiega perché abbiamo bisogno di due passi iniziali.

$$\begin{aligned}
 S_{n+1,k} &= S_{n,k} + S_{n,k-1} = \binom{n}{k} + \binom{n}{k-1} \\
 &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!} \\
 &= \frac{n!((n-k+1) + k)}{k!(n-k+1)!} = \frac{n!(n+1)}{k!(n+1-k)!} = \frac{(n+1)!}{k!(n+1-k)!} \\
 &= \binom{n+1}{k}
 \end{aligned}$$

□

1.1.4 Disposizioni

Definizione 1.1.5. Una disposizione $D_{n,k}$ significa il numero di modi per "prendere" k oggetti ordinati da un insieme di n elementi.

Ovviamente avremo

$$D_{n,k} = S_{n,k} \cdot \text{Perm}(k) = \frac{n!}{(n-k)!k!} \cdot k! = \frac{n!}{(n-k)!} \quad (1.4)$$

1.2 Probabilità Condizionata

Esempio 1.2.1. Lancio due dadi sommando il risultato, qual è $P(\geq 10)$ sapendo che il primo ha fatto almeno 3?

Sappiamo che $P(\text{Somma} \geq 10) = 6/36 = 1/6$

Poniamo il vincolo che il lancio del primo dado risulti almeno ≥ 3 . Allora se vediamo i possibili risultati vediamo che i casi favorevoli sono 6 e quelli possibili sono 24, quindi

$$P(\text{Somma} \geq 10 \mid \text{Primo dado} \geq 3) = 6/24 = 1/4$$

Definizione 1.2.1. Ponendo Ω = gli eventi possibili; La **probabilità condizionata** che succeda A sapendo B si indica con:

$$\begin{aligned}
 P(A|B) &= \frac{\text{casi favorevoli}}{\text{casi possibili}} = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|}{|\Omega|} \cdot \frac{|\Omega|}{|B|} \\
 P(A|B) &= \frac{P(A \cap B)}{P(B)}
 \end{aligned} \quad (1.5)$$

Esempio 1.2.2. Nel lancio di un dado, la probabilità di ottenere ≤ 4 sapendo che è uscito un numero pari è

$$\begin{aligned}
P(\leq 4|\text{pari}) &= \frac{P(\leq 4|\text{pari})}{P(\text{pari})} \\
P(\text{pari}) &= 3/6 = 1/2 \\
P(\leq 4 \cap \text{pari}) &= 2/6 \\
\implies P(\leq 4|\text{pari}) &= \frac{2/6}{1/2} = 2/3
\end{aligned}$$

Definizione 1.2.2. Definiamo il **complementare** di un evento, ovvero $A^C = \Omega \setminus A$. La probabilità di un complementare è $P(A^C) = 1 - P(A) = P(\Omega) - P(A)$

In generale, dati due eventi A, B con $A \cap B \neq \emptyset$ si ha che la probabilità dell'unione è $P(A \cup B) = P(A) + P(B)$.

Definizione 1.2.3. Due eventi A e B si dicono **indipendenti** se $P(A | B) = P(A)$.

Dalla definizione di probabilità condizionata si ottiene che se A e B sono indipendenti vale $P(A \cap B) = P(A) \cdot P(B)$

1.3 Esercizi

Esercizio 1.3.1. Terno al lotto: Giocando 5 numeri al lotto (estrazione da 1 a 90) calcolare la probabilità di ottenere un terno esatto e più di un terno.

Se vogliamo ottenere un terno esatto i casi possibili sono $\binom{90}{5}$ (I modi di estrarre 5 palline dall'urna). I casi favorevoli saranno $S_{5,3} \cdot S_{85,2} = \binom{5}{3} \cdot \binom{85}{2}$ (ovvero i modi in cui si possono scegliere 3 numeri tra i 5 estratti e 2 numeri dagli altri).

La probabilità di ottenere un terno esatto sarà quindi

$$P(\text{terno esatto}) = \frac{\binom{5}{3} \cdot \binom{85}{2}}{\binom{90}{5}} \simeq \frac{1}{1230}$$

Per ottenere almeno un terno i casi favorevoli sono

- terno: $\binom{5}{3} \binom{85}{2}$
- quaterna: $\binom{5}{4} \binom{85}{1}$
- cinquina: 1

La probabilità di ottenere almeno un terno sarà data dalla somma delle probabilità corrispondenti a terno, quaterna e cinquina:

$$P(\text{almeno un terno}) = \frac{\binom{85}{2} \binom{5}{3}}{\binom{90}{5}} + \frac{\binom{85}{1} \binom{5}{4}}{\binom{90}{5}} + 1$$

Esercizio 1.3.2. Probabilità del gioco di Monty Hall Nel gioco televisivo di Monty Hall il partecipante deve scegliere una fra tre porte, una di esse contiene un premio mentre le altre due contengono rispettivamente due capre. Dopo la scelta del giocatore iniziale il presentatore apre una delle due porte contenenti una capra. Al giocatore conviene cambiare porta o mantenere quella scelta in origine?

Ipotesi Se scelgo una porta e la mantengo vinco solo se il premio era nella porta che ho scelto
 $\implies P = 1/3$

Ipotesi Se scelgo una porta e la cambio avrò $P = 2/3$

1	2	3
x	x	\$
x	\$	x
\$	x	x

Tabella 1.1: Gioco di Monty Hall

Esercizio 1.3.3. Dado rosso e dado nero

Tiriamo due dadi, uno rosso ed uno nero. Calcolare la probabilità che il dado rosso risulti 3 sapendo che sul dado nero è uscito 2:

$$P(R = 3 \mid N = 2) = \frac{P(R = 3 \cap N = 2)}{P(N = 2)} = \frac{1/36}{6/36} = \frac{1}{6}$$

$$P(R = 3) = 1/6$$

Ne otteniamo che $P(A|B) = P(A) \implies A, B$ sono indipendenti.

Capitolo 2

Spazio Probabilizzato

2.1 Lo Spazio Probabilizzato

Definizione 2.1.1. L'insieme delle parti di A , indicato con $\mathcal{P}(A)$, è dato da tutti i sottoinsiemi che posso costruire a partire dagli elementi di A . Ad esempio:

$$A = \{0, 1\}$$
$$\mathcal{P}(A) = \{\emptyset, \{0, 1\}, \{0\}, \{1\}\}$$

Definizione 2.1.2. Un insieme $F \subseteq \mathcal{P}(A)$ chiuso rispetto a intersezione, unione e complementare, ovvero

$$A, B \in F \implies \begin{cases} A \cup B \\ A \cap B \\ A^C, B^C \end{cases} \in F$$

Si chiama *algebra*.

Se è chiuso rispetto all'unione numerabile di insiemi, ovvero se $A_1, \dots, A_n \subset F$ allora $\bigcup_{n \in \mathbb{N}} A_n \in F$, allora F si dice σ -Algebra (o tribù).

Uno **spazio probabilizzato** è un ente matematico che serve per introdurre una definizione più rigorosa e più flessibile di probabilità rispetto a quella usata fino ad adesso.

Definizione 2.1.3. Uno spazio probabilizzato è definito come una terna:

$$(\Omega, F, P) \tag{2.1}$$

Dove Ω è l'insieme degli eventi elementari, ovvero tutti i risultati possibili, ad esempio in un lancio di un dado $\Omega = \{1, 2, 3, 4, 5, 6\}$, F è una σ -algebra contenuta nelle parti di Ω e la probabilità P è una funzione definita come

$$P : F \rightarrow [0, 1]$$
$$P(\Omega) = 1 \tag{2.2}$$
$$\forall i \neq j . A_i \cap A_j = \emptyset \implies P\left(\bigcup_n A_n\right) = \sum_n P(A_n)$$

Spesso, se Ω è un insieme finito, allora si prende $F = \mathcal{P}(\Omega)$ e

$$P(A \subset F) = \frac{\#A}{\#\Omega}$$

ritrovando il concetto intuitivo del primo capitolo, ma definito in maniera rigorosa.

Proposizione 2.1.1. *Dagli assiomi di probabilità dati sopra si possono dimostrare le seguenti proprietà:*

$$\begin{aligned} P(A^C) &= 1 - P(A) \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ A \subseteq B &\implies P(A) \leq P(B) \end{aligned} \tag{2.3}$$

Esempio 2.1.1. Lotteria di De' Finetti: Si assiste all'estrazione di un numero $n \in \mathbb{N}$ casuale. Supponiamo che ogni numero abbia la stessa probabilità di essere estratto.

Dato un altro naturale $m \in \mathbb{N}$

$$p_n = P(n) = \text{La probabilità di estrarre il numero } n$$

Vogliamo che $0 \leq p_n \leq 1$ e anche $p_m = p_n \forall n = m$. Quindi $1 = P(\Omega) = P\left(\bigcup_{n \in \mathbb{N}} n\right) = \sum_{n \in \mathbb{N}} p_n$.
Assumendo $p_n = 0 \forall n$ allora $\sum_n p_n = 0$. Se $p_n = c > 0, \forall n$ allora $\sum_n p_n = \sum_n c = +\infty$. In nessuno dei due casi è possibile che $P(\Omega) = \sum_{n \in \mathbb{N}} p_n = 1$, quindi la lotteria di De Finetti non si può modellizzare con l'ipotesi che ogni numero sia equiprobabile.

Si noti che nello stesso esempio, se si prende come $F = \{\emptyset, \{\text{numeri pari}\}, \{\text{numeri dispari}\}, \mathbb{N}\}$, supponendo sempre che ogni numero abbia la stessa probabilità di venire estratto, si può dare una buona definizione di probabilità almeno all'estrazione di un numero pari o dispari:

$$\begin{aligned} P(\emptyset) &= 0 \\ P(\{\text{numeri pari}\}) &= 1/2 \\ P(\{\text{numeri dispari}\}) &= 1/2 \\ P(\mathbb{N}) &= 1/2 \end{aligned}$$

Questo è un esempio in cui è utile prendere una σ -algebra che non coincida con le parti di A

Definizione 2.1.4. Densità di Probabilità discreta: Una successione $\{p_n\}_{n \in \mathbb{N}}$ con $p_n \in \mathbb{R}$ con $p_n \geq 0$ e $\sum_{n \in \mathbb{N}} p_n = 1$ è detta densità di probabilità discreta.

Se $\{p_n\}$ è una densità di probabilità discreta, allora su $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ si può definire la probabilità $P(n) = p_n$. In tal modo $(\mathbb{N}, \mathcal{P}(\mathbb{N}), P)$ diventa uno spazio probabilizzato.

Definizione 2.1.5. Ultimo assioma di probabilità: Se $P(B) \neq 0$, si definisce la probabilità condizionata come

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

e si dice che A e B sono indipendenti se $P(A | B) = P(A)$.

Questo ultimo assioma rende rigoroso quello che avevamo visto con la probabilità intuitiva nel capitolo precedente. La definizione di indipendenza a prima vista sembra strana: da una parte il ruolo di A e B sembra simmetrico (si dicono indipendenti entrambi), ma nella formula il ruolo di A e B non sembra intercambiabile. In effetti anche nella formula precedente i ruoli si possono scambiare, come dimostra la proposizione seguente.

Proposizione 2.1.2. *Se $P(A | B) = P(A)$ allora $P(B | A) = P(B)$*

Dimostrazione. Supponiamo che entrambi gli eventi non siano impossibili. Si ha

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

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

Allora

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

da cui segue

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

□

Si può usare la definizione equivalente di eventi indipendenti

Definizione 2.1.6. A e B sono indipendenti se $A \cap B = P(A) \cdot P(B)$

Quest'ultima definizione si generalizza bene al caso di molti eventi indipendenti.

Definizione 2.1.7. Gli eventi A_1, \dots, A_n si dicono indipendenti se per ogni scelta di indici i_1, \dots, i_k vale

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_k})$$

Esempio 2.1.2. A, B indipendenti $\implies \{(A, B^C), (A^C, B^C), (A^C, B)\}$ indipendenti

Dimostrazione. Dimostriamo solo che A e B^C sono indipendenti, le altre sono analoghe

$$A = (A \cap B) \cup (A \cap B^C) \implies$$

$$P(A) = P(A \cap B) + P(A \cap B^C) = P(A) \cdot P(B) + P(A \cap B^C) \implies$$

$$P(A \cap B^C) = P(A) - P(A) \cdot P(B) = P(A) [1 - P(B)] = P(A) \cdot P(B^C)$$

□

Esercizio 2.1.3. Lancio due dadi, uno rosso ed uno nero. Definiamo lo spazio probabilizzato con $\Omega = \{(r, n), \text{ dove } r = 1, 2, 3, 4, 5, 6; n = 1, 2, 3, 4, 5, 6\}$, $F = \mathcal{P}(\Omega)$, e la probabilità intuitiva, es $P(n = 1) = \frac{1}{36}$.

Calcoliamo la probabilità che il rosso sia 3 sapendo che rosso + nero fa 6

$$P(r = 3 | r + n = 6) = \frac{P(r = 3 \cap r + n = 6)}{P(r + n = 6)} = \frac{\frac{1}{36}}{\frac{5}{36}} = \frac{1}{5}$$

Probabilità che il rosso sia pari sapendo che rosso + nero fa 6

$$P(r = \text{pari} | r + n = 6) = \frac{P(r \text{ pari} \cap r + n = 6)}{P(r + n = 6)} = \frac{\frac{2}{36}}{\frac{5}{36}} = \frac{2}{5}$$

Esercizio 2.1.4. Gioco di Monty Hall: Riprendendo il gioco delle tre porte definiamo lo spazio probabilizzato: Formalizzo di aver scelto la porta 3. $\Omega = (x, y)$ dove $x = 1, 2, 3$ è la porta vincente e $y = 1, 2$ è la porta perdente che è stata aperta dal presentatore. Gli eventi impossibili saranno $P(1, 1) = 0, P(2, 2) = 0$

$$\begin{aligned}
P(x=1) &= P(x=2) = P(x=3) = \frac{1}{3} \\
P(x=1, y=2) &= \frac{1}{3} \\
P(x=2, y=1) &= \frac{1}{3} \\
P(x=3, y=1) &= P(x=3, y=2) = \frac{1}{6}
\end{aligned}$$

Quindi $P(y=1) = P(y=2) = \frac{1}{2}$

Se scelgo la porta 3, suppongo venga aperta la 2. Se non cambio e vinco ($x=3$) allora

$$P(x=3 \mid y=2) = \frac{P((3,2))}{P(y=2)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Se scelgo la porta 3, suppongo venga aperta la 1 e cambio allora:

$$P(x=1 \mid y=2) = \frac{P((1,2))}{P(y=2)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{2}{3} \quad (2.4)$$

2.2 Formula di fattorizzazione

Definizione 2.2.1. Dato uno spazio probabilizzabile (Ω, F) , una famiglia di insiemi $B_1, \dots, B_n \in F$, con $n \in \mathbb{N}$ è detta una partizione finita di Ω se se $\forall i, j$ con $j \neq i$ allora $B_i \cap B_j = \emptyset$ e se $\bigcup_{i=1}^n B_i = \Omega$.

Lemma 2.2.1. Sia $\{B_i\}_{i=1, \dots, n}$ partizione finita di Ω e sia $P(B_i) > 0 \forall i$. Allora si avrà

$$P(A) = \sum_{i=1}^n P(A \mid B_i) \cdot P(B_i) \quad (2.5)$$

Dimostrazione.

$$P(A) = \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(A \mid B_i) \cdot P(B_i)$$

□

Definizione 2.2.2. Condizionamento Ripetuto: Dati A_1, \dots, A_n eventi, allora

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2) \cdot \dots \cdot P(A_n \mid A_1 \cap \dots \cap A_{n-1})$$

2.3 Formula di Bayes

Lemma 2.3.1. Dati due eventi A, B con probabilità non nulla $P(A) > 0$ e $P(B) > 0$ allora

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \quad (2.6)$$

Dimostrazione.

$$\begin{aligned} P(A | B) &= \frac{P(A \cap B)}{P(B)} \\ P(B | A) &= \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} \\ \Rightarrow P(A \cap B) &= P(A | B) \cdot P(B) = P(B | A) \cdot P(A), \\ \Rightarrow P(A | B) &= \frac{P(B | A) \cdot P(A)}{P(B)}. \end{aligned}$$

□

2.4 Esercizi

Esercizio 2.4.1. Un produttore di vino produce due vini (bianco B e rosso R) e vende in Francia (F) e Germania (G). Le vendite totali sono $1/3$ per la Francia e $2/3$ per la Germania. Le richieste della Francia sono per $3/4$ vino bianco e per $1/4$ di vino rosso. Le richieste della Germania si dividono equamente tra vino bianco e vino rosso. Utilizzando la formula di partizione troviamo la probabilità che una richiesta (senza sapere da chi viene fatta) sia vino bianco.

$$P(B) = P(B | G) \cdot P(G) + P(B | F) \cdot P(F) = \frac{1}{2} \cdot \frac{2}{3} + \frac{3}{4} \cdot \frac{1}{3} = \frac{7}{12}$$

Esercizio 2.4.2. Abbiamo 3 livelli di preparazione di degli studenti iscritti ad un esame: Ottimo, Buono e Scarso. Un esito dell'esame è Promosso o Respinto.

$$\begin{aligned} P(\text{Promosso} | \text{Ottimo}) &= 0.995 \\ P(\text{Promosso} | \text{Scarso}) &= 0.3 \\ P(\text{Promosso} | \text{Buono}) &= 0.8 \end{aligned}$$

Uno studente prova l'esame e viene respinto. Qual è la probabilità che avesse una preparazione scarsa (ovvero $P(\text{Scarso} | \text{Respinto})$)? Prima calcoliamo la probabilità di essere respinti.

$$P(R) = P(R | O) \cdot P(O) + P(R | B) \cdot P(B) + P(R | S) \cdot P(S) = 0.302$$

Senza informazioni aggiuntive $P(O) = P(B) = P(S) = \frac{1}{3}$. La probabilità di essere respinto è $P(R) = 0.302$ quindi

$$P(S | R) = \frac{P(R | S) \cdot P(S)}{P(R)} = \frac{0.7 \cdot 1/3}{0.302} = 0.773$$

Esercizio 2.4.3. Qual è la probabilità che lo studente aveva di avere una preparazione scarsa, sapendo che è stato respinto e sapendo che le probabilità dei voti sono:

$$P(O) = \frac{1}{6}, P(B) = \frac{2}{3}, P(S) = \frac{1}{6}$$

Calcoliamo, come prima $P(R) = 0.005 \cdot \frac{1}{6} + 0.2 \cdot \frac{2}{3} + 0.7 \cdot \frac{1}{6} \approx 0.25$. Abbiamo quindi che

$$P(S | R) = \frac{0.7 \cdot 1/6}{0.25} \approx 0.466$$

Esercizio 2.4.4. La probabilità di ammalarsi di un soggetto a rischio (R) è 0.2, mentre la probabilità di ammalarsi di un soggetto non a rischio (N) è 0.006. Il 15% della popolazione è di soggetti a rischio. Un malato si denota con M mentre uno sano con S . Vogliamo sapere

1. $P(\text{Soggetto casuale sia malato}) =$

$$P(M) = P(M | R) \cdot P(R) + P(R | N) \cdot P(N) = 0.35$$

$$P(M) = 0.2 \cdot 0.15 + 0.006 \cdot 0.85 = 0.35$$

2. $P(\text{Soggetto malato fosse a rischio}) =$

$$P(R | M) = \frac{P(M | R) \cdot P(R)}{P(M)} = \frac{0.2 \cdot 0.15}{0.35} = 0.855$$

3. $P(\text{Soggetto soggetto sano sia a rischio}) =$

$$P(R | S) = \frac{P(S | R) \cdot P(R)}{P(S)} = \frac{(1 - 0.2) \cdot 0.15}{(1 - 0.35)} = 0.124$$

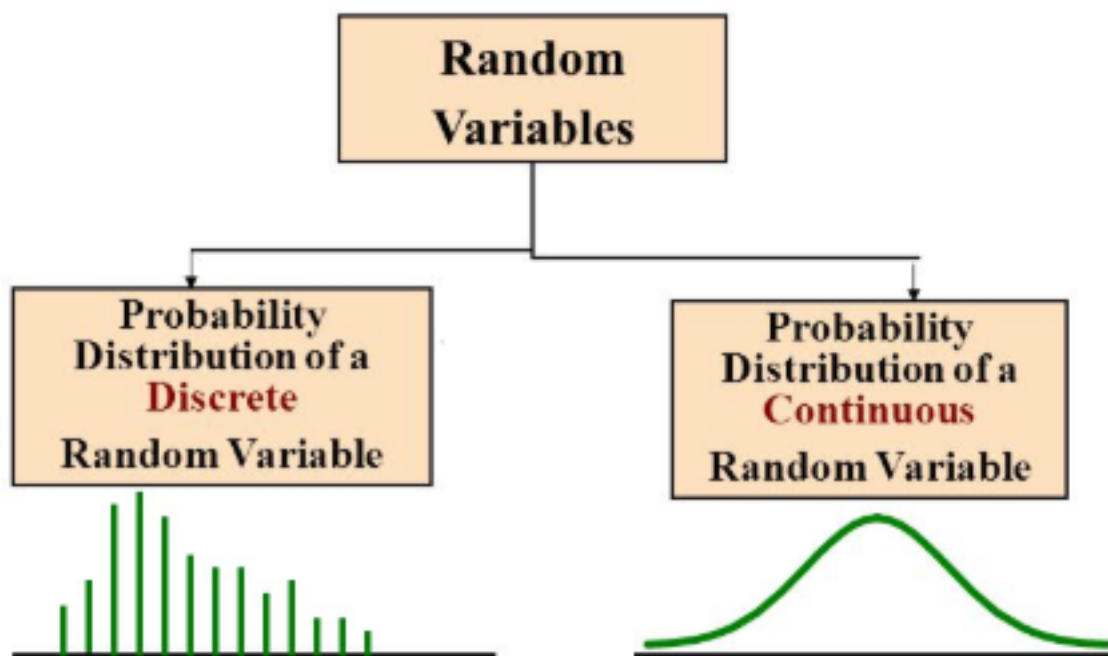
Nota. La probabilità che l'evento A^c (A complementare) si verifichi sapendo B è $P(A^c | B) = 1 - P(A | B)$ mentre la probabilità di A sapendo B^c è $P(A | B^c) \neq 1 - P(A | B)$.

Definizione 2.4.1. Prendendo S = soggetti sani, M = soggetti malati, T^- = test negativo, T^+ = test positivo. La **specificità** di un test è $P(T^- | S)$. Una specificità alta implica pochi falsi positivi. La **sensibilità** è $P(T^+ | M)$. Una sensibilità alta implica pochi falsi negativi.

Capitolo 3

Variabili Aleatorie

Figura 3.1: Tipi di variabili casuali



3.1 Variabili Aleatorie Discrete

Una variabile aleatoria è una funzione che può assumere diversi valori in dipendenza da qualche fenomeno casuale. Il risultato del lancio di un dado, o la vincita legata a tale risultato, ad esempio, sono variabili aleatorie.

Definizione 3.1.1. Prendiamo uno spazio probabilizzabile (Ω, F) . Una variabile aleatoria discreta è una funzione $X : \Omega \rightarrow \mathbb{R}$, che assume valori in un sottoinsieme finito o numerabile $\{a_1, \dots, a_k, \dots\} \subset \mathbb{R}$ e tale che $\forall j$ la controimmagine di a_j sia un elemento della σ -algebra, ovvero che

$$X^{-1}(a_j) = \{\omega \in \Omega, X(\omega) = a_j\} \in F.$$

Definizione 3.1.2. Data la probabilità di tutti gli eventi posso definire la densità di probabilità associata ad X : nello spazio probabilizzato (Ω, F, P) la probabilità che la variabile aleatoria assuma il valore a_j sarà $p_j = P(X = a_j) = P(X^{-1}(a_j))$. La successione $\{p_j\}$ viene detta *densità di probabilità di X* .

Dato che X è una funzione valgono le seguenti proprietà

1. $\forall j$ gli insiemi $X^{-1}(a_j)$ sono tutti disgiunti.
2. La loro unione copre Ω

Proposizione 3.1.1. *La successione $\{p_j\}$ sopra definita è effettivamente una densità di probabilità*

Dimostrazione. Ovviamente abbiamo $p_j \geq 0$ per ogni j . Inoltre vale $\sum_{j=1}^k p_j = 1$, infatti

$$1 = P(\Omega) = P\left(\bigcup_j X^{-1}(a_j)\right) = \sum_j P(X^{-1}(a_j))$$

□

Sia $X : \Omega \rightarrow \{a_1, \dots, a_k, \dots\}$ una variabile aleatoria definita su uno spazio probabilizzabile (Ω, F) , sia $\{p_j\}_j$ la densità di probabilità di X . Allora si può dotare (Ω, F) della probabilità indotta da X definita come $P(X = a_j) = p_j$.

Esempio 3.1.1. Preso uno spazio probabilizzabile (Ω, F) e una variabile aleatoria $X : \Omega \rightarrow \{a_1, \dots, a_k\}$, supponendo che i numeri a_j siano ordinati, sia p_j la densità di probabilità, come posso ricostruire, ad esempio $P(X \leq a_3)$?

$$P(X \leq a_3) = P(X = a_1) + P(X = a_2) + P(X = a_3) = p_1 + p_2 + p_3$$

Esempio 3.1.2. Voglio contare quanti 6 escono in 10 lanci di dadi.

Sia $\Omega = \{(1, 2, 3, 4, 5, 6)\}^{10}$, ovvero tutte le possibili parole di 10 elementi composte dai numeri da 1 a 6. Ad ogni lancio, ho $\frac{1}{6}$ di probabilità di ottenere 6 e $\frac{5}{6}$ di ottenere gli altri numeri. Definiamo la variabile aleatoria $X : \Omega \rightarrow \{0, 1, 2, \dots, 9, 10\}$ come il conteggio dei risultati dei lanci dove ottengo 6. Qual è la probabilità di ottenere 3 lanci dove ho fatto 6? Deve uscire 3 volte il numero 6, con probabilità $1/6$, e 7 volte un numero diverso da 6, con probabilità $5/6$. Inoltre non conta l'ordine, quindi devo moltiplicare per i modi in cui si prendono 3 oggetti (i tre dadi che hanno fatto 6) da un insieme di 10 elementi (tutti i lanci). Abbiamo quindi

$$P(X = 3) = \binom{10}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^7$$

3.2 Variabili Aleatorie notevoli

Definizione 3.2.1. Legge di Bernoulli: Faccio un esperimento, il risultato positivo ha probabilità p , mentre il risultato negativo ha probabilità $1 - p$. Definisco $\Omega = \{\text{successo}, \text{insuccesso}\}$. Una variabile aleatoria Bernoulliana è definita come

$$\begin{aligned} X : \Omega &\rightarrow \{0, 1\} \\ X(\text{successo}) &= 1 \\ X(\text{insuccesso}) &= 0. \end{aligned}$$

Ovviamente avremo la densità di X data da $p_0 = p$, $p_1 = 1 - p$.

Definizione 3.2.2. Legge Binomiale Sia k il conteggio dei successi di n esperimenti ripetuti, tutti indipendenti. Abbiamo quindi che $B(n, p) : \Omega = \{\text{successo}, \text{insuccesso}\}^n \rightarrow \{0, \dots, n\}$. Abbiamo che la densità di probabilità Binomiale $p_k = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ (abbiamo k successi con prob. p , $n - k$ insuccessi di prob. $1 - p$ e non conta l'ordine).

Siamo sicuri che p_k sia una densità? Sappiamo che $p_k \geq 0$. Vediamo che $1 = \sum_k p_k$, infatti, ricordando che $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$ (binomio di Newton), abbiamo che

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

Esempio 3.2.1. Somma di Variabili Aleatorie Lancio due dadi, uno rosso e uno nero, avremo quindi $\Omega = \{(R, N) : R, N = 1, \dots, 6\} = \{(1, \dots, 6)\}^2$. Definisco due variabili aleatorie, X per il dado rosso dove $X : (R, N) \rightarrow R$ e la variabile $Y : (R, N) \rightarrow N$. La densità per X sarà $p_j = \frac{1}{6} \forall j$ mentre la densità per Y sarà $q_j = \frac{1}{6} \forall j$. Definiamo $Z = X + Y$ conta la somma dei dadi.

Calcolare la densità di Z

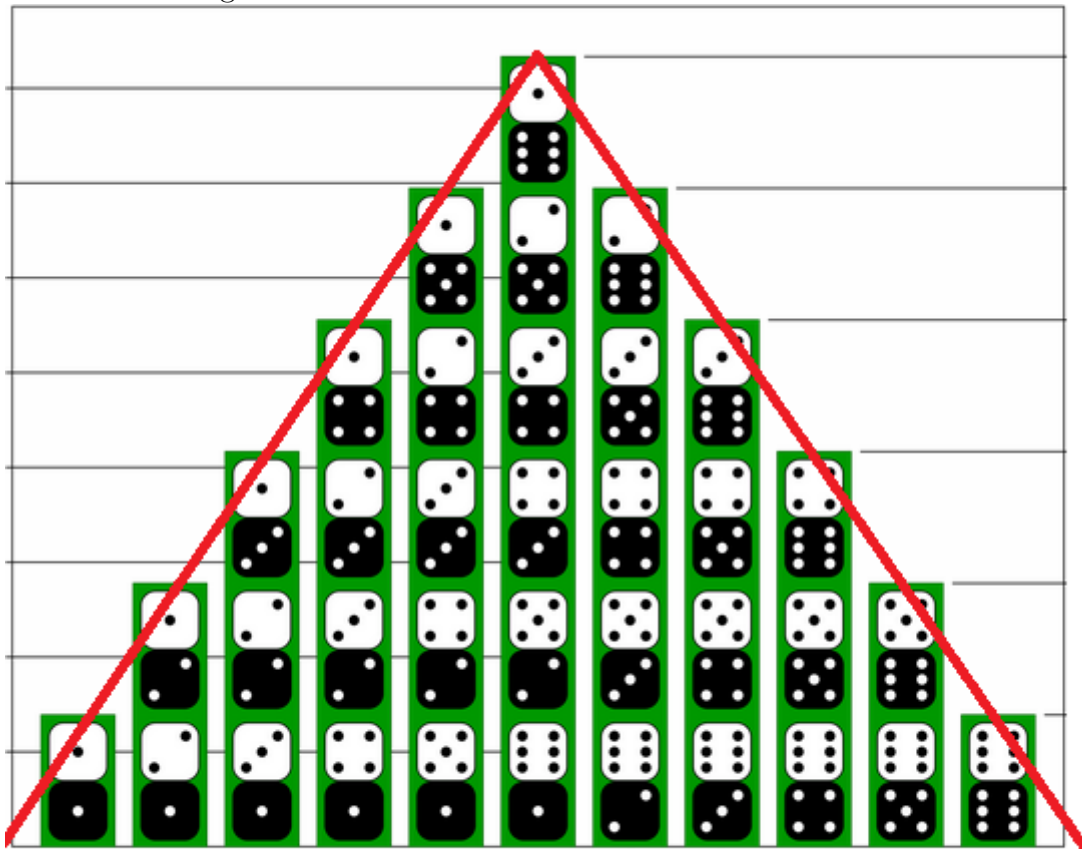
In questo caso X, Y sono indipendenti, quindi avremo la densità di Z detta t_n con $n = 2, \dots, 12$ data da

$$t_n = P(Z = n) = \sum_{i=1}^{n-1} P(X = i) \cdot P(Y = n - i)$$

$X + Y$	2	3	4	5	6	7	8	9	10	11	12
$p_j + q_j$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Tabella 3.1: Distribuzione discreta della somma del lancio di due dadi.

Figura 3.2: Distribuzione della somma del lancio di due dadi



Esempio 3.2.2. Calcolare $P(4 \leq Z \leq 6)$

$$P(4 \leq Z \leq 6) = P(Z = 4) + P(Z = 5) + P(Z = 6) = \frac{3}{36} + \frac{4}{36} + \frac{5}{36} = \frac{12}{36} = \frac{1}{3}$$

Definizione 3.2.3. Indipendenza di variabili aleatorie Due variabili aleatorie X_1, X_2 sono indipendenti se $\forall I_1, I_2, \subseteq \mathbb{R}$ intervalli o semirette si ha che

$$(P(X_1 \in I_1) \cap P(X_2 \in I_2)) = P(X_1 \in I_1) \cdot P(X_2 \in I_2)$$

Nell'esempio di prima X, Z sono dipendenti perché, dati $I_1 = [3, 4], I_2 = [1, 2]$, allora si ha che $P(X \in I_1) = P(X = 3, 4) = \frac{1}{3}$ e si ha anche $P(Z \in I_2) = P(Z = 1, 2) = P(Z = 2) = \frac{1}{36}$. D'altra parte se ottengo 3 o 4 con il primo dado la somma sarà sempre superiore a 5, quindi

$$P((X \in I_1) \cap (Z \in I_2)) = 0 \neq P(X \in I_1) \cdot P(Z \in I_2)$$

Definizione 3.2.4. Variabili Aleatorie Congiunte Due variabili aleatorie $X, Y : \Omega \rightarrow \mathbb{R}$ discrete sono congiunte quando si può calcolare $P(X = m \cap Y = n) = p_{n,m}$, ovvero una densità di probabilità $\{p_{n,m}\}_{n,m}$ con $(p_{n,m} \geq 0) \wedge (\sum_{n,m} p_{n,m} = 1)$.

Sapendo $p_{n,m}$ ricavo tutti i $P(X = m) = p_m^X$ e $P(Y = n) = q_n^Y$ nel modo seguente

$$\begin{aligned} p_m^X &= P(X = m) = P\left(X = m \cap \bigcup_n Y = n\right) \\ &= \sum_n P(X = m, Y = n) = \sum_n p_{n,m} \\ q_n^Y &= P(Y = n) = \sum_m p_{n,m} \end{aligned}$$

In generale non si può ricostruire dalle due densità di probabilità la densità della variabile aleatoria congiunta. Ad esempio, conoscendo p_n^X, q_n^Y cerco $p_{n,m} = P(X = m \cap Y = n) = P(X = m | Y = n) \cdot P(Y = n)$, ma non conosco necessariamente $P(X = m | Y = n)$. Se X, Y sono indipendenti allora $P(X = m | Y = n) = P(X = m)$ e vale il prodotto $p_{m,n} = p_n^X \cdot q_m^Y$. Si noti che in questo caso si verifica subito che $p_{m,n}$ è una densità di probabilità, infatti $p_{m,n} \geq 0$ perché sia p_n^X che q_m^Y lo sono; inoltre

$$\sum_m \sum_n p_{n,m} = \sum_m \sum_n p_n^X \cdot q_m^Y = \left(\sum_m p_m^X\right) \left(\sum_n q_n^Y\right) = 1$$

Definizione 3.2.5. Formula di Convoluzione

Tornando alla somma di due variabili aleatorie discrete, dati X, Y indipendenti e $Z = X + Y$, con $X, Y : \Omega \rightarrow \mathbb{N}$, se vogliamo calcolare $P(Z = n)$ e la densità di probabilità discreta $\{Z = n\}$ possiamo usare le seguenti formule (dette di convoluzione)

$$\begin{aligned} \{Z = n\} &= \bigcup_{i=0}^n \{X = i \cap Y = n - i\} \\ P(Z = n) &= \sum_i P(X = i \cap Y = n - i) = \sum_{i=0}^n P(X = i) \cdot P(Y = n - i) \end{aligned}$$

Ho definito $B(n, p)$ come i successi di n esperimenti che hanno successo con prob. p . In effetti vale la seguente formula

Proposizione 3.2.1. *Si ha che $B(n, p) = \sum_{i=1}^n \text{Bern}_i(p)$*

Dimostrazione. Caso base, per $n = 1$ ovviamente $B(1, p) = \text{Bern}(p)$. Vediamo il passo induttivo:

$$B(n, p) = \sum_{i=1}^n \text{Bern}_i(p) \implies B(n+1, p) = \sum_{i=1}^{n+1} \text{Bern}_i(p)$$

Quindi

$$\sum_{i=1}^{n+1} \text{Bern}_i(p) = \left(\sum_{i=1}^n \text{Bern}_i(p) \right) + \text{Bern}_{n+1}(p) = B(n, p) + \text{Bern}(p)$$

Introduciamo le densità per continuare la dimostrazione

$$\begin{aligned} P(B(n, p) + \text{Bern}(p) = k) &= P(B(n, p) = k \cap \text{Bern}(p) = 0) + P(B(n, p) = k-1 \cap \text{Bern}(p) = 1) \\ &= \binom{n}{k-1} p^{k-1} (1-p)^{n-k+1} \cdot p + \binom{n}{k} p^k (1-p)^{n-k} \cdot (1-p) \\ &= \left[\binom{n}{k-1} + \binom{n}{k} \right] p^k (1-p)^{n-k+1} = \binom{n+1}{k} p^k (1-p)^{n+1-k} = P(B(n+1, p) = k) \end{aligned}$$

□

Definizione 3.2.6. Variabile Geometrica Ripetiamo una successione esperimenti identici ed indipendenti con probabilità di successo $0 \leq p \leq 1$ fino ad ottenere il primo successo. $\text{Geom}(p)$ conta il numero di prove che abbiamo fatto. Ovvero $P(\text{Geom}(p) = k)$ è la probabilità di fare k esperimenti di cui i primi $k-1$ sono stati insuccessi e l'ultimo un successo. La densità di probabilità sarà quindi

$$p_k = P(\text{Geom}(p) = k) = (1-p)^{k-1} \cdot p$$

Ovviamente $p_k \geq 0$ e anche

$$\sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{k=1}^{\infty} (1-p)^{k-1} = \frac{p}{1-(1-p)} = 1$$

dalla somma della una serie geometrica. Quindi p_k è una densità di probabilità

Una variabile geometrica **non ha memoria**, ovvero la probabilità di avere un successo all'esperimento numero n dopo $n-1$ fallimenti ha la stessa probabilità avere un successo al primo esperimento. Infatti vale la seguente proposizione.

Proposizione 3.2.2. *Si ha*

$$P(\text{Geom}(p) = n+m | \text{Geom}(p) > n) = P(\text{Geom}(p) = m)$$

Dimostrazione. Notiamo che dire che abbiamo avuto successo alla prova $m+n$ già contiene l'informazione che abbiamo avuto almeno n insuccessi. Inoltre $\text{Geom}(p) > n$ significa avere avuto sicuramente n insuccessi, quindi $P(\text{Geom}(p) > n) = (1-p)^n$, quindi si ha

$$\begin{aligned} &P(\text{Geom}(p) = n+m | \text{Geom}(p) > n) = \\ &= \frac{P(\text{Geom}(p) = m+n \cap \text{Geom}(p) > n)}{P(\text{Geom}(p) > n)} = \frac{P(\text{Geom}(p) = m+n)}{P(\text{Geom}(p) > n)} \\ &= \frac{(1-p)^{m+n-1} \cdot p}{(1-p)^n} = (1-p)^{m-1} \cdot p = P(\text{Geom}(p) = m) \end{aligned}$$

□

Si può dimostrare sapendo che $\forall m. P(\text{Geom}(p) = n + m \cap P(\text{Geom}) > n) = P(\text{Geom}(p) = m + n)$

Definizione 3.2.7. Variabili Ipergeometriche

Siano dati r sfere rosse, b sfere bianche, n estrazioni senza reimbussolamento, k = numero di sfere rosse estratte, $H(b + r, r, n)$ conta il numero di sfere rosse estratte dopo n tentativi. Questa variabile si chiama variabile ipergeometrica. Le condizioni di esistenza per i parametri sono $(0 < n \leq b + r) \wedge (k \leq n) \wedge (k \leq r) \wedge (n - k \leq b)$ ovvero $\max(0, n - b) \leq k \leq \min(n, r)$.

Estrarre k sfere rosse in n estrazioni significa estrarne k rosse e $n - k$ bianche, senza tener conto dell'ordine, e i casi possibili sono $\binom{b+r}{n}$ (estrarre n sfere da un totale di $b + r$) Quindi la densità di probabilità è

$$P(H(b + r, r, n) = k) = \frac{\binom{r}{k} \binom{b}{n-k}}{\binom{b+r}{n}}$$

Esempio 3.2.3. Abbiamo 10 sfere rosse, 15 bianche e facciamo 7 estrazioni. Voglio contare il numero di sfere rosse estratte.

Senza reimbussolamento: $H(25, 10, 7)$

Con reimbussolamento: $B(7, 2/5)$ (infatti sono 7 esperimenti identici con prob. di successo pari a $10/25 = 2/5$)

Definizione 3.2.8. Binomiale Negativa (o di Pascal) Sia data una Bernoulliana di parametro p . Ripetiamo l'esperimento fino a che non ho n successi. Quanti sono i fallimenti ottenuti? Se la binomiale negativa, indicata con $NB(n, p)$ ha valore k significa che ho fatto $n + k$ prove, e ho avuto n successi e k fallimenti (in qualsiasi ordine) la densità di probabilità di una Binomiale Negativa quindi si definisce come

$$P(NB(n, p) = k) = p^n (1 - p)^k$$

Riassumendo: una variabile Binomiale conta i successi, una variabile Geometrica conta i fallimenti prima del primo successo e la Binomiale Negativa (NB) conta i fallimenti prima del successo n -esimo. In una Binomiale Negativa non conta l'ordine degli esperimenti (tranne l'ultimo risultato). Le prove totali prima di avere n successi sono $n + NB$.

Esempio 3.2.4. Lancio una moneta fino ad ottenere 3 croci (non consecutive). Qual è la probabilità di aver fatto esattamente 2 risultati testa? E di avere ottenuto almeno una testa?

$$P(NB(3, \frac{1}{2}) = 2) = \binom{3+2-1}{2} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = \frac{3}{16}.$$

La probabilità di ottenere almeno un risultato testa è complementare ad avere ottenuto solo croci, quindi è pari a $1 - P(NB(3, \frac{1}{2}) = 0) = 1 - \frac{1}{8} = \frac{7}{8}$

3.3 Valore Atteso

Consideriamo di voler calcolare la media pesata dei voti degli esami universitari. La media sarà per ogni esame i :

$$\sum_i \frac{(\text{voto})_i \cdot (\text{crediti})_i}{\sum_i \text{crediti}} = \sum_i (\text{voto})_i \cdot (\text{peso})_i$$

Se vogliamo definire la media, o il valore atteso di una variabile aleatoria discreta ragioniamo in maniera analoga, con il $(\text{peso})_i$ sarà dato dalla probabilità che X assuma il valore i .

Definizione 3.3.1. Media Pesata, Speranza o Valore Atteso Sia X una variabile aleatoria discreta. La media di X , detta anche **speranza** o **valore atteso** si definisce come

$$\mathbb{E}[X] = \sum_k k \cdot P(X = k) = \sum_k k \cdot p_k$$

Più in generale, data $f : \mathbb{R} \rightarrow \mathbb{R}$ per calcolare il valore atteso della variabile $f(X)$ si definisce

$$\mathbb{E}[f(X)] = \sum_k f(k) \cdot p_k$$

Osservazione $\sum_k p_k = 1$ **non implica che** $\sum_k kp_k$ sia convergente. Se $\sum_k kp_k$ non converge ad un numero allora si dice che la variabile non ha media.

Proposizione 3.3.1. *Se la variabile assume solo valori interi positivi allora*

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} P(X > k) \quad (3.1)$$

Dimostrazione. Se si scrive $P(X > k) = P(X = k+1) + P(X = k+2) + \dots$ per ogni k , e poi si inizia a sommare ci si accorge che p_0 non compare mai, p_1 compare una sola volta, p_2 2 volte e in generale p_k compare esattamente k volte. Quindi $\sum_k P(X > k) \sum_k kp_k = \mathbb{E}[X]$. \square

Proposizione 3.3.2. *La speranza di una Binomiale $B(n, p)$ vale $\mathbb{E}[B(n, p)] = np$.*

Dimostrazione. La densità di $B(n, p)$ è $p_k = \binom{n}{k} p^k (1-p)^{n-k} \forall k \in [0, n]$. Abbiamo che $kp_k = k \binom{n}{k} p^k (1-p)^{n-k} = n \binom{n-1}{k-1} p^k (1-p)^{n-k}$. Definiamo $h = k-1$.

$$\begin{aligned} \mathbb{E}[B(n, p)] &= \sum_{k=1}^n kp_k = \sum_{k=1}^n n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{h=0}^{n-1} \binom{n-1}{h} p^h (1-p)^{(n-1)-h} \\ &= np(p + (1-p))^{n-1} = np \end{aligned}$$

dove nell'ultimo passaggio abbiamo usato la formula del binomio di Newton. \square

Proposizione 3.3.3. *La speranza di una variabile geometrica $\text{Geom}(p)$ vale $\mathbb{E}[\text{Geom}(p)] = 1/p$.*

Dimostrazione. Usiamo la formula (3.1), perché altrimenti non è banale nemmeno dimostrare la convergenza della serie.

$$\mathbb{E}[\text{Geom}(p)] = \sum_{k=0}^{\infty} P(\text{Geom}(p) > k)$$

Come abbiamo già osservato $P(\text{Geom}(p) > k) = (1-p)^k$, quindi

$$\sum_{k=0}^{\infty} P(\text{Geom}(p) > k) = \sum_{k=0}^{\infty} (1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p}$$

\square

Esempio 3.3.1. Scommetto. Pago 1 euro. Lancio 3 dadi e guadagno 1 euro ogni 6 che esce. Rappresento il guadagno con una variabile binomiale $X = B(3, 1/6) - 1$. Abbiamo che $\mathbb{E}[X] = \mathbb{E}[B(3, 1/6)] - 1 = 3 \cdot 1/6 - 1 = -1/2$ (quindi non mi conviene scommettere!)

Definizione 3.3.2. Distribuzione di Poisson Nel caso di una variabile binomiale conosco p e n (numero esperimenti). In una distribuzione di Poisson si conosce una media μ di successi in un intervallo di osservazione τ , e si suppone che nell'intervallo di tempo si verifichino molte prove indipendenti tra di loro. In qualche senso (euristico) la variabile di Poisson di parametro μ rappresenta il limite di una binomiale $B(n, p)$ con $n \rightarrow \infty$. In effetti supponiamo che nell'intervallo τ avvengano n prove di parametro p tutte indipendenti tra di loro. Se abbiamo in media μ successi per intervallo τ , se n è molto grande si può supporre che $\mu = \mathbb{E}[B(n, p)] = np$. Per definire la Poisson allora proviamo a fare il limite di una densità binomiale.

$$\begin{aligned} \lim_{n \rightarrow \infty} P(B(n, p) = k) &= \lim_{n \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} = \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k} = \\ &= \lim_{n \rightarrow \infty} \frac{n!}{k!(n-k)!} \frac{1}{n^k} \frac{\mu^k}{\left(1 - \frac{\mu}{n}\right)^k} \cdot \left(1 - \frac{\mu}{n}\right)^n = \lim_{n \rightarrow \infty} \frac{n(n-1) \dots (n-k+1)}{n^k k!} \frac{\mu^k}{\left(1 - \frac{\mu}{n}\right)^k} \cdot \left(1 - \frac{\mu}{n}\right)^n = \\ &= \lim_{n \rightarrow \infty} \frac{1 \cdot \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right)}{k!} \cdot \frac{\mu^k}{\left(1 - \frac{\mu}{n}\right)^k} \cdot \left(1 - \frac{\mu}{n}\right)^n = \\ &= \frac{\mu^k}{k!} e^{-\mu} \end{aligned}$$

Con questo ragionamento (non rigoroso!) possiamo definire

$$P(\text{Poisson}(\mu) = k) = \frac{\mu^k}{k!} e^{-\mu} \quad (3.2)$$

La variabile di Poisson in qualche senso ci dice quanto è la probabilità che in un certo intervallo di tempo ci si discosti dal valore medio che ci si aspetta in uno stesso intervallo di tempo.

Esempio 3.3.2. Siamo nel secolo 1800, prendiamo l'esercito di Napoleone nel reparto della cavalleria. Ogni anno 12 cavalieri muoiono per incidente a cavallo. Voglio sapere la probabilità che nel 1861 siano morti 7 cavalieri.

$$P(\text{anno } 1861 | \text{sono morti } 7 \text{ cavalieri})$$

Utilizziamo la distribuzione di Poisson.

$$P(\text{Poisson}(12) = 7) = \frac{12^7}{7!} e^{-12} \approx 0.04$$

Proposizione 3.3.4. Linearità della media

Siano date X, Y variabili aleatorie, $\alpha, \beta \in \mathbb{R}$. Abbiamo che

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

Dimostrazione. Dimostriamo la proprietà nel caso $\alpha = \beta = 1$. La dimostrazione nel caso generale è assolutamente analoga. Supponiamo anche, per semplificare le notazioni, che X, Y assumano valori naturali. Come al solito, chiamiamo $p_{i,j} = P(X = i \cap Y = j)$, $p_i^X = P(X = i)$ e $q_j^Y = P(Y = j)$

$$\begin{aligned} \mathbb{E}[X + Y] &= \sum_{i,j} (i + j) p_{i,j} = \sum_{i,j} i p_{i,j} + \sum_{i,j} j p_{i,j} \\ &= \sum_i i \sum_j p_{i,j} + \sum_j j \sum_i p_{i,j} \\ &= \sum_i i p_i^X + \sum_j j q_j^Y = \mathbb{E}[X] + \mathbb{E}[Y] \end{aligned}$$

□

Proposizione 3.3.5. Valore atteso del prodotto Siano date X, Y variabili aleatorie **indipendenti**. Allora

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

Dimostrazione. Se X e Y sono indipendenti, allora abbiamo che $p_{i,j} = p_i^X \cdot q_j^Y$, quindi

$$\begin{aligned} \mathbb{E}[X \cdot Y] &= \sum_{i,j} ijP(X=i \cap Y=j) \\ &= \sum_{i,j} ijp_{i,j} = \sum_{i,j} ijp_i^X q_j^Y = \sum_i ip_i^X \cdot \sum_j jq_j^Y = \mathbb{E}[X] \cdot \mathbb{E}[Y] \end{aligned}$$

□

Definizione 3.3.3. Momenti di ordine superiore

Si definisce $\mathbb{E}[X^n]$ il momento di X di ordine n . Dalla definizione di valore atteso sappiamo, ad esempio $\mathbb{E}[X^2] = \sum_i i^2 p_i$. Attenzione: per quanto detto sopra, e visto che X non è indipendente da se stessa, in generale $\mathbb{E}[X^2] \neq (\mathbb{E}[X])^2$.

Definizione 3.3.4. Varianza

Sia X una variabile aleatoria e $\mathbb{E}[X] = \mu$ la sua media. La varianza di X si definisce come

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mu)^2] \\ &= \sum_i (i - \mu)^2 P(X=i) \end{aligned}$$

Proposizione 3.3.6. La varianza di X è anche $\mathbb{E}[X^2] - (\mathbb{E}[X])^2$

Dimostrazione.

$$\begin{aligned} \sum_i (i - \mu)^2 p_i &= \sum_i (i^2 - 2i\mu + \mu^2) p_i \\ &= \sum_i i^2 p_i - 2\mu \sum_i i p_i + \mu^2 \sum_i p_i \\ &= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 = \mathbb{E}[X^2] - \mu^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \end{aligned}$$

□

Diamo ora alcune disuguaglianze utili (senza dimostrarle)

Proposizione 3.3.7. Disuguaglianza di Hölder

$$\begin{aligned} \mathbb{E}[X \cdot Y] &\leq (E[X^p])^{\frac{1}{p}} (E[Y^q])^{\frac{1}{q}} \\ \text{con } \frac{1}{p} + \frac{1}{q} &= 1 \end{aligned} \tag{3.3}$$

In particolare, per $p = q = 2$,

$$\mathbb{E}[X \cdot Y] \leq \sqrt{E[X^2]E[Y^2]} \tag{3.4}$$

Proposizione 3.3.8. Disuguaglianza di Markov Sia $X \geq 0$ e $a > 0$. Vale

$$P(X > a) \leq \frac{\mathbb{E}[X]}{a} \tag{3.5}$$

Proposizione 3.3.9. Disuguaglianza di Chebishev Sia $X \geq 0$ e $a > 0$. Vale

$$P(|X - \mu| > a) \leq \frac{\text{Var}(X)}{a^2} \quad (3.6)$$

Della disuguaglianza di Chebishev diamo anche la dimostrazione.

Dimostrazione.

$$\begin{aligned} P(|X - \mu| > a) &= \sum_{n \text{ t.c. } |n - \mu| > a} P(X = n) \\ &\text{se } |n - \mu| > a \text{ allora } 1 \leq \frac{|n - \mu|^2}{a^2} \\ \Rightarrow \sum_{n \text{ t.c. } |n - \mu| > a} p_n &\leq \sum_{n \text{ t.c. } |n - \mu| > a} \frac{|n - \mu|^2}{a^2} p_n \leq \frac{1}{a^2} \sum_{n \in \mathbb{N}} |n - \mu|^2 p_n = \frac{\text{Var}(X)}{a^2} \end{aligned}$$

□

La varianza quindi misura, in qualche senso, quanto una variabile aleatoria si scosti dalla sua media.

Definizione 3.3.5. σ Deviazione Standard

Sia X variabile aleatoria. Allora definiamo la deviazione standard $\sigma(X) := \sqrt{\text{Var}(X)}$.

Proposizione 3.3.10. Data μ media di X , e dati $\alpha, \beta \in \mathbb{R}$ si ha allora che

$$\begin{aligned} \text{Var}(\alpha X + \beta) &= \alpha^2 \text{Var}(X) \\ \sigma(\alpha X + \beta) &= |\alpha| \sigma(X) \end{aligned}$$

Dimostrazione. Dimostriamo solo la prima formula, la seconda segue immediatamente.

$$\begin{aligned} \text{Var}(\alpha X + \beta) &= \mathbb{E}[(\alpha X + \beta)^2] - (\mathbb{E}[\alpha X + \beta])^2 \\ &= \mathbb{E}[\alpha^2 X^2 + 2\alpha\beta X + \beta^2] - (\alpha \mathbb{E}[X] + \beta)^2 \\ &= \alpha^2 \mathbb{E}[X^2] + 2\alpha\beta \mathbb{E}[X] + \beta^2 - \alpha^2 (\mathbb{E}[X])^2 - 2\alpha\beta \mathbb{E}[X] - \beta^2 \\ &= \alpha^2 (\mathbb{E}[X^2] - (\mathbb{E}[X])^2) = \alpha^2 \text{Var}(X) \end{aligned}$$

□

Definizione 3.3.6. Somma di varianza

Siano date X, Y variabili aleatorie indipendenti $\Rightarrow \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

Dimostrazione.

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[X \cdot Y] + \mathbb{E}[Y^2] - (\mathbb{E}[X])^2 - 2\mathbb{E}[X] \mathbb{E}[Y] - (\mathbb{E}[Y])^2 \\ &= \text{Var}(X) + \text{Var}(Y) + 2(\mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]) \end{aligned}$$

e se X, Y sono indipendenti si ha che $2(\mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]) = 0$

□

Esempio 3.3.3. Varianza di una Bernoulliana: $X = B(1, p)$

$$\begin{aligned} \mathbb{E}[X] &= p \\ \mathbb{E}[X]^2 &= p^2 \\ \text{Var}(X) &= p - p^2 = p(1 - p) \end{aligned}$$

Esempio 3.3.4. Varianza di una Binomiale

$X = B(n, p) \Rightarrow \text{Var}(X) = n \text{Var}(B(1, p)) = np(1 - p)$

Esempio 3.3.5. Varianza di una distribuzione di Poisson

$X = \text{Poisson}(\lambda) \Rightarrow \text{Var}(\text{Poisson}(\lambda)) = \lambda$ (senza dim)

3.4 Esercizi

Esercizio 3.4.1. Ho una scatola con 12 lampadine, 4 di esse sono fulminate. Ne prendo 2. La probabilità che siano entrambe funzionanti è $P(H(12, 4, 2) = 0) = \frac{\binom{4}{0}\binom{8}{2}}{\binom{12}{2}}$

Esercizio 3.4.2. Ho una moneta truccata. La probabilità che esca testa è $P_t = 0.55$ e la probabilità che esca croce è $P_c = 0.45$. Lancio la moneta dieci volte, qual è la probabilità che avvenga la sequenza testa-croce per la prima volta al lancio 9-10? Perché ciò sia possibile deve uscire una sequenza composta da $0 \leq h \leq 8$ lanci “croce” consecutivi e $9 - h$ lanci “testa” consecutivi, in modo da ottenere una sequenza formata da $C^h T^{9-h} C$. La probabilità è $P(C^h T^{9-h} C) = (0.45)^h \cdot (0.55)^{9-h} \cdot (0.45) = (0.45)^{h+1} \cdot (0.55)^{9-h}$. La probabilità dell’unione \bigcup_h delle stringhe sarà $P = \sum_{h=0}^8 (0.45)^{h+1} (0.55)^{9-h}$. Provare a svolgere l’esercizio usando la variabile geometrica. Provare a fare l’esercizio con una moneta equilibrata. Come si semplifica la formula?

Esercizio 3.4.3. Un ubriaco cammina in una strada in pendenza. Va in salita con probabilità $P(\text{salita}) = 1/4$ oppure in discesa con probabilità $P(\text{discesa}) = 3/4$. Ogni 10 secondi decide casualmente una direzione. Si muove lungo un asse X partendo dall’origine a velocità $\frac{1m}{10s}$. Qual è la posizione più probabile dopo 1 minuto? Introduciamo una variabile X = la posizione dopo 1 minuto. L’ubriaco si sposterà al massimo di 6 metri in salita o 6 metri in discesa, quindi $X \in \{-6, \dots, +6\}$. Introduciamo anche la variabile Y = il numero di volte che l’ubriaco cambia direzione verso la discesa. Y è una variabile binomiale Bernoulliana (conta il numero di “successi” in 6 esperimenti ripetuti) $\implies Y = B(6, 3/4)$. Abbiamo che $X = -1 \cdot Y + 1(6 - Y) = 6 - 2Y$. Per quanto detto

$$P_k^Y = P(B(6, 3/4) = k) = \binom{6}{k} \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{6-k}$$

k	0	1	2	3	4	5	6
P_k	$\frac{1}{4^6}$	$\frac{6 \cdot 3}{4^6}$	$\frac{15 \cdot 3^2}{4^6}$	$\frac{20 \cdot 3^3}{4^6}$	$\frac{15 \cdot 3^4}{4^6}$	$\frac{6 \cdot 3^5}{4^6}$	$\frac{3^6}{4^6}$

Tabella 3.2: Distribuzione della variabile Y

Il valore più probabile per Y è quindi 5, e di conseguenza il valore più probabile per X sarà -4.

Esempio 3.4.4. Prendo un seme di carte francesi $\{A, 2, \dots, 10, J, Q, K\}$ L’asso ha valore 11. I numeri da 2 a 10 hanno lo stesso valore del numero, le figure hanno valore 10. Estraggo una carta. Definiamo una variabile aleatoria X = punteggio. $X \in \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$. Abbiamo che $p_k = 0 \iff k < 2 \wedge k > 11$. Abbiamo anche che $p_k = 1/13 \iff k = 2, \dots, 11$ e $p_k = 4/13 \iff k = 10$

Esercizio 3.4.5. È più probabile fare almeno un 6 lanciando 4 dadi o almeno una coppia di 6 lanciando 25 volte una coppia di dadi?

Esercizio 3.4.6. Siano date due slot machines apparentemente identiche A, B . La probabilità di vincere sulla A è $\frac{1}{2}$. La probabilità di vincere sulla B è $\frac{1}{4}$. Calcolare $P(\text{aver giocato su } A \mid \text{aver vinto})$ [Sugg.: usare Bayes e fattorizzazione]

Esercizio 3.4.7. Data un'urna contenente 2 palline bianche e 5 nere. Se la prima estrazione è una pallina bianca, essa viene rimossa. Se invece è una pallina nera, la rimettiamo dentro e aggiungiamo altre 2 nere. Calcolare P (seconda estrazione sia una pallina nera)

Esercizio 3.4.8. In Finlandia il 70% delle ragazze sono Bionde, il 20% sono Rosse, il 10% sono More. Hanno gli occhi Scuri il 10% delle Bionde, il 25% delle Rosse e il 50% delle More. Conosco una ragazza (via email, quindi non ho foto) che dice di avere gli occhi scuri. Con che probabilità è bionda?

I dati che abbiamo sono quindi: $P(B) = 7/10$, $P(R) = 1/5$, $P(M) = 1/10$, $P(S | B) = 1/10$, $P(S | R) = 1/4$, $P(S | M) = 1/2$,

Utilizzando la formula di Bayes e di Fattorizzazione calcoliamo

$$\begin{aligned} P(B | S) &= \frac{P(S | B) \cdot P(B)}{P(S)} \\ &= \frac{P(S | B) \cdot P(B)}{P(S | B)P(B) + P(S | R)P(R) + P(S | M)P(M)} \approx 0.41 \end{aligned}$$

Esercizio 3.4.9. Ho 3 carte colorate sulla faccia e sul dorso. Una carta con la faccia rossa e il retro nero si scrive $\frac{R}{N}$. Le tre carte sono quindi $\frac{R}{N}$, $\frac{R}{R}$, $\frac{N}{N}$. Una di queste tre carte è sul tavolo e la faccia visibile è Rossa. Calcolare $P(\text{Faccia coperta} = R)$. Indichiamo con V la faccia Visibile e con C quella coperta.

$$P(C = R | V = R) = \frac{P(C = R \cap V = R)}{P(V = R)} = \frac{1/3}{1/2} = \frac{2}{3}$$

Esercizio 3.4.10. Siano dati due eventi $A, B \subset \Omega$. Abbiamo che $P(A) = \frac{3}{4}$ e abbiamo $P(B) = \frac{1}{3}$. Possono essere disgiunti? No. Perché la probabilità della loro unione è maggiore di uno $P(A \cup B) = P(A) + P(B) = \frac{3}{4} + \frac{1}{3}$, che è > 1 . Abbiamo che $A \cup B \subset \Omega$, ma $1 = P(\Omega) > P(A \cup B)$.

Si verifichi la disuguaglianza $\frac{1}{12} \leq P(A \cap B) \leq \frac{1}{3}$.

Sappiamo che

$$\begin{aligned} A \cap B &\subseteq A \text{ e } A \cap B \subseteq B \\ \implies (P(A \cap B) \leq P(A)) \text{ e } (P(A \cap B) \leq P(B)) \\ \implies P(A \cap B) &\leq \min\{P(A), P(B)\} \end{aligned}$$

Quindi che $P(A \cap B) \leq \frac{1}{3}$. Verifichiamo ora la prima parte della disuguaglianza $A = (A \cap B) \cup (A \cap B^C)$ e gli insiemi $(A \cap B)$ e $(A \cap B^C)$ sono disgiunti. Inoltre $A \cap B^C \subset B^C$ Quindi

$$\begin{aligned} P(A \cap B) &= P(A) - P(A \cap B^C) \geq P(A) - P(B^C) \\ &= P(A) - (1 - P(B)) = P(A) + P(B) - 1 = \frac{1}{3} + \frac{3}{4} - 1 = \frac{1}{12} \end{aligned}$$

Esercizio 3.4.11. Siano X e Y variabili aleatorie congiunte $X, Y : \Omega \rightarrow \{0, 1\}$. Sia $p_{i,j}$ la densità di probabilità della variabile congiunta, con $p_{0,0} = 0,51$, $p_{1,0} = 0,02$, $p_{1,1} = 0,46$.

1. Si determini $p_{0,1}$ [R: $p_{0,1} = 0,01$]

2. Si trovino le densità di X e di Y [R: $p_0^X = 0,52$, $p_1^X = 0,48$ e $q_0^X = 0,53$, $q_1^X = 0,47$]

3. Si calcoli $\mathbb{E}[X]$ e $\mathbb{E}[Y]$ [R: $\mathbb{E}[X] = 0,48$ e $\mathbb{E}[Y] = 0,47$]
4. Le due variabili sono indipendenti? [R: No, ad es. $p_{0,1} \neq p_0^X \cdot q_1^Y$]
5. Si calcoli $P(X = 1 | Y = 1)$ [R: $P(X = 1 | Y = 1) = P(Y = 1 | X = 1) / P(X = 1) \simeq 0,95$]

Esercizio 3.4.12. Sia X una Bernoulliana di parametro $1/2$ e sia Y una variabile aleatoria a valori in $\{1, 2, 3\}$ tale che $P(Y = 1 | X = 0) = 0,1$; $P(Y = 2 | X = 0) = 0,4$; $P(Y = 3 | X = 0) = 0,5$; $P(Y = 1 | X = 1) = 0,5$; $P(Y = 2 | X = 1) = 0,4$; $P(Y = 3 | X = 1) = 0,1$.

1. Trovare la distribuzione congiunta di X, Y e la distribuzione di Y [R: $P(Y = 1, X = 0) = P(Y = 1 | X = 0) / P(X = 0) = 0,05$ ecc.]
2. Calcolare $\mathbb{E}[Y]$ e $\text{Var}(Y)$ [R: $\mathbb{E}[Y] = 2$, $\text{Var}(Y) = 0,6$]
3. Calcolare $\mathbb{E}[X]$ e $\mathbb{E}[XY]$ [R: $\mathbb{E}[X] = 0,5$, $\mathbb{E}[XY] = 0,8$]
4. Dire se X e Y sono indipendenti [R: No]
5. Trovare la densità di probabilità di $Z = X/Y$ [R: $P(Z = 0) = 0,5$; $P(Z = 1) = 0,25$; $P(Z = 1/2) = 0,2$; $P(Z = 1/3) = 0,05$]

Esercizio 3.4.13. Si lancia un dado a 4 facce, con valori $N = \{0, 1, 2, 3\}$. Il risultato ci dice quante volte dobbiamo lanciare una moneta. La variabile k conta il numero totale di teste uscite.

1. Determinare $p_{N,k}$ [R: $p_{N,k} = 0$ se $k > N$ (non posso avere più teste di quante volte ho lanciato la moneta) e $p_{N,k} = \binom{N}{k} \frac{1}{2^{N+2}}$ se $k \leq N$]
2. Calcolare $P(N = 2 | k = 3)$ e $P(N = 3 | k = 1)$ [R: $P(N = 2 | k = 3) = 0$; $P(N = 3 | k = 1) = 3/11$]

Esercizio 3.4.14. Siano $X : \Omega \rightarrow \{-1, 0, 1\}$ e $Y : \Omega \rightarrow \{0, 1\}$ variabili congiunte, con $p_{1,1} = p_{1,0} = p_{0,0} = p_{-1,1} = 1/10$; $p_{0,1} = p_{-1,0} = 3/10$.

1. Trovare p_i^X e q_j^Y .
2. Dire se X e Y sono indipendenti
3. Trovare la distribuzione di X^2 ed di XY
4. Trovare valor medio e varianza di X, Y, X^2, XY .

3.5 Esercitazione del 29/10/19

Lezione tenuta da Maurizio Pratelli.

Esercizio 3.5.1. Ci sono 3 monete indistinguibili, delle quali due sono truccate. La probabilità che esca testa è nell'ordine $1/4$, $1/2$ e $3/4$. Si sceglie una moneta casuale, si lancia 5 volte e si ottiene testa 4 volte: qual è la probabilità di aver scelto la terza moneta?

Utilizziamo la formula di Bayes. Definiamo le 3 monete A_1, A_2, A_3 . Sappiamo anche che $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$. Definiamo l'evento B come "esce 4 volte su 5 testa".

$$\begin{aligned}
 P(A_3 | B) &= \frac{P(B | A_3)}{P(B | A_1) + P(B | A_2) + P(B | A_3)} \\
 P(B | A_1) &= \binom{5}{4} \left(\frac{1}{4}\right)^4 \cdot \frac{3}{4} \\
 P(B | A_2) &= \binom{5}{4} \left(\frac{1}{2}\right)^5 \\
 P(B | A_3) &= \binom{5}{4} \left(\frac{3}{4}\right)^4 \cdot \frac{1}{4}
 \end{aligned}$$

Esercizio 3.5.2. Si prende un giorno a caso in un anno non bisestile ed è un mercoledì: in quel mese ci sono esattamente 4 mercoledì. Qual è la probabilità che quel giorno sia di febbraio?

Definiamo

- A_1 = "Mese di 28 giorni". $P(A_1) = \frac{1}{12}$
- A_2 = "Mese di 30 giorni". $P(A_2) = \frac{4}{12}$
- A_3 = "Mese di 31 giorni". $P(A_3) = \frac{7}{12}$
- B = in quelle mese ci sono 4 mercoledì.

$$\begin{aligned}
 P(A_1 | B) &= P(B | A_1) \frac{P(A_1)}{P(B)} \\
 &= \frac{P(B | A_1) P(A_1)}{P(B | A_1) P(A_1) + P(B | A_2) P(A_2) + P(B | A_3) P(A_3)} \\
 P(B | A_1) &= 1 \text{ (a febbraio ci sono sicuramente 4 mercoledì)} \\
 P(B | A_2) &= 1 - \frac{2}{7} = \frac{5}{7} \\
 P(B | A_3) &= \frac{4}{7}
 \end{aligned}$$

Esercizio 3.5.3. Si lancia un dado equilibrato finché il numero 6 esce per 2 volte (non necessariamente consecutive). Qual è la probabilità che anche il penultimo lancio fosse un 6?

Sappiamo la definizione di variabile geometrica (conta il numero) di esperimenti con probabilità p falliti necessari prima di ottenere un successo: $X = \text{Geomp}$.

$$P(X = k) = (1 - p)^{k-1} p \quad \forall k$$

Definiamo Y = tentativi fino al secondo successo. I valori possibili sono 2, 3, 4...

Nota. Serie geometriche da ricordare

$$\begin{aligned}
 |a| < 1 &\implies \sum_{h=0}^{\infty} a^h = \frac{1}{1-a} \\
 &\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x
 \end{aligned}$$

Risolviamo l'esercizio.

$$\begin{aligned}
 P(Y = k) &= (k-1)(1-p)^{k-1} \cdot p^2 \\
 P(\text{Avere due successi consecutivi e prima nessuno}) &= \sum_{k=2}^{\infty} (1-p)^{k-1} p^2 = p^2 \sum_{h=0}^{\infty} (1-p)^h \\
 &= \frac{p^2}{1-(1-p)} = p = \frac{1}{6}
 \end{aligned}$$

Esercizio 3.5.4. Data una variabile aleatoria X per la quale si ha $\mathbb{E}[X] = 1$, $\text{Var}(X) = 2$ e si ha che $\mathbb{E}[X^4] = 10$. Quanto vale la varianza di X ?

$$\begin{aligned}
 \text{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\
 \mathbb{E}[X^2] &= 2 + 1^2 = 3 \\
 \text{Var}(X^2) &= \mathbb{E}[X^4] - (\mathbb{E}[X^2])^2 = 1
 \end{aligned}$$

Esercizio 3.5.5. Si lancia per 15 volte un dado equilibrato e indichiamo con X la somma dei numeri ottenuti. Quanto vale $\mathbb{E}[X]$? Qual è la varianza di X ?

$$\begin{aligned}
 X &= X_1 + X_2 + \dots + X_{15} \\
 X \text{ ha valori possibili } &\{15, 16, 17, \dots, 90\} \\
 \mathbb{E}[X] &= \mathbb{E}[X_1] + \dots + \mathbb{E}[X_{15}] = 15\mathbb{E}[X_1] \\
 \text{Var}(X) &= 15 \cdot \text{Var}(X_1) \\
 \mathbb{E}[X_1] &= \frac{1+2+3+4+5+6}{6} = \frac{7}{2} \\
 \mathbb{E}[X_1^2] &= \frac{\sum_{k=1}^6 k^2}{6} = \frac{91}{6}
 \end{aligned}$$

Esercizio 3.5.6. (Roulette Russa) Paolo, Andrea e Giacomo sparano a turno in questo ordine con una pistola a tamburo a 6 colpi, nella quale è presente un solo proiettile, finché il colpo non viene esploso. Si considerino queste due modalità:

- a) il tamburo viene ruotato una volta sola all'inizio
- b) il tamburo viene ruotato prima di ogni colpo

Quale delle due modalità è più conveniente per Giacomo, e quando e quanto è vantaggioso poter sparare per primo?

Ogni giocatore ha $1/3$ di probabilità di vincere. Consideriamo il caso *b*.

$$\begin{aligned}
 P(\text{vince Paolo}) &= \frac{1}{6} + \left(\frac{5}{6}\right)^3 \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^6 \cdot \frac{1}{6} + \dots = \frac{1}{6} \\
 &= \frac{1}{6} \cdot \sum_{h=0}^{\infty} \left(\left(\frac{5}{6}\right)^2\right)^h = \frac{1}{6} \cdot \frac{1}{1 - \frac{5^2}{6^2}} = \frac{36}{216 - 125} = a
 \end{aligned}$$

$$P(\text{vince Andrea}) = \frac{5}{6} \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^4 \cdot \frac{1}{6} = \frac{5}{6}a$$

$$P(\text{vince Giacomo}) = \left(\frac{5}{6}\right)^2 a$$

Esercizio 3.5.7. Un candidato affronta una test a risposte multiple, con 5 domande. Ogni domanda ha 4 risposte delle quali una sola è esatta: ogni risposta esatta è valutata un punto e ogni risposta errata è penalizzata con $-1/4$. Consideriamo un candidato del tutto impreparato che risponde a caso. Quale punteggio ottiene in media? Qual è la probabilità di ottenere un punteggio di 2.5

$$\begin{aligned}
 X &= \text{risposte esatte} = B\left(5, \frac{1}{4}\right) \\
 Y &= \text{punteggio} = X - \frac{1}{4}(5 - X) = \frac{5}{4}X - \frac{5}{4} \\
 &\quad \text{Media del punteggio} \\
 \mathbb{E}[Y] &= \frac{5}{4} \cdot \mathbb{E}[X] - \frac{5}{4} = \frac{5}{4} \cdot \frac{5}{4} - \frac{5}{4} = \frac{5}{16}
 \end{aligned}$$

Risolviamo il secondo punto. La probabilità di ottenere un punteggio di 2.5.

$$\begin{aligned}
 Y &= 2.5 \\
 \frac{5}{4}X - \frac{5}{4} &= 2.5 \\
 P(Y = 2.5) &= P(X = 3) = \binom{5}{3} \cdot \left(\frac{1}{4}\right)^3 \cdot \left(\frac{3}{4}\right)^2
 \end{aligned}$$

Capitolo 4

Catene di Markov

4.1 Catene di Markov e Processi Stocastici

Definizione 4.1.1. Processi Stocastici Spesso abbiamo bisogno di rappresentare quantità incerte che cambiano nel tempo. Possiamo rappresentarle con famiglie di variabili aleatorie indicizzate mediante un parametro, spesso corrispondente al “tempo”

Una famiglia di variabili aleatorie $\{X_t\}_{t \in \mathcal{T}}$ dove $\mathcal{T} \subseteq \mathbb{R}$ e che assumono tutti valori nello stesso insieme E è detta **processo stocastico**. L'insieme E è detto spazio degli stati del processo, mentre l'insieme \mathcal{T} è detto insieme dei tempi. Considereremo sempre gli insiemi degli stati e dei tempi *discreti* (numerabili) e molto spesso finiti. L'insieme dei tempi può essere un intervallo $\mathcal{T} = [0, T]$. Ad esempio, insiemi \mathcal{T} validi possono essere: $\mathbb{N}, \mathbb{Z}, \{0, 1, 2, \dots, n\}, \{t_1, t_2, \dots, t_n\}$. Dato un processo stocastico $\{X_t\}_{t \in \mathcal{T}}$ le variabili aleatorie $X_t \in E$ sono dette marginali del processo. Le leggi delle marginali di due processi potrebbero coincidere, pur essendo i due processi molto diversi.

Esempio 4.1.1. Consideriamo le estrazioni da un'urna contenente 10 palline rosse (R) e 3 palline blu (B). Prendiamo in considerazione il colore della pallina alla prima, seconda, terza, ecc. estrazione. Il fenomeno è rappresentabile con una famiglia di variabili aleatorie.

$$X_1, X_2, X_3, \dots, X_n \in \{\text{rossa, blu}\}$$

Considero tre tipi diversi di estrazione

1. estraggo senza reimbussolare
2. estraggo e reimbussolo la pallina estratta
3. estraggo la prima, estraggo la seconda e reimbussolo la prima, estraggo la terza e reimbussolo la seconda, ecc. In pratica ad ogni estrazione nell'urna ci sono tutte le palline escluso l'ultima estratta

Calcolare, nei tre casi, $P(3a \text{ estratta} = R | 1a \text{ estratta} = B, 2a \text{ estratta} = R)$ nei tre casi.

Nel caso a) abbiamo $P(3a = R | 1a = B, 2a = R) = P(3a = R) = 10/3$;

nel caso b) $P(3a = R | 1a = B, 2a = R) = 9/11$;

nel caso c) $P(3a = R | 1a = B, 2a = R) = P(3a = R | 2a = R) = 9/12$

In tutti i tre casi ogni marginale X_t si trova nello stesso modo:

$$P(X_t = R) = \frac{\text{palline rosse presenti nell'urna}}{\text{palline totali presenti nell'urna}}$$

ma la legge specifica ottenuta dipende dall'istante t in cui si fa l'estrazione e - a seconda del tipo di processo - dalla storia delle estrazioni precedenti.

Definizione 4.1.2. Processo di Markov: Un processo aleatorio è detto di Markov se l'evento all'istante $k + 1$ dipende solo dall'esito dell'evento k ma non da quelli precedenti. In formule, se e_k rappresenta il valore assunto dalla variabile X_k , il processo $\{X_t\}_{t \in \mathbb{N}}$ è di Markov se

$$P(X_{k+1} = e_{k+1} | X_k = e_k \cap X_{k-1} = e_{k-1} \cap \dots \cap X_0 = e_0) = P(X_{k+1} = e_{k+1} | X_k = e_k).$$

In altri termini si può anche dire

Un processo è di Markov se conoscendo il presente, passato e futuro sono indipendenti.

Esempio 4.1.2. Tornando all'esempio 4.1.1 dell'estrazione dall'urna un'estrazione con reimmissione è sicuramente un processo di Markov, se l'estrazione è senza reimmissione il processo **non** è di Markov. Il motivo è che tutta la sequenza di palline estratte è necessaria per conoscere il contenuto esatto dell'urna (l'informazione passata non può essere trascurata). Anche nel terzo caso si ha un processo di Markov, perché le estrazioni non sono indipendenti tra di loro, ma ogni estrazione dipende solo dalla precedente.

Nota. Se un processo è di Markov per conoscerlo interamente bastano le probabilità di transizione $P(X_{k+1} = j | X_k = i)$ (oltre alla marginale X_0).

Definizione 4.1.3. Un process di Markov $X_{ii \in \mathbb{N}}$ è **omogeneo** se le probabilità di transizione non dipendono dall'istante k , ovvero:

$$\forall i, j \in \mathbb{N}, P(X_{k+1} = j | \{X_k = i\}) = P(X_1 = j | \{X_0 = i\})$$

Esempio 4.1.3. Nell'esempio 4.1.1, a) e c) sono processi di Markov omogenei.

Esempio 4.1.4. Modifichiamo l'esempio 4.1.1 nel modo seguente: oltre a quanto detto, ad ogni estrazione aggiungiamo una pallina nera nell'urna. Allora a) e c) restano processi di Markov, ma non sono più omogenei

Nota. Se un processo è di Markov omogeneo per conoscerlo interamente bastano le probabilità della prima transizione $P(X_1 = j | X_0 = i)$ (oltre alla marginale X_0).

Definizione 4.1.4. Matrice di Transizione: Fissato un ordinamento degli stati di un processo di Markov omogeneo a stati finito, si definisce Matrice di transizione la matrice $Q = (q_{ij})_{ij}$ dove

$$Q_{ij} = Q_{i \rightarrow j} := P(X_1 = j | I \cap \{X_0 = i\})$$

Questa definizione ci permette di collezionare le probabilità di transizione in una singola matrice.

Definizione 4.1.5. Catena di Markov: Un processo di Markov omogeneo $\{X_i\}_{i=0, \dots, n}$ a stati finiti (o discreti) è detto **Catena di Markov**.

Si possono visualizzare le Catene di Markov, data una matrice di transizione, con un grafo orientato analogo agli automi a stati finiti. Ad ogni stato $i \in E$ facciamo corrispondere un nodo, e ad ogni probabilità di transizione $Q_{i \rightarrow j}$ strettamente positiva facciamo corrispondere un arco (i, j) . Non si disegnano gli archi delle probabilità di transizione nulle. La rappresentazione con i grafi non indica nulla sulle leggi marginali della catena.

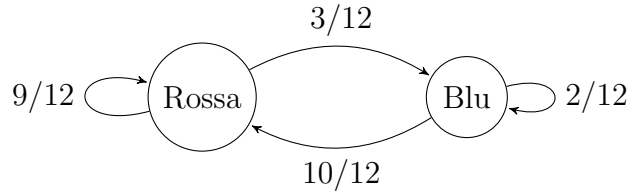
Esempio 4.1.5. Scriviamo la matrice di transizione per l'esempio 4.1.1 nei casi a) e c). Se associamo al rosso lo stato 1 e al blu lo stato 2 abbiamo

$$\text{a): } Q = \begin{pmatrix} 10/13 & 3/13 \\ 10/13 & 3/13 \end{pmatrix} \quad \text{c): } Q = \begin{pmatrix} 9/12 & 3/12 \\ 10/12 & 2/12 \end{pmatrix}$$

Si noti che la somma delle righe della matrice deve dare 1 (rappresenta una densità di probabilità).

Esempio 4.1.6. Scriviamo per l'esempio 4.1.1 nel caso c), il grafo associato alla catena di Markov

Figura 4.1: Catena di Markov

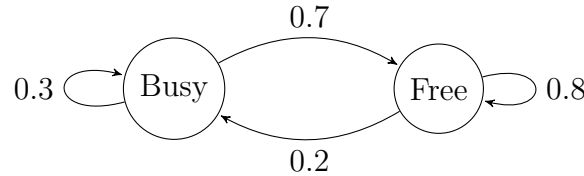


Si noti che in questo caso la somma dei valori *uscenti* da un nodo deve dare 1.

Esempio 4.1.7. All'interno di una CPU abbiamo due stati, **busy** (nodo 1) e **free** (nodo 2).

$$Q = \begin{pmatrix} 0,3 & 0,7 \\ 0,2 & 0,8 \end{pmatrix}$$

Figura 4.2: Catena di Markov



4.2 Calcolo Algebrico su catene di Markov

Proposizione 4.2.1. Calcolo del Marginale di una Catena di Markov Sia data la distribuzione di probabilità di trovarsi in uno stato iniziale $P(X_0 = j) \forall j$. Definiamo il vettore riga $v = (P(X_0 = j))_j$. Sia $Q = (q_{ji})_{ji}$ la matrice di transizione associata ad una catena di Markov, ovvero $q_{ji} = q_{j \rightarrow i} = P(X_1 = i | X_0 = j)$.

Allora possiamo calcolare la legge marginale $P(X_k = i)$ come:

$$P(X_k = i) = (v \cdot Q^k)_i = (v \cdot Q \cdot Q \cdot \dots \cdot Q)_i.$$

Dimostrazione. Dimostriamo per induzione. Per $k = 1$ abbiamo

$$\begin{aligned} P(X_1 = i) &= \sum_j (P(X_1 = i | X_0 = j))P(X_0 = j) = \\ &= \sum_j q_{j \rightarrow i} v_j = \sum_j v_j q_{j \rightarrow i} = (v \cdot Q)_i. \end{aligned}$$

Nel passo induttivo, supponiamo di conoscere la k -esima marginale $(P(X_k = j))_j$, che indico per comodità con il vettore $v^k = (P(X_k = j))_j$, e che, per ipotesi induttiva valga $v^k = v \cdot Q^k$. Allora per la $(k + 1)$ -esima marginale abbiamo

$$\begin{aligned} P(X_{k+1} = i) &= \sum_j (P(X_{k+1} = i | X_k = j))P(X_k = j) = \\ &= \sum_j q_{j \rightarrow i} v_j^k = \sum_j v_j^k q_{j \rightarrow i} = (v^k \cdot Q)_i = (v \cdot Q^k \cdot Q)_i = (v \cdot Q^{k+1})_i \end{aligned}$$

□

A volte, conosciamo con certezza lo stato iniziale di un sistema, ad esempio sappiamo con sicurezza che $X_0 = j$. In quel caso il vettore iniziale è dato da $v = e_j = \{0, 0, 0, \dots, 0, 1, \dots, 0, 0, \}$ dove l'unico valore 1 è in posizione j -esima). In quel caso per il calcolo della k -esima marginale si ha $P(X_k = j \mid X_0 = i) = (e_i Q^k)_j = (Q^k)_{ij}$.

Allora anche l'insieme delle $(Q^k)_{ij}$ al variare di j è una densità di probabilità e quindi deve valere $\sum_i (Q^k)_{ij} = 1$ per ogni j .

Esempio 4.2.1. Riprendendo l'esempio 4.1.1, caso c):

$$Q = \begin{pmatrix} \frac{9}{12} & \frac{3}{12} \\ \frac{10}{12} & \frac{2}{12} \end{pmatrix}$$

Distribuzione di X_0 : $v = (\frac{10}{13}, \frac{3}{13})$.

$$\text{Marginale } X_1: v \cdot Q = (\frac{10}{13}, \frac{3}{13}) \begin{pmatrix} \frac{9}{12} & \frac{3}{12} \\ \frac{10}{12} & \frac{2}{12} \end{pmatrix} = (\frac{120}{13 \cdot 12}, \frac{36}{13 \cdot 12}) = (\frac{10}{13}, \frac{3}{13})$$

$$\text{Si ha che } P(X_{10} = B) = (v \cdot Q^{10})_2 = ((v \cdot Q)Q^9)_2 = (v \cdot Q^9)_2 = \dots = (v)_2 = \frac{3}{13}$$

Calcolo dei valori attesi Data $(X)_k$ una Catena di Markov. Q matrice di transizione, $f: \mathbb{R} \rightarrow \mathbb{R}$ funzione reale, voglio calcolare $\mathbb{E}[f(X_k) \mid X_0 = i]$ (ovvero so che all'istante X_0 il sistema si trova nello stato i).

Prendiamo ad esempio il caso $k = 1$

$$\begin{aligned} \mathbb{E}[f(X_1) \mid X_0 = i] &= \sum_j f(j) \cdot P(X_1 = j \mid X_0 = i) \\ &= \sum_j f(j) \cdot q_{ij} = \sum_j q_{ij} f(j) = (Q \cdot \vec{f})_i \\ &\quad \text{dove } \vec{f} = (f(j))_j \end{aligned}$$

Proposizione 4.2.2. In generale si ha che

$$\mathbb{E}[f(X_k) \mid X_0 = i] = (Q^k \cdot \vec{f})_i$$

Se non conosco con certezza lo stato iniziale allora indichiamo, come prima, $(P(X_0) = i)_i = v$ e vale

$$\mathbb{E}[f(X_k \mid X_0)] = v \cdot Q^k \cdot \vec{f}$$

(con $(X_k \mid X_0)$ si intende la k -esima marginale conoscendo la distribuzione di probabilità iniziale, mentre con $(X_k \mid X_0 = i)$ si intende la k -esima marginale sapendo che lo stato iniziale vale con certezza i)

La dimostrazione è simile alla dimostrazione precedente, e quindi la omettiamo.

D'ora in poi indicheremo semplicemente con f il vettore \vec{f} , per alleggerire la notazione.

Esercizio 4.2.2. Prendiamo ancora l'esempio 4.1.1, caso c), e scommettiamo nel seguente modo: se esce una pallina rossa perdo 1 euro. Se esce la blu guadagno 5 euro. Diciamo che scommettiamo sulla seconda estrazione. Vediamo se conviene giocare nel caso in cui sappiamo che alla prima estrazione è uscita una pallina rossa, e nel caso in cui non conosciamo la prima estrazione.

Dobbiamo quindi calcolare, rispettivamente $\mathbb{E}[f(X_1) | X_0 = R]$ e $\mathbb{E}[f(X_1) | X_0]$. Il vettore f sarà $f = (-1, 5)$, calcoliamo intanto

$$Q \cdot f = \begin{pmatrix} \frac{9}{12} & \frac{3}{12} \\ \frac{10}{12} & \frac{2}{12} \end{pmatrix} \begin{pmatrix} -1 \\ 5 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}$$

Quindi nel caso in cui $X_0 = R$ il vettore iniziale è $v = (1, 0)$, quindi $\mathbb{E}[f(X_1) | X_0 = R] = (1, 0) \cdot Q \cdot f = 1/2$, mentre se non si conosce il risultato della prima estrazione il vettore iniziale è $v = (10/13, 3/13)$, quindi abbiamo $\mathbb{E}[f(X_1) | X_0 = R] = (10/13, 3/13) \cdot Q \cdot f = 10/26$.

Catene stazionarie e stati di equilibrio È importante lo studio degli equilibri di un sistema, ovvero degli stati per cui il sistema non varia nel tempo (e, in certe condizioni, a cui il sistema tende nel tempo anche se si parte da uno stato che non è di equilibrio). Per le catene di Markov la ricerca di questi stati non è difficile.

Definizione 4.2.1. Distribuzione invariante Sia data una catena di Markov omogenea, con matrice di transizione Q . La **distribuzione invariante** per tale catena è un vettore $\mu = (\mu_i)_i$ tale che

$$\begin{cases} \mu_i \geq 0 \\ \sum_i \mu_i = 1 \\ \vec{\mu} \cdot Q = \vec{\mu} \end{cases}$$

Nota. Si osservi che μ^τ è l'autovettore di Q^τ riferito all'autovalore 1, quindi se una distribuzione invariante per Q esiste, basta risolvere $(Q^\tau - Id)\mu^\tau = 0$.

Definizione 4.2.2. Catena Stazionaria

Una Catena di Markov $(X_k)_k$ è una **catena stazionaria** se **tutte** le sue marginali sono uguali, ovvero se $P(X_k = i) = pX_0 = i$ per ogni i, k .

Se la catena $(X_k)_k$ è stazionaria, se prendiamo il vettore della prima marginale $\mu = (pX_0 = i)_i$ abbiamo che $\mu \cdot Q^k = \mu$ per ogni k . Ma allora $\mu = (pX_0 = i)_i$ è una distribuzione invariante per Q . Viceversa, se abbiamo una distribuzione invariante μ per una certa catena di Markov con matrice di transizione Q , allora tale catena è stazionaria, infatti

$$\mu \cdot Q^k = (\mu \cdot Q) \cdot Q^{k-1} = \mu \cdot Q^{k-1} = \dots = \mu \cdot Q = \mu.$$

Quindi

$$(X_k)_k \text{ è una catena di Markov stazionaria} \Leftrightarrow (pX_0 = i)_i \text{ è una sua distribuzione invariante}$$

Esempio 4.2.3. Abbiamo visto nell'esempio 4.2.1 che il vettore $(\frac{10}{13}, \frac{3}{13})$ è una distribuzione invariante per la catena di Markov dell'esempio 4.1.1, caso c). Siamo stati fortunati o questo processo è una catena invariante indipendentemente dal numero di biglie blu e rosse?

Supponiamo di avere quindi $N = B + R$ biglie totali, di cui R Rosse e B Blu. Il primo nodo sono le biglie rosse ed il secondo quelle blu, al solito. Le probabilità di transizione, e la matrice

Q saranno quindi

$$\begin{aligned} q_{1,1} &= P(X_1 = R \mid X_0 = R) = \frac{R-1}{N-1} \\ q_{1,2} &= P(X_1 = B \mid X_0 = R) = \frac{B}{N-1} \\ q_{2,1} &= P(X_1 = R \mid X_0 = B) = \frac{R}{N-1} \\ q_{2,2} &= P(X_1 = B \mid X_0 = B) = \frac{B-1}{N-1} \\ Q &= \begin{pmatrix} \frac{R-1}{N-1} & \frac{B}{N-1} \\ \frac{R}{N-1} & \frac{B-1}{N-1} \end{pmatrix} \end{aligned}$$

Verifichiamo che lo stato iniziale $\mu = \left(\frac{R}{N}, \frac{B}{N}\right)$ è una distribuzione invariante.

$$\begin{aligned} \mu Q &= \left(\frac{R}{N}, \frac{B}{N}\right) \begin{pmatrix} \frac{R-1}{N-1} & \frac{B}{N-1} \\ \frac{R}{N-1} & \frac{B-1}{N-1} \end{pmatrix} \\ &= \left(\frac{R(R+B-1)}{N(N-1)}, \frac{B(R+B-1)}{N(N-1)}\right) = \left(\frac{R}{N}, \frac{B}{N}\right) \end{aligned}$$

Definizione 4.2.3. Matrice di Transizione regolare Una matrice di transizione Q si dice regolare se per qualche k si ha che $(Q^k)_{ij} > 0$.

Se Q è regolare e v è uno stato iniziale qualsiasi allora $v \cdot Q^k$ tende alla distribuzione invariante per la catena di Markov. Quindi, in questo caso, da qualsiasi stato iniziale si converge verso una distribuzione che possiamo calcolare a priori. Quindi il comportamento di questo sistema diventa prevedibile su tempi lunghi.

Esempio 4.2.4. All'interno di una CPU abbiamo due stati, **busy** (nodo 1) e **free** (nodo 2).

$$Q = \begin{pmatrix} 0,3 & 0,7 \\ 0,2 & 0,8 \end{pmatrix}$$

Cerco μ distribuzione invariante. Sappiamo che μ è autovettore di autovalore 1:

$$\begin{aligned} \mu Q = \mu &\iff Q^T \mu^T = \mu^T \iff (Q^T - I) \mu^T = 0 \\ Q^T - I &= \begin{pmatrix} 0,3 & 0,2 \\ 0,7 & 0,8 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -0,7 & 0,2 \\ 0,7 & -0,2 \end{pmatrix} \\ \begin{pmatrix} -0,7 & 0,2 \\ 0,7 & -0,2 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = 0 &\implies \begin{cases} -0,7\mu_1 + 0,2\mu_2 = 0 \\ 0,7\mu_1 - 0,2\mu_2 = 0 \\ \mu_1 + \mu_2 = 1 \\ \mu_1 \geq 0; \mu_2 \geq 0 \end{cases} \end{aligned}$$

Risolvendo il sistema si ottiene

$$\mu = \left(\frac{2}{9}, \frac{7}{9}\right)$$

Si noti che la matrice di transizione dell'esercizio precedente è regolare. Quindi se partiamo da un qualsiasi stato, per tempi molto lunghi lo stato del sistema si avvicinerà alla distribuzione invariante. In altre parole, dopo abbastanza tempo la probabilità di trovare la CPU occupata sarà circa di $2/9$, mentre la probabilità di trovarla occupata sarà circa di $7/9$.

4.3 Esercizi

Esercizio 4.3.1. Passeggiata Aleatoria Mi muovo nell'asse X casualmente partendo da 0. Al minuto k lancio una moneta. Se esce testa mi muovo a destra, se esce croce mi muovo a sinistra. Voglio ottenere la posizione al minuto k . Disegnare la catena di Markov come grafo.

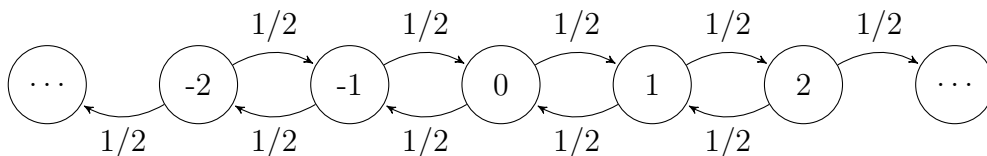
$$Y_k = \text{Bern} \left(\frac{1}{2} \right) \quad (\text{lancio della moneta})$$

$$Y_k \in \{-1, +1\}$$

$$X_k = \text{posizione}$$

$$\begin{cases} X_0 = 0 \\ X_{k+1} = X_k + Y_k \end{cases}$$

Figura 4.3: Catena di Markov della passeggiata aleatoria



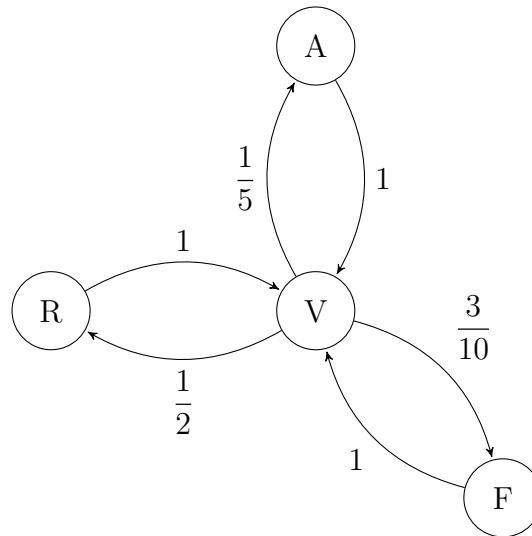
Esercizio 4.3.2. Invece di prendere una passeggiata aleatoria, immaginiamo di voler schematizzare una passeggiata di un ubriaco che si muove senza meta, quindi in modo aleatorio, ma che è restio a cambiare direzione. Quindi se all'istante k è andato a sinistra, all'istante $k+1$ la direzione sinistra sarà più probabile della destra. La sua posizione è una catena di Markov?

No, perché dipende dalla posizione all'istante precedente e dalla direzione. La scelta di andare a destra o sinistra, invece, può essere rappresentata da una Catena di Markov, perché dipende soltanto dall'ultima scelta fatta.

Esercizio 4.3.3. Vogliamo simulare un essere vivente elementare con un automa. I suoi stati sono (1) **relax**, (2) **vigile**, (3) **fuga**, (4) **attacca**. Sappiamo che in presenza di qualche segnale esterno dallo stato di relax l'animale passa allo stato vigile, da qui può tornare allo stato di relax (con probabilità $1/2$), oppure decidere di attaccare (con probabilità $1/5$) o fuggire (con probabilità $3/10$). Scrivere il grafo e la matrice di transizione legata a questo processo.

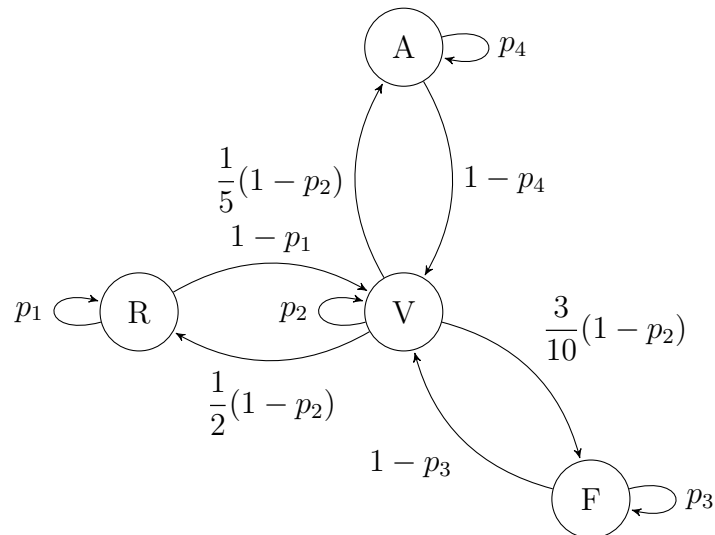
Questo modello non è del tutto soddisfacente, perché, ad esempio, il nostro automa uscirà dallo stato di relax a intervalli prestabiliti. In natura invece il cambiamento di stato di un animale dipende dagli stimoli esterni, che non sono ad intervalli regolari. Una maniera di introdurre quindi un tempo in una catena di Markov che non coincida necessariamente con gli intervalli a cui si studia il processo, si può introdurre per ogni stato i una probabilità p_i che lo stato vada in se stesso. Provare a scrivere la matrice di transizione per l'automa così modificato.

Figura 4.4: Catena di Markov dell'automa cellulare



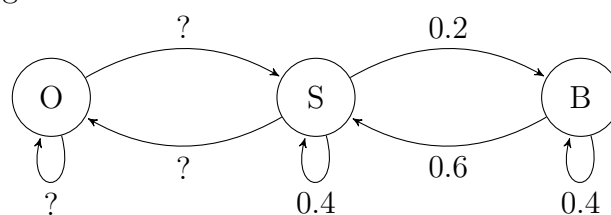
$$Q = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 3 & 1 \\ \frac{1}{2} & 0 & \frac{10}{10} & \frac{1}{5} \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Figura 4.5: Catena di Markov dell'automa cellulare con tempo



Esercizio 4.3.4. Abbiamo una CPU con 3 stati: (1) **Off**, (2) **Stand By**, (3) **Busy**

Figura 4.6: Catena di Markov della CPU a 3 stati



1. Completare gli archi del grafo. Si possono completare sapendo che la somma degli archi uscenti da un nodo dev'essere 1. La matrice Q sarà

$$Q = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.2 & 0.4 & 0.4 \\ 0 & 0.6 & 0.4 \end{pmatrix}$$

2. Calcolare $P(X_1 = O \mid X_0 = O)$ e $P(X_2 = O \mid X_0 = O)$. Abbiamo $P(X_1 = O \mid X_0 = O) = q_{00} = 0.8$. Calcoliamo ora $P(X_2 = O \mid X_0 = O)$. Il vettore corrispondente allo stato iniziale $X_0 = O$ è $v = (1, 0, 0)$, quindi

$$P(X_2 = O \mid X_0 = O) = (v \cdot Q \cdot Q)_1 = (0.68, \dots, \dots)_1 = 0.68$$

3. Supponiamo di avere una funzione costo tale che $c(O) = 0$, $c(S) = 5$ e $c(B) = 10$. Calcolare $\mathbb{E}[c(X_k) \mid X_0 = O]$ per $k = 1, 2$. Per $k = 1$ possiamo procedere in modo elementare

$$\begin{aligned} \mathbb{E}[c(X_1) \mid X_0 = O] &= 0 \cdot P(X_1 = O \mid X_0 = O) \\ &+ 5 \cdot P(X_1 = S \mid X_0 = O) + 10 \cdot P(X_1 = B \mid X_0 = O) \\ &= 5 \cdot P(X_1 = S \mid X_0 = O) = 5 \cdot 0.2 = 1 \end{aligned}$$

Per $k = 2$ invece conviene usare la formula vista precedentemente. Il vettore associato al costo è $f = (0, 5, 10)$ quindi

$$\mathbb{E}[c(X_2) \mid X_0 = O] = (Q^2 \cdot f)_1 = v \cdot Q^2 \cdot f = 2$$

4. Calcolare la varianza $\text{Var}(c(X_1) \mid X_0 = O)$. Abbiamo

$$\begin{aligned} c^2 &= (0, 25, 100) \\ \mathbb{E}[c^2(X_1 \mid X_0 = O)] &= 0 \cdot P(X_1 = O \mid X_0 = O) \\ &+ 25 \cdot P(X_1 = S \mid X_0 = O) + 100 \cdot P(X_1 = B \mid X_0 = O) \\ &= 25 \cdot 0.2 = 5 \end{aligned}$$

quindi

$$\text{Var}(c(X_1) \mid X_0 = O) = \mathbb{E}[c^2(X_1 \mid X_0 = O)] - (\mathbb{E}[c(X_1) \mid X_0 = O])^2 = 5 - 1 = 4$$

5. Calcolare μ distribuzione invariante e $\mathbb{E}[c(X_1) | \mu]$

(Calcoliamo μ)

$$(Q^T - I)\mu^T = 0$$

$$\mu = (\mu_1, \mu_2, \mu_3)$$

$$(Q^T - I)\mu^T = \begin{pmatrix} -0.2 & 0.2 & 0 \\ 0.2 & -0.6 & 0.6 \\ 0 & 0.4 & -0.6 \end{pmatrix} \mu^T = 0$$

$$\Rightarrow \begin{cases} -0.2\mu_1 + 0.2\mu_2 = 0 \\ 0.2\mu_1 - 0.6\mu_2 + 0.6\mu_3 = 0 \\ 0.4\mu_2 - 0.6\mu_3 = 0 \\ \mu_1 + \mu_2 + \mu_3 = 1 \end{cases} \Rightarrow \begin{cases} \mu_1 = \mu_2 \\ \mu_3 = \frac{2}{3}\mu_2 \\ \mu_1 + \mu_2 + \mu_3 = 1 \end{cases}$$

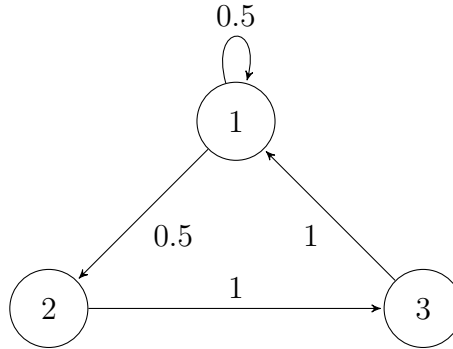
$$\Rightarrow \mu = \left(\frac{3}{8}, \frac{3}{8}, \frac{1}{4}\right)$$

(Calcoliamo $\mathbb{E}[c(X_1) | \mu]$)

$$\mathbb{E}[c(X_1) | \mu] = \mu \cdot Q \cdot f = 4.375$$

Esercizio 4.3.5. Dato il grafo di una Catena di Markov

Figura 4.7: Catena di Markov



1. Trovare Q matrice di transizione

$$Q = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

2. Scrivere le marginali X_1, X_2, X_3 sapendo che $X_0 = 1$

$$v = (1, 0, 0)$$

$$P(X_1 | X_0 = 1) = v \cdot Q = \left(\frac{1}{2}, \frac{1}{2}, 0\right)$$

$$P(X_2 | X_0 = 1) = v \cdot Q^2 = \left(\frac{1}{2}, \frac{1}{2}, 0\right) Q = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right)$$

$$P(X_3 | X_0 = 1) = v \cdot Q^3 = \left(\frac{1}{2}, \frac{1}{2}, 0\right) Q^2 = \left(\frac{5}{8}, \frac{1}{8}, \frac{1}{4}\right)$$

3. Calcolare la distribuzione invariante μ e $\mathbb{E}[\mu]$

$$(Q^T - I)\mu^T = \begin{pmatrix} -0.5 & 0 & 1 \\ 0.5 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}$$

$$\Rightarrow \begin{cases} -\mu_1 + 2\mu_3 = 0 \\ \mu_1 - 2\mu_2 = 0 \\ 2\mu_2 - 2\mu_3 = 0 \end{cases} \Rightarrow \begin{cases} \mu_1 = 2\mu_3 \\ \mu_1 = 2\mu_2 \\ \mu_1 + \mu_2 + \mu_3 = 1 \end{cases} \Rightarrow \mu = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$$

$$\Leftrightarrow \begin{cases} P(X_0 = 1) = \frac{1}{2} \\ P(X_0 = 2) = \frac{1}{4} \\ P(X_0 = 3) = \frac{1}{4} \end{cases}$$

$$\mathbb{E}[\mu] = 1 \cdot P(X_0 = 1) + 2 \cdot P(X_0 = 2) + 3 \cdot P(X_0 = 3) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} = \frac{7}{4}$$

4. Se la catena è stazionaria calcolare $P(X_1 = 1 \mid X_3 = 1)$. Utilizziamo la formula di Bayes.

$$P(X_1 = 1 \mid X_3 = 1) = P(X_3 = 1 \mid X_1 = 1) \cdot \frac{P(X_1 = 1)}{P(X_3 = 1)}$$

$$\text{La catena è stazionaria: } P(X_3 = 1) = P(X_1 = 1)$$

$$\Rightarrow P(X_1 = 1 \mid X_3 = 1) = P(X_3 = 1 \mid X_1 = 1)$$

$$= P(X_2 = 1 \mid X_0 = 1) = (1, 0, 0)Q^2 = \frac{1}{2}$$

Parte II

Variabili Aleatorie con Densità e
Introduzione alla Statistica
Lezioni di Maurizio Pratelli

Capitolo 5

Variabili Aleatorie con densità (continue)

5.1 Svolgere integrali multipli

Definizione 5.1.1. Integrale Multiplo

Un integrale multiplo è un integrale definito di una funzione a più di una variabile reale. Gli integrali di una funzione $f(x, y)$ su una regione di \mathbb{R}^2 sono detti **integrali doppi**, quelli di una funzione $f(x, y, z)$ su una regione di \mathbb{R}^3 sono detti **integrali tripli**.

Proprio come l'integrale definito di una funzione positiva ad una variabile rappresenta l'area della regione sottostante al grafico della funzione, l'integrale doppio di una funzione positiva a due variabili reali rappresenta il volume della regione fra la superficie definita dalla funzione (sul piano cartesiano tridimensionale dove $z = f(x, y)$) e il piano che contiene il dominio dell'integrale. Se sono presenti più variabili, un integrale multiplo produrrà ipervolumi di funzioni multidimensionali. L'integrazione multipla di una funzione ad n variabili $f(x_1, x_2, \dots, x_n)$ su un dominio D è rappresentata da segni di integrale nidificati, nell'ordine inverso di esecuzione (l'integrale più a sinistra è calcolato per ultimo), seguita dalla funzione e i differenziali nell'ordine proprio (il segno di integrale più a sinistra corrisponde al differenziale più a destra). Il dominio di integrazione è rappresentato su ogni integrale, o simbolicamente abbreviato attraverso un simbolo di variabile sull'integrale più interno.

Definizione 5.1.2. Regola pratica per gli integrali doppi

$$\iint_{\mathbb{R}^2} f(x, y) dx dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx = \int_{\mathbb{R}} dx \int_{\mathbb{R}} f(x, y) dy = \int_{\mathbb{R}} dy \int_{\mathbb{R}} f(x, y) dx$$

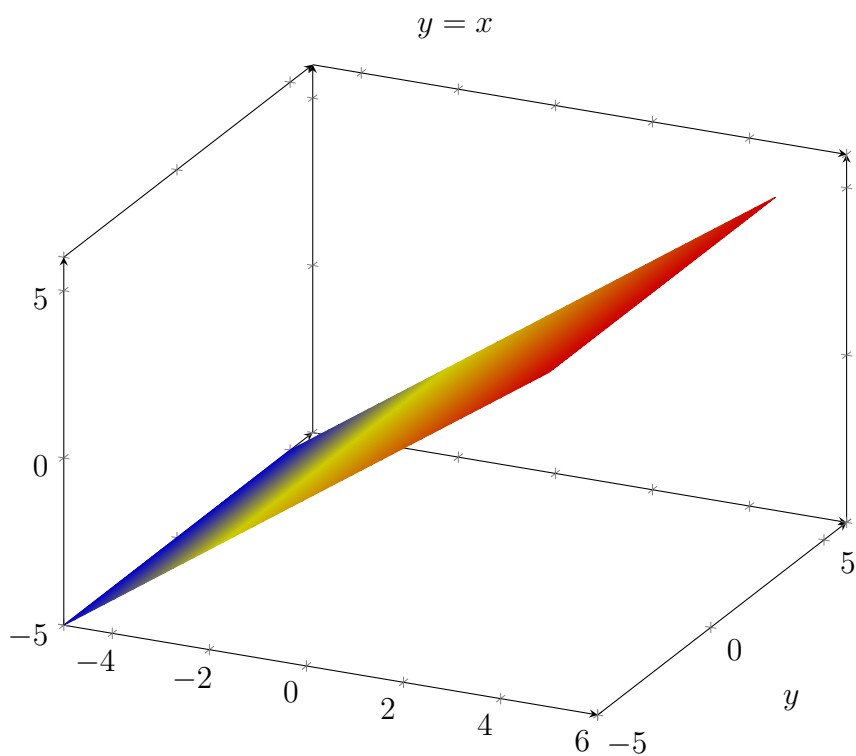
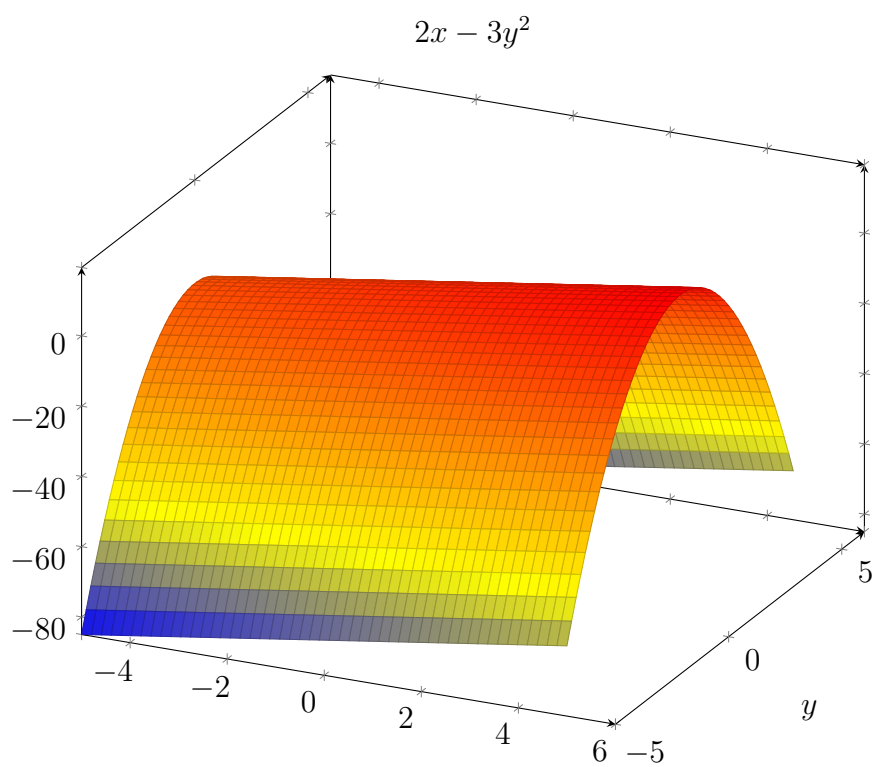
Esempio 5.1.1. Svolgiamo l'integrale doppio sul triangolo $T = 0 \leq y \leq x \leq 1$ di $f(x, y) = x$

$$\iint_T x dy dx = \int_0^1 \int_0^x x dy dx = \int_0^1 dx \int_0^x x dy = \int_0^1 x^2 dx = \frac{1}{3}$$

Che coincide, scambiando l'ordine di integrazione con

$$\int_0^1 dy \int_y^1 x dx = \frac{1}{2} \int_0^1 (1 - y^2) dy = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}$$

Esempio 5.1.2. Svolgiamo l'integrale doppio $\iint_R 2x - 3y^2 dx dy$ sul rettangolo $R : -1 \leq x \leq 1, 0 \leq y \leq 2$.

Figura 5.1: Grafico della funzione $y = x$ su \mathbb{R}^2 Figura 5.2: Grafico della funzione reale a due variabili $2x - 3y^2$

$$\begin{aligned}\iint_R 2x - 3y^2 dx dy &= \int_0^2 \left(\int_{-1}^1 2x - 3y^2 dx \right) dy = \\ &= \int_0^2 -6y^2 dy = \left(-6 * \frac{8}{3} \right) = -16\end{aligned}$$

5.2 Introduzione alle Variabili Aleatorie con Densità

Definizione 5.2.1. Variabile Aleatoria Abbiamo visto nella parte precedente del corso le variabili aleatorie. Una variabile aleatoria è una funzione definita come

$$X : (\Omega, \mathbb{F}, P) \rightarrow \mathbb{R}$$

Permette di trasportare le probabilità dai sottoinsiemi di Ω ai sottoinsiemi di \mathbb{R} , ovvero

$$\begin{aligned}P_X(A) &= P(X^{-1}(A)) \\ A &\subseteq \mathbb{R} \\ X^{-1}(A) &\subseteq \Omega\end{aligned}$$

Le variabili aleatorie possono essere di **discrete**, **con densità** o **più generali**. Una variabile aleatoria è discreta se la sua immagine è finita o numerabile, ovvero

$$p(x_i) = P(X = x_i)$$

Definizione 5.2.2. Variabile Aleatoria con Densità

X ha densità se esiste una funzione $f : \mathbb{R} \rightarrow [0, +\infty)$ (la densità) integrabile, tale che

$$\begin{aligned}P(X \in A) &= \int_A f(x) dx \\ P(a \leq X \leq b) &= \int_a^b f(x) dx\end{aligned}$$

Perché f sia densità una condizione necessaria è che $\int_{\Omega} f(x) dx = 1$, dove in generale, $\Omega = \mathbb{R}$ se la funzione di densità è ad una variabile.

Il simbolo tilde (\sim) si usa per indicare che una variabile aleatoria ha una data distribuzione di probabilità, ad esempio $X \sim \Gamma(k, \theta)$

Definizione 5.2.3. Funzione di ripartizione (cumulative distribution function)

È una funzione definita come $F : \mathbb{R} \rightarrow [0, 1]$ È definita come

$$F_X(x) = P(X \leq x)$$

Se ne inferisce che la probabilità che la variabile aleatoria con densità X risieda nell'intervallo semichiuso $(a, b]$ con $a \leq b$ è quindi

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$$

Sulle **variabili discrete** è tipicamente discontinua, definita come

$$F_X(t) = \sum_{x_i \leq t} P(X = x_i) = \sum_{x_i \leq t} p(x_i)$$

Funzione di ripartizione su variabili con densità Sulle **variabili con densità** è tipicamente continua e può essere espressa come l'integrale della funzione di densità di probabilità della variabile come segue:

$$F_X(t) = \int_{-\infty}^t f(x)dx$$

Definizione 5.2.4. Proprietà di una funzione di ripartizione

- $s < t \implies F_X(s) \leq F_X(t)$
- $\lim_{x \rightarrow -\infty} F(x) = 0$
 $\lim_{x \rightarrow +\infty} F(x) = 1$
- $F(x) = \lim_{y \rightarrow x^+} F(y)$

5.3 Proprietà principali delle variabili aleatorie con densità

Definizione 5.3.1. Indipendenza di variabili con densità

La definizione di indipendenza di due variabili con densità è identica al caso in cui le variabili siano discrete, ma dev'essere data parlando solo di sottoinsiemi misurabili di \mathbb{R} . Due variabili aleatorie con densità si dicono tra loro indipendenti se, \forall coppia A, B di sottoinsiemi *misurabili* di \mathbb{R} , risulta:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Si osserva facilmente che due variabili aleatorie X, Y sono indipendenti $\iff \forall$ coppia di numeri reali s, t , risulta

$$P(X \leq s, Y \leq t) = F_X(s)F_Y(t)$$

Dove F_X e F_Y sono le rispettive funzioni di ripartizione di X e Y .

Definizione 5.3.2. Assenza di memoria delle variabili aleatorie con densità esponenziale

$$P(X > 1 + t \mid X > 1) = P(X > t)$$

Dimostrazione.

$$\begin{aligned} P(X > t) &= P(X \in [t, +\infty)) = \int_t^{+\infty} \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x} \Big|_t^{+\infty} = e^{-\lambda t} \\ \frac{P(X > t+1 \cap X > 1)}{P(X > 1)} &= \frac{e^{-\lambda(t+1)}}{e^{-\lambda}} = e^{-\lambda t} \end{aligned}$$

□

Definizione 5.3.3. Formula della convoluzione

La **convoluzione** è un'operazione fra due funzioni ad una variabile che consiste nell'integrare il prodotto tra la prima e la seconda traslata un certo valore. In probabilità si usa la convoluzione per ottenere la distribuzione di una somma di due variabili aleatorie con densità.

$$(f \star g)(x) = \int_{\mathbb{R}} f(x-y)g(y)dy$$

Siano date due variabili aleatorie: X con densità $f(x)$ e Y con densità $g(x)$. Sia data la coppia di v.a. (X, Y) con densità $g(x)$. ne ottiene che:

$$X \text{ ha densità } f(x) = \int_{-\infty}^{+\infty} h(x, y)dy; \quad Y \text{ ha densità } g(x) = \int_{-\infty}^{+\infty} h(x, y)dx$$

Se X e Y sono indipendenti e $Z = X + Y$ allora, la densità di Z detta $a(z)$ sarà data dalla formula di convoluzione:

$$a(z) = (g \star f)(z) = \int_{-\infty}^{+\infty} f(x)g(z-x)dx = (f \star g)(z) = \int_{-\infty}^{+\infty} g(y)f(z-y)dy$$

Viene rispettata la commutatività della somma.

Dimostrazione.

$$P(X \leq x) = P(X \leq x, -\infty \leq Y \leq +\infty)$$

La variabile Y è libera mentre la X è limitata dal valore di x . Ne segue che:

$$P(X \leq x, -\infty \leq Y \leq +\infty) = \int_{-\infty}^x \int h(t, y)dydt$$

$$f(x) = \frac{d}{dx} \int_{-\infty}^x \int h(t, y)dydt = \int_{-\infty}^{+\infty} h(x, y)dy$$

Per quanto riguarda $Z = X + Y$ si ha che:

$$P(X + Y \leq Z) = \int_{-\infty}^{+\infty} f(x) \left(\int_{-\infty}^{z-x} g(y)dy \right) dx = A(z)$$

$$a(z) = \int_{-\infty}^{+\infty} f(x)g(z-x)dx$$

□

Definizione 5.3.4. Prodotto di variabili aleatorie con densità a valori positivi

Siano date X e Y variabili aleatorie con corrispondenti densità $f(x)$ e $g(y)$ a valori positivi e sia $Z = XY$ il prodotto di esse. Per calcolare la densità $a(z)$ di Z , si calcola prima la funzione di ripartizione $A(z)$.

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \text{qualche valore} & x > 0 \end{cases}$$

$$g(y) = \begin{cases} 0 & y \leq 0 \\ \text{qualche valore} & y > 0 \end{cases}$$

$$A(z) = P(Z \leq z) = \begin{cases} 0 & z \leq 0 \\ \text{qualche valore} & z > 0 \end{cases}$$

La definizione precedente di $A(z)$ ha senso perché X e Y assumono valori positivi. Altrimenti, avrei dovuto distinguere il caso $z \leq 0 \implies P(Z \geq z) = 1 - A(z)$. Si ha quindi che (per $z > 0$)

$$A(z) = P(XY \leq z) = \int_0^{+\infty} dx \int_0^{\frac{z}{x}} f(x)g(y)dy$$

Si nota che z compare in questo integrale. Deriviamo adesso $G(z)$

$$g(z) = \frac{dG(z)}{dz} = \int_0^{+\infty} \frac{1}{x} f(x) g\left(\frac{z}{x}\right) dy$$

Definizione 5.3.5. Speranza di variabili aleatorie con densità

Si dice che X ha speranza matematica finita (o valore atteso) se

$$\int_{-\infty}^{+\infty} |x| f(x) dx < +\infty$$

E si dice **speranza matematica** di X il numero

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

Definizione 5.3.6. Momenti di una variabile aleatoria con densità

Il momento k -esimo di una v.a. con densità, se finito, è dato da

$$\begin{aligned} \mathbb{E}[X^k] &= \int_{-\infty}^{+\infty} x^k f(x) dx \\ \exists \mathbb{E}[X^k] &\implies \exists \mathbb{E}[X^m] \quad 1 \leq m < k \end{aligned}$$

Definizione 5.3.7. Proprietà della speranza

- **Regola della somma (linearità):**

$$\begin{aligned} \mathbb{E}[X + Y] &= \mathbb{E}[X] + \mathbb{E}[Y] \\ \mathbb{E}[\alpha X] &= \alpha \mathbb{E}[X] \end{aligned}$$

- **Regola del prodotto:** Se X e Y sono indipendenti allora

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

- **Monotonia:**

$$X \geq 0 \wedge \mathbb{E}[X] \geq 0 \wedge X \leq Y \implies \mathbb{E}[X] < \mathbb{E}[Y]$$

Dimostrazione. Dimostriamo la monotonia del valore atteso di una variabile aleatoria con densità che assume solo valori positivi:

$$\begin{aligned} X \geq 0 &\implies f(x) = \begin{cases} 0 & x \leq 0 \\ \text{qualche valore} & x > 0 \end{cases} \implies \exists \mathbb{E}[X] \\ \mathbb{E}[X] &= \int_{-\infty}^0 x f(x) dx + \int_0^{+\infty} x f(x) dx \geq 0 \end{aligned}$$

□

Definizione 5.3.8. Varianza di una variabile aleatoria con densità

Se X ha momento secondo ($\mathbb{E}[X^2]$) finito, allora si definisce, come nel caso discreto, la varianza

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])^2 f(x) dx = \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_{-\infty}^{+\infty} x^2 f(x) dx - \left(\int_{-\infty}^{+\infty} x f(x) dx \right)^2 \end{aligned}$$

Definizione 5.3.9. Covarianza

La covarianza è un numero che fornisce una misura di quanto due variabili aleatorie varino insieme, ovvero della loro dipendenza:

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

Due variabili aleatorie hanno covarianza = 0 se sono indipendenti. Si nota anche che

$$\begin{aligned}\text{Covar}(X, X) &= \text{Var}(X) \\ \text{Var}(X + Y) &= \text{Var}(X) + \text{Var}(Y) + 2\text{Covar}(X, Y)\end{aligned}$$

Definizione 5.3.10. Disuguaglianze di Markov e Chebishev

Le disuguaglianze di Markov e Chebishev sono identiche nel caso due variabili aleatorie siano discrete o con densità. Dimostriamo la disuguaglianza di Markov.

Proposizione 5.3.1.

$$\begin{aligned}P(X > a) &\leq \frac{\mathbb{E}[X]}{a} \\ \mathbb{E}[X] &\geq aP(X > a)\end{aligned}$$

Dimostrazione.

$$\begin{aligned}X \geq 0 \wedge a > 0 &\implies P(X > a) a < \mathbb{E}[X] \\ X \geq 0 &\implies f(x) = \begin{cases} 0 & x \leq 0 \\ \text{qualche valore} & x > 0 \end{cases} \implies \exists \mathbb{E}[X] \\ \mathbb{E}[X] &= \int_0^{+\infty} x f(x) dx = \int_0^a x f(x) dx + \int_a^{+\infty} x f(x) dx \\ &\geq \int_a^{+\infty} x f(x) dx \geq \int_a^{+\infty} a f(x) dx\end{aligned}$$

Nell'integrale, x varia fra a e $+\infty$, se al posto di x sostituisco a , se ne ottiene una quantità minore poiché a è il minimo valore che la variabile x sulla quale integriamo assume.

$$= a \int_a^{+\infty} f(x) dx = aP(X > a)$$

□

5.4 Principali tipologie di densità

Definizione 5.4.1. Densità uniforme su $[a, b]$

Sia data X con densità $f(x)$

$$f(x; a, b) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{altrimenti} \end{cases}$$

Il valore atteso è $\frac{a+b}{2}$. La varianza è $\frac{(b-a)^2}{12}$

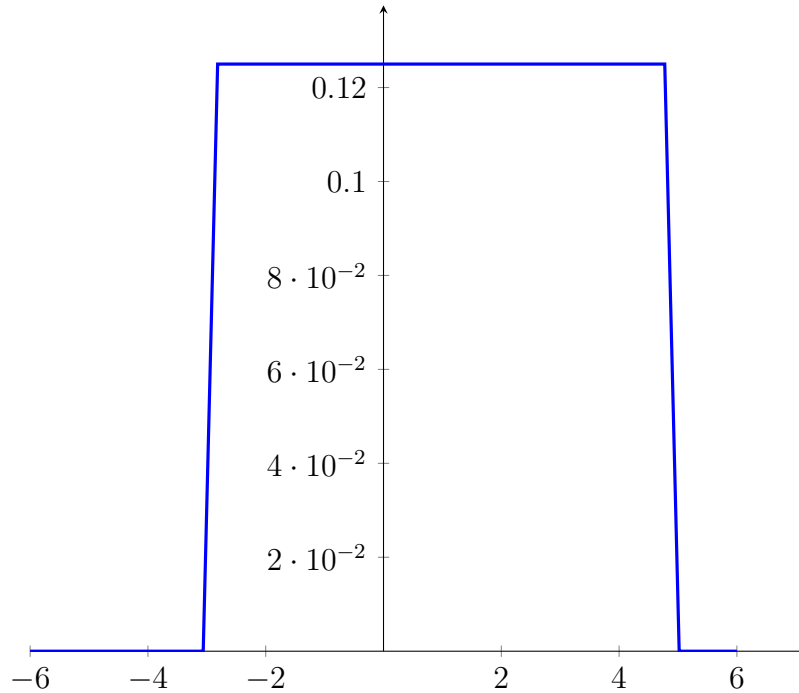


Figura 5.3: Densità di probabilità uniforme di parametri $a = -3$ e $b = 5$

La funzione di ripartizione F è definita come

$$F(x) = \int_{-\infty}^x f(t)dt = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

F è la funzione di ripartizione in tutti i punti in cui f è continua

$$f(x) = \frac{dF(x)}{dx}$$

Definizione 5.4.2. Densità esponenziale di parametro $\lambda > 0$

Sia data la densità

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Il valore atteso è $\frac{1}{\lambda}$. La speranza è $\frac{1}{\lambda^2}$

La funzione di ripartizione F è

$$F(x) = \int_{-\infty}^x f(t)dt = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$

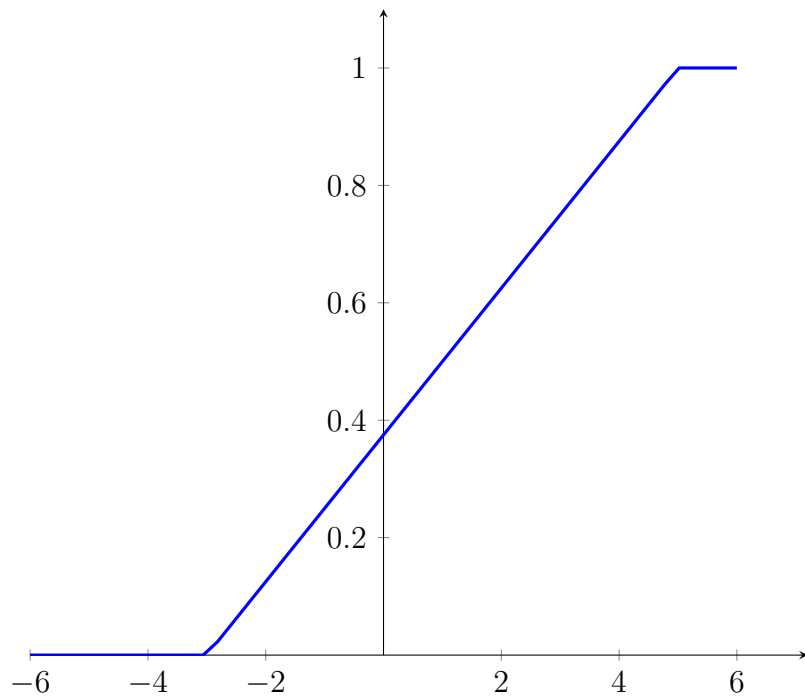


Figura 5.4: Funzione di ripartizione della densità di probabilità uniforme di parametri $a = -3$ e $b = 5$

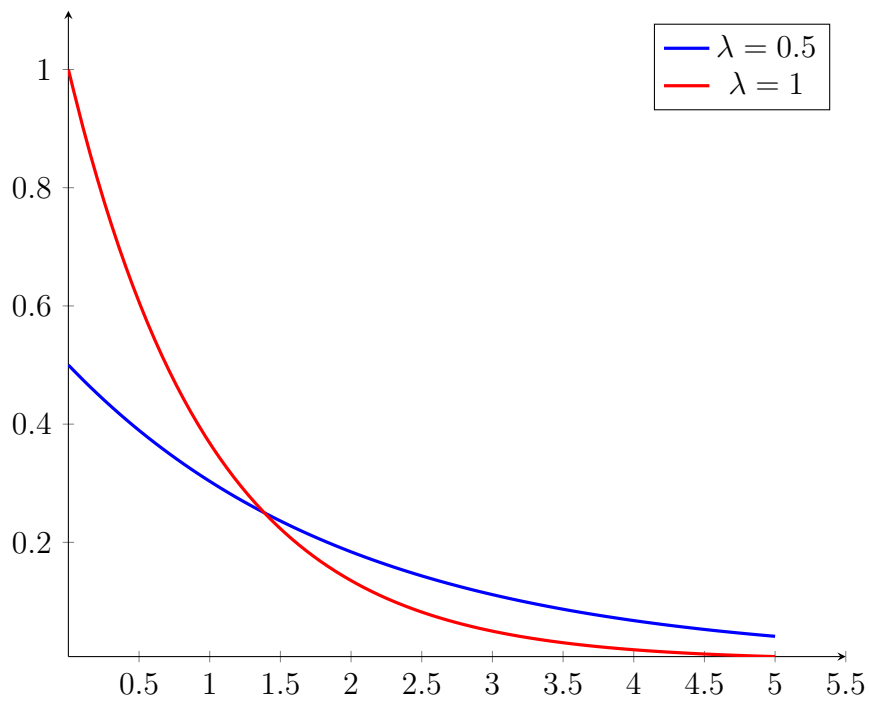


Figura 5.5: Densità di probabilità esponenziale

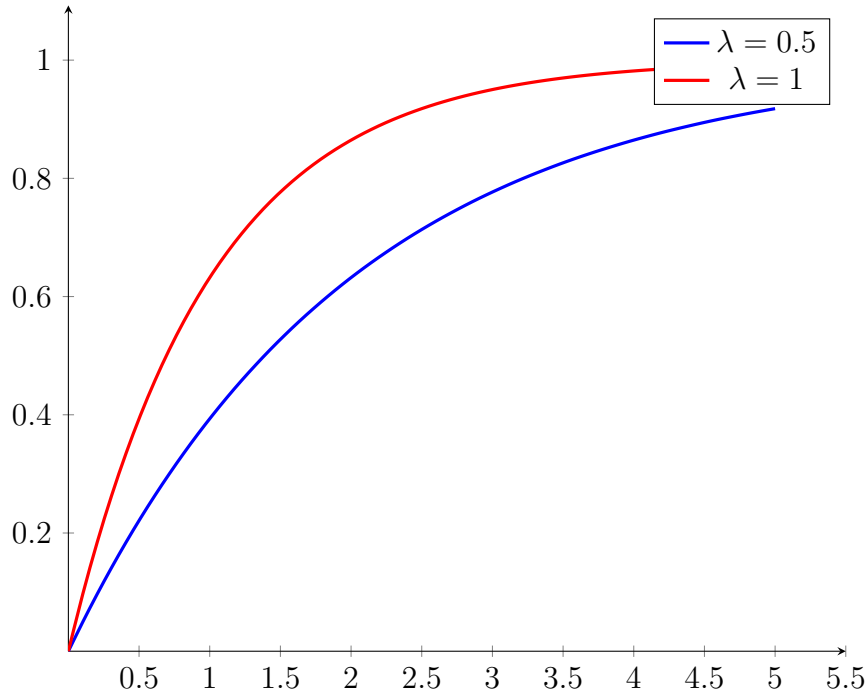


Figura 5.6: $P(X \leq x)$ (funzione di ripartizione) di un'esponenziale di parametro $\lambda = 0.5$ (blu) e $\lambda = 1$ (rosso)

Nota. Se X ha densità

$$P(X = x) = \int_x f(t) dt = 0$$

Definizione 5.4.3. Densità di Cauchy

La densità di Cauchy (detta anche distribuzione di Cauchy o distribuzione di Lorentz) è una particolare funzione di densità che descrive nel piano euclideo l'intersezione tra l'asse delle ascisse ed una retta passante per un punto fissato ed inclinata ad un angolo che segue la distribuzione continua uniforme. I momenti di una distribuzione di Cauchy non sono definiti.

$$f(x) = \frac{1}{\pi(1+x^2)}$$

Proposizione 5.4.1. *La distribuzione di Cauchy è una densità di probabilità.*

Dimostrazione.

$$\begin{aligned} \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1}{1+x^2} dx &= \frac{1}{\pi} \lim_{M \rightarrow +\infty} \int_{-M}^M \frac{dx}{1+x^2} \\ &= \frac{1}{\pi} \lim_{M \rightarrow +\infty} (\arctan(M) - \arctan(-M)) = \frac{1}{\pi} \left(\frac{\pi}{2} + \frac{\pi}{2} \right) = 1 \end{aligned}$$

□

Proposizione 5.4.2. *La densità di Cauchy non ha momenti.*

Dimostrazione.

$$\begin{aligned} \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{|x|}{1+x^2} dx &= \frac{1}{\pi} \int_0^{+\infty} \frac{2x}{1+x^2} dx = \frac{1}{\pi} [\log(1+x^2)]_0^{+\infty} = -\frac{1}{\pi} \cdot 0 + \frac{1}{\pi} \cdot +\infty = +\infty \\ &\Rightarrow \nexists \mathbb{E}[X] \end{aligned}$$

□

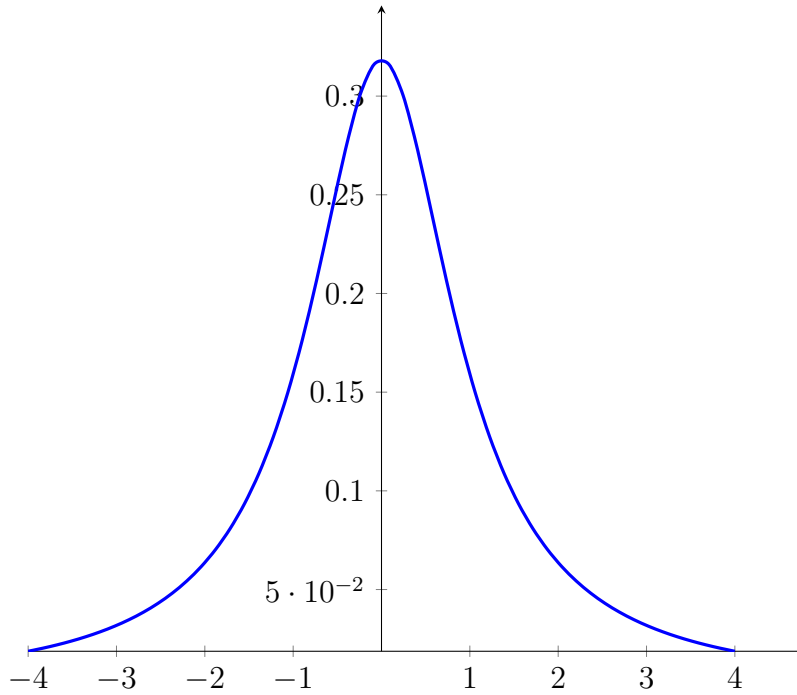


Figura 5.7: Densità di probabilità di Cauchy

Definizione 5.4.4. Funzione Gamma Γ

La funzione Γ è un'estensione della funzione fattoriale ai numeri complessi. È definita su tutti i numeri complessi con parte reale > 0 . Per definire la **densità di probabilità** Γ , utilizzeremo la funzione Γ definita solo per i numeri reali positivi.

$$r > 0 \implies \Gamma(r) = \int_0^{+\infty} x^{r-1} e^{-x} dx$$

Proposizione 5.4.3. *Sui reali positivi, la funzione $\Gamma(r+1) = r\Gamma(r)$. Sui numeri naturali, allora $\Gamma(n+1) = n!$.*

Dimostrazione.

$$\begin{aligned} \Gamma(r+1) &= \int_0^{+\infty} x^r e^{-x} dx = [-x^r e^{-x}]_0^{+\infty} + \int_0^{+\infty} r x^{r-1} e^{-x} dx \\ &= \lim_{x \rightarrow +\infty} (-x^r e^{-x}) - (0e^0) + r \int_0^{+\infty} x^{r-1} e^{-x} dx \\ &\quad \text{osservando che } \lim_{x \rightarrow +\infty} (-x^r e^{-x}) \text{ tende a } 0 \\ \Gamma(r+1) &= r \int_0^{+\infty} x^{r-1} e^{-x} dx = r\Gamma(r) \end{aligned}$$

□

Completiamo la definizione della funzione Γ :

$$n \in \mathbb{N} \implies \Gamma(n) = \begin{cases} \Gamma(n) = \Gamma(1) = 1 & n = 1 \\ \Gamma(n) = (n-1)! & n > 1 \end{cases}$$

Definizione 5.4.5. Distribuzione Γ

Dopo aver definito la funzione Γ , possiamo definire la densità di probabilità per la distribuzione Γ . I due parametri della densità vengono detti forma (r) e scala (λ) della distribuzione.

$$\Gamma(r, \lambda) = f(x) = \begin{cases} \left(\frac{1}{\Gamma(r)}\right) \lambda^r x^{r-1} e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Il **valore atteso** è $\frac{r}{\lambda}$, la **varianza** è $\frac{r}{\lambda^2}$.

Si nota come una densità esponenziale di parametro λ sia uguale ad una densità Γ di parametri $r = 1, \lambda = \lambda$. Infatti, le densità esponenziali e le densità χ^2 sono casi speciali della distribuzione Γ

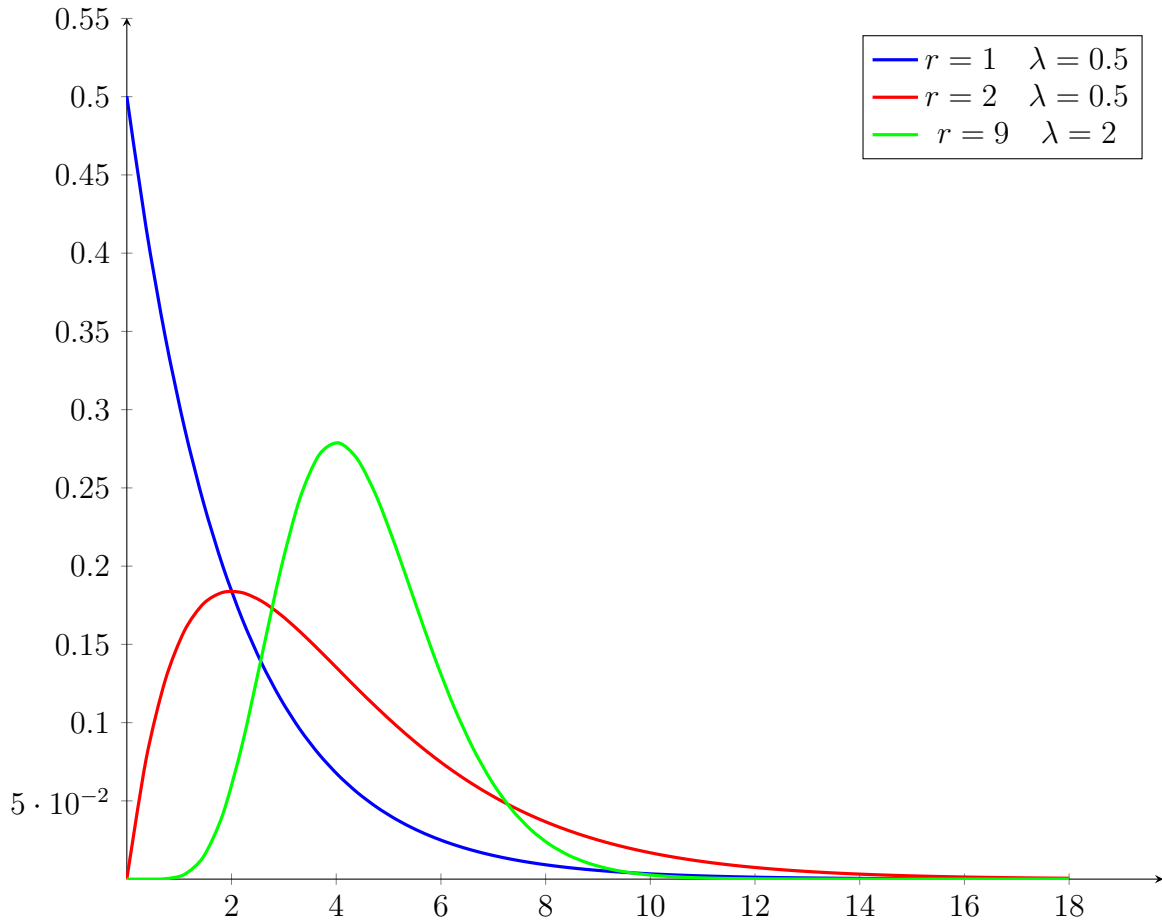


Figura 5.8: Densità di probabilità Γ

Proposizione 5.4.4. *La densità $\Gamma(r, \lambda)$ è una densità di probabilità*

Dimostrazione. Il prodotto di numeri positivi è positivo, quindi la funzione di densità $f \geq 0$. Rimane da dimostrare che $\int_{-\infty}^{+\infty} f(x)dx = 1$ per dimostrare che la funzione data da $\Gamma(r, \lambda)$ è una di probabilità. Conosciamo intanto $\int_{-\infty}^0 f(x)dx = 0$.

$$\begin{aligned} \int_0^{+\infty} \lambda^r x^{r-1} e^{-\lambda x} dx &= \int_0^{+\infty} (\lambda x)^{r-1} e^{-\lambda x} dx = \\ \text{Risolvendo per sostituzione: } t &= \lambda x \implies dt = \lambda dx \\ &= \int_0^{+\infty} t^{r-1} e^{-t} dt = \Gamma(r) \\ \implies \int_{-\infty}^{+\infty} f(x)dx &= \frac{1}{\Gamma(r)} \Gamma(r) = 1 \end{aligned}$$

□

Definizione 5.4.6. Formula dei momenti della distribuzione Γ

Definiamo una formula per calcolare il momento n -esimo di una variabile aleatoria con densità data dalla distribuzione Γ .

$$\begin{aligned} X &\sim \Gamma(r, \lambda) \implies \\ \mathbb{E}[X] &= \frac{\Gamma(r+1)}{\Gamma(r)\lambda} = \frac{\Gamma(r)(r)}{\Gamma(r)\lambda} = \frac{r}{\lambda} \\ \mathbb{E}[X^2] &= \frac{\Gamma(r+2)}{\Gamma(r)\lambda^2} = \frac{r(r+1)}{\lambda^2} \end{aligned}$$

Ed in generale:

$$X \sim \Gamma(r) \wedge n \in \mathbb{N} \implies \mathbb{E}[X^n] = \frac{\Gamma(r+n)}{\Gamma(r)\lambda^n}$$

Dimostrazione.

$$\begin{aligned} \mathbb{E}[X^n] &= \frac{1}{\Gamma(r)} \int_0^{+\infty} x^n \lambda^r x^{r-1} e^{-\lambda x} dx \\ &= \frac{1}{\Gamma(r)} \cdot \frac{1}{\lambda^n} \int_0^{+\infty} \lambda^{r+n} x^{n+r-1} e^{-\lambda x} dx \\ &= \frac{\Gamma(r+n)}{\Gamma(r)\lambda^n} \end{aligned}$$

□

Definizione 5.4.7. Somma (convoluzione) di variabili aleatorie indipendenti con distribuzione Γ

$$\begin{aligned} X &\sim \Gamma(r_X, \lambda) \wedge Y \sim \Gamma(r_Y, \lambda) \wedge X, Y \text{ indipendenti} \\ &\implies (X + Y) \sim \Gamma(r_X + r_Y, \lambda) \end{aligned}$$

Definizione 5.4.8. Distribuzione normale (Gaussiana)

La distribuzione Gaussiana ha una funzione di densità a forma di campana. È utilizzata per rappresentare variabili aleatorie a valori reali che sono prodotte dalla somma di tanti piccoli risultati. Ad esempio, la distribuzione normale viene utilizzata per modellare l'altezza della popolazione, perché l'altezza può essere il risultato di tanti piccoli fattori genetici e ambientali.

Sia data la densità $f(x; \mu, \sigma^2)$ dove μ è il parametro che rappresenta il valore atteso e σ^2 rappresenta la varianza. Si indica comunemente con $N(\mu, \sigma^2)$. Dal grafico si nota che la campana si sposta orizzontalmente al variare di μ , mentre al variare della varianza σ^2 la campana "cambia forma".

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

Definizione 5.4.9. Somma (convoluzione) di due variabili aleatorie con densità gaussiana

Siano date due variabili aleatorie indipendenti: $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, la loro somma è data quindi da $Z = X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

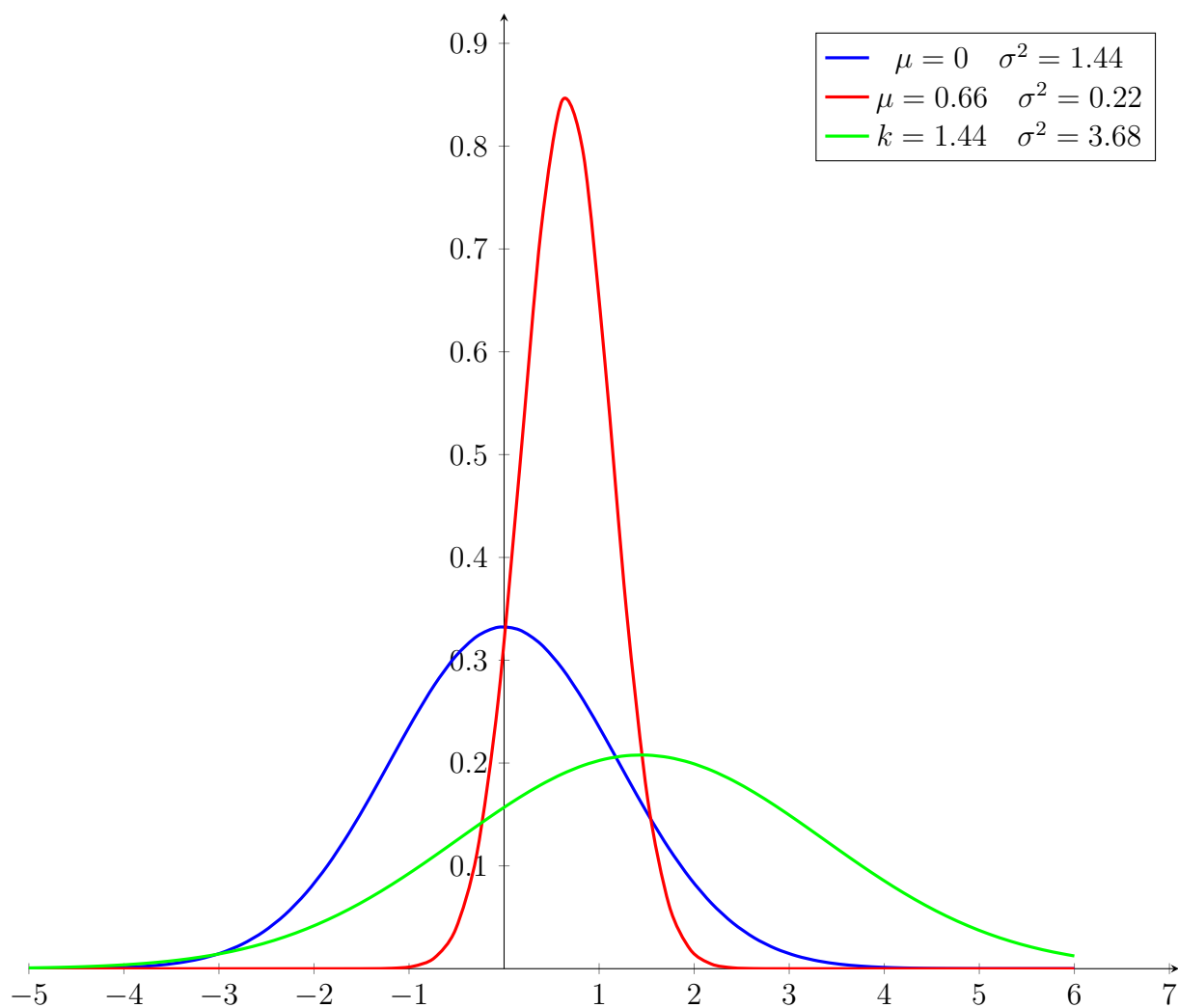


Figura 5.9: Densità di probabilità Gaussiana

Definizione 5.4.10. Densità Gaussiana Standard

La densità Gaussiana $N(0, 1)$ di parametri $\mu = 0$ e $\sigma^2 = 1$ viene detta **densità Gaussiana standard**. Si usa comunemente la lettera greca ϕ (phi) per indicare la funzione di densità e la controparte maiuscola Φ per indicare la funzione di ripartizione (CDF).

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

La funzione di ripartizione è:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-(t^2/2)} dt$$

Proposizione 5.4.5. *Una variabile $X \sim N(0, 1)$ con densità Gaussiana standard ha tutti i momenti.*

Dimostrazione.

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |x|^n e^{-(x^2/2)} dx < +\infty$$

Sapendo che $e^{-(x^2/2)}$ tende a 0 per $x = +\infty$ e per $x = -\infty$ □

Definizione 5.4.11. Momenti n -esimi di una Gaussiana standard

Per calcolare il **momento n -esimo** di una gaussiana standard definiamo una funzione $\text{Mo}(n)$ come

- Se n è dispari:

$$n = 2h + 1 \implies \text{Mo}(n) = \mathbb{E}[X^{2h+1}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2h+1} e^{-(x^2/2)} dx = 0$$

- Se n è pari:

$$n = 2h \implies \text{Mo}(n) = \mathbb{E}[X^{2h}] = (2h-1)\mathbb{E}[X^{2h-2}] = (n-1)\text{Mo}(n-2)$$

Ad esempio, $\mathbb{E}[X^2] = 1$ e $\mathbb{E}[X^4] = 3 \cdot \mathbb{E}[X^2] = 3$.

Definizione 5.4.12. Varianza di una Gaussiana standard

Dalla definizione 5.4.11 si ottiene che la varianza di una variabile aleatoria $X \sim N(0, 1)$ è

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1 - 0^2 = 1$$

Definizione 5.4.13. Conversione di una variabile aleatoria da Gaussiana qualsiasi a Gaussiana standard

Proposizione 5.4.6. *Si può sempre pensare una variabile aleatoria $X \sim N(\mu, \sigma^2)$ nella forma $\sigma X + m$ con $X \sim N(0, 1)$.*

Dimostrazione. Siano dati:

$$\mu \in \mathbb{R}, \quad \sigma > 0, \quad X \sim (N(0, 1) = f(x)), \quad Y = \sigma X + \mu, \quad Y \sim g(y)$$

$$y = \sigma x + \mu \implies x = \frac{y - \mu}{\sigma}$$

$$\implies g(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y - \mu}{\sigma}\right)^2}$$

$$\implies Y \sim N(\mu, \sigma^2) \implies \mathbb{E}[Y] = \mu \wedge \text{Var}(Y) = \sigma^2$$

□

Definizione 5.4.14. Somma (convoluzione) di due variabili aleatorie con densità gaussiana standard

Siano date due variabili aleatorie indipendenti: $X \sim N(0, 1)$, $Y \sim N(0, 1)$. Sappiamo che $X + Y = Z \sim N(0, 2)$, cioè ha densità $g(z) = \frac{1}{2\sqrt{\pi}}e^{-(z^2/4)}$

Dimostrazione.

$$\begin{aligned}
 g(z) &= \int_{-\infty}^{+\infty} f_1(z-x)f_2(x)dx \\
 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-(z-x)^2/2}e^{-(x^2/2)}dx \\
 &= \frac{e^{-(z^2/4)}}{2\pi} \int_{-\infty}^{+\infty} e^{((z/\sqrt{2})-\sqrt{2}x)^2/2}dx
 \end{aligned}$$

□

Tavola della funzione di ripartizione della distribuzione $N(0,1)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998
4.1	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99999	0.99999
4.2	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999
4.3	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999
4.4	0.99999	0.99999	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

5.5 Teoremi e leggi

Nota. Si usa l'acronimo inglese **i.i.d.** (independent and identically distributed) per indicare una collezione di variabili aleatorie con la stessa distribuzione di probabilità, ma tutte mutualmente indipendenti fra di loro.

Definizione 5.5.1. Convergenza in probabilità

Data una successione di variabili aleatorie $(X_n)_{n \in \mathbb{N}}$ si dice che essa **converge in probabilità** alla variabile aleatoria X , in simboli $X_n \xrightarrow{p} X$, se $\forall \epsilon > 0$ vale

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0 \text{ o equivalentemente}$$

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \epsilon) = 1$$

Vediamo un criterio comodo:

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = c \quad \wedge \quad \lim_{n \rightarrow \infty} \text{Var}(X_n) = 0$$

$$\implies X_n \text{ converge a } c$$

Dimostrazione. Ricordiamo la disuguaglianza di Chebishev

$$P(|X - \mathbb{E}[X]| > t) \leq \frac{\text{Var}(X)}{t^2}$$

e ricordiamo che

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

Se sostituiamo c a $\mathbb{E}[X]$ nella disuguaglianza di Chebishev si ottiene:

$$P(|X - c| > t) \leq \frac{\mathbb{E}[(X - c)^2]}{t^2}$$

$$\implies 0 \leq P(|X_n - c| > \epsilon) \leq \frac{\mathbb{E}[(X_n - c)^2]}{\epsilon^2}$$

Ne segue che

$$\begin{aligned} \frac{\mathbb{E}[(X_n - c)^2]}{\epsilon^2} &= \frac{\mathbb{E}[(X_n - \mathbb{E}[X_n] + \mathbb{E}[X_n] - c)^2]}{\epsilon^2} \\ &= \frac{\mathbb{E}[(X_n - \mathbb{E}[X_n])^2] + \mathbb{E}[(\mathbb{E}[X_n] - c)^2] + 2\mathbb{E}[(X_n - \mathbb{E}[X_n])(\mathbb{E}[X_n] - c)]}{\epsilon^2} \\ &= \frac{\text{Var}(X_n) + \mathbb{E}[(\mathbb{E}[X_n] - c)^2] + 2\mathbb{E}[(X_n - \mathbb{E}[X_n])(\mathbb{E}[X_n] - c)]}{\epsilon^2} \end{aligned}$$

Sappiamo che, per ipotesi $\text{Var}(X_n) = 0$ e che la speranza di una costante $c = c$.

$$\frac{\mathbb{E}[(X_n - c)^2]}{\epsilon^2} = \frac{0 + 0 + 2(0)}{\epsilon^2}$$

$$\implies 0 \leq P(|X_n - c| > \epsilon) \leq 0$$

□

Definizione 5.5.2. Media empirica

Data una collezione di variabili aleatorie i.i.d., si dice **media empirica** (o campionaria)

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

Nota. $\mathbb{E}[X_i] = \mu \implies \mathbb{E}[\overline{X}] = \mu$

Definizione 5.5.3. Legge (debole) dei Grandi Numeri

Sia data $(X_n)_{n \in \mathbb{N}}$ successione di variabili aleatorie i.i.d con media $\mu = \mathbb{E}[X_i]$ e varianza $\sigma^2 = \text{Var}(X_i)$ (valido $\forall i \in \mathbb{N} \wedge 1 \leq i \leq n$) si ha che

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

converge a μ in probabilità, per $n \rightarrow \infty$

Esempio 5.5.1. Consideriamo ad esempio il lancio di una moneta:

$$\begin{aligned} X_i &= \begin{cases} 0 & \text{se il risultato è croce} \\ 1 & \text{se il risultato è testa} \end{cases} \\ \mathbb{E}[X_i] &= \text{probabilità che esca testa} \\ \implies \frac{X_1 + \dots + X_n}{n} &= \left(\frac{\text{numero di risultati testa}}{\text{numero di risultati croce}} \right) \xrightarrow{p} \mathbb{E}[X_i] \\ \frac{\mathbb{E}[X_1 + \dots + X_n]}{n} &= \frac{n\mu}{n} = \mu \end{aligned}$$

Il **Teorema del Limite Centrale** (CLT o TLC) afferma che, in alcune situazioni, dato un numero di variabili aleatorie, la loro somma propriamente normalizzata tende approssimativamente ad una distribuzione normale, anche se le variabili aleatorie sommate non sono originariamente Gaussiane. Definiamo il teorema formalmente, senza però fornirne una dimostrazione.

Definizione 5.5.4. Teorema del Limite Centrale

Siano date X_1, \dots, X_n variabili aleatorie **i.i.d.** (collezione detta anche campionamento casuale di dimensione n), il quale valore atteso è $\mathbb{E}[X_i] = \mu$ e la varianza è $\text{Var}(X_i) = \sigma^2 > 0$ (valido $\forall i \in \mathbb{N} \wedge 1 \leq i \leq n$), si ha allora che

$$\lim_{n \rightarrow +\infty} P \left(a \leq \frac{X_1, \dots, X_2 - n\mu}{\sigma\sqrt{n}} \leq b \right) = \Phi(b) - \Phi(a) \quad (5.1)$$

Dove $\Phi(x)$ è la CDF (funzione di ripartizione) della distribuzione Gaussiana.

Definizione 5.5.5. Conclusioni pratiche del Teorema del Limite Centrale

Con un numero elevato di variabili aleatorie nel campionamento casuale, ovvero almeno $n > 50$, (con $n > 80$ l'approssimazione è ottima), allora la loro somma è approssimativamente normale.

$$\frac{X_1, \dots, X_2 - n\mu}{\sigma\sqrt{n}} \approx Z \sim N(0, 1)$$

Definizione 5.5.6. Convergenza in Distribuzione

Siano date X_1, X_2, \dots variabili aleatorie, non necessariamente i.i.d., con relative funzioni di ripartizione (CDF) F_1, F_2, \dots e sia X variabile aleatoria con CDF F continua, si dice che $\{X_n\}_{n \in \mathbb{N}}$ converge in distribuzione ad X se, punto a punto per ogni x vale

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x) \quad \forall x$$

Nota. Se ho X_1, X_2, \dots variabili aleatorie i.i.d. ed ho $E[X_i] = \mu$ ed anche $\text{Var}(X_i) = \sigma^2$, allora posso applicare il teorema del limite centrale per trovare a cosa converge in distribuzione la successione di variabili aleatorie. Vale quindi

$$Z_n = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \quad \text{e sia} \quad Z \sim N(0, 1)$$

$$\implies \{Z_n\}_{n \in \mathbb{N}} \text{ converge in distribuzione a } Z$$

(equivale in simboli a)

$$P(Z_n \leq x) \xrightarrow{d} \Phi(x)$$

Esempio 5.5.2. Siano dati $Z \sim N(0, 1)$, $\alpha \in \mathbb{R}$, $\alpha \in (0, 1)$, trovare z .

$$P(Z \leq z) = \alpha \implies \Phi(z) = \alpha \implies z = \Phi^{-1}(\alpha)$$

va applicata la tavola $N(0, 1)$ al contrario

5.6 Accenni di Inferenza Statistica

Definizione 5.6.1. Inferenza statistica

Partendo da un campione si ottengono informazioni su una popolazione (ad esempio sondaggi elettorali). Un **Campione Statistico** è una collezione di variabili aleatorie X_1, \dots, X_n indipendenti ed identicamente distribuite (i.i.d.) con distribuzione di probabilità solo parzialmente conosciuta.

Definizione 5.6.2. Varianza campionaria

$$S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}$$

Nota. $\mathbb{E}[S^2] = \sigma^2$

5.7 Esercizi

Esercizio 5.7.1. Consideriamo una variabile (X, Y) avente densità

$$f(x, y) = \begin{cases} e^{-x} & \text{sull'insieme } 0 < y < x \\ 0 & \text{altrove} \end{cases}$$

1. Provare che la funzione sopra scritta è effettivamente una densità e calcolare le densità marginali delle variabili X, Y . Proseguiamo controllando che valgano le condizioni della definizione 5.2.2, controllando che $\int_{\mathbb{R}^2} f(x, y) dx dy = 1$.

$$\begin{aligned} \int_0^{+\infty} \int_0^x e^{-x} dx dy &= \int_0^{+\infty} dx \int_0^x e^{-x} dy \\ &= \int_0^{+\infty} e^{-x} \cdot y \Big|_0^x dx = \int_0^{+\infty} x e^{-x} dx = 1 \end{aligned}$$

2. Le variabili X e Y sono indipendenti? Controlliamo verificando che valgano le condizioni nella definizione 5.3.1. Si verifica facilmente osservando che $f(x, y) \neq f(x)f(y)$.
3. Calcolare la densità di $Z = (X + Y)$. Calcoliamo $G(z) = P(Z \leq z)$

$$G(z) = \int_0^{z/2} \left(\int_y^{z-y} e^{-x} dx \right) dy = \int_0^{z/2} - (e^{-z+y} - e^{-y}) dy$$

E deriviamo

$$g(z) = \begin{cases} 0 & z \leq 0 \equiv (x + y \leq 0) \\ \frac{dG(z)}{dz} = e^{-z/2} - e^{-z} & z > 0 \equiv (x + y > 0) \end{cases}$$

Esercizio 5.7.2. Siano X, Y due numeri scelti a caso e in modo indipendente fra 0 e 1, e sia $Z = XY$ il loro prodotto, calcolare la densità di Z . Abbiamo che $Z \in [0, 1]$ e $G(z) = P(Z \leq z) = P(XY \leq z)$.

$$\begin{aligned} P(XY \leq z) &= \int_z^1 \left(\int_0^{z/x} dy \right) dx + z = \\ &= z + \int_z^1 \frac{z}{x} dx = z(1 - \log(x)) \end{aligned}$$

Ne otteniamo che

$$G(x) = \begin{cases} 0 & z \leq 0 \\ 1 & z \geq 1 \\ z(1 - \log z) & 0 < z < 1 \end{cases} \implies g(z) = \begin{cases} 0 & z \leq 0 \vee z \geq 1 \\ -\log z & 0 < z < 1 \end{cases}$$

