Appunti di Calcolo Probabilità e Statistica

Alessandro Cheli - Prof. Ghimenti - Prof.ssa Chiodaroli

A.A 2019-2020

Indice

1 Probabilità Discreta e Condizionata **1**
- 1.1 Probabilità Discreta e Formule Combinatorie 1
  - 1.1.1 Permutazioni di n elementi 1
  - 1.1.2 Coefficiente Binomiale 2
  - 1.1.3 Disposizioni 2
- 1.2 Probabilità Condizionata 2
- 1.3 Esercizi 3

2 Spazio Probabilizzato **5**
- 2.1 Lo Spazio Probabilizzato 5
- 2.2 Formula di fattorizzazione 7
- 2.3 Formula di Bayes 8
- 2.4 Esercizi 8

3 Variabili Aleatorie **11**
- 3.1 Variabili Aleatorie Discrete 11
- 3.2 Leggi su Variabili Aleatorie 12
- 3.3 Valore Atteso 16
- 3.4 Esercizi 19

4 Catene di Markov **23**
- 4.1 Catene di Markov e Processi Stocastici 23
- 4.2 Calcolo Algebrico su catene di Markov 26
- 4.3 Esercizi 27

Capitolo 1

Probabilità Discreta e Condizionata

1.1 Probabilità Discreta e Formule Combinatorie

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

La probabilità (discreta) di un evento è definita come

$$\mathrm{P}\left(\text{evento}\right) = \frac{\#\text{ casi favorevoli}}{\#\text{ casi possibili}} \tag{1.1}$$

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 di ottenere un numero ≥ 4 lanciando 2 dadi.

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

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

$$\begin{cases} \text{Se pesco un Asso } \implies 4 \cdot 53 \\ \text{Se non pesco un Asso } \implies 4 \cdot 50 \end{cases}$$

1.1.1 Permutazioni di n elementi

Definizione 1.1.2. 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.

Le permutazioni di un insieme di n elementi sono definite come

$$\mathrm{Perm}(n) = n! \tag{1.2}$$

Dimostrazione. Dimostrando per induzione, i casi base sono $\text{Perm}(0) = 1$ e $\text{Perm}(1) = 1$

Il passo induttivo sarà

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

□

1.1.2 Coefficiente Binomiale

Definizione 1.1.3. Il coefficiente binomiale è un numero intero non negativo definito dalla seguente formula, è analogo alla proposizione "Come scegliere k oggetti da un insieme di n elementi"

$$S_{n,k} = \binom{n}{k} = \frac{n!}{(n-k)!k!} \tag{1.3}$$

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

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

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

Il secondo passo iniziale è

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

Procedendo per passo induttivo:

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

□

1.1.3 Disposizioni

Definizione 1.1.4. Una disposizione $D_{n,k}$ significa il numero di modi per "prendere" k oggetti ordinati da un insieme di n elementi.

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

1.2 Probabilità Condizionata

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

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

Poniamo il vincolo che il lancio del primo dado risulti almeno ≥ 3

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

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

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

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 $A^C = \Omega \backslash A$. La probabilità di un complementare è $P\left(A^C\right) = 1 - P(A) = P(\Omega) - P(A)$

La probabilità di un intersezione di eventi è $P(A \cap B) = P(A) \cdot P(B)$

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 $S_{5,3} \cdot S_{85,2} = \binom{5}{3} \cdot \binom{85}{2}$

La probabilità di ottenere un terno esatto sarà quindi

$$P(\text{terno esatto}) = \frac{\binom{5}{3} \cdot \binom{85}{2}}{\binom{90}{5}} = \frac{85 * 84 * 5}{\binom{90}{5}}$$

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 ed il dado nero risulti 2:

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

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

In generale, dati due eventi A, B con $A \cap B \neq 0$ la probabilità dell'unione è $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$, allora $\mathrm{P}(\Omega) = 1$ dove $\Omega =$ tutti gli eventi.

Capitolo 2

Spazio Probabilizzato

2.1 Lo Spazio Probabilizzato

Definizione 2.1.1. Uno **spazio probabilizzato** è un costrutto matematico che modella un processo del mondo reale o "esperimento", consistente in degli stati che occorrono casualmente. Viene costruito su una situazione o esperimento particolare, Uno spazio probabilizzato è definito come una terna:

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

Ω è l'insieme degli eventi elementari, ovvero tutti i risultati possibili, ad esempio in un lancio di un dado $\Omega = \{1, 2, 3, 4, 5, 6\}$

$\mathcal{P}(A)$, ovvero le parti di A, sono tutti gli insiemi che posso costruire a partire dagli elementi di A. Ad esempio:

$$\begin{aligned} A &= \{0, 1\} \\ \mathcal{P}(A) &= \{0, \{0, 1\}, \{0\}, \{1\}\} \end{aligned}$$

$F \subseteq \mathcal{P}(\Omega)$ è un sottoinsieme delle parti di omega, chiuso rispetto a intersezione, unione e complementare.

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

Se $A_1, \ldots, A_n \subset F$ allora $\bigcup\limits_{n \in \mathbb{N}} A_n \in F$

Se F comprende queste proprietà si dice che è una σ Algebra (tribù)

La probabilità P è una funzione definita come

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

Dato Ω insieme finito, allora $F = \mathcal{P}(\Omega)$ allora la probabilità sarà

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

Dimostrazione. La terna (Ω, F, P) si può dimostrare.

$$P(A^C) = 1 - P(A) \iff \begin{cases} A \cap A^C \neq \emptyset \\ A \cup A^C = \Omega \\ P(\Omega) = 1 \end{cases} \tag{2.3}$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$A \subseteq B \implies P(A) \leq P(B)$$

□

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 $\sum_n p_n = \sum_n c = +\infty$

Ciò significa che nella lotteria di De' Filetti è impossibile che ogni numero sia equiprobabile perché P non è definibile. Abbiamo dimostrache che non è possibile che $\mathrm{P}(\Omega) = \sum_{n\in\mathbb{N}} p_n = 1$, essendo $\mathbb{N}$ insieme infinito.

Definizione 2.1.2. Densità di Probabilità: Definiamo $\{p_n\}_{n\in\mathbb{N}}$ con $p_n \in \mathbb{R}$ come funzione, detta densità di probabilità come $p_n \geq 0$ e $\sum_{n\in\mathbb{N}} p_n = 1$

Se $\{p_n\}$ è una densità di probabilità discreta $\implies (\mathbb{N}, \mathcal{P}(\mathbb{N}), P)$ è uno spazio probabilizzato. Vale anche per eventi non equiprobabili.

Definizione 2.1.3. Indipendenza degli eventi: Sia dato $P(A \mid B) \equiv \dfrac{P(A \cap B)}{P(B)}$. Di conseguenza, se $P(A \mid B) = P(A)$ due eventi A, B sono indipendenti. Un esempio è il lancio di due dadi, il primo dado non influenzerà in alcun modo il risultato del secondo, per questo gli eventi del lancio di due dadi A, B sono indipendenti.

È vero quindi che $P(A \mid B) = P(A) \implies P(B \mid A) = P(B)$? Sì se A, B sono indipendenti.

$$\begin{gathered} P(A \mid B) = \frac{P(A \cap B)}{P(B)} = P(A), \text{ se } P(B) \neq 0 \\ P(B \mid A) = \frac{P(B \cap A)}{P(A)} = P(B), \text{ se } P(A) \neq 0 \\ P(A \cap B) = P(A) \cdot P(B) = P(B \cap A) \end{gathered} \tag{2.4}$$

Se $A_1, \ldots, A_n$ sono indipendenti:

$$\forall i_1, \ldots, i_k.k \leq n \implies P(A) \cap (A_{i_1}, \ldots, A_{i_k}) = P(A_{i_1}) \cdot \cdots \cdot P(A_{i_k})$$

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

Dimostrazione.

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

A e B^C sono indipendenti □

Esercizio 2.1.3. Lancio due dadi, uno rosso ed uno nero. $\Omega = (r, n)$ dove $r = 1, 2, 3, 4, 5, 6$ e $n = 1, 2, 3, 4, 5, 6$ Definiamo lo spazio probabilizzato con $F = \mathcal{P}(\Omega), \omega \in \Omega$. Ad esempio $\mathrm{P}(n) = \frac{1}{36}$

Probabilità che il rosso sia 3 sapendo che rosso + nero fa 6

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

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

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

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 $\mathrm{P}(1, 1) = 0, \mathrm{P}(2, 2) = 0$

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

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

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

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

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

$$\mathrm{P}(x = 1 \mid y = 2) = \frac{\mathrm{P}((1, 2))}{\mathrm{P}(y = 2)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3} \tag{2.5}$$

2.2 Formula di fattorizzazione

Supponiamo di avere una famiglia di insiemi $B_1, \ldots, B_n$ con $n \in \mathbb{N}$ che è detta una partizione finita di Ω (insieme fondamentale). Voglio che $\forall i. B_i \in F$ e che $\forall i \forall j \neq i. B_i \cap B_j = \emptyset$ e anche che $\bigcup_{i=1}^{n} B_i = \Omega$

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

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i) \tag{2.6}$$

Dimostrazione.

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

□

Definizione 2.2.1. Condizionamento Ripetuto: Dati $A_1, \ldots, A_n$ eventi, allora

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

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*

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \tag{2.7}$$

Dimostrazione.

$$\begin{aligned}
\mathrm{P}(A \mid B) &= \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}, \text{ se } \mathrm{P}(B) \neq 0,\\
\mathrm{P}(B \mid A) &= \frac{\mathrm{P}(B \cap A)}{\mathrm{P}(A)}, \text{ se } \mathrm{P}(A) \neq 0,\\
\Rightarrow \mathrm{P}(A \cap B) = \mathrm{P}(A \mid B) \cdot \mathrm{P}(B) &= \mathrm{P}(B \mid A) \cdot \mathrm{P}(A),\\
\Rightarrow \mathrm{P}(A \mid B) &= \frac{\mathrm{P}(B \mid A) \cdot \mathrm{P}(A)}{\mathrm{P}(B)}, \text{ se } \mathrm{P}(B) \neq 0.
\end{aligned}$$

□

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 sono $1/3$ per la Francia e $1/3$ per la Germania. $3/4$ delle richieste dalla Francia sono vino bianco. $1/4$ delle richieste dalla Francia sono vino rosso. $1/2$ delle richieste dalla Germania sono vino bianco. $1/2$ delle richieste dalla Germania sono vino rosso.

Utilizzando la formula di partizione troviamo la probabilità che una richiesta sia vino bianco.

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

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.

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

Uno studente prova l'esame e viene respinto. Qual'è la probabilità che aveva di avere una preparazione scarsa? P (Scarso | Respinto). Prima calcoliamo la probabilità di essere respinti.

$$\mathrm{P}(R) = \mathrm{P}(R \mid O) \cdot \mathrm{P}(O) + \mathrm{P}(R \mid B) \cdot \mathrm{P}(B) + \mathrm{P}(R \mid S) \cdot \mathrm{P}(S) = 0.302$$

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

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

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:

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

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

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

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 sono soggetti a rischio. Un malato si denota con M mentre uno sano con S. Vogliamo sapere

1. P (Soggetto casuale sia malato) =

   $\mathrm{P}(M) = \mathrm{P}(M \mid R) \cdot \mathrm{P}(R) + \mathrm{P}(R \mid N) \cdot \mathrm{P}(N) = 0.35$

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

2. P (Soggetto malato fosse a rischio) =

   $\mathrm{P}(R \mid M) = \dfrac{\mathrm{P}(M \mid R) \cdot \mathrm{P}(R)}{\mathrm{P}(M)} = \dfrac{0.2 \cdot 0.15}{0.35} = 0.855$

3. P (Soggetto soggetto sano sia a rischio) =

   $\mathrm{P}(R \mid S) = \dfrac{\mathrm{P}(S \mid R) \cdot \mathrm{P}(R)}{\mathrm{P}(S)} = \dfrac{(1 - 0.2) \cdot 0.15}{(1 - 0.35)} = 0.124$

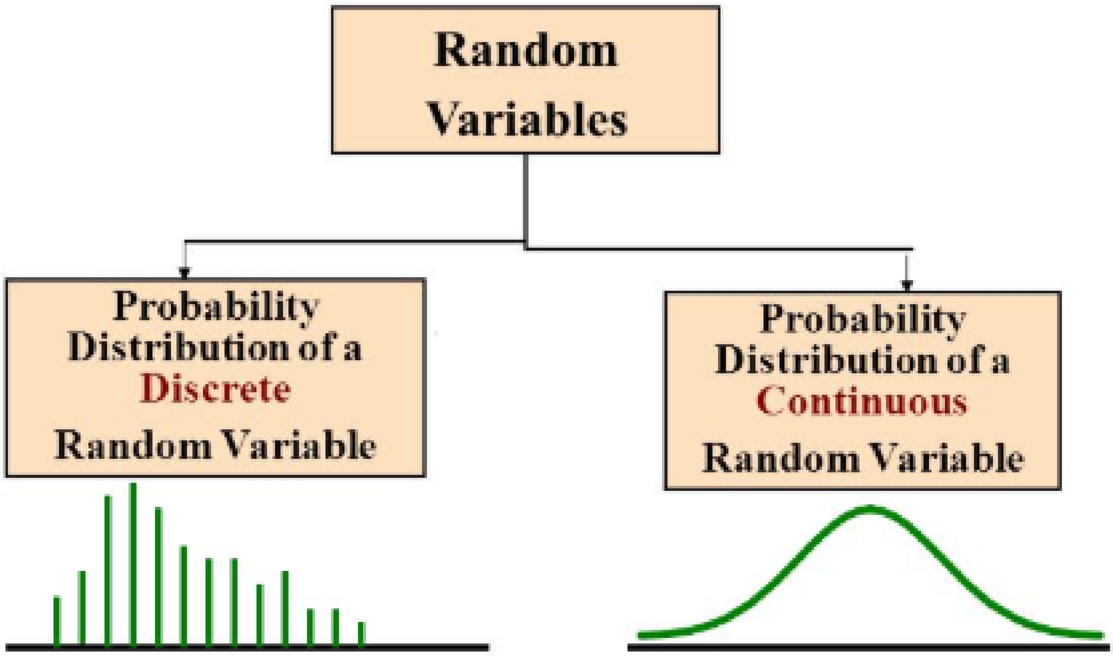
Nota. La probabilità che l'evento A^c (A complementare) si verifichi sapendo B è $\mathrm{P}(A^c \mid B) = 1 - \mathrm{P}(A \mid B)$ mentre la probabilità di A sapendo B^c è $\mathrm{P}(A \mid B^c) \neq 1 - \mathrm{P}(A \mid B)$.

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

Capitolo 3

Variabili Aleatorie

Figura 3.1: Tipi di variabili casuali



3.1 Variabili Aleatorie Discrete

Definizione 3.1.1. Una variabile aleatoria (casuale o stocastica) discreta è una variabile che può assumere diversi valori in dipendenza da qualche fenomeno casuale. Il risultato del lancio di un dado, ad esempio, è una variabile aleatoria discreta.

Prendiamo uno **spazio probabilizzabile** (Ω, F) e una variabile aleatoria discreta $X : \Omega \to \mathbb{R}$, che è una funzione continua non surgettiva. I valori di X devono essere un sottoinsieme finito di $\mathbb{R}$ ovvero $\{a_1, \ldots, a_k\}$. Vogliamo anche che $\forall j \in [1, k]$ sia vero $X^{-1}(a_j) = \{\omega \in \Omega, X(\omega) = a_j\} \in F$. Utilizziamo la funzione inversa per ottenere gli elementi di Ω su cui possiamo definire la probabilità.

Con la probabilità di tutti gli eventi definisco la densità di probabilità. Nello spazio probabilizzato $(\Omega, F, \mathcal{P})$ la probabilità che la variabile aleatoria assuma il valore a_j sarà $p_j = \mathrm{P}(X = a_j) = \mathrm{P}(X^{-1}(a_j))$, che viene detta densità di probabilità. Varranno quindi le seguenti proprietà

1. $\forall j.\, X^{-1}(a_j)$ sono tutti insiemi disgiunti.

2. Essi coprono tutto Ω

Vale che $\sum_{j=1}^{k} p_j = 1$. Poiché:

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

Sia $X : \Omega \to \{a_1, \ldots, a_k\}$ una variabile aleatoria, e sia la densità di probabilità $p_j \geq 0$ e anche $\sum_{j=1}^{k} p_j = 1$, allora $\mathrm{P}(X = a_j) = p_j$.

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

$$\mathrm{P}(X \leq a_3) = \mathrm{P}((X = a_1) \cup (X = a_2) \cup (X = a_3)) = \mathrm{P}(X = a_1) + \mathrm{P}(X = a_2) + \mathrm{P}(X = a_3)$$

Esempio 3.1.1. 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 $X : \Omega \to \{0, 1, 2, \ldots, 9, 10\}$ come il conteggio dei risultati dei lanci dove ottengo 6. Qual'è la probabilità di ottenere 3 lanci dove ho fatto 6?

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

3.2 Leggi su Variabili Aleatorie

Definizione 3.2.1. Legge di Bernoulli:

Faccio un esperimento, il risultato positivo ha probabilità p, mentre il risultato negativo ha probabilità $1 - p$. Sia lo spazio $\Omega = \begin{cases} \text{successo} \to 1 \\ \text{insuccesso} \to 0 \end{cases}$. Una variabile aleatoria Bernoulliana è definita come $X : \Omega \to \begin{cases} p_1 = p \\ p_0 = 1 - p \end{cases}$

Definizione 3.2.2. Legge Binomiale Sia k il conteggio dei successi di n esperimenti, abbiamo quindi che $B(n, p) = X_i\{(0, 1)\}^n \to \{0, \ldots, n\}$. Abbiamo che la densità di probabilità Binomiale $p_k = \mathrm{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$.

p_k è una densità? Sappiamo che $p_k \geq 0$ e sappiamo che $1 = \sum_k p_k$, con il binomio di Newton possiamo dimostrare che $(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$. Proseguendo, abbiamo che

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

Definizione 3.2.3. Somma di Variabili Aleatorie Lancio due dadi, uno rosso e uno nero, avremo quindi $\Omega = \{R, N\} = \{(1, 6)\}^2$. Definisco due variabili aleatorie, X per il dado rosso dove $X : (R, N) \to R$ e la variabile $Y : (R, N) \to N$. La densità per X sarà $p_j = \frac{1}{6} \forall j$ mentre la densità per Y sarà $q_j = \frac{1}{6} \forall j$ Avremo che $Z = X + Y$ conta la somma dei dadi.

Esempio 3.2.1. Calcolare la densità di Z

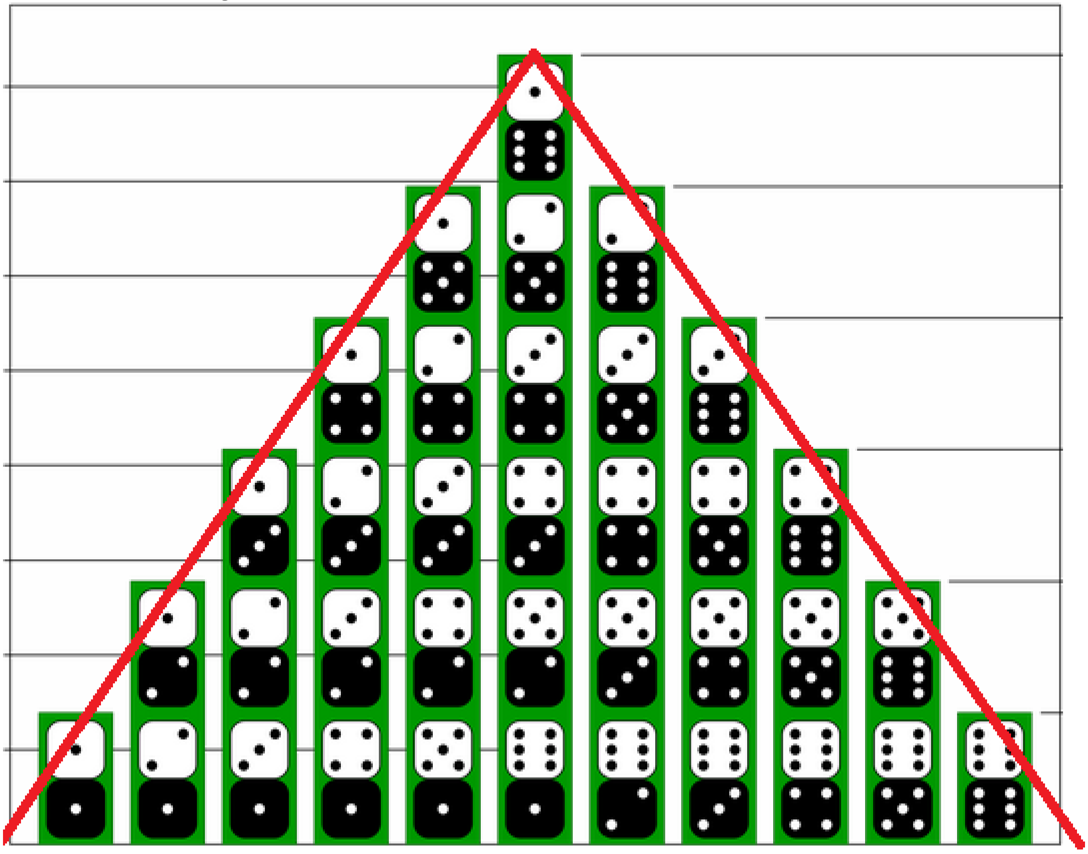
In questo caso X, Y sono indipendenti, quindi avremo la densità di Z detta t_Z.

$$\forall n \in [2, 12]. \left(\mathrm{P}\left(Z = n\right) = \sum_{i=1}^{n-1} \mathrm{P}\left(X = i\right) \cdot \mathrm{P}\left(Y = n - i\right) \right)$$

$X + Y$	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 $\mathrm{P}\left(4 \leq Z \leq 6\right)$

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

Definizione 3.2.4. 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

$$\left(\mathrm{P}\left(X_1 \in I_1\right) \cap \mathrm{P}\left(X_2 \in I_2\right)\right) = \mathrm{P}\left(X_1 \in I_1\right) - \mathrm{P}\left(X_2 \in I_2\right)$$

Nell'esempio di prima X, Z e Y, Z sono dipendenti perché dati $I_1 = [1, 2], I_2 = [3, 4]$ allora si ha che $\mathrm{P}(X \in I_1) = \mathrm{P}(X = 1, 2) = \frac{1}{3}$ e si ha anche $\mathrm{P}(Z \in I_2) = \mathrm{P}(2 = 3, 4) = \frac{5}{36}$. Ne otteniamo che:

$$\mathrm{P}((X \in I_1) \cap (Z = 3, 4)) = \mathrm{P}(X = 1, 2, Z = 3, 4) = \frac{4}{12}$$

Definizione 3.2.5. Variabili Aleatorie Congiunte Due variabili aleatorie $X, Y : \Omega \to \mathbb{R}$ discrete sono congiunte quando si può calcolare $\mathrm{P}(X = m \cap Y = n) = p_{n,m}$, 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 $\mathrm{P}(X = m) = p_m$ e $\mathrm{P}(Y = n) = q_n$ con

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

Non si può ricostruire dalle due probabilità la variabile aleatoria congiunta. Ad esempio, conoscendo p_n, q_n cerco $p_n, m = \mathrm{P}(X = m \cap Y = n) = \mathrm{P}(X = m | Y = n) \cdot \mathrm{P}(Y = n)$. Se X, Y sono indipendenti allora $P(X = m \mid Y = n) = \mathrm{P}(X = m)$ e vale il prodotto $p_{m,n} = p_n \cdot p_m$.

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

Definizione 3.2.6. Formula di Convoluzione

Tornando alla somma di due variabili aleatorie discrete, dati X, Y indipendenti e $Z = X + Y$, con $X, Y : \Omega \to \mathbb{N}$, dobbiamo calcolare $\mathrm{P}(Z = n)$ e la densità di probabilità discreta $\{Z = n\}$

$$\{Z = n\} = \bigcup_{i=0}^{n} \{X = 1 \cap Y = n - i\}$$

$$\mathrm{P}(z = n) = \sum_i \mathrm{P}(X = i \cap Y = n - 1) = \sum_{i=n}^{n} \mathrm{P}(X = i) \cdot \mathrm{P}(Y = n - i)$$

Teorema 3.2.1. ***Rapporto fra Bernoulli e Binomiale***

Sommando n esperimenti di p dove conto i successi ottengo la binomiale $B(n, p)$, quindi $\sum_{i=1}^{n} Bern_i(p) = B(n, p)$

Dimostrazione. Caso base, per $n = 1 \implies B(1, p) = \mathrm{Bern}(p)$. Come passo induttivo abbiamo

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

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

Introduciamo le densità per continuare la dimostrazione

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

□

Definizione 3.2.7. Variabile Geometrica Data una successione di esperimenti ripetuti con probabilità di successo $0 \leq p \leq 1$. Lo ripeto fino ad ottenere un successo. $\mathrm{Geom}(p)$ conta il numero di prove necessarie. Ovvero $\mathrm{P}\left(\mathrm{Geom}(p)=k\right)=$ la probabilità di fare k esperimenti ed avere un successo dall'ultimo. La densità di probabilità sarà $p_k=\mathrm{P}\left(\mathrm{Geom}(p)=k\right)=(1-p)^{k-1}\cdot p$ con $p_k \geq 0$ e anche $\sum_{k=1}^{\infty}(1-p)^{k-1}p=p\sum_{k=1}^{\infty}(1-p)^{k-1}$, che è una serie geometrica. Sarà quindi equivalente a $p\sum_{i=0}^{n}(1-p)^i=\dfrac{p}{1-(1-p)}=1$

Osservazione: consideriamo la serie geometrica $\sum_{i=0}^{\infty}q^i$. Sappiamo che $(1-q)\sum_{i=0}^{\infty}q^i=1$ perché $(1-q)(1+q+q^2+q^3\dots)=(1-q+q-q^2+q^2+\dots)$, semplificando i termini rimane 1.

Una variabile geometrica **non ha memoria**, l'esperimento numero n ha la stessa probabilità degli altri esperimenti:

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

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

Definizione 3.2.8. 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. È detta variabile ipergeometrica. Sappiamo che $(0<n\leq b+r)\wedge(k\leq n)\wedge(k\leq r)\wedge(n-k\leq b)$.

Definiamo la densità di probabilità

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

Definizione 3.2.9. 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? Sappiamo che 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 $n+\mathrm{NB}$. Sapendo che per avere n successi e k fallimenti, la probabilità di una Binomiale Negativa è definita come

$$
\mathrm{P}\left(\mathrm{NB}\left(n,p\right)=k\right)=p^n(1-p)^k
$$

Esempio 3.2.3. Lancio una moneta fino ad ottenere 3 croci (non consecutive). Qual'è la probabilità di aver fatto esattamente 2 risultati testa? $\mathrm{P}\left(\mathrm{NB}\left(3,\frac{1}{2}\right)=2\right)=\binom{3+2-1}{2}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2=\frac{3}{16}$.

La probabilità di ottenere almeno un risultato testa è $\mathrm{P}\left(\mathrm{NB}\left(3,\frac{1}{2}\right)=0\right)=1-\frac{1}{8}=\frac{7}{8}$

3.3 Valore Atteso

Definizione 3.3.1. Media Pesata, Speranza o Valore Atteso Consideriamo di voler calcolare la media pesata dei voti degli esami universitari. La media sarà per ogni esame i:

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

Definiamo X variabile aleatoria discreta $X \subseteq \mathbb{R}$. La media di X è detta **speranza** o **valore atteso** e si denota con

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

Per calcolare la media di una funzione $f : \mathbb{R} \to \mathbb{R}$:

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

Osservazione $\sum_k p_k = 1$ non implica che $\sum_k kp_k$ sia convergente. Se non converge ad un numero allora la variabile non ha media. Se la variabile assume solo valori positivi allora

$$\mathbb{E}[X] = \sum_{k>0} \mathrm{P}(X > k) = \sum_{k>0} (\mathrm{P}(X = k+1) + \mathrm{P}(X = k+2) + \dots)$$

Definizione 3.3.2. Speranza di una Binomiale $B(n,p)$. La distribuzione $p_k = \binom{n}{k} p^k (1-p)^{n-k} \forall k \in [0,n]$. 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 $h = k - 1$.

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

Definizione 3.3.3. Speranza di una variabile geometrica $\mathrm{Geom}(p)$, che conta il numero di successi prima di un successo:

$$\mathbb{E}[\mathrm{Geom}(p)] = \sum_{k=0}^{\infty} k(1-p)^k p = \sum_k \mathrm{P}(\mathrm{Geom}(p) > k)$$

Prendiamo in caso

$$\sum_k \mathrm{P}(\mathrm{Geom}(p) > k) = \sum_{h=k+1}^{\infty} (1-p)^(h-1)p = \sum_k (1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p}$$

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 $\mathbb{E}[X] = \mathbb{E}[B(3, 1/6)] - 1 = 3 \cdot 1/6 - 1 = -1/2$

Definizione 3.3.4. 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. Definiamo un intervallo τ, suddiviso in n sottointervalli. Abbiamo μ successi. Se n è grande abbastanza tale che in ogni intervallo avviene 1 evento, generati da n esperimenti indipendenti allora $\mu = \mathbb{E}[B(n,p)] = np$. Ne segue che:

$$p = \frac{\mu}{n} \mathrm{P}\left(B(n,p) = k\right) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k} =$$

$$\frac{n!}{k!(n-k)!} \frac{1}{n^k} \frac{\mu^k}{\left(1-\frac{\mu}{n}\right)^k} \cdot \left(1 - \frac{\mu}{n}\right)^n =$$

$$\frac{n^k \cdot \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{k-1}{n}\right)}{k!} \cdot \frac{\mu^k}{\left(1-\frac{\mu}{n}\right)^k} \cdot \left(1 - \frac{\mu}{n}\right)^n$$

Ne otteniamo che

$$\mathrm{P}\left(\mathrm{Poisson}(\mu) = k\right) = \frac{\mu^k}{k!} e^{-\mu} \tag{3.1}$$

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.

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

Utilizziamo la distribuzione di Poisson.

$$\mathrm{P}\left(\mathrm{Poisson}(12) = 7\right) = \frac{12^7}{7} e^{-12} \approx 0.04$$

Definizione 3.3.5. Linearità della media
Siano date X, Y variabili aleatorie, $\alpha, \beta \in \mathbb{R}$. Abbiamo che

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

Definizione 3.3.6. Valore atteso della somma

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] \tag{3.2}$$

Dimostrazione.

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

□

Definizione 3.3.7. Valore atteso del prodotto Siano date X, Y variabili aleatorie **indipendenti**.

$$\begin{aligned}
\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y] &= \sum_{i,j} i \cdot j \cdot \mathrm{P}(X = i \cap Y = j) \\
&= \sum_{i,j} i \cdot j p_{ij} = \sum_{i,j} ij p_i^X p_j^Y = \sum_i i p_i^X \cdot \sum p_j^Y
\end{aligned}$$

Definizione 3.3.8. Ordinamento

Sia $\mathbb{E}[X^n]$ ordinamento di ordine n. Dato un $\mathbb{E}[f(x)]$, ad esempio $\mathbb{E}[X^n] = \sum_i i^n p_i$ Oppure ad esempio $\mathbb{E}[X^2] = \sum_i i^2 \mathrm{P}(X = i)$. Ne segue che $\mathbb{E}[X^2] \neq (\mathbb{E}[X])^2$.

Definizione 3.3.9. Varianza

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

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

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

Dimostrazione.

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

□

Definizione 3.3.10. Disuguaglianza di Hölder

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

Definizione 3.3.11. Disuguaglianza di Markov

$$\begin{aligned}
X &\geq 0 \wedge a > 0 \\
\mathrm{P}(X > a) &< \frac{\mathbb{E}[X]}{a}
\end{aligned} \tag{3.4}$$

Definizione 3.3.12. Disuguaglianza di Chebichev

$$\begin{aligned}
X &\geq 0 \wedge a > 0 \\
\mathrm{P}(|X - \mu| > a) &\leq \frac{\mathrm{Var}(X)}{a^2}
\end{aligned}$$

Dimostriamo la disuguaglianza di Chebichev in quanto ci sarà utile nel resto del corso

Dimostrazione.

$$\mathrm{P}\left(|X-\mu|>a\right)=\sum_{n}\mathrm{P}\left(X=n\right)=\sum_{|X-\mu|>n}p_n$$
$$\text{se } |X-\mu|>a \text{ allora } 1<\frac{|X-\mu|^2}{q^2}$$
$$\implies \sum_{|X-\mu|>n}p_n\leq\sum_{|X-\mu|>n}\frac{|X-\mu|^2}{a^2}p_n\leq\frac{1}{a^2}\sum_{n}|n-\mu|^2p_n$$

□

Definizione 3.3.13. σ Deviazione Standard
Sia X variabile aleatoria. Allora abbiamo che $\sigma(X)=\sqrt{\mathrm{Var}\left(X\right)}$. Data μ media di X, e dati $\alpha,\beta\in\mathbb{R}$ si ha allora che

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

Definizione 3.3.14. Somma di varianza
Siano date X,Y variabili aleatorie indipendenti $\implies \mathrm{Var}\left(X+Y\right)=\mathrm{Var}\left(X\right)+\mathrm{Var}\left(Y\right)$

Dimostrazione.

$$\begin{aligned}\mathrm{Var}\left(()\,X+Y\right)&=\mathbb{E}\left[(X+Y)^2\right]-(\mathbb{E}\left[X+Y\right])^2\\&=\mathbb{E}\left[X^2\right]-2\mathbb{E}\left[X\cdot Y\right]+\mathbb{E}\left[Y^2\right]\\&=(\mathbb{E}\left[X\right])^2-2\mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]-(\mathbb{E}\left[Y\right])^2\\&=\mathrm{Var}\left(X\right)+\mathrm{Var}\left(Y\right)+2(\mathbb{E}\left[XY\right]-\mathbb{E}\left[X\right]\mathbb{E}\left[Y\right])\\ \text{Si ha che } 2(E[XY]&-\mathbb{E}\left[X\right]\mathbb{E}\left[Y\right])=0 \text{ se X,Y sono indipendenti.}\end{aligned}$$

□

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

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

Esempio 3.3.4. $X=B(n,p)\implies \mathrm{Var}\left(X\right)=nB(1,p)=np(1-p)$

Esempio 3.3.5. $X=\mathrm{Poisson}(\lambda)\implies \mathrm{Var}\left(\mathrm{Poisson}(\lambda)\right)=\lambda$

3.4 Esercizi

Esercizio 3.4.1.

Dati $r=10, 15=b, 7=n$ abbiamo che $\mathrm{P}\left(H(25,15,7)=3\right)$ è $\dfrac{\binom{15}{3}\binom{10}{7-3}}{\binom{25}{7}}$

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

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 $0 \le h \le 8$ lanci "croce" consecutivi e $9-h$ lanci "testa" consecutivi, in modo da ottenere una sequenza formata da $C^hT^{9-h}C$. La probabilità è $\mathrm{P}\left(C^hT^{9-h}C\right) = (0.45)^h \cdot (0.55)^{9-h} \cdot (0.45) = (0.45)^{h+1} \cdot (0.55)^{9-h}$. La probabilità dell'unione $\bigcup_h$ delle stringhe sarà $P = \sum_h (0.45)^{h+1}(0.55)^{9-h}$. Svolgere l'esercizio con una variabile geometrica.

Esercizio 3.4.2. Un ubriaco cammina in salita con probabilità $\mathrm{P}(\text{salita}) = 1/4$ oppure in discesa con probabilità $\mathrm{P}(\text{discesa}) = 3/4$. 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 $X \in [-6,+6]$. 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) $\Longrightarrow$ $Y = B(6,3/4)$. Abbiamo quindi che $X = -1\cdot Y + 1(6-Y) = 6-2Y$.

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

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 X

Esempio 3.4.3. Prendo un seme di carte francesi $\{A,2,\ldots,10,J,Q,K\}$ 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 $p_k = 1/13 \iff k = 2,\ldots,11$ e $p_k = 4/13 \iff k = 10$

Esercizio 3.4.4. È più probabile fare almeno un 6 lanciando 4 dadi o almeno una coppia di 6 lanciando 25 volte una coppia di dadi?

Esercizio 3.4.5. Siano date due slot machine apparentemente identiche A, B. La probabilità di vincere sulla A è $\mathrm{P}(\text{vincere sulla } A) = \frac{1}{2}$. La probabilità di vincere sulla B è $\mathrm{P}(\text{vincere sulla } B) = \frac{1}{4}$. Calcolare $\mathrm{P}(\text{aver giocato su } A \mid \text{aver vinto})$

Esercizio 3.4.6. 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 $\mathrm{P}(\text{seconda estrazione sia una pallina nera})$

Esercizio 3.4.7. 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 che dice di avere gli occhi scuri. Con che probabilità è bionda?

Utilizzando la formula di Bayes e di Fattorizzazione calcoliamo

$$\mathrm{P}(B \mid S) = \frac{\mathrm{P}(S \mid B) \cdot \mathrm{P}(B)}{\mathrm{P}(S)}$$
$$= \frac{PS \mid B \cdot \mathrm{P}(B)}{\mathrm{P}(S \mid B)\,\mathrm{P}(B) + \mathrm{P}(S \mid R)\,\mathrm{P}(R) + \mathrm{P}(S \mid M)\,\mathrm{P}(M)} \approx 0.41$$

Esercizio 3.4.8. 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 $\frac{R}{N}, \frac{R}{R}, \frac{N}{N}$ e sono sul tavolo coperte. Una è scoperta e la faccia visibile è Rossa. Calcolare $\mathrm{P}(\text{Faccia coperta } = R)$. Indichiamo con V la carta Visibile e con C la carta coperta.

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

Esercizio 3.4.9. Siano dati $A, B \subset \Omega$. Abbiamo che $\mathrm{P}(A) = \frac{3}{4}$ e abbiamo $\mathrm{P}(B) = \frac{1}{3}$. Possono essere disgiunti? No. Perché la probabilità della loro unione è maggiore di uno $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) = \frac{3}{4} + \frac{1}{3}$, che è > 1. Abbiamo che $A \cup B \subset \Omega$, ma $1 = \mathrm{P}(\Omega) > \mathrm{P}(A \cup B)$.

È vera la disuguaglianza $\frac{1}{12} \leq \mathrm{P}(A \cap B) \leq \frac{1}{3}$?

$$\begin{aligned} A \cap B \subseteq A \wedge A \cap B \subseteq B \\ (\mathrm{P}(A \cap B) \leq \mathrm{P}(A)) \wedge (\mathrm{P}(A \cap B) \leq \mathrm{P}(B))) \\ \implies \mathrm{P}(A \cap B) \leq \min\{\mathrm{P}(A), \mathrm{P}(B)\} \end{aligned}$$

Sappiamo quindi che $\mathrm{P}(A \cap B) \leq \frac{1}{3}$. Verifichiamo ora la prima parte della disuguaglianza

$$\begin{aligned} A = (A \cap B) \cup (A \cap B^C) \implies \mathrm{P}(A) = \mathrm{P}(A \cap B) + \mathrm{P}\left(A \cap B^C\right) \\ \mathrm{P}(A \cap B) = \mathrm{P}(A) - \mathrm{P}(A \cap B^c) \geq \mathrm{P}(A) - \mathrm{P}(B^c) \\ \mathrm{P}(A) - \mathrm{P}(B^c) = \mathrm{P}(A) - (1 - \mathrm{P}(B)) \\ = \mathrm{P}(A) + \mathrm{P}(B) - 1 = \frac{1}{3} + \frac{3}{4} - 1 = \frac{1}{12} \end{aligned}$$

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 aleatore 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 tutte 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: $\mathbb{N}, \mathbb{Z}, \{0, 1, 2, \ldots, n\}, \{t_1, t_2, \ldots, t_n\}$. 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 palline rosse e palline blu. Prendiamo in considerazione il colore della pallina alla prima, seconda, terza, ecc. estrazione. Il fenomeno è rappresentabile con una famiglia di variabili aleatorie.

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

I due processi cambiano radicalmente se le estrazioni sono con reimmissione della pallina o senza, ma sappiamo che le marginali hanno tutte le stesse leggi rispetto a $\mathrm{P}(\cdot \mid \Omega)$

$$\mathrm{P}(X_k = \text{rossa} \mid \Omega) = \frac{\#\text{ palline rosse}}{\#\text{ palline totali}}$$

Esempio 4.1.2. Assumiamo che le variabili $X_1, \ldots, X_n \in \{\text{rossa}, \text{blu}\}$ siano tutte indipendenti (rispetto a Ω). Supponiamo di conoscere il numero di palline totali ed il numero di palline rosse inizialmente. Supponiamo di aver fatto $k < n$ estrazioni e di conoscere il loro esito esatto. Poniamo ad esempio che siano state tutte rosse. Qual'è la probabilità che all'estrazione $k + 1$ otteniamo una pallina rossa?

$$\begin{aligned}\mathrm{P}(X_{k+1} = rossa \mid \Omega \cap \{X_1 = \text{rossa}, \ldots, X_k = \text{rossa}\}) \\ = P(X_{k+1} = \text{rossa} \mid \Omega) \text{ (Per indipendenza)} \\ = \frac{\#\text{ palline rosse}}{\#\text{ palline totali}}\end{aligned}$$

L'ipotesi di indipendenza probabilistica significa che non siamo capaci di "imparare" dal passato.

Definizione 4.1.2. Proprietà di Markov: Nei processi di Markov le informazioni ottenibili dal "passato" (la storia del processo fino al presente) possono essere utili a fare inferenza sullo stato futuro. In realtà costituiscono la classe più semplice in cui tutta la storia passata può essere trascurata ai fini di fare inferenza sul futuro.

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

Dato un processo a tempi $\mathcal{T}$ e stati E discreti. Un processo $\{X_{t_i}\}_{i=0,\dots,n}$ con $t_0 < t_1 < \ldots < t_n$ è detto **di Markov** (rispetto a $\mathrm{P}(\cdot \mid I)$) se, presi qualunque $k \in \{1, \ldots, n-1\}$ e $A_0, A_i, A_{k+1} \subseteq E$ vale la seguente proprietà

$$\begin{aligned}\mathrm{P}\left(X_{t_k+1} \in A_{k+1} \mid I \cap \{X_{t_k} \in A_k\} \cap \{X_{t_{k-1}} \in A_{k-1}\} \cap \{X_0 \in A_0\}\right) = \\ = \mathrm{P}\left(X_{t_{k+1}} \in A_{k+1} \mid I \cap \{X_{t_k} \in A_k\}\right)\end{aligned}$$

La proprietà di Markov permette di semplificare molto (ma non troppo) un modello probabilistico. L'indipendenza probabilistica va sempre considerata come un'ipotesi che introduciamo nel modello.

Più in generale, una variabile aleatoria del processo $X_{t_{k+1}}$ deve dipendere soltanto da X_{t_k}.

Nota. Dato che E è un insieme discreto, si potrebbe dimostrare che è sufficiente verificare la prorietà di Markov su insiemi A_k costituiti da singoli punti $A_k = \{i_k\}$, in modo tale che la condizione $\{X_{t_k} \in A_k\}$ diventi $\{X_{t_k} = i_k\}$

Esempio 4.1.3. Tornando all'esempio dell'estrazione dall'urna (di cui conosciamo il contenuto), 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).

Esempio 4.1.4. Vediamo un esempio di processo di Markov basato dalle estrazioni dall'urna contenente R palline rosse e B palline blu ($N = R + B$). Quando estraiamo la prima pallina, la teniamo all'esterno dell'urna. Successivamente estraiamo la seconda pallina, reinseriamo la prima pallina estratta e mescoliamo l'urna. Si procede poi tenendo fuori sempre l'ultima pallina estratta. Poniamo $X_k =$ il colore della pallina estratta all'estrazione k. La proprietà di Markov vale (l'informazione di tutta la sequenza di estrazioni non è rilevante eccetto l'ultima). Dati $i, j \subseteq E$, le probabilità di transizione

$$\mathrm{P}\left(X_{t_{k+1}} = j \mid I \cap X_{t_k} = i\right)$$

potrebbero in generale dipendere da k. Studieremo il caso in cui queste non dipendono da k per semplificare, e per semplificare ulteriormente assumiamo che $t_k = k$.

Definizione 4.1.3. Un process di Markov $X_{i i=0,\dots,n}$ è **omogeneo** se le probabilità di transizione non dipendono da $k \in \{0, 1, \ldots, n-1\}$, ovvero:

$$\forall i, j \in E \,.\, \mathrm{P}(X_{k+1} = j \mid I \cap \{X_k = i\}) = \mathrm{P}(X_1 = j \mid I \cap \{X_0 = i\})$$

Definizione 4.1.4. Matrice di Transizione: Questa definizione ci permette di collezionare le probabilità di transizione in una singola matrice di transizione

$$\forall i, j \in E \,.\, Q_{ij} = Q_{i \to j} := \mathrm{P}(X_1 = j \mid I \cap \{X_0 = i\})$$

Nota. Per scrivere una matrice di trasizione Q bisogna fissare un ordinamento degli stati E. Una volta fissato questo ordinamento va seguito in tutto il problema.

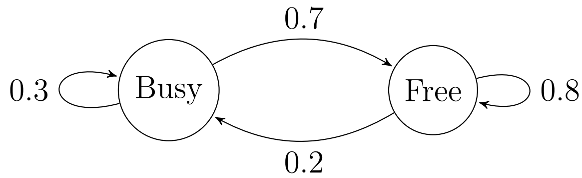
$$\sum_{j\in E} Q_{i\to j} = \sum_{j\in E} \mathrm{P}\left(X_1 = j \mid I \cap \{X_0 = i\}\right) = \mathrm{P}\left(X_1 \in E \mid I \cap \{X_0 = i\}\right) = 1$$

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 $Q_{i\to j}$ 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. 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 4.1: Catena di Markov



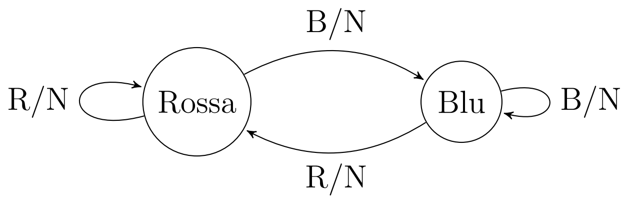
Esempio 4.1.6. Consideriamo un urna contenente R palline rosse e B palline blu, in tutto $N = R+B$ palline. Effettuiamo estrazioni con reimmissione. Abbiamo già visto che la proprietà di Markov vale. Le probabilità di transizione sono molto semplici da calcolare grazie all'indipendenza delle variabili $\{X_k\}$.

$$\begin{aligned}
\mathrm{P}\left(X_{k+1} = \text{rossa} \mid \Omega \cap \{X_k = \text{rossa}\}\right) = \mathrm{P}\left(X_{k+1} = \text{rossa} \mid \Omega\right) &= \frac{R}{N} \\
\mathrm{P}\left(X_{k+1} = \text{blu} \mid \Omega \cap \{X_k = \text{rossa}\}\right) = \mathrm{P}\left(X_{k+1} = \text{blu} \mid \Omega\right) &= \frac{B}{N} \\
\mathrm{P}\left(X_{k+1} = \text{rossa} \mid \Omega \cap \{X_k = \text{blu}\}\right) = \mathrm{P}\left(X_{k+1} = \text{rossa} \mid \Omega\right) &= \frac{R}{N} \\
\mathrm{P}\left(X_{k+1} = \text{blu} \mid \Omega \cap \{X_k = \text{blu}\}\right) = \mathrm{P}\left(X_{k+1} = \text{blu} \mid \Omega\right) &= \frac{B}{N}
\end{aligned}$$

La catena di Markov sugli stati $E = \{\text{rossa}, \text{blu}\}$ è omogenea con matrice di transizione.

$$Q = \begin{pmatrix} R/N & B/N \\ R/N & B/N \end{pmatrix}$$

Figura 4.2: Catena di Markov



4.2 Calcolo Algebrico su catene di Markov

Calcolo del Marginale di una Catena di Markov $\mathrm{P}(X_0 = j)\,\forall j$ è un vettore che chiamiamo $v = (\mathrm{P}(X_0 = j))$, è lo stato iniziale. Definiamo $Q(q_{ji})$ matrice di transizione. $q_{ji} = \mathrm{P}(X_1 = i \mid X_0 = j)$. Definiamo anche $q_{j\to i} = \mathrm{P}(X_1 = i \mid X_0 = j)$. Ne otteniamo che $\mathrm{P}(X_k = i) = (v \cdot Q^k)_i = (v_i = Q \cdot Q \cdot \ldots \cdot Q)_i$. Per correttezza, $v \cdot Q$ è la legge di X_1. Calcoliamo X_1.

$$\mathrm{P}(X_1 = i) \cdot \sum_j (\mathrm{P}(X_1 = i \mid X_0 = j))$$

$$\mathrm{P}(X_1 = j) = \sum_j q_{j\to i} v_j = \sum_j v_j q_{j\to i} = (v \cdot Q_i)$$

A volte, si può assegnare lo stato iniziale, ad esempio $X_0 = j \iff v = \{0, 0, 0, \ldots, 0, 1, \ldots, 0, 0, \} = e_j$ (significa che vi è un 1 in posizione j). Si ha che:

$$(e_i Q^k) \cdot (\mathrm{P}(X_k = l \mid X_0 = i)) = (e_i Q^k)_l = (Q^k)_{il}$$

Esempio 4.2.1. Riprendendo l'esempio delle palline rosse e blu:

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

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

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

Si ha che $\mathrm{P}(X_{10} = B) = (v \cdot Q^{10})_B = ((v \cdot Q)Q^9)_B = (v)_B = \dfrac{3}{13}$

Calcolo dei valori attesi Data $(X)_k$una Catena di Markov. Q matrice di transizione, $f : \mathbb{R} \to \mathbb{R}$ funzione reale, si ha che $\mathbb{E}[f(X_k) \mid X_0 = i]$. Prendiamo il caso $k = 1$

$$\begin{aligned} \mathbb{E}[f(X_1] \mid X_0 = i) &= \sum_j f(j) \cdot \mathrm{P}(x_1 = j \mid X_0 = i) \\ &= \sum_j f(j) \cdot q_{ij} = (Q \cdot f)_i \\ \vec{f} &= (f(j))_j \end{aligned}$$

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

Se non conosco lo stato iniziale devo conoscere la distribuzione $(P(X_0) = i)_i = v)$

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

Definizione 4.2.1. Distribuzione invariante

Cerchiamo di capire se uno stato di una catena di Markov è uno stato limite. La **distribuzione invariante** è un vettore $\vec{\mu} = (\mu_i)_i$ per Q matrice di transizione se

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

Ovvero $\vec{mu}^T$ è l'autovettore dell'autovalore 1 per Q^T

Per trovare μ^T si risolve $(Q^T - Id)\mu^T = 0$

Definizione 4.2.2. Catena Stazionaria
Una Catena di Markov $(x_k)_k$ è una **catena stazionaria** se **tutte** le marginali sono uguali:

$$\begin{cases} \exists \mu \text{ distribuzione invariante} \\ \mu Q^k = \mu \forall k \end{cases}$$

Ovvero se $\mathrm{P}(X_0) = \mathrm{P}(X_1) = \ldots = \mathrm{P}(X_k)$

Definizione 4.2.3. Matrice di Transizione regolare Una matrice di transizione Q si dice regolare se $\forall k \,.\, (Q^k)_{ij} > 0$. Se Q è regolare e v è uno stato iniziale qualsiasi allora vQ^k tende ad una qualche distribuzione limite e invariante.

4.3 Esercizi

Esercizio 4.3.1. Ho un'urna con $N = B + R$ biglie ($(2)B =$ Blu, $(1)R =$ Rosse)

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

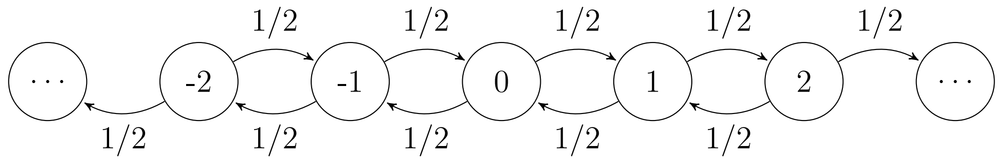
Lo stato iniziale $\mu = \left(\frac{R}{N}, \frac{B}{N}\right)$. Vogliamo sapere se μ è invariante.

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

Esercizio 4.3.2. 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.

$$Y_k = \text{Bern}\left(\frac{1}{2}\right) \text{ (lancio della moneta)}$$
$$Y_k \in \{-1, +1\}$$
$$X_k = \text{posizione}$$
$$\begin{cases} x_0 = 0 \\ x_{k+1} = x_k + y_k \end{cases}$$

Figura 4.3: Catena di Markov della passeggiata aleatoria



Esercizio 4.3.3. Un ubriaco è restio a cambiare direzione. Se al momento k è andato a sinistra, per $k+1$ la sinistra è più probabile della destra. La sua posizione è una catena di Markov? No, perché dipende dalla posizione all'istante precedente e dalla direzione.

Esercizio 4.3.4. 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:

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

$$Q^T - I = \begin{pmatrix} 0.3 & 0.2 \\ 0.7 & 0.8 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -0.7 & 0.2 \\ 0.7 & -0.2 \end{pmatrix}$$

$$\begin{pmatrix} -0.7 & 0.2 \\ 0.7 & -0.2 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = 0 \implies \begin{cases} \mu_1 + \mu_2 = 1 - 0.7\mu_1 + 0.2\mu_2 = 0 \\ 0.7\mu_1 - 0.2\mu_2 = 0 \\ \mu_1 \geq 0; \mu_2 \geq 0 \end{cases}$$

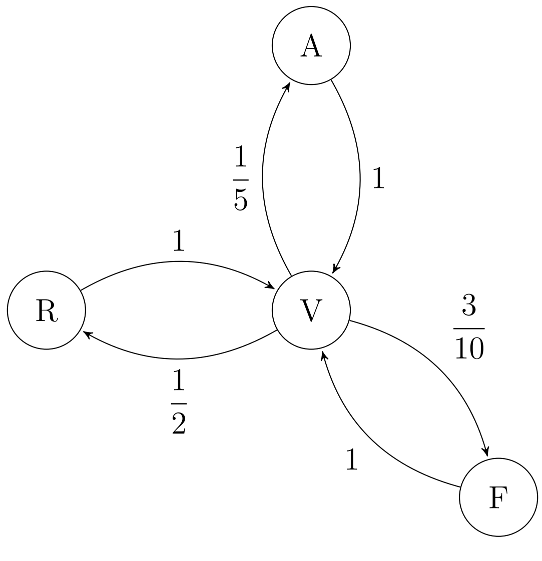
(Risolvendo il sistema si ottiene)

$$\mu_1 = \frac{0.2}{0.9} = \frac{2}{9}$$

$$\mu_2 = \frac{7}{9}$$

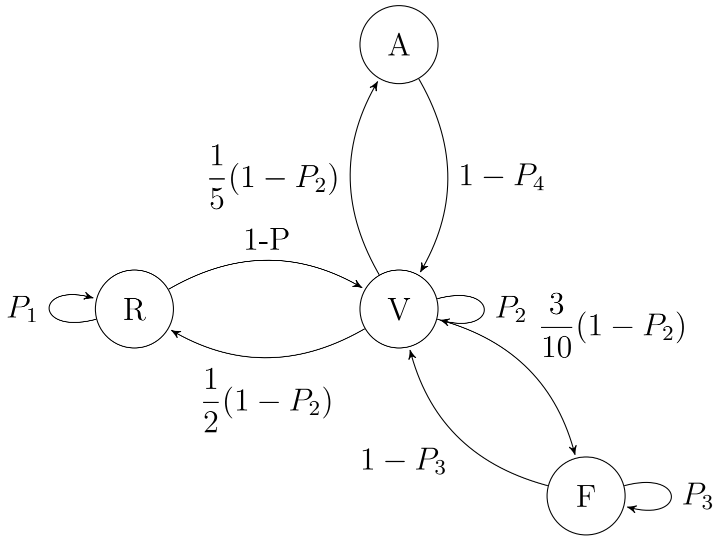
Esercizio 4.3.5. Vogliamo simulare un essere vivente elementare in un automa cellulare. I suoi stati sono (1) **relax**, (2) **vigile**, (3) **fuga**, (4) **attacca**

Figura 4.4: Catena di Markov dell'automa cellulare



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

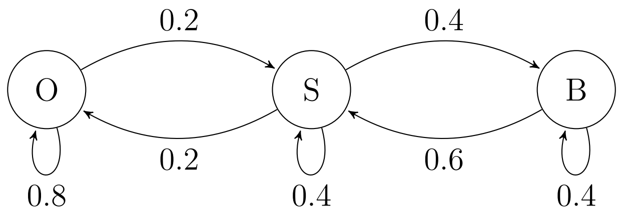
Figura 4.5: Catena di Markov dell'automa cellulare con tempo



Esercizio 4.3.6. content

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.

2. Calcolare $\mathrm{P}(X_1 = O \mid X_0 = O)$ e $\mathrm{P}(X_2 = O \mid X_0 = O)$. Abbiamo

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

$$q_{00} \to \mathrm{P}(X_1 = O \mid X_0 = O) = 0.8$$

$$\text{Calcoliamo ora } \mathrm{P}(X_2 = 0 \mid X_0 = 0)$$

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

$$v \cdot Q \cdot Q = (0.68, \ldots, \ldots) \text{ (Marginale della legge } X_2)$$

3. Trovare i costi $c \to 0$ per O, $c \to 5$ per S, $c \to 10$ per B. Calcolare $\mathbb{E}\left[c(X_k) \mid X_0 = O\right]$ per $k = 1, 2$. Per $k = 1$ si ha che

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

Per $k = 2$ si ha che

$$\begin{aligned} f = (0, 5, 10) \\ \mathbb{E}\left[c(X_2) \mid X_0 = O\right] = (Q^2 \cdot f)_1 = vQ^2 f = \ldots = 2 \end{aligned}$$

4. Calcolare la varianza

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

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

$$\text{(Calcoliamo } \mu\text{)}$$

$$\begin{aligned} (Q^T - I)\mu^T = 0 \\ \mu = (\mu_1, \mu_2, \mu_3) \\ (Q^T - I)\mu^T = \begin{pmatrix} -0.2 & 0.2 & 0 \\ 0.2 & -0.6 & 0.6 \\ 0 & 0.4 & -0.6 \end{pmatrix} \mu^T = 0 \end{aligned}$$

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

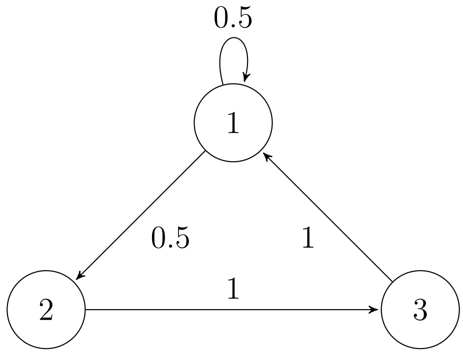
$$\Longrightarrow \mu = \left(\frac{3}{8}, \frac{3}{8}, \frac{1}{4}\right)$$

$$\text{(Calcoliamo } \mathbb{E}\left[c(X_1) \mid \mu\right]\text{)}$$

$$\begin{aligned} \mathbb{E}\left[c(X_1) \mid \mu\right] = \mu \cdot Q \cdot f \\ \text{(si può calcolare anche con) } \mathbb{E}\left[c(X_1) \mid \mu\right] = 0 \cdot \mathrm{P}\left(X_1 = O \mid \mu\right) \\ +5 \cdot \mathrm{P}\left(X_1 = S \mid \mu\right) + 10 \cdot \mathrm{P}\left(X_1 = B \mid \mu\right) \\ \text{(Si prosegue per fattorizzazione)} \end{aligned}$$

Esercizio 4.3.7. 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 $X_0 = 1$

$$v = (1, 0, 0)$$
$$\mathrm{P}\left(X_1 \mid X_0 = 1\right) = v \cdot Q = \left(\frac{1}{2}, \frac{1}{2}, 0\right)$$
$$\mathrm{P}\left(X_2 \mid X_0 = 1\right) = v \cdot Q^2 = \left(\frac{1}{2}, \frac{1}{2}, 0\right) Q = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right)$$
$$\mathrm{P}\left(X_3 \mid X_0 = 1\right) = v \cdot Q^3 = \left(\frac{1}{2}, \frac{1}{2}, 0\right) Q^2 = \left(\frac{5}{8}, \frac{1}{8}, \frac{1}{4}\right)$$

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

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

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

$$\iff \begin{cases} \mathrm{P}\left(X_0 = 1\right) = \frac{1}{2} \\ \mathrm{P}\left(X_0 = 2\right) = \frac{1}{4} \\ \mathrm{P}\left(X_0 = 3\right) = \frac{1}{4} \end{cases}$$

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

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

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

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

$$\implies \mathrm{P}\left(X_1 = 1 \mid X_3 = 1\right) = \mathrm{P}\left(X_3 = 1 \mid X_1 = 1\right)$$
$$= \mathrm{P}\left(X_2 = 1 \mid X_0 = 1\right) = (1, 0, 0)Q^2 = \frac{1}{2}$$