

The background is a complex, abstract composition of various geometric shapes and vibrant colors. It features a mix of reds, oranges, yellows, blues, and purples, creating a dynamic and somewhat chaotic visual field. The shapes are layered and overlapping, giving a sense of depth and movement. The overall style is reminiscent of a modern, expressive painting or a digital collage.

# Dispense di Calcolo delle Probabilità

Emanuele Izzo, 0307385

Corso di laurea magistrale Informatica  
Università Tor Vergata, Facoltà di Scienze MM.FF.NN.  
12/10/2021

Documento realizzato in  $\text{\LaTeX}$



# Indice

<b>I</b>	<b>Calcolo delle probabilità</b>	<b>3</b>
<b>1</b>	<b>Introduzione</b>	<b>4</b>
1.1	Gli eventi e la probabilità . . . . .	4
1.2	Unione, intersezione e dipendenza tra eventi . . . . .	5
1.2.1	Intersezione di eventi . . . . .	5
1.2.2	Union bound . . . . .	5
1.2.3	Principio di inclusione-esclusione . . . . .	6
1.3	Indipendenza e dipendenza tra eventi . . . . .	7
1.3.1	Formula delle probabilità totali . . . . .	8
1.3.2	Formula di Bayes . . . . .	8
<b>2</b>	<b>Variabili aleatorie discrete e media</b>	<b>9</b>
2.1	Le variabili aleatorie discrete . . . . .	9
2.2	La media . . . . .	10
2.2.1	Linearità della media . . . . .	11
2.2.2	Disuguaglianza di Jensen . . . . .	12
2.2.3	Disuguaglianza di Markov . . . . .	13
2.3	La media condizionata . . . . .	14
<b>3</b>	<b>Momenti, varianze e deviazioni</b>	<b>16</b>
3.1	Il momento $k$ -esimo . . . . .	16
3.2	La varianza e la covarianza . . . . .	16
3.2.1	Disuguaglianza di Chebyshev . . . . .	19
<b>4</b>	<b>Funzioni generatrici e Chernoff bounds</b>	<b>21</b>
4.1	Le funzioni generatrici dei momenti . . . . .	21
4.2	I Chernoff Bounds . . . . .	23
4.2.1	Processo di Poisson . . . . .	23
4.2.2	Stime di processi di Poisson con Chernoff Bounds . . . . .	24
<b>II</b>	<b>Distribuzioni notevoli</b>	<b>28</b>
<b>5</b>	<b>Distribuzione di Bernoulli</b>	<b>29</b>
5.1	Definizione . . . . .	29
5.2	Media . . . . .	29
5.3	Varianza . . . . .	29
5.4	Funzione generatrice dei momenti . . . . .	30

<b>6</b>	<b>Distribuzione binomiale</b>	<b>31</b>
6.1	Definizione . . . . .	31
6.2	Media . . . . .	31
6.3	Varianza . . . . .	32
6.4	Funzione generatrice dei momenti . . . . .	33
<b>7</b>	<b>Distribuzione geometrica</b>	<b>35</b>
7.1	Definizione . . . . .	35
7.2	Media . . . . .	36
7.3	Varianza . . . . .	37
7.4	Funzione generatrice dei momenti . . . . .	38
7.5	Mancanza di memoria della distribuzione geometrica . . . . .	38
<b>8</b>	<b>Distribuzione di Poisson</b>	<b>40</b>
8.1	Definizione . . . . .	40
8.2	Media . . . . .	40
8.3	Varianza . . . . .	41
8.4	Funzione generatrice dei momenti . . . . .	41
<b>III</b>	<b>Catene di Markov</b>	<b>42</b>
<b>9</b>	<b>Introduzione</b>	<b>43</b>
9.1	Le catene di Markov . . . . .	43
9.2	Comunicazione tra due stati e irriducibilità . . . . .	46
9.3	Gli stati ricorrenti e transienti . . . . .	47
9.4	La periodicità e l'aperiodicità per stati e catena . . . . .	48
<b>10</b>	<b>Distribuzioni stazionarie</b>	<b>49</b>
<b>11</b>	<b>Passeggiata aleatoria su grafi non orientati</b>	<b>54</b>
<b>IV</b>	<b>Esercizi</b>	<b>56</b>
<b>A</b>	<b>Esercizi sul calcolo delle probabilità</b>	<b>57</b>
<b>B</b>	<b>Esercizi sulle catene di Markov</b>	<b>58</b>

# Parte I

## Calcolo delle probabilità

# Capitolo 1

## Introduzione

### 1.1 Gli eventi e la probabilità

**Definizione:** Uno **spazio di probabilità** è una terna  $(\Omega, \mathcal{F}, P)$  dove:

- $\Omega$  è un insieme non vuoto (che rappresenta l'insieme dei possibili risultati del fenomeno aleatorio in esame).
- $\mathcal{F}$  è una famiglia di sottoninsiemi di  $\Omega$  detti **eventi**.
- $P$  è una *misura di probabilità*.

In genere si suppone che la famiglia  $\mathcal{F}$  abbia proprietà ragionevoli, cioè che  $\Omega \in \mathcal{F}$  e che presenta una *chiusura rispetto alle operazioni insiemistiche*: ciò permette di identificare l'insieme  $\mathcal{F}$  come una  **$\sigma$ -algebra**. Nel caso in cui  $\Omega$  sia un insieme discreto, cioè finito o numerabile, possiamo presupporre che  $\mathcal{F}$  sia l'**insieme delle parti** di  $\Omega$ .

**Definizione:** Una **misura di probabilità**  $P$  è una funzione  $P : \mathcal{F} \rightarrow [0, \infty)$  tale che  $P(\Omega) = 1$  e, per ogni successione di eventi  $\{E_n\}_n$  disgiunti a due a due (cioè  $E_i \cap E_j = \emptyset$  per  $i \neq j$ ), si ha  $P(\cup_n E_n) = \sum_n P(E_n)$ .

Tramite il teorema appena definito è possibile dimostrare le seguenti affermazioni:

- $\emptyset \in \mathcal{F}$  e  $P(\emptyset) = 0$ .
- Per ogni scelta di  $E, F \in \mathcal{F}$  tali che  $E \subset F$  si ha che  $P(E) \leq P(F)$ .
- Per ogni scelta di  $E \in \mathcal{F}$  si ha che  $P(E) \in [0, 1]$  ed inoltre si ha che  $P(E) = 1 - P(E^C)$  e  $P(E^C) = 1 - P(E)$ .
- Per ogni famiglia finita di eventi  $\{E_n\}_n$  disgiunti a due a due, si ha che  $P(\cup_n E_n) = \sum_n P(E_n)$ .
- Per ogni scelta di  $E, F \in \mathcal{F}$  si ha che  $P(E) = P(E \cap F) + P(E \cap F^C)$ .

Notare che nel caso in cui  $\Omega = \cup_n \{\omega_n\}$ , per  $n$  che appartiene ad un insieme finito o numerabile, l'insieme dei valori  $\{P(\{\omega_n\})\}_n$  consente di determinare il valore di  $P(E)$  per ogni  $E \in \mathcal{F}$ ; infatti:

$$E = \cup_{n:\omega_n \in E} \{\omega_n\} \text{ implica } P(E) = \sum_{n:\omega_n \in E} P(\{\omega_n\})$$

Questo perché  $E = \cup_{n:\omega_n \in E} \{\omega_n\}$  è un'unione di insieme disgiunti a due a due. Spesso nei risultati generati che vengono forniti si sottintende di avere uno spazio di probabilità  $(\Omega, \mathcal{F}, P)$  assegnato, non necessariamente con  $\Omega$  finito o numerabile.

## 1.2 Unione, intersezione e dipendenza tra eventi

### 1.2.1 Intersezione di eventi

**Teorema (1.1):** Per ogni scelta di  $E_1, E_2 \in \mathcal{F}$  si ha  $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

**Teorema (1.1):** Osserviamo che  $E_1 \cup E_2 = (E_1 \cap E_2^C) \cup (E_1 \cap E_2) \cup (E_1^C \cap E_2)$ , la quale è l'unione di eventi disgiunti a due a due, pertanto:

$$P(E_1 \cup E_2) = P(E_1 \cap E_2^C) + P(E_1 \cap E_2) + P(E_1^C \cap E_2)$$

Allora, sommando e sottraendo  $P(E_1 \cap E_2)$ , si ha:

$$P(E_1 \cup E_2) = \underbrace{P(E_1 \cap E_2^C) + P(E_1 \cap E_2)}_{P(E_1)} + \underbrace{P(E_1 \cap E_2) + P(E_1^C \cap E_2)}_{P(E_2)} - P(E_1 \cap E_2)$$

### 1.2.2 Union bound

**Lemma (1.1a):** Per ogni famiglia finita o successione  $\{E_n\}_n$  di elementi di  $\mathcal{F}$  si ha  $P(\cup_n E_n) \leq \sum_n P(E_n)$

**Lemma (1.1a):** Consideriamo i seguenti eventi  $\{\hat{E}_n\}_n$ :

$$\hat{E}_1 = E_1$$

$$\hat{E}_n = E_n \cap (E_1 \cup E_2 \cup \dots \cup E_{n-1})^C \text{ per } n \geq 2$$

Allora si ha  $\cup_n E_n = \cup_n \hat{E}_n$  e gli eventi  $\{\hat{E}_n\}_n$  sono disgiunti a due a due. Quindi,

essendo anche per costruzione  $\hat{E}_n \subset E_n$  per ogni scelta di  $n$ , si ha:

$$P(\cup_n E_n) = P(\cup_n \hat{E}_n) = \sum_n P(\hat{E}_n) \leq \sum_n P(E_n)$$

### 1.2.3 Principio di inclusione-esclusione

**Lemma (1.1b):** Per ogni famiglia finita o successione  $\{E_n\}_n$  di elementi di  $\mathcal{F}$  si ha che:

$$\begin{aligned} P(E_1 \cup E_2 \cup \dots \cup E_n) &= \sum_i P(E_i) - \sum_{i < j} P(E_i \cap E_j) + \\ &+ \sum_{i < j < k} P(E_i \cap E_j \cap E_k) + \dots + (-1)^{n-1} P(E_1 \cap E_2 \cap \dots \cap E_n) \end{aligned}$$

**Lemma (1.1b):** Procediamo per induzione su  $n$ :

- Per  $n = 2$  il risultato è vero per il lemma (1.2a).
- Per  $n \geq 3$  applichiamo il passo induttivo prendendo che la formula sia vera per  $n - 1$ . Con riferimento al caso  $n = 2$  per  $E_1 \cup E_2 \cup \dots \cup E_{n-1}$  e  $E_n$  si ha:

$$\begin{aligned} P(E_1 \cup E_2 \cup \dots \cup E_{n-1} \cup E_n) &= \\ &= P(E_1 \cup E_2 \cup \dots \cup E_{n-1}) + P(E_n) - P((E_1 \cup E_2 \cup \dots \cup E_{n-1}) \cap E_n) = \\ &= P(E_1 \cup E_2 \cup \dots \cup E_{n-1}) + P(E_n) - P(\cup_{i=1}^{n-1} (E_i \cap E_n)) \end{aligned}$$

A questo punto si usa la formula per il caso  $n - 1$  per il primo e il terzo termine. Quindi si ha:

$$\begin{aligned} P(E_1 \cup E_2 \cup \dots \cup E_{n-1} \cup E_n) &= \\ &= \sum_i P(E_i) - \sum_{i < j} P(E_i \cap E_j) + \sum_{i < j < k} P(E_i \cap E_j \cap E_k) + \dots \\ &\quad \dots + (-1)^{n-1-1} P(\cap_{i=1}^{n-1} E_i) + P(E_n) + \\ &\quad - \{ \sum_i P(E_i \cap E_n) - \sum_{i < j} P(E_i \cap E_j \cap E_n) + \dots \\ &\quad \dots + (-1)^{n-1-1} P(\cap_{i=1}^n (E_i \cap E_n)) \} = \\ &= \sum_i P(E_i) - \sum_{i < j} P(E_i \cap E_j) + \sum_{i < j < k} P(E_i \cap E_j \cap E_k) + \dots \\ &\quad \dots + (-1)^{n-1-1} P(\cap_{i=1}^{n-1} E_i) + P(E_n) + \\ &\quad - \{ \sum_i P(E_i \cap E_n) - \sum_{i < j} P(E_i \cap E_j \cap E_n) + \dots \\ &\quad \dots + (-1)^{n-1-1} P(\cap_{i=1}^n E_i) \} \end{aligned}$$



A questo punto non resta che combinare i seguenti termini:

$$\begin{aligned}
& \sum_i P(E_i) \text{ e } P(E_n) \\
& \sum_{i < j} P(E_i \cap E_j) \text{ e } \sum_i P(E_i \cap E_n) \\
& \sum_{i < j < k} P(E_i \cap E_j \cap E_k) \text{ e } \sum_{i < j} P(E_i \cap E_j \cap E_n) \\
& \dots \\
& - (-1)^{n-1-1} P(E_1 \cap E_2 \cap \dots \cap E_{n-1} \cap E_n) = \\
& = (-1)^{n-1} P(E_1 \cap E_2 \cap \dots \cap E_{n-1} \cap E_n)
\end{aligned}$$

### 1.3 Indipendenza e dipendenza tra eventi

**Definizione:** Siano due eventi  $E, F \in \mathcal{F}$ , essi sono **indipendenti** se  $P(E \cap F) = P(E)P(F)$ .

In generale, data un'arbitraria famiglia di eventi  $\{E_n\}_n$ , abbiamo eventi indipendenti se, presa una qualsiasi sottofamiglia finita con almeno due eventi  $E_{i_1}, E_{i_2}, \dots, E_{i_k}$  con  $k \geq 2$ , si ha che:

$$P(E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_k}) = P(E_{i_1})P(E_{i_2})\dots P(E_{i_k})$$

**Definizione:** Dati due eventi  $E, F \in \mathcal{F}$  tali che  $P(F) \neq 0$ , la **probabilità condizionata** di  $E$  sapendo che si è verificato  $F$  è definita come segue:

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

Spesso viene usata la terminologia semplificata "*probabilità condizionata di  $E$  dato  $F$* ". In generale è come se si passasse da  $\Omega$  a  $F$  e quindi si normalizza dividendo per  $P(F)$  in modo che  $P(\Omega|F) = 1$ . Se  $P(F) \neq 0$ , allora  $E$  ed  $F$  sono **indipendenti** se e solo se  $P(E|F) = P(E)$ . È possibile usare la formula della probabilità condizionata per calcolare la probabilità dell'intersezione di due eventi:

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

Questa uguaglianza vale anche se  $P(F) = 0$  perché. anche se non possiamo definire  $P(E|F)$ , possiamo dire che:

$$P(E \cap F) = P(E|F) \underbrace{P(F)}_{=0} = 0$$

### 1.3.1 Formula delle probabilità totali

**Teorema (1.2) :** Sia  $\{E_n\}_n$  una famiglia finita o una successione di eventi disgiunti a due a due, e tali che  $\cup_n E_n = \Omega$  (ossia  $\{E_n\}_n$  è una partizione finita o numerabile). Allora, per ogni  $B \in \mathcal{F}$ , si ha che:

$$P(B) = \sum_n P(B|E_n)P(E_n)$$

**Teorema (1.2):** Osservando che gli eventi  $\{B \cap E_n\}_n$  sono disgiunti a due a due, si ha:

$$\begin{aligned} P(B) &= P(B \cap \Omega) = P(B \cap (\cup_n E_n)) = \\ &= P(\cup_n (B \cap E_n)) = \\ &= \sum_n P(B \cap E_n) = \\ &= \sum_n P(B|E_n)P(E_n) \end{aligned}$$

### 1.3.2 Formula di Bayes

**Teorema (1.3):** Sia  $\{E_n\}_n$  una famiglia finita o una successione di eventi disgiunti a due a due, e tali che  $\cup_n E_n = \Omega$  (ossia  $\{E_n\}_n$  è una partizione finita o numerabile). Supponendo che  $P(B) \neq 0$ , si ha che:

$$P(E_j|B) = \frac{P(B|E_j)P(E_j)}{\sum_n P(B|E_n)P(E_n)} \text{ per ogni indice } j$$

**Teorema (1.3):** Si tratta di una verifica diretta, usando la definizione della probabilità condizionata e la formula delle probabilità totali:

$$P(E_j|B) = \frac{P(B \cap E_j)}{P(B)} = \frac{P(B|E_j)P(E_j)}{\sum_n P(B|E_n)P(E_n)} \text{ per ogni indice } j$$

# Capitolo 2

## Variabili aleatorie discrete e media

### 2.1 Le variabili aleatorie discrete

**Definizione:** Una **variabile aleatoria** è un'applicazione  $X : \Omega \rightarrow \mathbb{R}$  con opportune proprietà (le controimmagini di intervalli finiti appartengono alla  $\sigma$ -algebra). Una variabile aleatoria è **discreta** se l'insieme dei valori assunti dalla variabile aleatoria è finito o numerabile.

Per come è definita una variabile aleatoria discreta, per ogni  $a \in \mathbb{R}$  l'evento:

$$\{X = a\} := \{\omega \in \Omega : X(\omega) = a\}$$

È costituito da un insieme finito o numerabile di punti, e quindi si ha:

$$P(X = a) = \sum_{\omega \in \Omega : X(\omega) = a} P(\{\omega\})$$

**Definizione:** Due variabili aleatorie discrete  $X, Y$  (definite su uno stesso spazio di probabilità) sono **indipendenti** se:

$$P(\{X = x\} \cap \{Y = y\}) = P(X = x)P(Y = y) \text{ per ogni } x, y \in \mathbb{R}$$

La definizione si estende al caso più di due variabili aleatorie: precisamente, date  $\{X_i\}_{i \in I}$ , queste sono indipendenti se, per ogni sottoinsieme finito  $\{i_1, i_2, \dots, i_k\} \subset I$  con  $k \geq 2$ , si ha che:

$$\begin{aligned} P(\{X_{i_1} = x_{i_1}\} \cap \{X_{i_2} = x_{i_2}\} \cap \dots \cap \{X_{i_n} = x_{i_n}\}) &= \\ &= P(X_{i_1} = x_{i_1})P(X_{i_2} = x_{i_2}) \dots P(X_{i_n} = x_{i_n}) \end{aligned}$$

Nel caso in cui si ha  $I = \{1, 2, \dots, n\}$ , basta chiedere:

$$P(\{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \dots \cap \{X_n = x_n\}) = P(X_1 = x_1)P(X_2 = x_2) \dots P(X_n = x_n)$$

## 2.2 La media

**Definizione:** Una variabile aleatoria discreta  $X$  ha **valore atteso** o **media** finita se:

$$\sum_{x \in X} |x| P(X = x) < \infty$$

In tal caso il valore atteso è uguale a:

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

Possiamo notare che se  $X$  assume un numero finito di valori, la condizione per l'esistenza del valore atteso è sicuramente verificata. È possibile alterare la definizione di media per estenderla anche nel caso in cui la condizione  $\sum_{x \in X} |x| P(X = x) < \infty$  non sia soddisfatta.

**Definizione:** Sia  $X$  una variabile aleatoria a valori non negativi, allora:

$$\mathbb{E}[X] = \sum_{n \geq 1} P(X \geq n)$$

**Definizione:** Partendo dal secondo membro abbiamo che:

$$\sum_{n \geq 1} P(X \geq n) = \sum_{k \geq 1} \sum_{n \geq 1} P(X = k)$$

Allora, scambiando opportunamente l'ordine delle sommatorie (ciò è possibile poiché si hanno termini non negativi), si ha:

$$\begin{aligned} \sum_{n \geq 1} P(X \geq n) &= \sum_{k \geq 1} \sum_{n=1}^k P(X = k) = \\ &= \sum_{k \geq 1} P(X = k) \underbrace{\sum_{n=1}^k 1}_k = \\ &= \sum_{k \geq 1} k P(X = k) = \\ &= \sum_{k \geq 0} k P(X = k) = \\ &= \mathbb{E}[X] \end{aligned}$$

### 2.2.1 Linearità della media

**Teorema (2.1):** Siano  $X_1, X_2, \dots, X_n$  variabili aleatorie discrete definite su uno stesso spazio di probabilità, e con valore atteso finito. Allora, per ogni  $c_1, c_2, \dots, c_n \in \mathbb{R}$ , la variabile aleatoria  $\sum_{i=1}^n c_i X_i$  ha speranza matematica finita e si ha:

$$\mathbb{E}\left[\sum_{i=1}^n c_i X_i\right] = \sum_{i=1}^n c_i \mathbb{E}[X_i]$$

**Teorema (2.1):** La condizione di valore atteso finito è soddisfatta poiché

$$\begin{aligned} \sum_{x_1, x_2, \dots, x_n} |c_1 x_1 + c_2 x_2 + \dots + c_n x_n| P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) &\leq \\ &\leq \sum_{i=1}^n |c_i| \underbrace{\sum_{x_i} |x_i| P(X_i = x_i)}_{< \infty} < \infty \end{aligned}$$

Inoltre si ha:

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^n c_i X_i\right] &= \sum_{x_1, x_2, \dots, x_n} (c_1 x_1 + c_2 x_2 + \dots + c_n x_n) P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \\ &= c_1 \sum_{x_1, x_2, \dots, x_n} x_1 P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) + \\ &\quad + c_2 \sum_{x_1, x_2, \dots, x_n} x_2 P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) + \dots + \\ &\quad + c_n \sum_{x_1, x_2, \dots, x_n} x_n P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \end{aligned}$$

In generale, per ogni  $i \in \{1, 2, \dots, n\}$ , nella  $i$ -esima sommatoria si ha:

$$\sum_{x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_i = x_i)$$

Quindi, ritornando alla quantità da calcolare, si ha:

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^n c_i X_i\right] &= c_1 \sum_{x_1} x_1 P(X_1 = x_1) + c_2 \sum_{x_2} x_2 P(X_2 = x_2) + \dots + \\ &\quad + c_n \sum_{x_n} x_n P(X_n = x_n) = \sum_{i=1}^n c_i \mathbb{E}[X_i] \end{aligned}$$

La formula appena dimostrata non solo è valida anche nel caso di variabili aleatorie non indipendenti, ma è estendibile anche ad una famiglia di variabili aleatorie nume-

riche  $\{X_i\}_{i \geq 1}$ , supponendo che le costanti  $\{c_i\}_{i \geq 1}$  scelte arbitrariamente siano tali che  $\sum_{i \geq 1} |c_i| \mathbb{E}[X_i] < \infty$ .

**Teorema (2.2):** Siano  $X_1, X_2, \dots, X_n$  variabili aleatorie indipendenti, con valore atteso finito. Allora, la variabile aleatoria  $\prod_{i=1}^n X_i$ , cioè  $X_1 \cdot X_2 \cdot \dots \cdot X_n$ , ha valore atteso finito e si ha:

$$\mathbb{E}\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n \mathbb{E}[X_i]$$

**Teorema (2.2):** Supponiamo che il valore atteso del prodotto  $\prod_{i=1}^n X_i$  sia finito. Allora si ha:

$$\begin{aligned} \mathbb{E}\left[\prod_{i=1}^n X_i\right] &= \sum_{x_1, x_2, \dots, x_n} \prod_{i=1}^n x_i P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \\ &= \sum_{x_1, x_2, \dots, x_n} \prod_{i=1}^n x_i P(X_1 = x_1) P(X_2 = x_2) \dots P(X_n = x_n) = \\ &= \sum_{x_1, x_2, \dots, x_n} x_1 P(X_1 = x_1) x_2 P(X_2 = x_2) \dots x_n P(X_n = x_n) = \\ &= \underbrace{\sum_{x_1} x_1 P(X_1 = x_1)}_{\mathbb{E}[X_1]} \underbrace{\sum_{x_2} x_2 P(X_2 = x_2)}_{\mathbb{E}[X_2]} \dots \underbrace{\sum_{x_n} x_n P(X_n = x_n)}_{\mathbb{E}[X_n]} = \\ &= \prod_{i=1}^n \mathbb{E}[X_i] \end{aligned}$$

## 2.2.2 Disuguaglianza di Jensen

**Teorema (2.4):** Sia  $X$  una variabile aleatoria (non necessariamente discreta). Sia  $f : \mathbb{R} \rightarrow \mathbb{R}$  una funzione convessa. Allora  $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$ .

**Teorema (2.4):** Vogliamo considerare il caso restrittivo in cui  $f$  ammetta la possibilità di avere una formula di Taylor del secondo ordine, cioè:

$$f(x) = f(y) + f'(y)(x - y) + \frac{f''(c)}{2}(x - y)^2$$

Con  $c$  punto medio opportuno che appartiene all'intervallo  $[x, y]$ . Si vuole far riferimento a tale formula per  $y = \mathbb{E}[X]$ . Inoltre si sfrutta anche il fatto che, per

convessità di  $f$ , si ha  $f''(c) \geq 0$ . Quindi:

$$f(x) = f(\mathbb{E}[X]) + f'(\mathbb{E}[X])(x - \mathbb{E}[X]) + \frac{f''(c)}{2}(x - \mathbb{E}[X])^2$$

Da cui, essendo  $\frac{f''(c)}{2}(x - \mathbb{E}[X])^2 \geq 0$ , si ha:

$$f(x) \geq f(\mathbb{E}[X]) + f'(\mathbb{E}[X])(x - \mathbb{E}[X])$$

Notare che nel teorema sopra si fa riferimento ad una funzione convessa. Si dice che una funzione è *convessa* se per ogni  $x_1, x_2 \in \mathbb{R}$  e per ogni  $\lambda \in [0, 1]$  si ha che  $f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$ , ossia che la corda che congiunge i due punti distinti  $(x_1, f(x_1)), (x_2, f(x_2))$  è sopra al grafico della funzione  $f$  tra  $x_1$  e  $x_2$ . Un esempio di possibile funzione  $f$  convessa è  $f(x) = x^k$  con  $k$  intero positivo pari (per ogni  $x \in \mathbb{R}$ ).

### 2.2.3 Disuguaglianza di Markov

**Teorema (2.5):** Sia  $X$  una variabile aleatoria non negativa, allora:

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a} \text{ per ogni } a > 0$$

**Teorema (2.5):** Per ogni  $a > 0$  consideriamo la variabile aleatoria  $1_{X \geq a}$ , ossia la funzione indicatrice dell'evento  $\{X \geq a\}$ , definita come segue:

$$1_{X \geq a}(\omega) \begin{cases} 1 & \text{se } X(\omega) \geq a \\ 0 & \text{se } X(\omega) < a \end{cases}$$

Allora:

$$1_{X \geq a} \leq \frac{X(\omega)}{a} \text{ per ogni } \omega \in \Omega$$

Infatti:

- Se  $X(\omega) < a$  si ha  $0 \leq \frac{X(\omega)}{a}$  (ciò è dato dal fatto che  $X$  è una variabile aleatoria non negativa).
- Se  $X(\omega) \geq a > 0$  si ha  $1 \leq \frac{X(\omega)}{a}$ .

A questo punto, se prendiamo il valore atteso membro a membro, si ha:

$$\mathbb{E}[1_{X \geq a}] \leq \mathbb{E}\left[\frac{X}{a}\right] \rightarrow P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

## 2.3 La media condizionata

Possiamo dire che  $P(\cdot|A)$ , dato un qualsiasi evento  $A$ , è una misura di probabilità (per essere più precisi,  $P(\cdot|A)$  è una normalizzazione di  $P(\cdot)$  rispetto all'evento  $A$ ). È possibile quindi definire la media di una variabile aleatoria  $X$  condizionata rispetto ad un evento  $A$  utilizzando le probabilità condizionate.

**Definizione:** Una variabile aleatoria discreta  $X$  ha **valore atteso condizionato** o **media condizionata** rispetto all'evento  $A$  finita se:

$$\sum_{x \in X} |x| P(X = x|A) < \infty$$

In tal caso il valore atteso è uguale a:

$$\mathbb{E}[X|A] = \sum_{x \in X} x P(X = x|A)$$

In molti casi l'evento condizionante fa riferimento ad un'altra variabile aleatoria.

**Teorema (2.6):** Siano  $X, Y$  due variabili aleatorie definite su uno stesso spazio di probabilità. Allora se  $X$  ha speranza matematica finita, vale la seguente relazione:

$$\mathbb{E}[X] = \sum_{y: P(Y=y)>0} \mathbb{E}[X|Y=y] P(Y=y)$$

**Teorema (2.6):** Consideriamo il secondo membro, tenendo in conto l'ipotesi di valore atteso finito per  $X$ :

$$\begin{aligned} \sum_{y: P(Y=y)>0} \mathbb{E}[X|Y=y] P(Y=y) &= \sum_{y: P(Y=y)>0} \left( \sum_x x P(X=x|Y=y) \right) P(Y=y) = \\ &= \sum_x x \sum_{y: P(Y=y)>0} P(X=x|Y=y) P(Y=y) = \\ &= \sum_x x \sum_{y: P(Y=y)>0} P(X=x, Y=y) = \\ &= \sum_x x P(X=x) = \mathbb{E}[X] \end{aligned}$$

Osserviamo che il risultato appena ottenuto si può estendere ad una partizione di eventi generica al più numerabile  $\{A_k\}_{k \in I}$ , e non necessariamente a quella dei valori assunti dalla variabile aleatoria discreta  $Y$ :

$$\mathbb{E}[X] = \sum_{k: P(A_k)>0} \mathbb{E}[X|A_k] P(A_k)$$



Si può inoltre dire che la linearità vista per il valore atteso vale anche per le medie condizionate. Talvolta la media condizionata viene vista come una variabile aleatoria in funzione della variabile aleatoria condizionante la media. Precisamente, se  $Y$  assume un insieme di valori  $\{y\}_{y \in I}$  con  $I$  finito o numerabile, la variabile aleatoria  $\mathbb{E}[X|Y]$  è definita come segue:

$$\forall y \in I, \mathbb{E}[X|Y] = \mathbb{E}[X|Y = y] \text{ quando } Y = y$$

**Teorema (2.7):** Siano  $X, Y$  due variabili aleatorie definite su uno stesso spazio di probabilità. Allora se  $X$  ha speranza matematica finita, vale la seguente relazione:

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

# Capitolo 3

## Momenti, varianze e deviazioni

### 3.1 Il momento $k$ -esimo

**Definizione:** Si definisce **momento  $k$ -esimo** della variabile aleatoria  $X$  il valore atteso della variabile aleatoria  $X^k$  e quindi  $\mathbb{E}[X^k]$ .

### 3.2 La varianza e la covarianza

**Definizione:** Si definisce **varianza** della variabile aleatoria  $X$  il momento secondo della variabile aleatoria  $X - \mathbb{E}[X]$ , supponendo che  $X$  abbia valore atteso finito. La varianza di una variabile aleatoria  $X$  è pertanto definita come:

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

La varianza, essendo un valore atteso di una variabile aleatoria non negativa (essendo un quadrato) può essere finita o uguale a  $+\infty$ . Talvolta diremo che la varianza è ben definita quando  $X^2$  ha valore atteso finito. Nel caso in cui la varianza di  $X$  è ben definita, si definisce con **scarto quadratico medio** di  $X$  la seguente quantità:

$$\sigma(X) := \sqrt{Var[X]}$$

È possibile inoltre ricavare una formula alternativa per il calcolo della varianza:

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

A livello di informazioni, la varianza (e allo stesso modo lo scarto quadratico) è un grandezza non negativa che indica quanto la distribuzione di una variabile aleatoria  $X$  sia concentrata rispetto al suo valore atteso  $\mathbb{E}[X]$ . La concentrazione cresce al decrescere

della varianza, che in effetti assume il suo valore minimo zero se e solo se la variabile aleatoria  $X$  è costante (in altri termini se  $(P(X = c) = 1$  per qualche  $c \in \mathbb{R}$ , e in tal caso  $\mathbb{E}[X] = c$ ); Al contrario la varianza non ha un valore massimo. Ciò significa che la probabilità che  $X$  assuma valori "lontani da  $\mathbb{E}[X]$ " non può essere troppo grande se la varianza è "piccola".

**Definizione:** Siano  $X, Y$  due variabili aleatorie definite su uno stesso spazio di probabilità, si definisce **covarianza** (delle due variabili aleatorie) il valore atteso di  $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])$  (supponendo che tale variabile aleatoria abbia valore atteso finito). La covarianza di due variabili aleatorie  $X, Y$  è pertanto definita come:

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

Notare che la covarianza è ben definita quando  $X^2$  e  $Y^2$  hanno valore atteso finito. A differenza della varianza, la covarianza può assumere valori negativi. È possibile inoltre ricavare una formula alternativa per il calcolo della covarianza:

$$\begin{aligned} Cov(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \\ &= \mathbb{E}[XY - Y\mathbb{E}[X] - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]] = \\ &= \mathbb{E}[XY] - \mathbb{E}[Y\mathbb{E}[X]] - \mathbb{E}[X\mathbb{E}[Y]] + \mathbb{E}[\mathbb{E}[X]\mathbb{E}[Y]] = \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] = \\ &= \mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \end{aligned}$$

È possibile osservare due proprietà interessanti della covarianza:

- $Cov(X, Y) = Cov(Y, X)$ .
- $Cov(X, X) = Var[X]$ .

**Teorema (3.1):** Siano  $X_1, X_2, \dots, X_n$  variabili aleatorie definite su uno stesso spazio di probabilità. Allora, nel caso in cui  $X_1 + X_2 + \dots + X_n$  abbia varianza finita (questo accade se  $X_1^2, X_2^2, \dots, X_n^2$  hanno valore atteso finito), si hanno le seguenti formule tutte equivalenti tra loro:

$$\begin{aligned} Var[X_1 + X_2 + \dots + X_n] &= \sum_{i=1}^n Var[X_i] + 2 \sum_{j:j>i} Cov(X_i, X_j) \\ Var[X_1 + X_2 + \dots + X_n] &= \sum_{i=1}^n Var[X_i] + 2 \sum_{j:j<i} Cov(X_i, X_j) \\ Var[X_1 + X_2 + \dots + X_n] &= \sum_{i=1}^n Var[X_i] + \sum_{j:j \neq i} Cov(X_i, X_j) \end{aligned}$$

**Teorema (3.1):** Si ha:

$$\begin{aligned}
Var[X_1 + X_2 + \dots + X_n] &= \\
&= \mathbb{E}[(X_1 + X_2 + \dots + X_n - \mathbb{E}[X_1 + X_2 + \dots + X_n])^2] = \\
&= \mathbb{E}[(X_1 + X_2 + \dots + X_n - \mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n])^2] = \\
&= \mathbb{E}\left[\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right)^2\right] = \\
&= \mathbb{E}\left[\sum_{i=1}^n (X_i - \mathbb{E}[X_i])^2 + \sum_{j:j \neq i} (X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right] = \\
&= \mathbb{E}\left[\sum_{i=1}^n (X_i - \mathbb{E}[X_i])^2\right] + \mathbb{E}\left[\sum_{j:j \neq i} (X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right] = \\
&= \sum_{i=1}^n \mathbb{E}[(X_i - \mathbb{E}[X_i])^2] + \sum_{j:j \neq i} \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] = \\
&= \sum_{i=1}^n Var[X_i] + \sum_{j:j \neq i} Cov(X_i, X_j)
\end{aligned}$$

Notare che abbiamo verificato solo la terza formula: questo perchè le tre formule sono equivalenti per la simmetria della covarianza (è possibile dimostrare le altre due formule modificando leggermente la dimostrazione).

**Teorema (3.2):** Siano  $X_1, X_2, \dots, X_n$  variabili aleatorie definite su uno stesso spazio di probabilità. Allora, nel caso in cui  $X_1 + X_2 + \dots + X_n$  abbia varianza finita (questo accade se  $X_1^2, X_2^2, \dots, X_n^2$  hanno valore atteso finito), allora:

$$Var[X_1 + X_2 + \dots + X_n] = \sum_{i=1}^n Var[X_i]$$

Nel caso in cui si abbia una delle seguenti condizioni:

- $Cov(X_i, X_j) = 0$  per ogni  $i, j \in [1, n]$  con  $i \neq j$ .
- $X_i$  e  $X_j$  sono indipendenti per ogni  $i, j \in [1, n]$  con  $i \neq j$ .

**Teorema (3.2):** La dimostrazione per la prima condizione è banale, poiché si tratta di un caso speciale del teorema precedente. Per dimostrare che la seconda condizione porta allo stesso risultato basta mostrare che essa implica la prima condizione. Supponiamo che  $X_1, X_2, \dots, X_n$  siano indipendenti a due a due; allora, per ogni  $i, j \in [1, n]$  con  $i \neq j$ ,  $X_i$  e  $X_j$  sono indipendenti, e pertanto:

$$\mathbb{E}[X_i X_j] = \mathbb{E}[X_i] \mathbb{E}[X_j]$$

Che equivale a dire:

$$\text{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \mathbb{E}[X_j] = 0$$

**Teorema (3.3):** Siano  $X_1, X_2, \dots, X_n$  variabili aleatorie definite su uno stesso spazio di probabilità. Allora, nel caso in cui  $X_1 + X_2 + \dots + X_n$  abbia varianza finita (questo accade se  $X_1^2, X_2^2, \dots, X_n^2$  hanno valore atteso finito), allora:

$$\text{Var}[X_1 + X_2 + \dots + X_n] = \sum_{i=1}^n \text{Var}[X_i]$$

Nel caso in cui si  $X_1, X_2, \dots, X_n$  sono indipendenti.

Quest'ultimo teorema è un caso speciale che deriva direttamente dal teorema precedente (basta solo pensare al fatto che, se  $X_1, X_2, \dots, X_n$  sono indipendenti, allora lo sono anche a due a due).

### 3.2.1 Disuguaglianza di Chebyshev

**Teorema (3.4):**  $X$  una variabile aleatoria con valore atteso finito, allora:

$$P(|X - \mathbb{E}[X]| \geq a) \leq \frac{\text{Var}[X]}{a^2} \text{ per ogni } a > 0$$

**Teorema (3.4):** per ogni  $a > 0$  abbiamo la seguente uguaglianza tra eventi:

$$\{|X - \mathbb{E}[X]| \geq a\} = \{(X - \mathbb{E}[X])^2 \geq a^2\}$$

Quindi:

$$P(|X - \mathbb{E}[X]| \geq a) = P((X - \mathbb{E}[X])^2 \geq a^2)$$

A questo punto si applica la disuguaglianza di Markov alla variabile aleatoria  $(X - \mathbb{E}[X])^2$ , con  $a^2$  al posto di  $a$ :

$$P((X - \mathbb{E}[X])^2 \geq a^2) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{a^2} = \frac{\text{Var}[X]}{a^2}$$

**Teorema (3.5):**  $X$  una variabile aleatoria con valore atteso finito, allora:

- Se  $X$  ha anche varianza finita, si ha:

$$P(|X - \mathbb{E}[X]| \geq t\sigma(X)) \leq \frac{1}{t^2} \text{ per ogni } t > 0$$

- Si ha:

$$P(|X - \mathbb{E}[X]| \geq t\mathbb{E}[X]) \leq \frac{\text{Var}[X]}{t^2\mathbb{E}^2[X]} \text{ per ogni } t > 0$$

# Capitolo 4

## Funzioni generatrici e Chernoff bounds

### 4.1 Le funzioni generatrici dei momenti

**Definizione:** Si definisce **funzione generatrice dei momenti** (o **FGM**) di una variabile aleatoria  $X$  la seguente funzione:

$$M_X : \mathbb{R} \rightarrow (0, \infty) \cup \{\infty\}, M_X(t) := \mathbb{E}[e^{tX}]$$

In effetti tale funzione può assumere il valore infinito per certi valori di  $t$ . In ogni modo si ha sempre  $M_X(0) = 1$ , inoltre possiamo dire quanto segue:

- Se  $P(X \geq 0) = 1$ ,  $M_X$  è crescente e quindi  $M_X(t) \leq M_X(t) = 1 < \infty$  per ogni  $t \leq 0$ .
- Se  $P(X \leq 0) = 1$ ,  $M_X$  è decrescente e quindi  $M_X(t) \leq M_X(t) = 1 < \infty$  per ogni  $t \geq 0$ .
- Se  $X = 0$  si ha  $M_X(t) = 1$  per ogni  $t \in \mathbb{R}$ .
- Se  $X$  è limitata, cioè  $P(a \leq X \leq b) = 1$  per qualche  $a, b \in \mathbb{R}$ , allora  $M_X(t) < \infty$  per ogni  $t \in \mathbb{R}$ . Non vale però il viceversa: esistono variabili aleatorie per cui si ha  $M_X(t) < \infty$  per ogni  $t \in \mathbb{R}$ , ma non sono limitate. Ad esempio, consideriamo la variabile aleatoria di Poisson di parametro  $\lambda > 0$ , la sua funzione generatrice sarà:

$$\begin{aligned} M_X(t) &= \sum_{n \geq 0} e^{tn} P(X = n) = \\ &= \sum_{n \geq 0} e^{tn} \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= e^{-\lambda} \sum_{n \geq 0} \frac{(e^t \lambda)^n}{n!} = \\ &= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)} \end{aligned} \quad \text{per ogni } t \in \mathbb{R}$$

**Teorema (4.1):** Supponiamo che  $M_X$  sia finita in un intorno dell'origine (cioè esiste  $\delta > 0$  tale che  $M_X(t) < \infty$  per  $t \in (-\delta, \delta)$ ). Allora in tal caso si può derivare rispetto a  $t$  sotto il segno del valore atteso e si ha:

$$M_X^{(n)}(0) = \mathbb{E}[X^n] \text{ per ogni } n$$

**Teorema (4.1):** Supponiamo di poter derivare rispetto a  $t$  sotto il segno del valore atteso. Allora si vede che:

$$M_X^{(n)}(t) = \mathbb{E}[e^{tX} X^n] \text{ per ogni } n$$

L'uguaglianza sopra si ottiene semplicemente ponendo  $t = 0$ .

**Teorema (4.2):** Se due variabili aleatorie  $X$  e  $Y$  hanno una funzione generatrice dei momenti coincidente in un intorno dell'origine, cioè esiste un  $\delta > 0$  tale che  $M_X(t) = M_Y(t)$  (a valori finiti) per  $t \in (-\delta, \delta)$ , allora  $X$  e  $Y$  hanno la stessa distribuzione.

**Teorema (4.3):** Siano  $X_1, X_2, \dots, X_n$  variabili aleatorie indipendenti, allora:

$$M_{X_1+X_2+\dots+X_n} = \prod_{i=1}^n M_{X_i}(t)$$

**Teorema (4.3):** Osserviamo che:

$$M_{X_1+X_2+\dots+X_n} = \mathbb{E}[e^{t(X_1+X_2+\dots+X_n)}] = \mathbb{E}\left[\prod_{i=1}^n e^{tX_i}\right] = \prod_{i=1}^n \mathbb{E}[e^{tX_i}] = \prod_{i=1}^n M_{X_i}(t)$$

Questo è possibile poiché le variabili aleatorie  $e^{tX_1}, e^{tX_2}, \dots, e^{tX_n}$  sono indipendenti e poiché è vera per ogni  $t \in \mathbb{R}$  per cui  $M_{X_1+X_2+\dots+X_n} < \infty$  (in realtà si verifica per ogni  $t \in \mathbb{R}$  anche quando  $M_{X_1+X_2+\dots+X_n} = \infty$ ).

Possiamo dire che la somma di variabili aleatorie ha una certa distribuzione se riconosciamo l'espressione della sua funzione di generazione dei momenti (a patto che sia finita in un intorno dell'origine). Questo consente di trovare più facilmente la distribuzione della somma delle variabili aleatorie (a anche di calcolare i suoi momenti) rispetto a quello che si potrebbe fare a partire dai calcoli sulle distribuzioni delle singole variabili aleatorie.



## 4.2 I Chernoff Bounds

**Definizione:** Sia  $X$  una variabile aleatoria e sia  $a \in \mathbb{R}$ , allora:

- La stima della coda della distribuzione di  $X$  a  $+\infty$  è pari a:

$$P(X \geq a) \leq \frac{M_X(t)}{e^{ta}} \text{ per ogni } t > 0$$

Infatti si ha  $\{X \geq a\} = \{e^{tx} \geq e^{ta}\}$ , e pertanto:

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}} = \frac{M_X(t)}{e^{ta}}$$

A questo punto si tratta di scegliere  $t$  in maniera opportuna in modo da ottimizzare il secondo membro, per ottenere:

$$P(X \geq a) \leq \inf_{t>0} \frac{M_X(t)}{e^{ta}}$$

- La stima della coda della distribuzione di  $X$  a  $-\infty$  è pari a:

$$P(X \leq a) \leq \frac{M_X(t)}{e^{ta}} \text{ per ogni } t > 0$$

Infatti si ha  $\{X \leq a\} = \{e^{tx} \geq e^{ta}\}$ , e pertanto:

$$P(X \leq a) = P(e^{tX} \geq e^{ta}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}} = \frac{M_X(t)}{e^{ta}}$$

A questo punto si tratta di scegliere  $t$  in maniera opportuna in modo da ottimizzare il secondo membro, per ottenere:

$$P(X \leq a) \leq \inf_{t>0} \frac{M_X(t)}{e^{ta}}$$

### 4.2.1 Processo di Poisson

**Definizione:** Si definisce **processo di Poisson** una sequenza di variabili aleatorie  $X_1, X_2, \dots, X_n$  indipendenti, Bernoulliane e di parametri  $p_1, p_2, \dots, p_n \in [0, 1]$  rispettivamente che non sono necessariamente uguali.

Consideriamo la variabile aleatoria:

$$X = \sum_{i=1}^n X_i, \text{ con valore atteso } \mu := \mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n p_i$$

Nel caso in cui  $p_1 = p_2 = \dots = p_n$  si parla di **processo Bernoulliano** e  $X$  ha distribuzione binomiale. Consideriamo ora le funzioni generatrici dei momenti delle variabili aleatorie

$X_1, X_2, \dots, X_n$ ; avremo che:

$$\begin{aligned} M_{X_i}(t) &= \mathbb{E}[e^{tX_i}] = \\ &= e^{t \cdot 1} P(X_i = 1) + e^{t \cdot 0} P(X_i = 0) = \\ &= e^t p_i + 1 \cdot (1 - p_i) = 1 + p_i(e^t - 1) \end{aligned} \quad \text{per ogni } t \in \mathbb{R}$$

Allora si ha:

$$M_X(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n (1 + p_i(e^t - 1))$$

Inoltre, poiché  $e^x \geq 1 + x$  per ogni  $x \in \mathbb{R}$  (sapendo che si ha l'uguaglianza solo per  $x = 0$ ), possiamo dire che:

$$M_X(t) \leq \prod_{i=1}^n p_i^{p_i(e^t - 1)} = e^{\sum_{i=1}^n p_i(e^t - 1)} = e^{\mu(e^t - 1)}$$

Questa disuguaglianza diventa un'uguaglianza nei seguenti casi:

- Se  $p_1 = p_2 = \dots = p_n = 0$ , che equivale a dire  $\mu = 0$ ; in quel caso si ha  $X = 0$  e pertanto si verifica che  $M_X(t) = e^{\mu(e^t - 1)} = 1$  per ogni  $t \in \mathbb{R}$ .
- Qualunque sia il valore di  $p_1, p_2, \dots, p_n$  e quindi qualunque sia il valore di  $\mu$ , per  $t = 0$  si ha che  $M_X(0) = e^{\mu(e^0 - 1)} = 1$ .

## 4.2.2 Stime di processi di Poisson con Chernoff Bounds

**Teorema (4.4):** Consideriamo lo schema dei processi di Poisson. Allora abbiamo le seguenti stime:

- Per ogni  $\delta > 0$  si ha:

$$P(X \geq (1 + \delta)\mu) \leq \left(\frac{e^\delta}{(1 + \delta)^{1 + \delta}}\right)^\mu$$

- Per ogni  $\delta \in (0, 1]$  si ha:

$$P(X \geq (1 + \delta)\mu) \leq e^{-\frac{\mu\delta^2}{3}}$$

- Sia  $\mu \neq 0$ , per ogni  $R \geq 6\mu$  si ha:

$$P(X \geq R) < 2^{-R}$$

**Teorema (4.4):** Sia  $t > 0$ , allora si ha:

$$P(X \geq (1 + \delta)\mu) = P(e^{tX} \geq e^{t(1 + \delta)\mu}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1 + \delta)\mu}}$$

Da cui segue che:

$$P(X \geq (1 + \delta)\mu) \leq \frac{e^{\mu(e^t - 1)}}{e^{t(1 + \delta)\mu}} = e^{\mu(e^t - 1 - t(1 + \delta))}$$

A questo punto bisogna ottimizzare rispetto a  $t$ . Consideriamo la funzione:

$$f_\delta(t) := e^t - 1 - t(1 + \delta)$$

E consideriamo la sua derivata prima:

$$\begin{aligned} f'_\delta(t) &= e^t - (1 + \delta) > 0 \\ e^t &> 1 + \delta \\ t &> \ln(1 + \delta) := t_0(\delta) \quad (\text{si osservi che } t_0(\delta) > 0) \end{aligned}$$

Allora:

$$\inf_{t > 0} f_\delta(t) = f_\delta(t_0(\delta)) = 1 + \delta - 1 - (1 + \delta) \ln(1 + \delta) = \delta - (1 + \delta) \ln(1 + \delta)$$

In conclusione si ottiene:

$$P(X \geq (1 + \delta)\mu) \leq e^{\mu(\delta - (1 + \delta) \ln(1 + \delta))} = \left( \frac{e^\delta}{(1 + \delta)^{1 + \delta}} \right)^\mu$$

Ottenendo quindi il primo risultato. Per  $\delta \in (0, 1]$ , inoltre, si ha che:

$$\frac{e^\delta}{(1 + \delta)^{1 + \delta}} \leq e^{-\frac{\delta}{3}} \leftrightarrow \delta - (1 + \delta) \ln(1 + \delta) \leq -\frac{\delta^2}{3}$$

Consideriamo la funzione:

$$g(\delta) := \delta - (1 + \delta) \ln(1 + \delta) + \frac{\delta^2}{3}$$

E verifichiamo che  $g(\delta) \leq 0$  per ogni  $\delta \in (0, 1]$ . Prendiamo la derivata prima e seconda di  $g$ :

$$\begin{aligned} g'(\delta) &= 1 - (\ln(1 + \delta) + (1 + \delta) \frac{1}{1 + \delta}) + \frac{2\delta}{3} = -\ln(1 + \delta) + \frac{2\delta}{3} \\ g''(\delta) &= -\frac{1}{1 + \delta} + \frac{2}{3} \end{aligned}$$

Allora  $g''(\delta) > 0$  se e solo se:

$$\begin{aligned} -\frac{1}{1 + \delta} &> -\frac{2}{3} \\ \frac{1}{1 + \delta} &< \frac{2}{3} \\ 1 + \delta &> \frac{3}{2} \\ \delta &> \frac{3}{2} - 1 \\ \delta &> \frac{1}{2} \end{aligned}$$

Quindi  $g'(\delta)$  cresce per  $\delta \in (\frac{1}{2}, 1)$  e decresce per  $\delta \in (0, \frac{1}{2})$ ; precisamente  $g'(\delta)$  va da  $g'(0) = 0$  a  $g'(\frac{1}{2}) < 0$  e poi crescere fino a  $g'(1) = -\ln 2 + \frac{2}{3} < 0$ ; in conclusione possiamo dire che  $g'(\delta) \leq g'(0) = 0$  per ogni  $\delta \in (0, 1)$ , da cui segue che  $g$  è non crescente in  $\delta \in (0, 1)$ , e quindi la  $g$  è massimizzata da  $g(0) = 0$  come volevamo, ottenendo in secondo risultato. Prendendo  $R + (1 + \delta)\mu$ , ricordando che  $\mu \neq 0$ , si ottiene  $\delta = \frac{R}{\mu} - 1$ ; inoltre  $R \geq 6\mu$  implica  $1 + \delta \geq 6$ . Allora si ha:

$$\begin{aligned} P(X \geq R) &\leq \left( \frac{e^{\frac{R}{\mu}-1}}{(1+\delta)^{1+\frac{R}{\mu}-1}} \right) = \\ &= \frac{e^{R-\mu}}{(1+\delta)^R} < \\ &< \frac{e^R}{(1+\delta)^R} = \\ &= \left( \frac{e}{1+\delta} \right)^R \leq \\ &\leq \left( \frac{e}{6} \right)^R < \\ &< \left( \frac{1}{2} \right)^R = 2^{-R} \end{aligned}$$

Dimostrando quindi il terzo risultato.

**Teorema (4.5):** Consideriamo lo schema dei processi di Poisson. Allora abbiamo le seguenti stime:

- Per ogni  $\delta \in (0, 1]$  si ha:

$$P(X \leq (1 - \delta)\mu) \leq \left( \frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right)^\mu$$

- Per ogni  $\delta \in (0, 1]$  si ha:

$$P(X \leq (1 - \delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

**Teorema (4.6):** Consideriamo lo schema dei processi di Poisson. Allora, per ogni  $\delta \in (0, 1]$  si ha:

$$P(|X - \mu| \geq \delta\mu) \leq 2e^{-\frac{2\mu\delta^2}{2}}$$

**Teorema (4.7):** Sia  $Z$  la variabile aleatoria definita come  $\sum_{i=1}^n Z_i$ , dove  $Z_1, Z_2, \dots, Z_n$  sono variabili aleatorie indipendenti e identicamente definite tali che  $P(Z_i = 1) = P(Z_i = -1) = \frac{1}{2}$  per ogni  $i \in [1, n]$ , allora:

$$P(Z \geq a) \leq e^{-\frac{a^2}{2n}} \text{ per ogni } a > 0$$

**Corollario (4.7a):** Sia  $Z$  la variabile aleatoria definita come  $\sum_{i=1}^n Z_i$ , dove  $Z_1, Z_2, \dots, Z_n$  sono variabili aleatorie indipendenti e identicamente definite tali che  $P(Z_i = 1) = P(Z_i = -1) = \frac{1}{2}$  per ogni  $i \in [1, n]$ , allora:

$$P(|Z| \geq a) \leq 2e^{-\frac{a^2}{2n}}$$

**Corollario (4.7b):** Sia  $Y$  la variabile aleatoria definita come  $\sum_{i=1}^n Y_i$ , dove  $Y_1, Y_2, \dots, Y_n$  sono variabili aleatorie indipendenti e identicamente definite con distribuzione di Bernoulli  $1_Y(\frac{1}{2})$ , allora:

- Per ogni  $a > 0$  si ha che:

$$P(Y \geq \mu + a) \leq e^{-\frac{2a^2}{2n}}$$

- Per ogni  $\delta > 0$  si ha che:

$$P(Y \geq (1 + \delta)\mu) \leq e^{-\delta^2 \mu}$$

**Corollario (4.7c):** Sia  $Y$  la variabile aleatoria definita come  $\sum_{i=1}^n Y_i$ , dove  $Y_1, Y_2, \dots, Y_n$  sono variabili aleatorie indipendenti e identicamente definite con distribuzione di Bernoulli  $1_Y(\frac{1}{2})$ , allora:

- Per ogni  $a \in (0, \mu)$  si ha che:

$$P(Y \geq \mu - a) \leq e^{-\frac{2a^2}{2n}}$$

- Per ogni  $\delta \in (0, 1)$  si ha che:

$$P(Y \geq (1 - \delta)\mu) \leq e^{-\delta^2 \mu}$$

# Parte II

## Distribuzioni notevoli

# Capitolo 5

## Distribuzione di Bernoulli

La **distribuzione di Bernoulli** è una distribuzione di probabilità su due soli valori: 0 ed 1, detti anche **fallimento** e **successo**. Più informalmente, può essere pensata come il modello per il set dei possibili risultati di un singolo esperimento che cerca di rispondere ad una domanda generale (ossia ad una domanda che accetta solo due risposte: una affermativa e una contraria).

### 5.1 Definizione

Una variabile aleatoria  $X$  ha distribuzione di Bernoulli se, preso  $p \in [0, 1]$ , si ha che:

$$P(X = k) = \begin{cases} p & \text{se } k = 1 \\ q = 1 - p & \text{se } k = 0 \\ 0 & \text{altrimenti} \end{cases}$$

In questo caso si scriverà che  $X \sim 1_X(p)$ .

### 5.2 Media

La media di una variabile aleatoria  $X$  con distribuzione di Bernoulli è pari a:

$$\mathbb{E}[X] = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = 1 \cdot p = p$$

### 5.3 Varianza

La varianza di una variabile aleatoria  $X$  con distribuzione di Bernoulli è pari a:

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \\ &= 1^2 \cdot P(X = 1) + 0^2 \cdot P(X = 0) - p^2 = \\ &= 1 \cdot p - p^2 = p - p^2 = \\ &= p(1 - p) = pq \end{aligned}$$

## 5.4 Funzione generatrice dei momenti

La funzione generatrice dei momenti di una variabile aleatoria  $X$  con distribuzione di Bernoulli è pari a:

$$\begin{aligned}M_X(t) &= \mathbb{E}[e^{tX}] = \\&= e^{t \cdot 1}P(X = 1) + e^{t \cdot 0}P(X = 0) = \\&= e^t p + (1 - p) = \\&= 1 - p + pe^t\end{aligned}$$



# Capitolo 6

## Distribuzione binomiale

La **distribuzione binomiale** è una distribuzione di probabilità discreta che descrive il numero di successi in un processo di Bernoulli, ovvero è definita sulla variabile aleatoria  $X = X_1 + X_2 + \dots + X_n$  che somma  $n$  variabili aleatorie indipendenti di uguale distribuzione di Bernoulli  $1_X(p)$ .

### 6.1 Definizione

Una variabile aleatoria  $X$  ha distribuzione binomiale se, presi  $n \in \mathbb{N}$  e  $p \in [0, 1]$ , si ha che:

$$P(X = k) = P(X_1 + X_2 + \dots + X_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

In questo caso si scriverà che  $X \sim \text{BIN}(n, p)$ .

### 6.2 Media

La media di una variabile aleatoria  $X$  con distribuzione binomiale è pari a:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=0}^n k \binom{n}{k} p^k (1 - p)^{n-k} = \\ &= \sum_{k=0}^n k \frac{n!}{k! (n-k)!} p^k (1 - p)^{n-k} = \\ &= \sum_{k=1}^n k \frac{n \cdot (n-1)!}{k \cdot (k-1)! (n-1-(k-1))!} p \cdot p^{k-1} (1 - p)^{n-1-(k-1)} = \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)! (n-1-(k-1))!} p^{k-1} (1 - p)^{n-1-(k-1)} = \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1 - p)^{n-1-(k-1)} \end{aligned}$$

Effettuando il cambio di indice  $j = k - 1$  si ha:

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

Questo risultato può essere ottenuto anche in un secondo modo: una distribuzione binomiale è la somma di  $n$  variabili aleatorie indipendenti con distribuzione di Bernoulli, e pertanto:

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = n\mathbb{E}[X] = np$$

### 6.3 Varianza

La varianza di una variabile aleatoria  $X$  con distribuzione binomiale è pari a:

$$\begin{aligned}Var[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \\ &= \sum_{k=0}^n k^2 P(X = k) - (np)^2 = \\ &= \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} - (np)^2 = \\ &= \sum_{k=0}^n (k^2 - k + k) \binom{n}{k} p^k (1-p)^{n-k} - (np)^2 = \\ &= \sum_{k=0}^n (k^2 - k) \binom{n}{k} p^k (1-p)^{n-k} + \underbrace{\sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}}_{= \mathbb{E}[X] = np} - (np)^2 = \\ &= \sum_{k=0}^n (k^2 - k) \binom{n}{k} p^k (1-p)^{n-k} + np - (np)^2\end{aligned}$$

Supponendo che  $n \geq 2$  (poichè per  $n = 1$  si avrebbe una Bernoulliana):

$$\begin{aligned}
Var[X] &= \sum_{k=0}^n (k^2 - k) \binom{n}{k} p^k (1-p)^{n-k} + np - (np)^2 = \\
&= \sum_{k=2}^n (k^2 - k) \binom{n}{k} p^k (1-p)^{n-k} + np - (np)^2 = \\
&= \sum_{k=2}^n k(k-1) \frac{n!}{k! (n-k)!} p^k (1-p)^{n-k} + np - (np)^2 = \\
&= \sum_{k=2}^n n(n-1) \frac{(n-2)!}{(k-2)! (n-k)!} p^{k-2+2} (1-p)^{n-k} + np - (np)^2 = \\
&= n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)! (n-2-(k-2))!} p^{k-2} (1-p)^{n-2-(k-2)} + np - (np)^2
\end{aligned}$$

Effettuando il cambio di indice  $j = k - 2$  si ha:

$$\begin{aligned}
Var[X] &= n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)! (n-2-(k-2))!} p^{k-2} (1-p)^{n-2-(k-2)} + np - (np)^2 = \\
&= n(n-1)p^2 \sum_{j=0}^{n-2} \frac{(n-2)!}{(j)! (n-2-j)!} p^j (1-p)^{n-2-j} + np - (np)^2 = \\
&= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j} + np - (np)^2 = \\
&= n(n-1)p^2 (p + 1-p)^{n-2} + np - (np)^2 = \\
&= n(n-1)p^2 + np - (np)^2 = \\
&= n^2p^2 - np^2 + np - n^2p^2 = \\
&= -np^2 + np = np(1-p)
\end{aligned}$$

Questo risultato può essere ottenuto anche in un secondo modo: una distribuzione binomiale è la somma di  $n$  variabili aleatorie indipendenti con distribuzione di Bernoulli, e pertanto:

$$Var[X] = \sum_{i=1}^n Var[X_i] = nVar[X] = np(1-p)$$

## 6.4 Funzione generatrice dei momenti

La funzione generatrice dei momenti di una variabile aleatoria  $X$  con distribuzione binomiale è pari a:

$$\begin{aligned}
M_X(t) &= \mathbb{E}[e^{tX}] = \\
&= \sum_{k=0}^n e^{kt} \binom{n}{k} p^k (1-p)^{n-k} = \\
&= \sum_{k=0}^n \binom{n}{k} (pe^t)^k (1-p)^{n-k} = (1-p + pe^t)^n
\end{aligned}$$

Questo risultato può essere ottenuto anche in un secondo modo: una distribuzione binomiale è la somma di  $n$  variabili aleatorie indipendenti con distribuzione di Bernoulli, e pertanto:

$$M_X(t) = M_{X_1+X_2+\dots+X_n}(t) = \prod_{i=1}^n M_{X_i}(t) = (1 - p + pe^t)^n$$

# Capitolo 7

## Distribuzione geometrica

La **distribuzione geometrica** è una distribuzione di probabilità discreta che descrive il numero di fallimenti che precedono il primo successo in un processo di Bernoulli avente  $n$  variabili aleatorie indipendenti di uguale distribuzione di Bernoulli  $1_X(p)$ .

### 7.1 Definizione

Una variabile aleatoria  $X$  ha distribuzione geometrica se, preso  $p \in [0, 1]$ , si ha che:

$$P(X = k) = P(X_1 = 0)P(X_2 = 0)\dots P(X_{k-1} = 0)P(X_k = 1) = (1-p)^{k-1}p \text{ con } k \in [1, \infty)$$

In questo caso si scriverà che  $X \sim \text{GEOM}(p)$ . Notare che la seguente distribuzione calcola la probabilità di effettuare  $k-1$  fallimenti su  $k$  prove, le quali precedono il primo successo al  $k$ -esimo tentativo. È possibile modificarla in modo da calcolare la probabilità di avere  $k$  fallimenti prima di ottenere il primo successo:

$$P(X = k) = P(X_1 = 0)P(X_2 = 0)\dots P(X_k = 0)P(X_{k+1} = 1) = (1-p)^k p \text{ con } k \in [0, \infty)$$

In questo caso si scriverà che  $X \sim \text{GEOM}^+(p)$ .

## 7.2 Media

La media di una variabile aleatoria  $X$  con distribuzione geometrica è pari a:

$$\begin{aligned}
 \mathbb{E}[X] &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p = \\
 &= p \sum_{k=1}^{\infty} k(1-p)^{k-1} = \\
 &= p \sum_{k=1}^{\infty} ((1-p)^k (-1))' = \\
 &= -p \sum_{k=1}^{\infty} ((1-p)^k)' = \\
 &= -p \left( \sum_{k=1}^{\infty} (1-p)^k \right)' = \\
 &= -p \left( \frac{(1-p)^1}{1 - (1-p)} \right)' = \\
 &= -p \left( \frac{1-p}{p} \right)' = -p \left( \frac{1}{p} - 1 \right)' = \\
 &= -p(p^{-1} - 1)' = \\
 &= -p \left( -\frac{1}{p^2} \right) = \frac{1}{p}
 \end{aligned}$$

È possibile calcolare la media utilizzando anche la seconda definizione di valore atteso data in precedenza. Possiamo notare che, per  $n \geq 1$ , i valori:

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

Sono tutti valori non negativi, e pertanto:

$$\mathbb{E}[X] = \sum_{n \geq 1} (1-p)^{n-1} = \frac{1}{1 - (1-p)} = \frac{1}{p}$$

## 7.3 Varianza

La varianza di una variabile aleatoria  $X$  con distribuzione geometrica è pari a:

$$\begin{aligned}
 Var[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \\
 &= \sum_{k \geq 1} k^2 P(X = k) - \left(\frac{1}{p}\right)^2 = \\
 &= \sum_{k \geq 1} k^2 (1-p)^{k-1} p - \frac{1}{p^2} = \\
 &= \sum_{k \geq 1} (k^2 - k + k) (1-p)^{k-1} p - \frac{1}{p^2} = \\
 &= \sum_{k \geq 1} (k^2 - k) (1-p)^{k-1} p + \underbrace{\sum_{k \geq 1} k (1-p)^{k-1} p}_{\mathbb{E}[X] = \frac{1}{p}} - \frac{1}{p^2} = \\
 &= \sum_{k \geq 1} (k^2 - k) (1-p)^{k-1} p + \frac{1}{p} - \frac{1}{p^2} = \\
 &= \sum_{k \geq 2} (k^2 - k) (1-p)^{k-1} p + \frac{1}{p} - \frac{1}{p^2} = \\
 &= p(1-p) \sum_{k \geq 2} (k^2 - k) (1-p)^{k-2} + \frac{1}{p} - \frac{1}{p^2}
 \end{aligned}$$

Possiamo utilizzare una serie notevole e le sue derivate per semplificare i calcoli. In particolare:

$$\begin{aligned}
 \sum_{k \geq 0} x^k &= \frac{1}{1-x} && \text{per ogni } |x| < 1 \\
 \sum_{k \geq 0} kx^{k-1} &= \sum_{k \geq 1} kx^{k-1} = \frac{1}{(1-x)^2} && \text{derivata prima} \\
 \sum_{k \geq 1} k(k-1)x^{k-2} &= \sum_{k \geq 2} k(k-1)x^{k-2} = \frac{2}{(1-x)^3} && \text{derivata seconda}
 \end{aligned}$$

Possiamo utilizzare l'ultima serie notevole trovata ponendo  $x = 1 - p$ :

$$\begin{aligned}
 Var[X] &= p(1-p) \sum_{k \geq 2} (k^2 - k) (1-p)^{k-2} + \frac{1}{p} - \frac{1}{p^2} = \\
 &= p(1-p) \frac{2}{(1-(1-p))^3} + \frac{1}{p} - \frac{1}{p^2} = \\
 &= p(1-p) \frac{2}{p^3} + \frac{1}{p} - \frac{1}{p^2} = \\
 &= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \\
 &= \frac{2 - 2p + p - 1}{p^2} = \frac{1-p}{p^2}
 \end{aligned}$$

## 7.4 Funzione generatrice dei momenti

La funzione generatrice dei momenti di una variabile aleatoria  $X$  con distribuzione geometrica è pari a:

$$\begin{aligned}
 M_X(t) &= \sum_{n \geq 1} e^{tn} P(X = n) = \\
 &= \sum_{n \geq 1} e^{tn} (1-p)^{n-1} p = \\
 &= \frac{p}{1-p} \sum_{n \geq 1} (e^t(1-p))^n = \\
 &= \begin{cases} \frac{p}{1-p} \frac{e^t(1-p)}{1 - e^t(1-p)} & \text{se } e^t(1-p) < 1 \\ \infty & \text{se } e^t(1-p) \geq 1 \end{cases} = \\
 &= \begin{cases} \frac{pe^t}{1 - e^t(1-p)} & \text{se } e^t(1-p) < 1 \\ \infty & \text{se } e^t(1-p) \geq 1 \end{cases}
 \end{aligned}$$

Possiamo notare che la condizione  $M_X(t) < \infty$  non è soddisfatta per alcuni valori di  $t$ ; in realtà questo accade nel caso particolare  $p = 1$  perché si ha:

$$M_X(t) = e^t \text{ per ogni } t \in \mathbb{R}$$

Comunque, per ogni  $0 < p \leq 1$ , possiamo dire che  $M_X(t)$  è crescente perché  $X$  è non negativa in accordo con quanto osservato, e la condizione  $M_X(t) < \infty$  è soddisfatta in un intorno dell'origine poiché  $-\log(1-p) > 0$ .

## 7.5 Mancanza di memoria della distribuzione geometrica

**Teorema (7.1):** Sia  $X$  una variabile aleatoria con distribuzione geometrica, allora si ha:

$$P(X = n + k | X > k) = P(X = n) \text{ per ogni } n, k > 1 \text{ interi}$$

**Teorema (7.1):** Calcoliamo la probabilità condizionata:

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

Osserviamo che:

$$P(X > k) = \sum_{i \geq k+1} (1-p)^{i-1} p = p \sum_{i \geq k+1} (1-p)^{i-1} = p \left( \frac{(1-p)^k}{1 - (1-p)} \right) = (1-p)^k$$



Da qui abbiamo che:

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

# Capitolo 8

## Distribuzione di Poisson

La **distribuzione di Poisson** è una distribuzione di probabilità discreta che esprime le probabilità per il numero di eventi che si verificano successivamente ed indipendentemente in un dato intervallo di tempo, sapendo che mediamente se ne verifica un numero  $\lambda$ .

### 8.1 Definizione

Una variabile aleatoria  $X$  ha distribuzione di Poisson se, preso  $\lambda > 0$ , si ha che:

$$P(X = k) = \frac{\lambda^n}{n!} e^{-\lambda} \text{ per ogni } n \in \mathbb{N}$$

In questo caso si scriverà che  $X \sim \text{POISSON}(\lambda)$ .

### 8.2 Media

La media di una variabile aleatoria  $X$  con distribuzione di Poisson è pari a:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= \sum_{n=1}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= \lambda e^{-\lambda} \sum_{n=1}^{\infty} n \frac{\lambda^{n-1}}{(n-1)!} = \\ &= \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = \\ &= \lambda e^{-\lambda} e^{\lambda} = \lambda \end{aligned}$$

### 8.3 Varianza

La varianza di una variabile aleatoria  $X$  con distribuzione di Poisson è pari a:

$$\begin{aligned} Var[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \\ &= \sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} e^{-\lambda} - \lambda^2 = \\ &= \sum_{n=0}^{\infty} n(n-1) \frac{\lambda^n}{n!} e^{-\lambda} + \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} - \lambda^2 = \\ &= \lambda^2 e^{-\lambda} \sum_{n=2}^{\infty} \frac{\lambda^{n-2}}{(n-2)!} + \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} - \lambda^2 = \\ &= \lambda^2 e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} + \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} - \lambda^2 = \\ &= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} - \lambda^2 = \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

### 8.4 Funzione generatrice dei momenti

La funzione generatrice dei momenti di una variabile aleatoria  $X$  con distribuzione di Poisson è pari a:

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] = \\ &= \sum_{n=0}^{\infty} e^{tn} \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} = \\ &= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)} \end{aligned}$$

# Parte III

## Catene di Markov

# Capitolo 9

## Introduzione

### 9.1 Le catene di Markov

**Definizione:** Una famiglia di variabili aleatorie  $\{X_t : t \in I\}$  tutte a valori in un insieme discreto  $V$ , dove  $I$  è un intervallo di numeri interi non negativi (con  $0 \in I$  pensato come stato iniziale), è una **catena di Markov** se:

$$P(X_t = j | X_{t-1} = i_{t-1}, X_{t-2} = i_{t-2}, \dots, X_0 = i_0) = P(X_t = j | X_{t-1} = i_{t-1}) := p_{ij}^t$$

Per ogni scelta di  $j, i_{t-1}, i_{t-2}, \dots, i_0 \in V$  (e per ogni scelta ammissibile di  $t \in I$ ) per cui le probabilità condizionate sono ben definite.

Gli elementi di  $V$  vengono definiti **stati**, inoltre se i valori  $p_{ij}^t$  non dipendono da  $t$ , è possibile parlare di **catene di Markov omogenee** (e viene utilizzata la notazione semplificata  $p_{ij}$ ). Per rappresentare una catena di Markov è possibile utilizzare una **matrice di transizione**:

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}$$

Tale matrice è costituita da numeri non negativi, inoltre la somma degli elementi di ogni riga è uguale 1 poichè, per ogni  $i \in V$  (in riferimento alla riga  $i$ -esima) si ha:

$$\begin{aligned} \sum_{j \in V} p_{ij} &= \sum_{j \in V} P(X_t = j | X_{t-1} = i) = \\ &= P(\cup_{j \in V} \{X_t = j\} | X_{t-1} = i) = \\ &= P(X_t \in V | X_{t-1} = i) = 1 \end{aligned} \quad \text{per ogni } i \in V$$

Notare come non sia richiesto che la somma delle colonne sia pari ad 1. Nel caso in cui anche tale proprietà è rispettata, ossia:

$$\sum_{i \in V} p_{ij} = 1 \text{ per ogni } j \in V$$

Allora la matrice di transizione prende il nome di matrice di transizione **bistocastica**. È possibile definire inoltre la **distribuzione marginale ad un tempo** per ogni tempo

$t \in I$  come:

$$\bar{p}(t) = (p_i(t))_{i \in V} \text{ dove } p_i(t) := P(X_t = i) \text{ per ogni } i \in V$$

Tale valore può essere rappresentato come un vettore che rappresenta la probabilità di trovarsi in uno degli stati  $i$  al tempo  $t$ , avendo inoltre che:

$$\sum_{i \in V} p_i(t) = 1$$

Inoltre, se consideriamo la formula delle probabilità totali rispetto alla partizione degli eventi  $\{\{X_{t-1} = i\} : i \in V\}$ , si ha:

$$\underbrace{P(X_t = j)}_{=p_j(t)} = \sum_{i \in V} \underbrace{P(X_t = j | X_{t-1} = i)}_{=p_{ij}} \underbrace{P(X_{t-1} = i)}_{p_j(t-1)}$$

Quindi, se pensiamo ad un prodotto righe per colonne tra vettori e matrici, abbiamo la seguente relazione:

$$\bar{p}(t) = \bar{p}(t-1)P$$

Sia  $m \geq 1$  un intero, e consideriamo la matrice costituita dai valori:

$$p_{ij}^m := P(X_{t+m-1} = j | X_{t-1} = i) \text{ per ogni } i, j \in V$$

La seguente matrice è detta **matrice di transizione ad  $m$  passi**, ed è definita in tale modo:

$$P^{(m)} = \begin{pmatrix} p_{11}^{(m)} & p_{12}^{(m)} & \cdots & p_{1n}^{(m)} \\ p_{21}^{(m)} & p_{22}^{(m)} & \cdots & p_{2n}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1}^{(m)} & p_{n2}^{(m)} & \cdots & p_{nn}^{(m)} \end{pmatrix}$$

Tale matrice è costituita da numeri non negativi, inoltre la somma degli elementi di ogni riga è uguale 1 poichè, per ogni  $i \in V$  (in riferimento alla riga  $i$ -esima) si ha:

$$\begin{aligned} \sum_{j \in V} p_{ij}^{(m)} &= \sum_{j \in V} P(X_{t+m-1} = j | X_{t-1} = i) = \\ &= P(\cup_{j \in V} \{X_{t+m-1} = j\} | X_{t-1} = i) = \\ &= P(X_{t+m-1} \in V | X_{t-1} = i) = 1 \end{aligned}$$

Tale matrice è una generalizzazione di transizione presentata prima (ponendo  $m = 1$ , infatti, abbiamo che  $P^{(m)} = P^{(1)} = P$ ). Inoltre, per  $m \geq 2$ , consideriamo la partizione degli eventi  $\{\{X_{t+m-1} = h\} : h \in V\}$  ed alcuni calcoli standard sulle probabilità condizionate:

$$\begin{aligned} p_{ij}^{(m)} &= P(X_{t+m} = j | X_t = i) = \\ &= \sum_{h \in V} P(X_{t+m} = j, X_{t+m-1} = h | X_t = i) = \\ &= \sum_{h \in V} \frac{P(X_{t+m} = j, X_{t+m-1} = h, X_t = i)}{P(X_t = i)} = \\ &= \sum_{h \in V} \frac{P(X_{t+m} = j, X_{t+m-1} = h, X_t = i)}{P(X_t = i)} \cdot \frac{P(X_{t+m-1} = h, X_t = i)}{P(X_t = i)} = \\ &= \sum_{h \in V} P(X_{t+m} = j | X_{t+m-1} = h, X_t = i) P(X_{t+m-1} = h | X_t = i) \end{aligned}$$

A questo punto sfruttiamo l'ipotesi che la Catena di Markov è omogenea:

$$\begin{aligned} P(X_{t+m} = j | X_{t+m-1} = h, X_t = i) &= P(X_{t+m} = j | X_{t+m-1} = h) = p_{hj} \\ P(X_{t+m-1} = h | X_t = i) &= p_{ih}^{(m-1)} \end{aligned}$$

Quindi, sostituendo nell'espressione ottenuta, otteniamo la seguente uguaglianza:

$$p_{ij}^{(m)} = \sum_{h \in V} p_{ih}^{(m-1)} p_{hj} \text{ per ogni } i, j \in V$$

Tale relazione ci consente di dire che vale la seguente uguaglianza:

$$P^{(m)} = P^{(m-1)} P$$

Allora, iterando la relazione, e tenendo conto che  $P^{(1)} = P$ , possiamo dire che, per ogni  $m \geq 1$ :

$$P^{(m)} = P^m, \text{ cioè } P^{(m)} = \underbrace{P \cdot P \cdot \dots \cdot P}_{m \text{ volte}}$$

Come potenza  $m$ -esima della matrice di transizione  $P$ . Per completezza, in conformità con le proprietà delle potenze, è utile considerare il caso  $m = 0$ , ponendo  $P^{(0)} = I$ , dove  $I$  è la matrice identità:

$$p_{ij}^{(0)} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

Possiamo anche dare una generalizzazione di  $\bar{p}(t)$ :

$$\begin{aligned} \bar{p}(t-m) P^{(m)} &= \bar{p}(t-m) \underbrace{P \cdot P \cdot \dots \cdot P}_{m \text{ volte}} = \\ &= \bar{p}(t-m+1) \underbrace{P \cdot P \cdot \dots \cdot P}_{m-1 \text{ volte}} = \\ &= \bar{p}(t-m+2) \underbrace{P \cdot P \cdot \dots \cdot P}_{m-2 \text{ volte}} = \\ &= \dots = \\ &= \bar{p}(t-m+m-1) P = \\ &= \bar{p}(t-m+m) = \bar{p}(t) \end{aligned}$$

Da qui possiamo dire che in generale il vettore riga  $\bar{p}(t)$  dipende dalla sola conoscenza della matrice di transizione  $P$  e dalla distribuzione iniziale  $\bar{p}(0)$ . È possibile rappresentare una catena di Markov omogenea anche attraverso un **grafo orientato pesato associato**, tale che:

- L'insieme dei nodi del grafo è  $V$ .
- Per ogni  $i, j \in V$ , si ha un arco da  $i$  a  $j$  se e solo se  $p_{ij} > 0$ , e tale valore rappresenta il peso dell'arco.

In questo modo possiamo dire che:

- Un evento del tipo:

$$\{X_t = i_0, X_{t+1} = i_1, \dots, X_{t+m-1} = i_{m-1}, X_{t+m} = i_m\}$$

Si può verificare se e solo se esiste un cammino attraverso gli archi del grafo orientato che congiunge in sequenza i nodi  $i_0, i_1, \dots, i_{m-1}, i_m$ ; inoltre la probabilità di tale evento è data dal prodotto dei pesi dei singoli archi che costituiscono il cammino.

- Ovviamente, per ogni  $i, j \in V$  e per ogni  $m \geq 1$ , la probabilità di andare da  $i$  a  $j$  in  $m$  passi è data dalla somma delle probabilità di tutte le sequenze di  $m$  stati che partono da  $i$  e finiscono con  $j$

## 9.2 Comunicazione tra due stati e irriducibilità

**Definizione:** Dati  $i, j \in V$ , si dice che " $i$  e  $j$  comunicano tra loro" ( $i \leftrightarrow j$ ) se esistono  $n, m \geq 0$  interi tali che  $p_{ij}^{(n)} > 0$  e  $p_{ji}^{(m)}$ .

Possiamo dire che  $i \leftrightarrow j$  se e solo se esistono dei cammini attraverso gli archi del grafo orientato da  $i$  a  $j$  e da  $j$  ad  $i$ . È possibile dimostrare che la relazione  $i \leftrightarrow j$  è una relazione di equivalenza:

- $i \leftrightarrow i$  per ogni  $i \in V$  (riflessività): basta prendere  $n = 0$  e  $m = 0$  e si ha  $p_{ii}^{(0)} = 1 > 0$ .
- Se  $i \leftrightarrow j$  allora  $j \leftrightarrow i$  (simmetria): infatti, se si ha  $p_{ij}^{(n)}$  e  $p_{ji}^{(m)}$ , allora basta scambiare il ruolo di  $n$  ed  $m$  e si ha anche  $p_{ij}^{(m)}$  e  $p_{ji}^{(n)}$ .
- Se  $i \leftrightarrow j$  e  $j \leftrightarrow h$ , allora  $i \leftrightarrow h$  (transitività): infatti, se si ha  $p_{ij}^{(n_1)}$  e  $p_{ji}^{(m_1)}$ , e  $p_{jh}^{(n_2)}$  e  $p_{hj}^{(m_2)}$ , allora:

$$p_{ih}^{(n_1+n_2)} = \sum_{r \in V} \underbrace{p_{ir}^{(n_1)} p_{rh}^{(n_2)}}_{\geq 0} \geq p_{ij}^{(n_1)} p_{jh}^{(n_2)} > 0$$

$$p_{hi}^{(m_1+m_2)} = \sum_{r \in V} \underbrace{p_{hr}^{(m_1)} p_{ri}^{(m_2)}}_{\geq 0} \geq p_{hj}^{(m_1)} p_{ji}^{(m_2)} > 0$$

Abbiamo quindi una partizione dell'insieme  $V$  costituito dalle classi di equivalenza rispetto alla relazione  $\leftrightarrow$ . La catena di Markov può andare da un elemento  $i$  di una classe ad un elemento  $j$  di un'altra classe (eventualmente in più passi), ma non può ritornare nella classe dell'elemento  $i$  (infatti, se fosse possibile, lo stato  $j$  sarebbe nella stessa classe di equivalenza di  $i$ ).

**Definizione** Si dice che una catena di Markov omogenea è **irriducibile** se  $V$  è costituito da un'unica classe di equivalenza rispetto a  $\leftrightarrow$ , e quindi tutti gli stati sono equivalenti tra loro rispetto a tale relazione.



Quindi, per ogni  $i, j \in V$ , è positiva la probabilità di andare da  $i$  a  $j$  e da  $j$  ad  $i$  con un numero finito di passi.

**Teorema (9.1):** C'è irriducibilità se e solo se il grafo orientato pesato associato è connesso.

### 9.3 Gli stati ricorrenti e transienti

**Definizione:** Sia  $r_{ij}^t$  la probabilità che, partendo dallo stato  $i$ , la catena raggiunga per la prima volta lo stato  $j$  dopo  $t$  transizioni, ossia:

$$r_{ij}^t = P(X_1 \neq j, X_2 \neq j, \dots, X_{t-1} \neq j, X_t = j | X_0 = i) \text{ per } i, j \in V \text{ e } t \geq 1$$

Si dice che  $i \in V$  è **ricorrente** se  $\sum_{t \geq 1} r_{ii}^t = 1$ , e che è **transiente** se  $\sum_{t \geq 1} r_{ii}^t < 1$ .

Possiamo dire quanto segue:

- Se la catena parte da un stato ricorrente  $i$ , certamente ci ritorna, quindi tale stato viene visitato infinite volte.
- Se la catena parte da uno stato transiente  $i$ , non è certo che ci torni. Possiamo dire che, ogni volta che la catena visita uno stato transiente  $i$ , ci ritorna con probabilità  $\bar{p}_i = \sum_{t \geq 1} r_{ii}^t < 1$  indipendentemente da quanto accaduto prima.

Inoltre, tutti gli stati in una classe di equivalenza di  $\leftrightarrow$  sono tutti dello stesso tipo.

**Definizione:** Sia  $h_{ij}^t$  il tempo medio di primo arrivo in  $j$  partendo da  $i$ , ossia:

$$h_{ij} = \begin{cases} \sum_{t \geq 1} t \cdot r_{ij}^t & \text{se } \sum_{t \geq 1} r_{ij}^t = 1 \\ \infty & \text{se } \sum_{t \geq 1} r_{ij}^t < 1 \end{cases}$$

Sia  $i \in V$  uno stato ricorrente, e sia  $h_{ii}$  il tempo medio di primo ritorno in  $i$ , allora si dice che  $i$  è **ricorrente positivo** se  $h_{ii} < \infty$ , e che  $i$  è **ricorrente nullo** se  $h_{ii} = \infty$ .

**Teorema (9.2):** Sia  $V$  finito, allora possiamo dire quanto segue:

- Almeno uno stato è ricorrente.
- Tutti gli stati ricorrenti sono ricorrenti positivi

## 9.4 La periodicità e l'aperiodicità per stati e catena

**Definizione:** Uno stato  $i$  è detto **periodico** se esiste  $\Delta_i > 1$  intero tale che  $P(X_{t+s} = i | X_t = i) = 0$  a meno che  $s$  sia divisibile per  $\Delta_i$ . Si dice che la catena di Markov omogenea è **periodica** se tutti gli stati sono periodici. Si usa il termine aperiodico se uno stato non è periodico. Analogamente si dice che la catena di Markov omogenea è **aperiodica** se tutti gli stati sono aperiodici.

Per quello che riguarda il valore  $\Delta_i$  possiamo definirlo come:

$$\Delta_i := \text{MCD}\{s \geq 1 : p_{ii}^{(s)} > 0\}$$

E definire tale valore come il periodo di  $i$ . Tramite ciò possiamo dire che  $i$  è aperiodico se  $\Delta_i = 1$ ; un problema da considerare però è che si potrebbe dover considerare l'MCD su di un insieme vuoto. In generale possiamo dire due cose sul valore  $\Delta_i$ :

- Se  $i$  è uno stato ricorrente, allora l'insieme  $\{s \geq 1 : p_{ii}^{(s)} > 0\}$  è non vuoto.
- Se la catena è irriducibile, allora tutti gli stati hanno stesso periodo, e pertanto  $\Delta_i$  assume lo stesso valore che non dipende da  $i \in V$ , rappresentando il periodo della catena.

**Definizione:** Uno stato aperiodico e ricorrente positivo è detto **ergodico**. Inoltre si dice che la catena di Markov omogenea è **ergodica** se tutti gli stati sono ergodici.

**Teorema (9.3):** Sia  $V$  finito, Supponiamo che la catena di Markov omogenea sia irriducibile e aperiodica. Allora è ergodica.

# Capitolo 10

## Distribuzioni stazionarie

**Definizione:** Consideriamo una catena di Markov omogenea su  $V$ , e con matrice di transizione  $P$ . Allora una distribuzione  $\bar{\pi} = (\pi_i)_{i \in V}$  su  $V$ , considerata come vettore riga (quindi con componenti non negative e a somma 1), è una **distribuzione stazionaria** se vale la seguente uguaglianza tra vettori riga:

$$\bar{\pi}P = \bar{\pi}$$

Se poniamo  $\bar{p}(0) = \bar{\pi}$ , usando ricorsivamente la formula  $\bar{p}(t) = \bar{p}(t-1)P$  si ha:

$$\begin{aligned}\bar{p}(1) &= \bar{p}(0)P = \bar{\pi}P = \bar{\pi} \text{ per } t = 1 \\ \bar{p}(2) &= \bar{p}(1)P = \bar{\pi}P = \bar{\pi} \text{ per } t = 2 \\ \bar{p}(3) &= \bar{p}(2)P = \bar{\pi}P = \bar{\pi} \text{ per } t = 3 \\ &\dots\end{aligned}$$

La condizione mostrata sopra è equivalente a chiedere che:

$$\bar{\pi}P^{(t)} = \bar{\pi} \text{ per ogni } t \geq 1$$

Infatti per  $t = 1$  ritroviamo la condizione sopra; ma se vale tale condizione possiamo dire che:

$$\bar{\pi}P^{(t)} = \bar{\pi}P^t = \bar{\pi} \underbrace{P \cdot P \cdot \dots \cdot P}_{t \text{ volte}} = \bar{\pi} \underbrace{P \cdot P \cdot \dots \cdot P}_{t-1 \text{ volte}} = \dots = \bar{\pi}P = \bar{\pi} \text{ per ogni } t \geq 1$$

**Teorema (10.1):** Sia  $V$  finito. Supponiamo che ci sia irriducibilità ed ergodicità. Allora:

- Esiste un'unica distribuzione stazionaria  $\bar{\pi} = (\pi_i)_{i \in V}$ .
- Per ogni  $i, j \in V$ , esiste  $\lim_{t \rightarrow \infty} p_{ij}^{(t)}$  e non dipende da  $j$ .
- Per ogni  $i, j \in V$ ,  $\lim_{t \rightarrow \infty} p_{ij}^{(t)} = \pi_i = \frac{1}{h_{ii}}$

**Lemma (10.1a):** Sia  $V$  finito. Supponiamo che ci sia irriducibilità ed ergodicità. Allora:

$$\lim_{t \rightarrow \infty} p_{ii}^{(t)} = \frac{1}{h_{ii}}$$

**Teorema (10.2):** Verrà utilizzato nella seguente dimostrazione il lemma 7.7a, il quale non verrà dimostrato. Consideriamo una generica coppia di stati  $i, j \in V$ . Allora, per ipotesi di irriducibilità, si ha  $\sum_{t \geq 1} r_{ij}^t = 1$ ; quindi, per  $\epsilon > 0$ , esiste  $t(\epsilon) \geq 1$  tale che:

$$1 - \epsilon \sum_{k=1}^{t_1(\epsilon)} r_{ij}^k \leq 1$$

Quindi per  $t \geq t_1(\epsilon)$  si ha anche:

$$0 \leq \sum_{k=t_1(\epsilon)+1}^t r_{ij}^k \leq \epsilon$$

Inoltre, per costruzione si ha:

$$p_{ij}^{(t)} = \sum_{k=1}^t r_{ij}^k p_{ii}^{(t-k)}$$

Ossia considerano tutti i possibili modi di andare da  $j$  ad  $i$  per la prima volta; se questo accade in  $k$  passi, nei rimanenti  $t - k$  passi si va da  $i$  ad  $i$ . Quindi, per  $t > t_1(\epsilon)$  si ha:

$$\sum_{k=1}^{t_1(\epsilon)} r_{ij}^k p_{ii}^{(t-k)} \leq \underbrace{\sum_{k=1}^{t_1(\epsilon)} r_{ij}^k p_{ii}^{(t-k)} + \sum_{k=t_1(\epsilon)+1}^t r_{ij}^k p_{ii}^{(t-k)}}_{=p_{ij}^{(t)}} \leq \sum_{k=1}^{t_1(\epsilon)} r_{ij}^k p_{ii}^{(t-k)} + \epsilon$$

E, facendo tendere  $t$  ad infinito, si ottiene:

$$\begin{aligned} \liminf_{t \rightarrow \infty} p_{ij}^{(t)} &\geq \liminf_{t \rightarrow \infty} \sum_{k=1}^{t_1(\epsilon)} r_{ij}^k p_{ii}^{(t-k)} = \sum_{k=1}^{t_1(\epsilon)} r_{ij}^k \frac{1}{h_{ii}} \geq \frac{1 - \epsilon}{h_{ii}} \\ \limsup_{t \rightarrow \infty} p_{ij}^{(t)} &\leq \limsup_{t \rightarrow \infty} \sum_{k=1}^{t_1(\epsilon)} r_{ij}^k p_{ii}^{(t-k)} + \epsilon = \sum_{k=1}^{t_1(\epsilon)} r_{ij}^k \frac{1}{h_{ii}} + \epsilon \leq \frac{1}{h_{ii}} + \epsilon \end{aligned}$$

In conclusione si ha:

$$\frac{1 - \epsilon}{h_{ii}} \leq \liminf_{t \rightarrow \infty} p_{ij}^{(t)} \leq \lim_{t \rightarrow \infty} p_{ij}^{(t)} \leq \frac{1}{h_{ii}} + \epsilon$$

E, per l'arbitrarietà di  $\epsilon$ , il secondo punto è dimostrato. Prendiamo ora il limite per  $t \rightarrow \infty$ , e per la prima parte della dimostrazione si ha:

$$1 = \lim_{t \rightarrow \infty} \sum_{i \in V} p_{ij}^{(t)} = \sum_{i \in V} \lim_{t \rightarrow \infty} p_{ij}^{(t)} = \sum_{i \in V} \frac{1}{h_{ii}}$$

Quindi il vettore riga  $\bar{\pi} = (\pi_i)_{i \in V}$  che abbiamo introdotto è effettivamente una distribuzione su  $V$ . Infine prendiamo la relazione che lega le matrici di transizione a  $t$  e a  $t + 1$  passi componente per componente, e prendiamo il limite per  $t \rightarrow \infty$  membro a membro, ottenendo

$$\begin{aligned} \lim_{t \rightarrow \infty} p_{ij}^{(t)} &= \lim_{t \rightarrow \infty} \sum_{k \in V} p_{ik}^{(t)} p_{kj} && \text{per ogni } i, j \in V \\ \lim_{t \rightarrow \infty} p_{ij}^{(t)} &= \sum_{k \in V} \lim_{t \rightarrow \infty} p_{ik}^{(t)} p_{kj} \\ \frac{1}{h_{ii}} &= \sum_{k \in V} \frac{1}{h_{kk}} p_{kj} && \text{per ogni } i \in V \end{aligned}$$

Quindi abbiamo dimostrato che vale l'uguaglianza  $\bar{\pi} = \bar{\pi}P$  e questo ci consente di dire che  $\bar{\pi} = (\pi_i)_{i \in V}$ , dove  $\pi_i = \frac{1}{h_{ii}}$  per ogni  $i \in V$ , è una distribuzione stazionaria. Sia  $\bar{\phi} = (\phi_i)_{i \in V}$  una generica distribuzione stazionaria, sappiamo che:

$$\bar{\phi} = \bar{\phi}P^{(t)}$$

Quindi, se scriviamo questa relazione componente per componente, si ha:

$$\phi_i = \sum_{k \in V} \phi_k p_{ki}^{(t)} \text{ per ogni } i \in V$$

Se prendiamo il limite per  $t \rightarrow \infty$  nel secondo membro, e se teniamo conto della prima parte della dimostrazione, si ottiene:

$$\phi_i = \sum_{k \in V} \phi_k \frac{1}{h_{ii}} = \frac{1}{h_{ii}} \underbrace{\sum_{k \in V} \phi_k}_{=1} = \frac{1}{h_{ii}} \text{ per ogni } i \in V$$

È possibile fare degli appunti sul teorema e sulla dimostrazione appena mostrati:

- Chiedere ergodicità e irriducibilità su  $V$  finito equivale a chiedere l'aperiodicità.
- Se  $V$  è finito, l'esistenza di almeno una distribuzione stazionaria è garantita in ogni caso.
- Senza la condizione di aperiodicità si può ancora dimostrare che la distribuzione stazionaria è unica; in tal caso però, indicando ancora l'unica distribuzione stazionaria con  $\bar{\pi} = (\pi_i)_{i \in V}$ , si ha:

$$\pi_i = \lim_{t \rightarrow \infty} \frac{\#\{k \in \{1, 2, \dots, t\} : X_k = i\}}{t}$$

- In generale, se  $V$  è finito, esiste almeno una classe di equivalenza rispetto a  $\leftrightarrow$  costituita da tutti stati ricorrenti. Inoltre, una volta che la catena raggiunge una di queste classi, ci resta per sempre; ognuna di queste classi consente di individuare una sottocatena irriducibile e con stati ricorrenti.

- Ne caso in cui  $V$  sia finito, si ha:

$$V = T \cup C_1 \cup C_2 \cup \dots \cup C_k \text{ per qualche } k \geq 1$$

Dove  $T$  è l'insieme degli stati transienti e  $C_1, C_2, \dots, C_k$  sono classi di equivalenza rispetto a  $\leftrightarrow$  costituite da elementi ricorrenti, e che quindi non comunicano tra loro; inoltre ogni distribuzione stazionaria  $\bar{\pi} = (\pi_i)_{i \in V}$  è una particolare combinazione lineare convessa delle distribuzioni stazionarie su  $C_1, C_2, \dots, C_k$  e con zero sugli stati transienti, cioè:

$$\bar{\pi} = \sum_{j=1}^k \alpha_j \bar{\pi}^{(j)}$$

Dove  $\alpha_1, \alpha_2, \dots, \alpha_k \geq 0$  e  $\sum_{j=1}^k \alpha_j = 1$ , e  $\bar{\pi}^{(j)} = (\pi_i^{(j)})_{i \in V}$  è l'unica distribuzione

stazionaria per la sottocatenata ristretta a  $C_j$  in cui si pone  $\pi_i^{(j)} = 0$  per  $i \notin C_j$ . Quindi, se  $V$  è finito, si ha un'unica distribuzione stazionaria  $\bar{\pi} = (\pi_i)_{i \in V}$  in tutti i casi con  $k = 1$ , ed inoltre  $\pi_i = 0$  se  $i$  è transiente.

- Supponiamo che  $V$  sia finito, e che ci sia irriducibilità ed ergodicità, possiamo dire che, qualunque sia la distribuzione iniziale  $\bar{p}(0) = (p_j(0))_{j \in V}$ , si ha:

$$\lim_{t \rightarrow \infty} \mathbb{E}[f(X_t)] = \sum_{i \in V} f(i) \pi_i \text{ per ogni } f : V \rightarrow \mathbb{R}$$

Infatti si ha:

$$\begin{aligned} \mathbb{E}[f(X_t)] &= \sum_{i,j \in V} f(i) P(X_t = j | X_0 = i) \\ &= \sum_{i,j \in V} f(i) P(X_t = i | X_0 = j) P(X_0 = j) \\ &= \sum_{i,j \in V} f(i) p_{ji}^{(t)} p_j(0) = \sum_{j \in V} p_j(0) \sum_{i \in V} f(i) p_{ji}^{(t)} \end{aligned}$$

E, prendendo il limite per  $t \rightarrow \infty$ , si ottiene:

$$\lim_{t \rightarrow \infty} \mathbb{E}[f(X_t)] = \sum_{j \in V} p_j(0) \sum_{i \in V} f(i) \lim_{t \rightarrow \infty} p_{ji}^{(t)} = \sum_{j \in V} p_j(0) \sum_{i \in V} f(i) \pi_i = \sum_{i \in V} f(i) \pi_i$$

**Teorema (10.3):** Sia  $P$  una matrice di transizione bistocastica su  $V$  finito. Allora, sia la distribuzione uniforme  $\bar{\pi} = (\pi_i)_{i \in V}$  definita come:

$$\pi_i = \frac{1}{\#V}$$

Possiamo dire che  $\bar{\pi} = (\pi_i)_{i \in V}$  è una distribuzione stazionaria.

**Teorema (10.4):** Sia  $\bar{\pi} = (\pi_i)_{i \in V}$  una distribuzione uniforme su  $V$  e supponiamo che sia reversibile, ossia:

$$\pi_i p_{ij} = \pi_j p_{ji} \text{ per ogni } i, j \in V$$

Possiamo dire che  $\bar{\pi} = (\pi_i)_{i \in V}$  è una distribuzione stazionaria.

**Teorema (10.4):** Si tratta di verificare il seguente sistema di equazioni:

$$\pi_i = \sum_{j \in V} \pi_j p_{ji} \text{ per ogni } i \in V$$

Questa verifica è molto semplice perché si ha:

$$\sum_{j \in V} \pi_j p_{ji} = \sum_{j \in V} \pi_j p_{ij} = \pi_j \underbrace{\sum_{j \in V} p_{ij}}_{=1} = \pi_i$$

Notare che l'uguaglianza  $\pi_i p_{ij} = \pi_j p_{ji}$  è sempre vera per  $i = j$ ; inoltre, se è vera per una coppia di indici, vale automaticamente anche per la coppia di indici in ordine inverso.

Possiamo fare alcuni appunti riguardo questo teorema:

- La condizione di distribuzione reversibile ha un'interpretazione legata alla proprietà di Markov con "il tempo in senso inverso".
- La condizione di distribuzione reversibile può essere più semplice da verificare rispetto a quella di distribuzione stazionaria.
- Questa proposizione si può considerare anche nei casi senza unicità della distribuzione stazionaria.

**Teorema (10.5):** Supponiamo di avere irriducibilità e aperiodicità. Allora abbiamo una delle due seguenti condizioni e, nel caso in cui  $V$  sia finito, vale la prima delle due:

- C'è ergodicità, per ogni  $i, j \in V$  esiste  $\lim_{t \rightarrow \infty} p_{ij}^{(t)}$  e non dipende da  $j$ , si ha un'unica distribuzione stazionaria  $\bar{\pi} = (\pi_i)_{i \in V}$ , e per ogni  $i, j \in V$  si ha  $\pi_i = \lim_{t \rightarrow \infty} p_{ij}^{(t)} > 0$ .
- Non esistono stati ricorrenti positivi, per ogni  $i, j \in V$  si ha  $\pi_i = \lim_{t \rightarrow \infty} p_{ij}^{(t)} = 0$  e non esistono distribuzioni stazionarie.

# Capitolo 11

## Passeggiata aleatoria su grafi non orientati

Con **Passeggiata aleatoria su grafi non orientati** si considera la seguente situazione:

- Il grafo ha un insieme finito  $V$  di nodi, ed è non orientato.
- Indichiamo con  $E$  l'insieme degli archi non orientati del graf e, per ogni  $i \in V$ , supponiamo di avere  $d_{out}(i) \geq 1$ , ossia esiste almeno un uscente da  $i$ .
- Si suppone di avere una particella che si muove a sul grafo e, ad ogni passo, indipendentemente da quel che è successo prima, si sceglie a caso uno dei  $d_{out}(i)$  archi uscenti dal nodo  $i$  (quindi si tratta di una situazione con scelta equiprobabile degli archi tra cui scegliere)

È importante nel calcolo dei valori  $D := \{d(i) : i \in V\}$  come vengono considerati gli archi uscenti da un nodo e rientrante nello stesso, definiti con il termine *ciclo*, rispetto al conteggio di  $d(i)$ :

- Possono essere considerati come archi doppi, ossia tale arco ha peso doppio rispetto agli altri archi nel conteggio di  $d(i)$ .
- Possono essere considerati come archi singoli, ossia tale arco ha peso uguale a tutti gli altri archi nel conteggio di  $d(i)$ .

Nel primo caso possiamo dire che esiste una relazione tra  $D$  ed  $E$ , infatti:

$$D = \sum_{i \in V} d(i)$$

È importante dire che si ha un grafo connesso (con almeno due elementi) se e solo se c'è irriducibilità.

**Teorema (11.1):** Consideriamo una passeggiata aleatoria su un grafo, e supponiamo che sia un grafo connesso (e con almeno due elementi). Allora si ha aperiodicità se e solo se esistono cammini costituiti da un numero dispari di archi da uno stato in se stesso (ossia si dice che il grafo non è *bipartito*).



**Teorema (11.1):** Per ipotesi c'è irriducibilità (perché il grafo è connesso), quindi tutti gli stati hanno lo stesso periodo dato dal numero:

$$\Delta = \text{MCD}\{s \geq 1 : p_{jj}^{(s)} > 0\}$$

Che non dipende da  $j$ . Inoltre, per come è fatto il grafo connesso, si ha  $p_{jj}^{(s)} > 0$  per  $s$  pari. Allora, se il grafo è bipartito (tutti i cammini da uno stato in se stesso è costituito da un numero pari di archi), si ha  $p_{jj}^{(s)} = 0$  per  $s$  dispari; allora si ha  $\Delta = 2$  e quindi c'è periodicità. Viceversa se il grafo non è bipartito, si ha un MCD di un insieme composto da numeri pari ed un numero dispari; allora si ha  $\Delta = 1$  e quindi c'è aperiodicità.

Una conseguenza interessante di questo teorema è che la presenza di un loop è condizione sufficiente per dire che c'è aperiodicità. Ciò è vero a prescindere da come viene considerato l'arco per il calcolo di  $d(i)$ .

**Teorema (11.22):** Consideriamo una passeggiata aleatoria su un grafo, e supponiamo che sia un grafo connesso. Allora abbiamo le seguenti affermazioni:

- La distribuzione  $\bar{\pi} = (\pi_i)_{i \in V}$ , dove:

$$\pi_i = \frac{d(i)}{D} \quad \text{per ogni } i \in V$$

È una distribuzione stazionaria.

- Supponiamo anche che il grafo sia connesso (e con almeno due elementi) e non bipartito, allora:

$$\lim_{t \rightarrow \infty} p_{ij}^{(t)} = \frac{d(i)}{D} = \frac{1}{h_{ii}} \quad \text{per ogni } i, j \in V$$

**Teorema (11.2):** Dimostriamo le due affermazioni separatamente:

- Si tratta di una verifica diretta; infatti, per ogni  $j \in V$ , si ha:

$$\sum_{i \in V} \pi_i p_{ij} = \sum_{i \in V} \frac{d(i)}{D} \frac{d(i, j)}{d(i)} = \frac{1}{D} \sum_{i \in V} d(i, j) = \frac{1}{D} \sum_{i \in V} d(j, i) = \frac{d(j)}{D} = \pi_j$$

- Per le ipotesi abbiamo una catena irriducibile (perché il grafo è connesso) e aperiodica, ovviamente a stati finiti. Allora, per ogni  $i, j \in V$ , abbiamo che:

$$\lim_{t \rightarrow \infty} p_{ji}^{(t)} = \pi_i = \frac{1}{h_{ii}}$$

Tenendo conto che l'unica distribuzione stazionaria è quella fornita al punto sopra.

## Parte IV

### Esercizi

# Appendice A

## Esercizi sul calcolo delle probabilità

## Appendice B

### Esercizi sulle catene di Markov