

# Probabilità e Statistica Inferenziale

v. 0.2.0

Alessandro Clerici Lorenzini

anno accademico 2020/21

## Indice

<b>1</b>	<b>Definizioni</b>	<b>3</b>
1.1	L'algebra degli eventi . . . . .	3
1.2	Assiomi di Kolmogorov . . . . .	4
1.2.1	Primo assioma . . . . .	4
1.2.2	Secondo assioma . . . . .	4
1.2.3	Terzo assioma . . . . .	4
1.3	Teoremi elementari . . . . .	4
1.3.1	Probabilità dell'evento complementare . . . . .	5
1.3.2	Probabilità dell'unione . . . . .	5
1.3.3	Probabilità dell'evento vuoto . . . . .	5
1.4	Spazi di probabilità . . . . .	6
1.4.1	Spazi equiprobabili . . . . .	6
<b>2</b>	<b>Probabilità condizionata</b>	<b>6</b>
2.1	Teorema delle probabilità totali . . . . .	7
2.1.1	Caso particolare . . . . .	7
2.1.2	Forma generale . . . . .	7
2.2	Teorema di Bayes . . . . .	8
2.3	Eventi indipendenti . . . . .	8
2.3.1	Proprietà . . . . .	8
2.3.2	Tripla di eventi indipendenti . . . . .	8
2.3.3	$n$ -upla di eventi indipendenti . . . . .	9
<b>3</b>	<b>Variabili aleatorie discrete</b>	<b>9</b>
3.1	Definizioni . . . . .	10
3.2	Funzione di ripartizione . . . . .	10
3.3	Funzione di massa di probabilità . . . . .	10
3.4	Valore atteso . . . . .	11
3.4.1	Linearità del valore atteso . . . . .	11
3.5	Varianza di una variabile aleatoria . . . . .	12
3.5.1	Proprietà della varianza . . . . .	12
3.6	Funzione di ripartizione congiunta . . . . .	13
3.7	Funzione di massa di probabilità congiunta . . . . .	13
3.8	Indipendenza . . . . .	14
3.9	Multivariati . . . . .	14

3.9.1	Funzione di ripartizione . . . . .	15
3.9.2	Funzione di massa di probabilità . . . . .	15
3.9.3	Indipendenza . . . . .	15
3.9.4	Valore atteso . . . . .	15
3.9.5	Altre proprietà . . . . .	16
3.10	Covarianza . . . . .	16
3.10.1	Proprietà . . . . .	16
3.10.2	Coefficiente di correlazione . . . . .	19
<b>4</b>	<b>Variabili aleatorie continue</b>	<b>20</b>
4.1	Valore atteso . . . . .	21
4.2	Disuguaglianza di Markov . . . . .	21
4.3	Disuguaglianza di Tchebyshev . . . . .	21
<b>5</b>	<b>Modelli</b>	<b>22</b>
5.1	Modello bernoulliano . . . . .	22
5.1.1	Funzione di massa di probabilità . . . . .	22
5.1.2	Valore atteso . . . . .	24
5.1.3	Varianza . . . . .	24
5.1.4	Funzione di ripartizione . . . . .	24
5.2	Modello binomiale . . . . .	24
5.2.1	Funzione di massa di probabilità . . . . .	24
5.2.2	Valore atteso . . . . .	24
5.2.3	Varianza . . . . .	25
5.2.4	Funzione di ripartizione . . . . .	25
5.2.5	Relazioni tra variabili binomiali . . . . .	25
5.3	Modello uniforme discreto . . . . .	25
5.3.1	Funzione di massa di probabilità . . . . .	25
5.3.2	Funzione di ripartizione . . . . .	26
5.3.3	Valore atteso . . . . .	26
5.3.4	Varianza . . . . .	26
5.4	Modello geometrico . . . . .	26
5.4.1	Funzione di massa di probabilità . . . . .	27
5.4.2	Valore atteso . . . . .	27
5.4.3	Varianza . . . . .	28
5.4.4	Funzione di ripartizione . . . . .	29
5.4.5	Assenza di memoria . . . . .	29
5.5	Modello di Poisson . . . . .	29
5.5.1	Funzione di massa di probabilità . . . . .	29
5.5.2	Valore atteso . . . . .	30
5.5.3	Varianza . . . . .	31
5.5.4	Approssimazione del modello binomiale . . . . .	31
5.6	Modello ipergeometrico . . . . .	32
5.6.1	Funzione di massa di probabilità . . . . .	32
5.6.2	Valore atteso . . . . .	32
5.6.3	Varianza . . . . .	33
5.7	Modello uniforme continuo . . . . .	34
5.7.1	Funzione di densità di probabilità . . . . .	34
5.7.2	Funzione di ripartizione . . . . .	34
5.7.3	Valore atteso . . . . .	35

5.7.4	Varianza . . . . .	35
5.8	Modello esponenziale . . . . .	35
5.8.1	Funzione di densità di probabilità . . . . .	35
5.8.2	Funzione di ripartizione . . . . .	36
5.8.3	Valore atteso . . . . .	36
5.8.4	Varianza . . . . .	37
5.8.5	Assenza di memoria . . . . .	37
5.9	Risultati notevoli sui modelli . . . . .	37
5.10	Modello gaussiano . . . . .	39
5.10.1	Funzione di densità di probabilità . . . . .	40
5.10.2	Funzione di ripartizione . . . . .	40
5.10.3	Valore atteso . . . . .	40
5.10.4	Varianza . . . . .	40
5.10.5	Distribuzione normale standard . . . . .	40
5.10.6	Risultati notevoli . . . . .	41
5.11	Risultati notevoli sui modelli . . . . .	42
5.12	Indici di variabili aleatorie . . . . .	42
5.13	Teorema centrale del limite . . . . .	42
5.13.1	Funzione cumulativa empirica . . . . .	42
<b>6</b>	<b>Statistica inferenziale</b>	<b>43</b>
6.1	Definizioni . . . . .	43
6.1.1	Stimatori non deviati . . . . .	43
6.2	Errore quadratico medio . . . . .	44
6.3	Stimatori consistenti . . . . .	44
6.4	Legge dei grandi numeri . . . . .	45
6.5	Problema dello scarto . . . . .	45
6.6	Processo di Poisson . . . . .	46

## Notazione

Alcune considerazioni sulle convenzioni di notazione scelte:

- per il valore atteso di  $X$  si usa  $\mathbb{E}[X]$ , e non  $\mathcal{E}(X)$ .
- $P(A, B) := P(A \wedge B)$  o talvolta  $P(A, B) := P(A \cap B)$

## 1 Definizioni

La definizione assiomatica della probabilità riassume in pochi punti quei concetti che si possono definire intuitivi nel calcolo delle probabilità.

### 1.1 L'algebra degli eventi

**Definizione 1.1.** Scelto un insieme  $\Omega$  detto spazio campionario o degli eventi, si dice esito un elemento  $\omega \in \Omega$  dell'insieme ed evento un suo sottoinsieme  $E \subseteq \Omega$ .

**Definizione 1.2.** Sia  $\mathcal{A} \subseteq 2^\Omega$  una collezione di sottoinsiemi di  $\Omega$ . Allora  $\mathcal{A}$  è un'algebra se

- $\Omega \in \mathcal{A}$
- $\forall E \subseteq \Omega \quad E \in \mathcal{A} \Rightarrow \bar{E} \in \mathcal{A} \quad \text{con } \bar{E} := \Omega \setminus E$
- $\forall E_1, \dots, E_n \in \mathcal{A} \quad \forall i \ E_i \in \mathcal{A} \Rightarrow \bigcup_{i=1}^n E_i \in \mathcal{A}$

$\mathcal{A}$  è una  $\sigma$ -algebra su  $\Omega$  se l'ultima condizione si può estendere a unioni numerabili qualsiasi.

## 1.2 Assiomi di Kolmogorov

La probabilità viene definita come una funzione di un'algebra degli eventi  $\mathcal{A}$  in  $\mathbb{R}$ :

$$P : \mathcal{A} \rightarrow \mathbb{R}$$

I seguenti assiomi, detti di Kolmogorov, decretano le proprietà che la probabilità rispetta:

### 1.2.1 Primo assioma

L'immagine di  $P$  è l'insieme  $[0, 1] \in \mathbb{R}$ . Equivalentemente, la probabilità di qualunque evento è compresa tra 0 e 1:

$$\forall E \in \mathcal{A} \quad 0 \leq P(E) \leq 1$$

### 1.2.2 Secondo assioma

La probabilità dello spazio campionario è 1:

$$P(\Omega) = 1$$

### 1.2.3 Terzo assioma

La probabilità dell'unione di eventi mutuamente esclusivi, cioè disgiunti (intuitivamente, il cui avvenire dell'uno esclude l'avvenire dell'altro), è uguale alla somma delle probabilità dei singoli:

$$\forall E_1, \dots, E_n \in \mathcal{A} \quad \forall i, j \ E_i \cap E_j = \emptyset \Rightarrow P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i)$$

## 1.3 Teoremi elementari

Dagli assiomi di Kolmogorov derivano alcune proprietà elementari facilmente dimostrabili.

### 1.3.1 Probabilità dell'evento complementare

**Definizione 1.3.** Dato un evento  $E \in \mathcal{A}$ , l'evento complementare è l'evento  $\bar{E} := \Omega \setminus E$ .

**Teorema 1.1** (probabilità dell'evento complementare). *Dato un evento  $E$ , se la probabilità di  $E$  è  $P(E)$ , la probabilità dell'evento complementare di  $E$  è  $1 - P(E)$ :*

$$\forall E \in \mathcal{A} \quad P(\bar{E}) = 1 - P(E)$$

*Dimostrazione.*

$$\left. \begin{aligned} E \cap \bar{E} &= \emptyset \\ E \cup \bar{E} &= \Omega \end{aligned} \right\} \quad \text{definizione di evento complementare}$$

$$\begin{aligned} 1 &= P(\Omega) && \text{secondo assioma} \\ &= P(E \cup \bar{E}) \\ &= P(E) + P(\bar{E}) && \text{terzo assioma} \end{aligned}$$

ergo:

$$P(\bar{E}) = 1 - P(E) \quad \square$$

### 1.3.2 Probabilità dell'unione

**Teorema 1.2.** *Dati due eventi  $E, F \in \mathcal{A}$ , la probabilità della loro unione è uguale alla somma delle loro probabilità meno la probabilità dell'intersezione:*

$$\forall E, F \in \mathcal{A} \quad P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

*Dimostrazione.* L'unione degli eventi è scrivibile come l'unione di due insiemi disgiunti:

$$E \cup F = E \cup (\bar{E} \cap F)$$

Passando alle probabilità:

$$\begin{aligned} P(E \cup F) &= P(E) + P(\bar{E} \cap F) \\ &= P(E) + P(\bar{E} \cap F) + P(E \cap F) - P(E \cap F) \\ &= P(E) + P((\bar{E} \cap F) \cup (E \cap F)) - P(E \cap F) && \text{terzo assioma} \\ &= P(E) + P(F) - P(E \cap F) \end{aligned} \quad \square$$

### 1.3.3 Probabilità dell'evento vuoto

La probabilità dell'evento vuoto ( $\emptyset$ ) è 0. Un evento con probabilità nulla viene detto evento impossibile.

$$P(\emptyset) = 0$$

*Dimostrazione.*

$$\begin{aligned} P(\Omega) &= 1 && \text{secondo assioma} \\ P(\emptyset) &= P(\bar{\Omega}) \\ &= 1 - P(\Omega) && \text{teorema 1.1} \\ &= 1 - 1 = 0 \end{aligned} \quad \square$$

## 1.4 Spazi di probabilità

**Definizione 1.4.** Uno spazio di probabilità è una tripla  $(\Omega, \mathcal{A}, P)$  composta da uno spazio campionario  $\Omega$ , un'algebra degli eventi  $\mathcal{A}$  e una funzione di probabilità  $P$ .

### 1.4.1 Spazi equiprobabili

**Definizione 1.5.** Uno spazio è equiprobabile se gli eventi elementari (cioè corrispondenti a singoletti) hanno probabilità costante  $p$ .

Un evento elementare è composto da un singolo esito, cioè è un singoletto dell'insieme delle parti dello spazio campionario.

**Teorema 1.3.** In uno spazio equiprobabile, la probabilità di ogni evento elementare  $e$  è uguale al reciproco del numero  $n$  degli eventi elementari (che sono ovviamente a due a due disgiunti):

$$P(e) = \frac{1}{n} \quad (1)$$

*Dimostrazione.*

$$\begin{aligned} 1 = P(\Omega) &= P\left(\bigcup_{i=1}^n e_i\right) && \text{secondo assioma} \\ &= \sum_{i=1}^n P(e_i) = np && \text{terzo assioma} \end{aligned}$$

Ergo:

$$\forall e_i \quad P(e_i) = p = \frac{1}{n} \quad \square$$

Gli eventi non elementari possono essere espressi come unione di eventi elementari.

**Teorema 1.4.** In uno spazio equiprobabile, dato un evento  $E = \{e_1, \dots, e_k\}$ :

$$P(E) = \frac{|E|}{n} \quad (2)$$

*Dimostrazione.*

$$P(E) = \sum_{i=1}^{|E|} P(e_i) = \frac{|E|}{n} \quad \square$$

## 2 Probabilità condizionata

La probabilità condizionata vede protagonisti un evento condizionato e un evento condizionante e, concettualmente, restringe lo spazio campionario all'evento condizionante.

**Definizione 2.1.** La probabilità condizionata da un evento  $F$  (evento condizionante) su un evento  $E$  (evento condizionato) viene definita, quando  $F$  non è impossibile (ossia per  $P(F) \neq 0$ ), come segue:

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

Dalla definizione di probabilità condizionata deriva la seguente **regola di fattorizzazione**:

$$P(E \cap F) = P(E | F) \cdot P(F) \quad (3)$$

## 2.1 Teorema delle probabilità totali

### 2.1.1 Caso particolare

Dati eventi  $E, F \in \mathcal{A}$ , l'evento  $E$  può essere scritto come la parte di  $E$  che non contiene elementi di  $F$ , ossia  $E \cap \bar{F}$ , unita disgiuntamente alla parte che ne contiene, ossia  $E \cap F$ :

$$\begin{aligned} (E \cap \bar{F}) \cup (E \cap F) &= E \cap (F \cup \bar{F}) && \text{proprietà distributiva} \\ &= E \cap \Omega = E \end{aligned}$$

$$\begin{aligned} (E \cap \bar{F}) \cup (E \cap F) &= E \cap (F \cup \bar{F}) && \text{proprietà distributiva} \\ &= E \cap \Omega = E \end{aligned}$$

Avendo espresso  $E$  come unione di eventi disgiunti, si può applicare il terzo assioma:

$$P(E) = P((E \cap \bar{F}) \cup (E \cap F)) = P(E \cap \bar{F}) + P(E \cap F)$$

e, per la regola (3) di fattorizzazione, assunto  $P(F) \neq 0 \wedge P(\bar{F}) \neq 0$ :

$$= P(E | \bar{F}) \cdot P(\bar{F}) + P(E | F) \cdot P(F)$$

### 2.1.2 Forma generale

**Teorema 2.1** (delle probabilità totali). *Data una partizione di  $\Omega$  composta da eventi non impossibili  $F_1, \dots, F_n$  e un evento  $E$ :*

$$P(E) = \sum_{i=1}^n P(E | F_i) \cdot P(F_i)$$

*Dimostrazione.* La dimostrazione avviene in analogia con il caso particolare. Questa volta, tuttavia, gli eventi della partizione sono disgiunti per ipotesi e non in quanto complementari:

$$P(E) = P\left(\bigcup_{i=1}^n (E \cap F_i)\right) = \sum_{i=1}^n P(E \cap F_i) = \sum_{i=1}^n P(E | F_i) \cdot P(F_i) \quad \square$$

## 2.2 Teorema di Bayes

Nelle stesse ipotesi del teorema delle probabilità totali si dimostra un altro teorema, detto di Bayes:

**Teorema 2.2** (di Bayes). *Data una partizione di  $\Omega$  composta da eventi non impossibili  $F_1, \dots, F_n$  e un evento  $E$ :*

$$P(F_j | E) = \frac{P(E | F_j) \cdot P(F_j)}{\sum_{i=1}^n P(E | F_i) \cdot P(F_i)}$$

*Dimostrazione.*

$$P(F_j | E) = \frac{P(F_j \cap E)}{P(E)} = \frac{P(E | F_j) \cdot P(F_j)}{\sum_{i=1}^n P(E | F_i) \cdot P(F_i)} \quad \square$$

## 2.3 Eventi indipendenti

**Definizione 2.2.** Eventi  $E$  e  $F$ , con  $P(F) > 0$ , si dicono indipendenti se  $P(E | F) = P(E)$

Equivalentemente, per definizione di probabilità condizionata

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

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

e ovviamente, se  $P(E) > 0$ :

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

### 2.3.1 Proprietà

**Teorema 2.3.** *Se  $E$  e  $F$  sono indipendenti, allora  $E$  e  $\bar{F}$  sono indipendenti.*

*Dimostrazione.*

$$E = (E \cap F) \cup (E \cap \bar{F})$$

$$\begin{aligned} P(E \cap \bar{F}) &= P(E) - P(E \cap F) && \text{terzo assioma} \\ &= P(E) - P(E) \cdot P(F) && \text{in quanto } E \text{ ed } F \text{ sono indipendenti} \\ &= P(E)(1 - P(F)) \\ &= P(E) \cdot P(\bar{F}) && \text{teorema 1.1} \end{aligned} \quad \square$$

### 2.3.2 Tripla di eventi indipendenti

**Definizione 2.3.** Tre eventi  $E$ ,  $F$ ,  $G$  sono indipendenti se

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



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

Per quanto riguarda  $E, F \cup G$ :

$$\begin{aligned}
 P(E \cap (F \cup G)) &= P((E \cap F) \cup (E \cap G)) && \text{distributiva} \\
 &= P(E \cap F) + P(E \cap G) - P((E \cap F) \cap (E \cap G)) && \text{teorema 1.2} \\
 &= P(E)P(F) + P(E)P(G) - P(E)P(F)P(G) && \text{indipendenza} \\
 &= P(E)(P(F) + P(G) - P(F)P(G))
 \end{aligned}$$

### 2.3.3 $n$ -upla di eventi indipendenti

**Definizione 2.4** ( $n$ -upla di eventi indipendenti). Eventi  $E_1, \dots, E_n$  si dicono indipendenti se e solo se comunque apportata su di essi una selezione di  $r$  eventi la probabilità dell'intersezione degli eventi selezionati è uguale al prodotto della probabilità dei singoli.

$$\forall r < n \in \mathbb{N} \forall 1 \leq \alpha_1 \leq \dots \leq \alpha_r \leq n \quad P\left(\bigcap_{i=1}^r E_{\alpha_i}\right) = \prod_{i=1}^r P(E_{\alpha_i})$$

*Esempio 2.3.1.* Dato un sistema in serie di dimensione  $n$ :

$$\begin{aligned}
 \forall i = 1, \dots, n \quad p_i &= P(\text{l}'i\text{-esimo componente funziona}) \\
 P(\text{il sistema funziona}) &= P(\text{tutti i componenti funzionano}) = \\
 &P\left(\bigcap_{i=1}^n \text{l}'i\text{-esimo componente funziona}\right)
 \end{aligned}$$

In un sistema in parallelo:

$$\begin{aligned}
 P(\text{il sistema funziona}) &= 1 - P(\text{il sistema non funziona}) = \\
 &1 - P(\text{tutti i componenti non funzionano}) = \\
 &1 - P\left(\bigcap_{i=1}^n \text{l}'i\text{-esimo componente non funziona}\right) = \\
 &1 - \prod_{i=1}^n P(\text{l}'i\text{-esimo componente non funziona}) = 1 - \prod_{i=1}^n (1 - p_i) =
 \end{aligned}$$

## 3 Variabili aleatorie discrete

**Definizione 3.1.** In uno spazio di probabilità  $(\Omega, \mathcal{A}, p)$ , una variabile aleatoria (o casuale) è una funzione<sup>1</sup>  $X : \Omega \rightarrow \mathbb{R}$ . L'immagine di una variabile aleatoria si dice supporto e si indica con  $D_X$ . Ogni elemento del supporto, cioè l'immagine per  $X$  di un elemento di  $\Omega$ , si dice specificazione. Una specificazione generica di  $X$  si indica tipicamente con  $x$ .

<sup>1</sup>in verità, la funzione deve soddisfare il criterio:  $\{\omega \mid X(\omega) \leq r\} \in \mathcal{A} \quad \forall r \in \mathbb{R}$

### 3.1 Definizioni

**Definizione 3.2.** Una variabile aleatoria è discreta se e solo se il suo supporto è numerabile.

*Esempio 3.1.1.* Se  $\Omega$  è l'insieme di esiti del lancio di due dadi, una possibile variabile aleatoria  $X$  è la somma degli esiti dei due dadi. In questo caso immagini di  $X$  non sono equiprobabili. Se invece per  $X$  si sceglie il valore del primo dado, le immagini sono equiprobabili.

**Definizione 3.3** (funzione indicatrice). Siano  $A$  e  $D$  insiemi tali che  $A \subseteq D$ . La funzione indicatrice (o caratteristica) di  $A$  rispetto a  $D$  è la funzione  $I : D \rightarrow \{0, 1\}$  che associa a un elemento di  $D$  il numero 1 se l'elemento appartiene ad  $A$ , o 0 altrimenti:

$$I_A(x) := \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

**Definizione 3.4** (funzione indicatrice di un evento). La funzione indicatrice di un evento  $A$  è la variabile aleatoria che ha specificazione 1 se l'evento  $A$  si verifica e 0 se l'evento non si verifica.

### 3.2 Funzione di ripartizione

**Definizione 3.5** (funzione di ripartizione). Data una variabile aleatoria  $X$ , la funzione di ripartizione (o di distribuzione cumulativa)  $F_X : \mathbb{R} \rightarrow [0, 1]$  è la funzione che associa a un valore  $x \in \mathbb{R}$  la probabilità che l'esito di  $X$  ne sia minore o uguale:

$$\forall x \in \mathbb{R} \quad F_X(x) := P(X \leq x)$$

La funzione di ripartizione si annulla quandunque l'input è minore della minima specificazione.

### 3.3 Funzione di massa di probabilità

**Definizione 3.6** (funzione di massa di probabilità). Data una variabile aleatoria discreta  $X$ , la funzione di massa di probabilità  $p_X : \mathbb{R} \rightarrow [0, 1]$  è la funzione che associa a un valore  $x \in \mathbb{R}$  la probabilità che l'esito di  $X$  sia uguale a  $x$ :

$$\forall x \in \mathbb{R} \quad p_X(x) = P(x = X)$$

La somma delle immagini della funzione di massa di probabilità per tutte le specificazioni di  $X$  è ovviamente uguale a 1:

$$\sum_i p_X(x_i) = 1 \tag{4}$$

La funzione di massa di probabilità di  $X$  si annulla nei punti di  $\mathbb{R}$  non appartenenti al supporto di  $X$ .

**Proprietà 3.3.1.** Per le variabili aleatorie discrete vale la seguente relazione tra la funzione di ripartizione e la funzione di massa di probabilità:

$$F_X(x) = \sum_{a \leq x} p_X(a)$$

con  $a \in D_X$ .

*Dimostrazione.*

$$\begin{aligned}
 F_X(x) &= P(X \leq x) && \text{definizione 3.5} \\
 &= P\left(\bigcup_{a \leq x} \{X = a\}\right) \\
 &= \sum_{a \leq x} P(X = a) && \text{unione di eventi disgiunti} \\
 &= \sum_{a \leq x} p_X(a) && \text{definizione 3.6}
 \end{aligned}$$

□

### 3.4 Valore atteso

Il valore atteso di una variabile aleatoria  $X$  è un indice di centralità delle specificazioni della variabile aleatoria.

**Definizione 3.7** (valore atteso). Data una variabile aleatoria discreta  $X$ , il valore atteso di  $X$  è la media delle specificazioni di  $X$  pesata con le rispettive probabilità:

$$\mathbb{E}[X] = \sum_i x_i P(X = x_i)$$

Talvolta si usa anche la notazione  $\mu = \mu_X = \mathbb{E}[X]$ .

In generale, non è detto che la sommatoria che definisce il valore atteso converga.

**Proprietà 3.4.1.** Il valore atteso della funzione indicatrice di un evento è uguale alla probabilità dell'evento:

$$\mathbb{E}[I_A] = P(A)$$

*Dimostrazione.* Applicando la definizione 3.7 di valore atteso e la 3.4 di funzione indicatrice di un evento:

$$\mathbb{E}[I_A] = 1 * P(A) + 0 * (1 - P(A)) = P(A)$$

□

#### 3.4.1 Linearità del valore atteso

**Proprietà 3.4.2.** Il valore atteso di una variabile aleatoria discreta  $X$  è lineare rispetto a una trasformazione lineare  $ax + b$  applicata alle specificazioni  $x_i$  di  $X$ :

$$D_Y = \{ax_i + b \mid x_i \in D_X\} \quad \Rightarrow \quad \mathbb{E}[Y] = a\mathbb{E}[X] + b$$

Si scrive anche  $Y = aX + b$ .

*Dimostrazione.* Poiché le probabilità per le rispettive specificazioni rimangono invariate:

$$\begin{aligned}
 \mathbb{E}[Y] &= \sum_i (ax_i + b)P(X = x_i) \\
 &= a \sum_i x_i P(X = x_i) + b \sum_i P(X = x_i) \\
 &= a \underbrace{\sum_i x_i p_X(x_i)}_{\mathbb{E}[x]} + b \underbrace{\sum_i p_X(x_i)}_1 \\
 &= a\mathbb{E}[x] + b
 \end{aligned}
 \tag{4} \text{ e definizione 3.7}$$

□

### 3.5 Varianza di una variabile aleatoria

**Definizione 3.8** (varianza di una variabile aleatoria). La varianza di una variabile aleatoria  $X$  è definita come segue:

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

**Definizione 3.9** (deviazione standard di una variabile aleatoria). La deviazione standard di una variabile aleatoria  $X$  è così definita:

$$\sigma(X) = \sigma_X := \sqrt{\text{var}(X)}$$

#### 3.5.1 Proprietà della varianza

**Proprietà 3.5.1.** La varianza di  $X$  può essere espressa equivalentemente come la differenza tra il valore atteso del quadrato di  $X$  e il valore atteso al quadrato di  $X$ :

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \tag{5}$$

*Dimostrazione.*

$$\begin{aligned}
 \text{var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] && \text{definizione 3.8} \\
 &= \mathbb{E}[X^2 - 2\mu X + \mu^2] && \text{sviluppando il quadrato} \\
 &= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 && \text{linearità del valore atteso} \\
 &= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 && \text{definizione di } \mu \\
 &= \mathbb{E}[X^2] - \mathbb{E}[X]^2
 \end{aligned}$$

□

**Proprietà 3.5.2.** La varianza della funzione indicatrice di un evento  $A$  è uguale al prodotto delle probabilità di  $A$  e del suo complementare:

$$\text{var}(I_A) = P(A)P(\bar{A})$$

*Dimostrazione.*

$$\begin{aligned}
 \text{var}(I_A) &= \mathbb{E}[I_A^2] - \mathbb{E}[I_A]^2 \\
 &= \mathbb{E}[I_A] - \mathbb{E}[I_A]^2 && \text{idempotenza di } I_A \\
 &= \mathbb{E}[I_A](1 - \mathbb{E}[I_A]) && \text{proprietà distributiva} \\
 &= P(A)P(\bar{A})
 \end{aligned}$$

□

**Proprietà 3.5.3.** *In una trasformazione lineare su una variabile aleatoria  $X$ , una traslazione non influisce sulla varianza; al contrario, una dilatazione si traduce in una dilatazione quadratica.*

$$\text{var}(aX + b) = a^2 \text{var}(X)$$

*Dimostrazione.*

$$\begin{aligned}
 \text{var}(aX + b) &= \mathbb{E}[(aX + b - \mathbb{E}[aX + b])^2] \\
 &= \mathbb{E}[(aX + b - a\mathbb{E}[X] - b)^2] \\
 &= \mathbb{E}[(a(X - \mathbb{E}[X]))^2] \\
 &= a^2 \mathbb{E}[(X - \mathbb{E}[X])^2] \\
 &= a^2 \text{var}(X)
 \end{aligned}$$

□

### 3.6 Funzione di ripartizione congiunta

**Definizione 3.10.** Data una coppia di variabili aleatorie  $X, Y$ , la funzione di ripartizione congiunta  $F_{X,Y} : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$  (o di distribuzione cumulativa congiunta) è la funzione che associa a una coppia di valori reali per  $x$  e  $y$  la probabilità che l'esito di  $X$  sia minore o uguale a  $x$  e l'esito di  $Y$  sia minore o uguale a  $y$ :

$$\forall x, y \in \mathbb{R} \quad F_{X,Y}(x, y) := P(X \leq x, Y \leq y)$$

Il limite per  $y$  che tende a  $+\infty$  della funzione di ripartizione congiunta di  $X$  e  $Y$  è uguale alla funzione di ripartizione di  $X$ . In questo caso si dice che  $F_X(x)$  è la distribuzione marginale rispetto a  $X$ :

$$\lim_{y \rightarrow +\infty} F_{X,Y}(x, y) = \lim_{y \rightarrow +\infty} P(X \leq x, Y \leq y) = P(X \leq x) = F_X(x) \quad (6)$$

Analogamente vale, ovviamente, per  $x \rightarrow +\infty$ .

### 3.7 Funzione di massa di probabilità congiunta

**Definizione 3.11.** Data una coppia di variabili aleatorie  $X, Y$ , la funzione di massa di probabilità congiunta  $p_{X,Y} : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$  è la funzione che associa a una coppia di valori reali per  $x$  e  $y$  la probabilità che l'esito di  $X$  sia uguale a  $x$  e quello di  $Y$  sia uguale a  $y$ :

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

Analogamente a quanto visto nella (6) con il limite, la somma per tutte le specificazioni  $y_j$  di  $Y$  di  $p_{X,Y}(x, y_j)$  è uguale alla massa di probabilità marginale rispetto a  $X$ :

$$\sum_j p_{X,Y}(x, y_j) = p_X(x) \quad (7)$$

### 3.8 Indipendenza

**Definizione 3.12.** Variabili aleatorie  $X$  e  $Y$  sono indipendenti se e solo se

$$\forall A, B \subseteq \mathbb{R} \quad P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

La definizione è equivalente alla seguente proprietà della funzione di massa di probabilità congiunta:

**Proprietà 3.8.1.** Variabili aleatorie  $X$  e  $Y$  sono indipendenti se e solo se

$$a, b \in \mathbb{R} \quad p_{X,Y}(a, b) = p_X(a)p_Y(b)$$

*Dimostrazione.* L'implicazione verso destra:

$$\begin{aligned} \forall A, B \subseteq \mathbb{R} \quad P(X \in A, Y \in B) &= P(X \in A)P(Y \in B) \\ \Rightarrow \\ a, b \in \mathbb{R} \quad p_{X,Y}(a, b) &= p_X(a)p_Y(b) \end{aligned}$$

è immediata dal momento che si possono scegliere sottoinsiemi di  $\mathbb{R}$  corrispondenti a singoletti contenenti  $a$  e  $b$  ( $A = \{a\}$  e  $B = \{b\}$ ).

L'implicazione inversa parte dalla ricostruzione di  $A$  e  $B$  come unioni di singoletti:

$$\begin{aligned} P(X \in A, Y \in B) &= \sum_{a \in A} \sum_{b \in B} p_{X,Y}(a, b) \\ &= \sum_{a \in A} \sum_{b \in B} p_X(a)p_Y(b) && \text{ipotesi} \\ &= \left( \sum_{a \in A} p_X(a) \right) \left( \sum_{b \in B} p_Y(b) \right) \\ &= P(X \in A)P(Y \in B) \end{aligned}$$

□

Si dimostra anche la seguente equivalenza:

**Proprietà 3.8.2.** Variabili aleatorie  $X$  e  $Y$  sono indipendenti se e solo se

$$a, b \in \mathbb{R} \quad F_{X,Y}(a, b) = F_X(a)F_Y(b)$$

### 3.9 Multivariati

I concetti visti per coppie di variabili aleatorie si estendono a vettori aleatori (più comunemente detti variabili aleatorie multivariate, o multivariati) di dimensione arbitraria:

### 3.9.1 Funzione di ripartizione

**Definizione 3.13.** Date  $n$  variabili aleatorie  $X_1, \dots, X_n$ , la funzione di ripartizione  $F_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow [0, 1]$  è la funzione che associa a una  $n$ -upla di valori reali  $x_1, \dots, x_n$  la probabilità che l'esito di  $X_i$  sia minore o uguale a  $x_i$  per ogni  $i$  da 1 a  $n$ :

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

### 3.9.2 Funzione di massa di probabilità

**Definizione 3.14.** Date  $n$  variabili aleatorie  $X_1, \dots, X_n$ , la funzione di massa di probabilità  $p_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow [0, 1]$  è la funzione che associa a una  $n$ -upla di valori reali  $x_1, \dots, x_n$  la probabilità che l'esito di  $X_i$  sia uguale a  $x_i$  per ogni  $i$  da 1 a  $n$ :

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$$

### 3.9.3 Indipendenza

**Definizione 3.15.** Variabili aleatorie  $X_1, \dots, X_n$  sono indipendenti se e solo se

$$\forall A_1, \dots, A_n \subseteq \mathbb{R} \quad P\left(\bigcap_{i=1}^n X_i \in A_i\right) = \prod_{i=1}^n P(X_i \in A_i)$$

### 3.9.4 Valore atteso

Il valore atteso per una funzione applicata a una coppia di variabili aleatorie discrete coinvolge tutte le combinazioni<sup>2</sup> di specificazioni delle due variabili:

$$\mathbb{E}[f(x, y)] = \sum_{i,j} f(x_i, y_j) P(X = x_i, Y = y_j)$$

**Proprietà 3.9.1.** Il valore atteso della somma di due variabili aleatorie discrete è uguale alla somma dei valori attesi delle stesse.

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

*Dimostrazione.*

$$\begin{aligned} \mathbb{E}[X + Y] &= \sum_i \sum_j (x_i + y_j) P(X = x_i, Y = y_j) \\ &= \sum_i \sum_j x_i P(X = x_i, Y = y_j) + \sum_i \sum_j y_j P(X = x_i, Y = y_j) \\ &= \sum_i x_i \sum_j P(X = x_i, Y = y_j) + \sum_j y_j \sum_i P(X = x_i, Y = y_j) \end{aligned}$$

Le due sommatorie interne sono probabilità marginali:

$$\begin{aligned} &= \sum_i x_i P(X = x_i) + \sum_j y_j P(Y = y_j) \\ &= \mathbb{E}[X] + \mathbb{E}[Y] \end{aligned}$$

□

---

<sup>2</sup>la sommatoria con doppio indice può essere scomposta in due sommatorie, una per ogni indice

Generalizzando:

**Proprietà 3.9.2.** *Il valore atteso della somma di variabili aleatorie discrete è uguale alla somma dei valori attesi delle stesse.*

$$\mathbb{E} \left[ \sum_i X_i \right] = \sum_i \mathbb{E} [X_i]$$

Di conseguenza, se si può esprimere una variabile  $X$  come la somma di variabili  $X_1, \dots, X_n$ , è sufficiente calcolare il valore atteso di ciascuna componente per calcolare quello della variabile iniziale.

### 3.9.5 Altre proprietà

Il seguente risultato è una generalizzazione della varianza, utile a valutare quanto un'approssimazione  $c$  di  $X$  è buona. La migliore approssimazione è  $c = \mathbb{E} [X]$ , in coerenza con l'interpretazione semantica di  $\mathbb{E} [X]$ .

**Proprietà 3.9.3.**

$$\mathbb{E} [(X - c)^2] = \text{var}(X) + (\mu - c)^2 \geq \text{var}(X)$$

*Dimostrazione.* Se  $\mu = \mathbb{E} [X]$

$$\begin{aligned} \mathbb{E} [(X - c)^2] &= \mathbb{E} [(X - \mu + \mu - c)^2] \\ &= \mathbb{E} [(X - \mu)^2 - 2(X - \mu)(\mu - c) + (\mu - c)^2] \\ &= \mathbb{E} [(X - \mu)^2] - \mathbb{E} [2(X - \mu)(\mu - c)] + (\mu - c)^2 \\ &= \mathbb{E} [(X - \mu)^2] - 2(\mu - c) \underbrace{\mathbb{E} [X - \mu]}_{=\mathbb{E}[X]-\mu=0} + (\mu - c)^2 \\ &= \mathbb{E} [(X - \mu)^2] + (\mu - c)^2 \end{aligned}$$

□

## 3.10 Covarianza

**Definizione 3.16.** Date due variabili aleatorie  $X$  e  $Y$ , la covarianza tra  $X$  e  $Y$  è così definita:

$$\text{cov}(X, Y) = \mathbb{E} [(X - \mu_X)(Y - \mu_Y)] \quad (8)$$

Ovviamente la covarianza è simmetrica, ossia  $\text{cov}(X, Y) = \text{cov}(Y, X)$ .

### 3.10.1 Proprietà

La seguente è una definizione equivalente per la covarianza tra due variabili aleatorie:

**Proprietà 3.10.1.** Date due variabili aleatorie  $X$  e  $Y$ , la covarianza tra  $X$  e  $Y$  è uguale al valore atteso del prodotto delle due meno il prodotto dei rispettivi valori attesi:

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



*Dimostrazione.* Sviluppando la (8):

$$\begin{aligned}\operatorname{cov}(X, Y) &= \mathbb{E}[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y] \\ &= \mathbb{E}[X, Y] - \underbrace{\mu_X \mathbb{E}[Y]}_{\mu_X \mu_Y} - \underbrace{\mu_Y \mathbb{E}[X]}_{\mu_X \mu_Y} + \mu_X \mu_Y \\ &= \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]\end{aligned}$$

□

**Proprietà 3.10.2.** *La covarianza è lineare rispetto alla dilatazione di una variabile per una costante:*

$$\operatorname{cov}(aX, Y) = a \operatorname{cov}(X, Y)$$

*Dimostrazione.* Applicando la proprietà 3.10.1:

$$\begin{aligned}\operatorname{cov}(aX, Y) &= \mathbb{E}[aXY] - \mathbb{E}[aX] \mathbb{E}[Y] \\ &= a(\mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]) \\ &= a \operatorname{cov}(X, Y)\end{aligned}$$

□

**Proprietà 3.10.3.**

$$\operatorname{cov}(X + Y, Z) = \operatorname{cov}(X, Z) + \operatorname{cov}(Y, Z)$$

*Dimostrazione.*

$$\begin{aligned}\operatorname{cov}(X + Y, Z) &= \mathbb{E}[(X + Y)Z] - \mathbb{E}[X + Y] \mathbb{E}[Z] \\ &= \mathbb{E}[XZ + YZ] - (\mathbb{E}[X] + \mathbb{E}[Y]) \mathbb{E}[Z] \\ &= \mathbb{E}[XZ] + \mathbb{E}[YZ] - \mathbb{E}[X] \mathbb{E}[Z] - \mathbb{E}[Y] \mathbb{E}[Z] \\ &= \operatorname{cov}(X, Z) + \operatorname{cov}(Y, Z)\end{aligned}$$

□

Generalizzando:

**Proprietà 3.10.4.**

$$\operatorname{cov}\left(\sum_i X_i, \sum_j Y_j\right) = \sum_i \sum_j \operatorname{cov}(X_i, Y_j)$$

*Dimostrazione.*

$$\begin{aligned}\operatorname{cov}\left(\sum_i X_i, \sum_j Y_j\right) &= \sum_i \operatorname{cov}\left(X_i, \sum_j Y_j\right) && \text{proprietà 3.10.3} \\ &= \sum_i \sum_j \operatorname{cov}(X_i, Y_j) && \text{proprietà 3.10.3}\end{aligned}$$

□

**Proprietà 3.10.5.** *La covarianza tra una variabile aleatoria  $X$  e se stessa è uguale alla varianza di  $X$ :*

$$\text{cov}(X, X) = \text{var}(X)$$

*Dimostrazione.*

$$\begin{aligned}\text{cov}(X, X) &= \mathbb{E}[(X - \mu)(X - \mu)] \\ &= \mathbb{E}[(X - \mu)^2] \\ &= \text{var}(X)\end{aligned}$$

□

**Proprietà 3.10.6.** *La varianza della somma di due variabili aleatorie  $X$  e  $Y$  è uguale alla somma delle rispettive varianze più il doppio della loro covarianza:*

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X, Y)$$

*Dimostrazione.* Usando la definizione alternativa (proprietà 3.5.1):

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

□

Generalizzando:

**Proprietà 3.10.7.**

$$\text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j)$$

**Teorema 3.1.** *Il valore atteso del prodotto di variabili aleatorie discrete indipendenti è uguale al prodotto dei rispettivi valori attesi:*

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

*Dimostrazione.*

$$\begin{aligned}\mathbb{E}[XY] &= \sum_i \sum_j x_i y_j P(X = x_i, Y = y_j) \\ &= \sum_i \sum_j x_i y_j P(X = x_i) P(Y = y_j) \\ &= \left( \sum_i x_i P(X = x_i) \right) \left( \sum_j y_j P(Y = y_j) \right) \\ &= \mathbb{E}[X] \mathbb{E}[Y]\end{aligned}$$

□

**Corollario 3.1.1.** *La covarianza di due variabili aleatorie discrete indipendenti  $X$  e  $Y$  è nulla:*

$$\text{cov}(X, Y) = 0$$

*Dimostrazione.*

$$\begin{aligned}\text{cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y] \\ &= \mathbb{E}[X] \mathbb{E}[Y] - \mathbb{E}[X] \mathbb{E}[Y] \\ &= 0\end{aligned}$$

□

**Corollario 3.1.2.** *Date variabili indipendenti  $X_1, \dots, X_n$ :*

$$\text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i)$$

*Dimostrazione.* Applicando la proprietà 3.10.6:

$$\begin{aligned}\text{var}\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n \text{var}(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{var}(X_i)\end{aligned}$$

□

**Proprietà 3.10.8.** *Date funzioni indicatrici di eventi, o in generale variabili aleatorie bernoulliane (ossia di specificazione binaria):*

$$\text{cov}(X, Y) > 0 \Rightarrow P(X = 1 \mid Y = 1) > P(X = 1)$$

*Dimostrazione.*

$$\begin{aligned}\text{cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y] \\ &= P(XY = 1) - P(X = 1)P(Y = 1) \\ &= P(X = 1, Y = 1) - P(X = 1)P(Y = 1)\end{aligned}$$

Se  $\text{cov}(X, Y) > 0$

$$\begin{aligned}P(X = 1, Y = 1) &> P(X = 1)P(Y = 1) \\ \frac{P(X = 1, Y = 1)}{P(Y = 1)} &> P(X = 1) \\ P(X = 1 \mid Y = 1) &> P(X = 1)\end{aligned}$$

□

### 3.10.2 Coefficiente di correlazione

**Definizione 3.17.** Il coefficiente di correlazione  $\rho$  è un indice di correlazione tra variabili aleatorie discrete ed è così definito:

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

## 4 Variabili aleatorie continue

**Definizione 4.1** (variabile aleatoria continua). Una variabile aleatoria  $X$  è continua se e solo se non è discreta, ossia il suo supporto non è numerabile.

**Definizione 4.2** (funzione di densità di probabilità). Data una variabile aleatoria  $X$ , la funzione di densità di probabilità di  $X$  è una funzione  $f_X : \mathbb{R} \rightarrow \mathbb{R}^+$  integrabile e tale che:

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

La probabilità di variabili aleatorie continue viene calcolata integrando la funzione di densità di probabilità:

$$\forall B \subseteq \mathbb{R}, P(X \in B) = \int_B f_X(x) dx$$

O, in un intervallo  $[a, b]$ :

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx$$

che è equivalente a quella dell'intervallo  $(a, b)$ .

La proprietà della definizione non è altro che la conseguenza del primo assioma di Kolmogorov esteso al continuo:

$$P(X \in \mathbb{R}) = \int_{-\infty}^{+\infty} f_X(x) dx = \int_{\mathbb{R}} f_X(x) dx = 1$$

Per variabili aleatorie continue non ha senso calcolare la probabilità che una variabile assuma uno specifico valore, infatti:

$$\forall a \in \mathbb{R} \quad P(X = a) = \int_a^a f_X(x) dx = 0$$

In un certo intervallo che approssima  $a$ , ossia  $[a - \frac{\varepsilon}{2}, a + \frac{\varepsilon}{2}]$ , invece:

$$P\left(a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2}\right) = \int_{a - \frac{\varepsilon}{2}}^{a + \frac{\varepsilon}{2}} f_X(x) dx \approx \varepsilon f_X(a)$$

Calcolando l'integrale della funzione di densità in un intervallo del tipo  $(-\infty, a)$  si calcola di fatto il valore della funzione di ripartizione in  $a$ :

$$\int_{-\infty}^a f_X(x) dx = P(X \leq a) = F_X(a)$$

Per il teorema fondamentale del calcolo integrale, da ciò deriva che la funzione di densità di probabilità è la derivata della funzione di ripartizione:

$$\frac{d}{dx} F_X(x) = f(x)$$

### 4.1 Valore atteso

**Definizione 4.3.** Il valore atteso per variabili aleatorie continue è così definito:

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

**Proprietà 4.1.1.** Per variabili aleatorie che non assumono specificazioni negative:

$$\forall x < 0 \quad f_X(x) = 0$$

vale

$$\int_0^{+\infty} 1 - F_X(x) dx = \mathbb{E}[X]$$

### 4.2 Disuguaglianza di Markov

**Teorema 4.1** (Disuguaglianza di Markov). Sia  $X$  una variabile aleatoria non negativa. Allora per ogni  $a$  reale positivo:

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

*Dimostrazione.*

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{+\infty} x f_X(x) dx \\ &= \int_0^{+\infty} x f_X(x) dx && \text{variabile non negativa} \\ &= \underbrace{\int_0^a x f_X(x) dx}_{\geq 0} + \int_a^{+\infty} x f_X(x) dx \\ &\geq \int_a^{+\infty} x f_X(x) dx \\ &\geq \int_a^{+\infty} a f_X(x) dx && \text{in quanto } x \geq a \\ &= a \int_a^{+\infty} f_X(x) dx \\ &= a P(X \geq a) \end{aligned}$$

ergo

$$\frac{\mathbb{E}[X]}{a} \geq P(X \geq a)$$

□

### 4.3 Disuguaglianza di Tchebyshev

La disuguaglianza di Tchebyshev dà una limitazione superiore alla probabilità che l'esito di una variabile aleatoria si discosti dal suo valore atteso di una quantità maggiore o uguale a una soglia scelta.

**Teorema 4.2** (Disuguaglianza di Tchebyshev). *Sia  $X$  una variabile aleatoria, con  $\mathbb{E}[X] = \mu$ ,  $\text{var}(X) = \sigma^2$ . Allora:*

$$\forall r > 0 \quad P(|X - \mu| \geq r) \leq \frac{\sigma^2}{r^2}$$

*Dimostrazione.*

$$|X - \mu| \geq r \Leftrightarrow (X - \mu)^2 \geq r^2$$

Passando alle probabilità:

$$\begin{aligned} P(|X - \mu| \geq r) &= P((X - \mu)^2 \geq r^2) \\ &\leq \frac{\mathbb{E}[(X - \mu)^2]}{r^2} && \text{disuguaglianza di Markov} \\ &= \frac{\text{var}(X)}{r^2} \end{aligned}$$

Ergo:

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

□

Un'interessante applicazione della disuguaglianza di Tchebyshev è quella che riguarda la deviazione standard, che esprime l'andamento della probabilità allontanandosi dal valore atteso di quantità ripetute della deviazione standard:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

## 5 Modelli

I modelli di distribuzione consistono in risultati notevoli corrispondenti a variabili aleatorie che rispettano determinate definizioni ricorrenti.

### 5.1 Modello bernoulliano

Il modello bernoulliano impone alle sue variabili aleatorie di avere specificazione binaria, cioè  $D_X = \{0, 1\}$  per qualunque variabile  $X$ .

Una variabile aleatoria bernoulliana di probabilità di successo ( $X = 1$ )  $p$  si indica con

$$X \sim B(p)$$

#### 5.1.1 Funzione di massa di probabilità

La funzione di massa di probabilità di una variabile aleatoria bernoulliana è

$$p_X(x) = P(X = x) = p^x(1-p)^{1-x}I_{\{0,1\}}(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \\ 0 & \text{altrimenti} \end{cases}$$

Modello	Parametri	F. di massa/densità	F. di ripartizione	V. atteso	Varianza
Bernoulli	$X \sim B(p)$	$p^x(1-p)^{1-x}I_{\{0,1\}}(x)$	$\begin{cases} 0 & x < 0 \\ 1-p & 0 \leq x < 1 \\ 1 & x \leq 1 \end{cases}$	$p$	$p(1-p)$
Binomiale	$X \sim B(n, p)$	$\binom{n}{i} p^i (1-p)^{n-i}$	$\begin{cases} \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i} & x \leq n \\ 1 & x > n \end{cases}$	$np$	$np(1-p)$
Uniforme discreto	$X \sim U(n)$	$\frac{1}{n} I_{\{1, \dots, n\}}(i)$	$\frac{\lfloor x \rfloor}{n} \cdot I_{\{1, \dots, n\}} + I_{\{n, \dots, +\infty\}}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
Geometrico	$X \sim G(p)$	$(1-p)^i p I_{\mathbb{N}}(i)$	$1 - (1-p)^{\lfloor x \rfloor + 1}$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Poisson	$X \sim P(\lambda)$	$e^{-\lambda} \frac{\lambda^i}{i!} I_{\mathbb{N}}(i)$	[non visto]	$\lambda$	$\lambda$
Ipergeometrico	$X \sim ?(N, M, n)$	$\frac{\binom{N}{i} \binom{M}{n-i}}{\binom{N+M}{n}}$	[non visto]	$np$	$\frac{NM}{(N+M)^2}$
Uniforme continuo	$X \sim U([a, b])$	$\frac{1}{b-a} I_{[a, b]}(x)$	$\frac{x-a}{b-a} \cdot I_{[a, b]}(x) + I_{(b, +\infty)}(x)$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Esponenziale	$X \sim E(\lambda)$	$\lambda e^{-\lambda x} I_{\mathbb{R}^+}(x)$	$1 - e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gaussiano	$X \sim N(\mu, \sigma)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy$	$\mu$	$\sigma^2$

Tabella 1: Tabella riassuntiva dei modelli di distribuzione.

### 5.1.2 Valore atteso

In analogia con la proprietà 3.4.1, il valore atteso di una variabile aleatoria bernoulliana è  $p$ :

$$\mathbb{E}[X] = p$$

E ovviamente

$$\mathbb{E}[X^2] = \mathbb{E}[X]$$

### 5.1.3 Varianza

In analogia con la proprietà 3.5.2, la varianza di una variabile aleatoria bernoulliana è  $p(1-p)$ :

$$\text{var}(X) = p(1-p)$$

### 5.1.4 Funzione di ripartizione

La funzione di ripartizione per variabili bernoulliane è ovviamente uguale a 0 per  $x < 0$ ,  $1-p$  per  $0 \leq x < 1$  e 1 per  $x \geq 1$ .

## 5.2 Modello binomiale

Il modello binomiale consiste in  $n$  ripetizioni di un esperimento bernoulliano di probabilità  $p$ . Una variabile aleatoria binomiale corrisponde al numero di successi tra gli  $n$  esperimenti.

$$X \sim B(n, p) \quad D_X = \{0, 1, \dots, n\}$$

### 5.2.1 Funzione di massa di probabilità

Per definizione:

$$p_X(i) = P(X = i)$$

Tale probabilità è l'intersezione degli eventi indipendenti che consistono nel successo dal primo all' $i$ -esimo e insuccesso dall' $i+1$ -esimo all' $n$ -esimo. In quanto probabilità di eventi indipendenti, essa può essere espressa come il prodotto delle singole probabilità. Ognuna delle probabilità di successo, per come è costruito il modello binomiale, è  $p$ , mentre ognuna delle probabilità di insuccesso è  $1-p$ . Ogni combinazione in cui  $i$  esperimenti hanno successo e  $n-i$  falliscono è valida, perciò:

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \quad (9)$$

Concorde con la (4), la somma delle immagini della funzione di massa di probabilità è 1:

$$\sum_{i=1}^n p_X(i) = \sum_{i=1}^n \binom{n}{i} p^i (1-p)^{n-i} = (p + 1-p)^n = 1$$

### 5.2.2 Valore atteso

Essendo ogni variabile aleatoria binomiale la somma di variabili aleatorie bernoulliane:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=0}^n X_i\right] = \sum_{i=0}^n \mathbb{E}[X_i] = \sum_{i=0}^n p = np \quad (10)$$



### 5.2.3 Varianza

Essendo le componenti bernoulliane indipendenti:

$$\text{var}(X) = \sum_{i=1}^n \text{var}(X_i) = \sum_{i=1}^n p(1-p) = np(1-p) \quad (11)$$

### 5.2.4 Funzione di ripartizione

Per calcolare con un'unica formula la funzione di ripartizione si aggiungono due funzioni indicatrici, che agiscono se  $x > n$ :

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= I_{(n, +\infty)}(x) + I_{[0, n]}(x) \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i} \\ &= \begin{cases} \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i} & x \leq n \\ 1 & x > n \end{cases} \end{aligned} \quad (12)$$

### 5.2.5 Relazioni tra variabili binomiali

Siano  $X_1$  e  $X_2$  due variabili aleatorie definite su modelli binomiali che differiscono solo per il numero di esperimenti:

$$\begin{aligned} X_1 &\sim B(n, p) & X_1 &= \sum_{i=1}^n X_{1,i} & X_{1,i} &\sim B(p) \quad \forall i \in \{1, \dots, n\} \\ X_2 &\sim B(m, p) & X_2 &= \sum_{j=1}^m X_{2,j} & X_{2,j} &\sim B(p) \quad \forall j \in \{1, \dots, m\} \end{aligned}$$

Se  $X_1$  e  $X_2$  sono indipendenti, allora:

$$X_1 + X_2 = \sum_{i=1}^n X_{1,i} + \sum_{j=1}^m X_{2,j} = \sum_{i=1}^{n+m} Y_i = Y$$

dove  $Y \sim B(n+m, p)$ .

## 5.3 Modello uniforme discreto

Nel modello uniforme discreto le variabili aleatorie consistono nell'esito di un esperimento da  $n$  esiti possibili equiprobabili:

$$X \sim U(n)$$

### 5.3.1 Funzione di massa di probabilità

Essendo gli  $n$  esiti equiprobabili, la funzione di massa di probabilità assume il valore  $\frac{1}{n}$  per tutti i valori dell'input compresi tra gli esiti (numerati qui da 1 a  $n$ ):

$$p_X(i) = P(X = i) = \frac{1}{n} I_{\{1, \dots, n\}}(i)$$

### 5.3.2 Funzione di ripartizione

$$\forall x \leq n \quad F_X(x) = P(X \leq x) = \sum_{i=1}^{\lfloor x \rfloor} P(X = i) = \sum_{i=1}^{\lfloor x \rfloor} \frac{1}{n} = \frac{\lfloor x \rfloor}{n}$$

Come per la (12), si aggiungono funzioni indicatrici per regolare il valore oltre  $n$ :

$$F_X(x) = \frac{\lfloor x \rfloor}{n} I_{[1,n]}(x) + I_{(n,+\infty)}(x) \quad (13)$$

### 5.3.3 Valore atteso

Banalmente, usando la definizione:

$$\mathbb{E}[X] = \sum_{i=1}^n i P(X = i) = \frac{1}{n} \sum_{i=1}^n i = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2} \quad (14)$$

### 5.3.4 Varianza

Usando la definizione equivalente di varianza e quanto appena calcolato:

$$\begin{aligned} \text{var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = && \text{proprietà 3.5.1} \\ &= \sum_{i=1}^n i^2 P(X = i) - \left(\frac{n+1}{2}\right)^2 && \text{valore atteso e (14)} \\ &= \frac{1}{n} \sum_{i=1}^n i^2 - \left(\frac{n+1}{2}\right)^2 && \text{ipotesi di equiprobabilità} \\ &= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 && \text{somma notevole} \\ &= (n+1) \left( \frac{2n+1}{6} + \frac{n+1}{4} \right) && \text{raccolgendo } n+1 \\ &= (n+1) \left( \frac{n-1}{12} \right) \\ &= \frac{n^2-1}{12} \end{aligned}$$

## 5.4 Modello geometrico

Una variabile aleatoria geometrica corrisponde al numero di insuccessi prima del primo successo in una sequenza di esperimenti bernoulliani con lo stesso parametro  $p$  e tra loro indipendenti.

Per  $p = 0$  non si ottiene mai un successo, pertanto la variabile geometrica non è definita. Per  $p = 1$  la variabile assume necessariamente il valore 0.

Il supporto di una variabile geometrica è l'insieme dei naturali.

$$X \sim G(p) \quad D_X = \mathbb{N}$$

### 5.4.1 Funzione di massa di probabilità

La funzione di massa di probabilità in  $i$  è uguale alla probabilità di insuccesso per ognuna delle  $i$  ripetizioni indipendenti per la probabilità di successo della  $i + 1$ -esima.

Come sempre si aggiunge una funzione indicatrice per aggiustare il dominio di  $F_X$ :

$$F_X(i) = P(X = i) = (1 - p)^i p I_{\mathbb{N}}(i) \quad (15)$$

La somma delle immagini della funzione di massa di probabilità converge, come dovrebbe, a 1:

$$\begin{aligned} \sum_{i=0}^{+\infty} P(X = i) &= \sum_{i=0}^{+\infty} p(1 - p)^i \\ &= p \sum_{i=0}^{+\infty} (1 - p)^i \\ &= p \frac{1}{1 - (1 - p)} \\ &= 1 \end{aligned}$$

### 5.4.2 Valore atteso

Tramite la definizione:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i=0}^{+\infty} iP(X = i) && \text{definizione 3.7} \\ &= \sum_{i=0}^{+\infty} ip(1 - p)^i && (15) \\ &= p(1 - p) \sum_{i=0}^{+\infty} i(1 - p)^{i-1} \\ &= -p(1 - p) \frac{d}{dx} \sum_{i=0}^{+\infty} (1 - p)^i && \text{derivata di } (1 - p)^i \text{ e della somma} \\ &= -p(1 - p) \frac{d}{dx} \frac{1}{p} && \text{serie geometrica di ragione } 1 - p \\ &= \frac{p(1 - p)}{p^2} = \frac{1 - p}{p} && \text{derivando} \end{aligned}$$

### 5.4.3 Varianza

Volendo usare la forma equivalente (5) di cui alla proprietà 3.5.1, si calcola innanzitutto il valore atteso del quadrato della variabile:

$$\begin{aligned}
 \mathbb{E}[X^2] &= \sum_{i=0}^{+\infty} i^2 p(1-p)^i \\
 &= p(1-p) \sum_{i=0}^{+\infty} i^2 (1-p)^{i-1} \\
 &= -p(1-p) \sum_{i=0}^{+\infty} i \frac{d}{dp} (1-p)^i && \text{derivata di } (1-p)^i \\
 &= -p(1-p) \frac{d}{dp} \sum_{i=0}^{+\infty} i(1-p)^i && \text{prodotto di una derivata per} \\
 &&& \text{una costante e derivata di} \\
 &&& \text{una somma} \\
 &= -p(1-p) \frac{d}{dp} (1-p) \sum_{i=0}^{+\infty} i(1-p)^{i-1} \\
 &= p(1-p) \frac{d}{dp} (1-p) \frac{d}{dp} \sum_{i=0}^{+\infty} (1-p)^i \\
 &= -p(1-p) \frac{d}{dp} \frac{1-p}{p^2} \\
 &= -p(1-p) \frac{-p^2 - 2p(1-p)}{p^4} \\
 &= (1-p) \frac{p + 2(1-p)}{p^2} \\
 &= \frac{(1-p)(2-p)}{p^2}
 \end{aligned}$$

Da cui:

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

#### 5.4.4 Funzione di ripartizione

$$\begin{aligned}
 F_X(n) &= P(X \leq n) = 1 - P(X > n) = \\
 &= 1 - \sum_{i=n+1}^{+\infty} P(X = i) \\
 &= 1 - \sum_{i=n+1}^{+\infty} p(1-p)^i \\
 &= 1 - p(1-p)^{n+1} \sum_{i=n+1}^{+\infty} (1-p)^{i-(n+1)} \quad \text{moltiplicando per } \frac{(1-p)^{n+1}}{(1-p)^{n+1}} \\
 &= 1 - p(1-p)^{n+1} \sum_{i=0}^{+\infty} (1-p)^i \quad \text{sostituzione: } i = i - (n+1) \\
 &= 1 - p(1-p)^{n+1} \frac{1}{1 - (1-p)} \\
 &= 1 - (1-p)^{n+1}
 \end{aligned}$$

Questo risultato è in realtà banale se si applica il concetto semantico alla variabile geometrica.

$$F_X(x) = P(X \leq x) = 1 - P(X > x) = 1 - (1-p)^{\lfloor x \rfloor + 1}$$

#### 5.4.5 Assenza di memoria

Come si può intuire, la probabilità di costante insuccesso all' $i + j$ -esimo esperimento non è condizionata dalla probabilità di costante insuccesso all' $i$ -esimo. Questo risultato prende il nome di assenza di memoria.

$$\begin{aligned}
 P(X \geq i+j \mid X \geq i) &= \frac{P(X \geq i+j \cap X \geq i)}{P(X \geq i)} \\
 &= \frac{P(X \geq i+j)}{P(X \geq i)} \\
 &= \frac{(1-p)^{i+j}}{(1-p)^i} \\
 &= (1-p)^j \\
 &= P(X \geq j)
 \end{aligned}$$

### 5.5 Modello di Poisson

$$X \sim P(\lambda) \quad D_X = \mathbb{N} \quad \lambda > 0$$

#### 5.5.1 Funzione di massa di probabilità

$$P_X(i) = P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!} I_{\mathbb{N}}(i)$$

Ancora una volta la somma delle immagini della funzione di massa di probabilità converge a 1:

$$\sum_{i=0}^{+\infty} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{i=0}^{+\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1 \quad (16)$$

### 5.5.2 Valore atteso

Tramite la definizione:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i=0}^{+\infty} iP(X=i) = \sum_{i=1}^{+\infty} iP(X=i) && \text{definizione 3.7} \\ &= \sum_{i=1}^{+\infty} ie^{-\lambda} \frac{\lambda^i}{i!} && (17) \end{aligned}$$

$$\begin{aligned} &= \lambda e^{-\lambda} \sum_{i=1}^{+\infty} \frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda e^{-\lambda} \sum_{i=0}^{+\infty} \frac{\lambda^i}{i!} && \text{sostituzione: } i = i - 1 \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda && (18) \end{aligned}$$

### 5.5.3 Varianza

Volendo usare la forma equivalente (5), si calcola innanzitutto il valore atteso di  $X^2$ :

$$\begin{aligned}
 \mathbb{E}[X^2] &= \sum_{i=1}^{+\infty} i^2 e^{-\lambda} \frac{\lambda^i}{i!} \\
 &= \sum_{i=1}^{+\infty} i e^{-\lambda} \frac{\lambda^i}{(i-1)!} \\
 &= \lambda \sum_{i=1}^{+\infty} i e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!} \\
 &= \lambda \sum_{i=1}^{+\infty} (i-1+1) e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!} \\
 &= \lambda \sum_{i=1}^{+\infty} \left( (i-1) e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!} + e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!} \right) \\
 &= \lambda \sum_{i=1}^{+\infty} (i-1) e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!} + \lambda \sum_{i=1}^{+\infty} e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!} \\
 &= \lambda \underbrace{\sum_{i=0}^{+\infty} i e^{-\lambda} \frac{\lambda^i}{i!}}_{\lambda} + \lambda \underbrace{\sum_{i=0}^{+\infty} e^{-\lambda} \frac{\lambda^i}{i!}}_1 \quad \text{con } i = i-1 \\
 &= \lambda^2 + \lambda \quad (17) \text{ e } (16)
 \end{aligned}$$

Ergo

$$\begin{aligned}
 \text{var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
 &= \lambda^2 + \lambda - \lambda^2 \\
 &= \lambda \quad (19)
 \end{aligned}$$

### 5.5.4 Approssimazione del modello binomiale

Il modello di Poisson è strettamente legato al modello binomiale. Infatti, se il prodotto dei parametri di una binomiale è costante, per  $n$  grandi essa è ben approssimata da una variabile di Poisson che ha come parametro tale prodotto:

$$X \sim B(n, p) \quad \text{con } np = \lambda$$

Per  $n \rightarrow +\infty$ :

$$\begin{aligned}
 P(X = i) &= \binom{n}{i} p^i (1-p)^{n-i} \\
 &= \binom{n}{i} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\
 &= \frac{n(n-1)\dots(n-i+1)}{i!} \cdot \frac{\lambda^i}{n^i} \left(1 - \frac{\lambda}{n}\right)^{n-i} \\
 &= \frac{n(n-1)\dots(n-i+1)}{n^i} \cdot \frac{\lambda^i}{i!} \left(1 - \frac{\lambda}{n}\right)^{n-i} \\
 &= \underbrace{\frac{n}{n}}_{\rightarrow 1} \cdot \underbrace{\frac{n-1}{n}}_{\rightarrow 1} \dots \underbrace{\frac{n-i+1}{n}}_{\rightarrow 1} \cdot \frac{\lambda^i}{i!} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow e^{-\lambda}} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-i}}_{\rightarrow 1} \\
 &\rightarrow \frac{\lambda^i}{i!} e^{-\lambda}
 \end{aligned}$$

## 5.6 Modello ipergeometrico

Il modello ipergeometrico descrive il classico problema dell'urna. Dati  $N$  oggetti funzionanti e  $M$  oggetti difettosi, sia  $n$  il numero di estrazioni senza reimmissione. La variabile aleatoria ipergeometrica  $X$  è il numero di oggetti funzionanti nelle  $n$  estrazioni.

$$X \sim ?(?)$$

Il modello è valido solo se  $P(X = 0) = 0$ .

### 5.6.1 Funzione di massa di probabilità

Il numero di casi possibili sono le combinazioni di  $n$  estrazioni senza reimmissione da un gruppo di  $N + M$ . Il numero di casi favorevoli si può calcolare applicando il principio fondamentale del calcolo combinatorio.

$$P(X = i) = \frac{\binom{N}{i} \binom{M}{n-i}}{\binom{N+M}{n}}$$

### 5.6.2 Valore atteso

Al fine di calcolare il valore atteso si sfrutta un approccio decomposizionale: si introducono  $n$  variabili aleatorie  $X_i$ , ognuna legata a un'estrazione, tali che

$$X_i = \begin{cases} 1 & \text{l}'i\text{-esimo oggetto estratto funziona} \\ 0 & \text{altrimenti} \end{cases}$$

Per tali variabili vale

$$P(X_i = 1) = \frac{N}{N+M} =: p = \mathbb{E}[X_i]$$



Da cui

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[\sum_{i=1}^n X_i\right] \\ &= \sum_{i=1}^n \mathbb{E}[X_i] \\ &= np\end{aligned}$$

### 5.6.3 Varianza

Applicando la proprietà 3.5.1 si calcola la varianza delle singole  $X_i$ :

$$\begin{aligned}\text{var}(X_i) &= \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 \\ &= \mathbb{E}[X_i](1 - \mathbb{E}[X_i]) \quad \text{idempotenza di } X_i \\ &= \frac{N}{N+M} + \frac{M}{N+M} \\ &= \frac{NM}{(N+M)^2}\end{aligned}$$

Essendo le variabili  $X_i$  non indipendenti, la varianza della loro somma non è uguale alla somma delle varianze. Si può comunque applicare la proprietà 3.10.7 e passare per le covarianze:

$$\begin{aligned}\text{cov}(X_i, X_j) &= \mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \mathbb{E}[X_j] \\ &= \mathbb{E}[X_i = 1 \cap X_j = 1] - \left(\frac{N}{N+M}\right)^2 \\ &= P(X_j = 1 \mid X_i = 1)P(X_i = 1) - \left(\frac{N}{N+M}\right)^2 \\ &= \frac{N-1}{N+M-1} \frac{N}{N+M} - \left(\frac{N}{N+M}\right)^2 \\ &= \frac{N}{N+M} \left(\frac{N-1}{N+M-1} - \frac{N}{N+M}\right) \\ &= \frac{-NM}{(N+M-1)(N+M)^2}\end{aligned}$$

Ergo

$$\begin{aligned}\text{var}(X) &= \sum_{i=1}^n \text{var}(X_i) + \sum_{i \neq j}^n \text{cov}(X_i, X_j) \\ &= n \frac{NM}{(N+M)^2} - n(n-1) \cdot \frac{-NM}{(N+M-1)(N+M)^2} \\ &= n \frac{NM}{(N+M)^2} \left(1 - (n-1) \frac{1}{N+M-1}\right) \\ &= np(1-p) \left(1 - \frac{n-1}{N+M-1}\right)\end{aligned}$$

Per  $N+M \rightarrow +\infty$  il modello si semplifica in un modello binomiale:

$$\rightarrow np(1-p)$$

## 5.7 Modello uniforme continuo

Il modello uniforme continuo estende al continuo i concetti visti alla sezione 5.3 per il modello uniforme discreto e viene determinato da un intervallo equivalentemente aperto o chiuso:

$$X \sim U([a, b])$$

### 5.7.1 Funzione di densità di probabilità

$$f(X)(x) = \frac{1}{b-a} I_{[a, b]}(x)$$

essendo la densità costante, per  $I \subseteq [a, b]$ :

$$P(X \in I) = \frac{|I|}{b-a}$$

dove  $|I|$  è la somma delle ampiezze  $p - q$  degli intervalli disgiunti  $[p, q]$  da cui  $I$  è composto.

Come da definizione, integrando nell'intero  $\mathbb{R}$  la funzione di densità di probabilità si ottiene 1:

$$\int_{-\infty}^{+\infty} f_X(x) dx = \int_a^b \frac{1}{b-a} dx = \frac{1}{b-a} x \Big|_a^b = 1$$

### 5.7.2 Funzione di ripartizione

Per definizione la funzione di ripartizione è la funzione integrale della funzione di densità:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \int_{-\infty}^x f_X(u) du \\ &= \int_a^x \frac{1}{b-a} du \\ &= \frac{1}{b-a} u \Big|_a^x \\ &= \frac{x-a}{b-a} \end{aligned}$$

Come sempre funzioni indicatrici aggiustano il risultato per punti non appartenenti all'intervallo:

$$F_X(x) = \frac{x-a}{b-a} I_{[a, b]}(x) + I_{(b, +\infty)}(x)$$

### 5.7.3 Valore atteso

Applicando la definizione:

$$\begin{aligned}
 \mathbb{E}[X] &= \int_a^b x f_X(x) dx \\
 &= \frac{1}{b-a} \int_a^b x dx \\
 &= \frac{1}{b-a} \left. \frac{x^2}{2} \right|_a^b \\
 &= \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} \\
 &= \frac{b+a}{2}
 \end{aligned}$$

### 5.7.4 Varianza

Come di consueto si intende applicare la proprietà 3.5.1, pertanto si calcola innanzitutto il valore atteso di  $X^2$ :

$$\begin{aligned}
 \mathbb{E}[X^2] &= \int_a^b x^2 f_X(x) dx \\
 &= \frac{1}{b-a} \int_a^b x^2 dx \\
 &= \frac{1}{b-a} \left. \frac{x^3}{3} \right|_a^b \\
 &= \frac{b^3 - a^3}{3(b-a)} \\
 &= \frac{a^2 + ab + b^2}{3}
 \end{aligned}$$

E infine:

$$\begin{aligned}
 \text{var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
 &= \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} \\
 &= \frac{(b-a)^2}{12}
 \end{aligned}$$

## 5.8 Modello esponenziale

Nel modello esponenziale la funzione di densità è esponenziale.

$$X \sim E(\lambda) \quad \lambda \in \mathbb{R}^+ \quad D_X = \mathbb{R}^+$$

### 5.8.1 Funzione di densità di probabilità

$$f_X(x) = \lambda e^{-\lambda x} I_{\mathbb{R}^+}(x)$$

Il modello esponenziale si usa per modellare il tempo che intercorre tra due eventi.

La funzione di densità rispetta la definizione, infatti:

$$\begin{aligned}
 \int_0^{+\infty} f_X(x) dx &= \int_0^{+\infty} \lambda e^{-\lambda x} dx \\
 &= \int_0^{+\infty} e^{-y} dy && \text{con } y = \lambda x \\
 &= -e^{-y} \Big|_0^{+\infty} \\
 &= 0 + e^{-0} \\
 &= 1
 \end{aligned}$$

### 5.8.2 Funzione di ripartizione

Applicando la definizione di funzione di ripartizione continua:

$$\begin{aligned}
 F_X(x) &= \int_0^x f_X(y) dy \\
 &= \int_0^x \lambda e^{\lambda y} dy \\
 &= \int_0^{\lambda x} e^{-z} dz && \text{con } z = \lambda y \\
 &= -e^{-z} \Big|_0^{\lambda x} \\
 &= -e^{-\lambda x} + e^0 \\
 &= 1 - e^{-\lambda x}
 \end{aligned}$$

Aggiungendo una funzione indicatrice:

$$F_X(x) = (1 - e^{-\lambda x}) I_{\mathbb{R}^+}(x) \quad (20)$$

### 5.8.3 Valore atteso

Applicando la definizione:

$$\begin{aligned}
 \mathbb{E}[X] &= \int_0^{+\infty} x f_X(x) dx \\
 &= \int_0^{+\infty} x \lambda e^{-\lambda x} dx \\
 &= -x e^{-\lambda x} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx && \text{per parti} \\
 &= \int_0^{+\infty} e^{-\lambda x} dx \\
 &= \frac{1}{\lambda} \underbrace{\int_0^{+\infty} \lambda e^{-\lambda x} dx}_1 \\
 &= \frac{1}{\lambda} && \text{proprietà di } f_X
 \end{aligned}$$

#### 5.8.4 Varianza

Volendo usare la forma equivalente (5), si calcola innanzitutto il valore atteso di  $X^2$ :

$$\begin{aligned}\mathbb{E}[X^2] &= \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx \\ &= -x^2 e^{-\lambda x} \Big|_0^{+\infty} + \int_0^{+\infty} 2x e^{-\lambda x} dx \\ &= 2 \int_0^{+\infty} x e^{-\lambda x} dx \\ &= \frac{2}{\lambda} \int_0^{+\infty} \lambda x e^{-\lambda x} dx \\ &= \frac{2}{\lambda} \mathbb{E}[X] = \frac{2}{\lambda^2}\end{aligned}$$

Da cui:

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

#### 5.8.5 Assenza di memoria

Le variabili di modello esponenziale godono della proprietà di assenza di memoria (già vista per il modello geometrico al paragrafo 5.4.5):

$$P(X > x) = 1 - F_X(x) = e^{-\lambda x}$$

Quindi:

$$\begin{aligned}P(X > s+t) &= e^{-\lambda(s+t)} \\ &= e^{-\lambda s} e^{-\lambda t} \\ &= P(X > s) P(X > t)\end{aligned}$$

Da cui

$$\begin{aligned}P(X > s) &= \frac{P(X > s+t)}{P(X > t)} \\ &= \frac{P(X > s+t \cap X > t)}{P(X > t)} \\ &= P(X > s+t \mid X > t)\end{aligned}$$

### 5.9 Risultati notevoli sui modelli

**Proprietà 5.9.1.** Siano  $X_1, \dots, X_n$  variabili aleatorie indipendenti e sia  $Y$  il massimo degli  $X_i$ , ossia  $Y := \max_i X_i$ . Allora:

$$F_Y(x) = \prod_{i=1}^n F_{X_i}(x)$$

*E per variabili indipendenti e identicamente distribuite (i.i.d.) secondo una funzione di ripartizione  $F$ :*

$$F_Y(x) = \prod_{i=1}^n F(x) = F(x)^n$$

*Dimostrazione.*

$$F_Y(x) = P(Y \leq x) = P(\max_i X_i \leq x) = P(\forall i X_i \leq x)$$

Dal momento che gli  $X_i$  sono indipendenti, l'ultima probabilità è uguale al prodotto delle singole:

$$= \prod_{i=1}^n P(X_i \leq x) = \prod_{i=1}^n F_{X_i}(x)$$

Nell'ulteriore ipotesi di variabili indipendenti e identicamente distribuite (i.i.d.) secondo una funzione di ripartizione  $F$ :

$$= \prod_{i=1}^n F(x) = F(x)^n$$

□

**Proprietà 5.9.2.** *Siano  $X_1, \dots, X_n$  variabili aleatorie indipendenti e sia  $Z$  il minimo degli  $X_i$ , ossia  $Z := \min_i X_i$ . Allora:*

$$F_Z(x) = 1 - \prod_{i=1}^n (1 - F_{X_i}(x))$$

*E per variabili indipendenti e identicamente distribuite (i.i.d.) secondo una funzione di ripartizione  $F$ :*

$$F_Z(x) = 1 - (1 - F(x))^n$$

*Dimostrazione.*

$$F_Z(x) = 1 - P(Z > x) = 1 - P(\min X_i > x) = 1 - P(\forall i X_i > x)$$

Dal momento che gli  $X_i$  sono indipendenti, l'ultima probabilità è uguale al prodotto delle singole:

$$= 1 - \prod_{i=1}^n P(X_i > x) = 1 - \prod_{i=1}^n (1 - F_{X_i}(x))$$

Nell'ulteriore ipotesi di variabili indipendenti e identicamente distribuite (i.i.d.) secondo una funzione di ripartizione  $F$ :

$$= 1 - \prod_{i=1}^n (1 - F(x)) = 1 - (1 - F(x))^n$$

□

**Proprietà 5.9.3.** Siano  $X_1, \dots, X_n$  variabili aleatorie indipendenti e sia  $Z$  il minimo degli  $X_i$ , ossia  $Z := \min_i X_i$ . Se  $X_i \sim E(\lambda_i)$  per ogni  $i$ , allora:

$$Z \sim E\left(\sum_{i=1}^n \lambda_i\right)$$

*Dimostrazione.* Essendo le variabili esponenziali le loro funzioni di ripartizione sono del tipo:

$$F_{X_i}(x) = 1 - e^{-\lambda_i x}$$

Per la proprietà 5.9.2:

$$\begin{aligned} F_Z(x) &= 1 - \prod_{i=1}^n (1 - F_{X_i}(x)) \\ &= 1 - \prod_{i=1}^n e^{-\lambda_i x} \\ &= 1 - e^{\sum_{i=1}^n -\lambda_i x} \\ &= 1 - e^{-x \sum_{i=1}^n \lambda_i} \end{aligned}$$

Chiamato  $\lambda = \sum_{i=1}^n \lambda_i$ , allora  $Z \sim E(\lambda)$ :

$$F_Z(x) = 1 - e^{-\lambda x}$$

□

**Proprietà 5.9.4.** Se  $X \sim E(\lambda)$  e  $Y := cX$  con  $c \in \mathbb{R}^+$ , allora  $Y$  è una variabile aleatoria esponenziale di parametro  $\frac{\lambda}{c}$ .

$$F_Y(x) = 1 - e^{-\frac{\lambda}{c} x}$$

*Dimostrazione.*

$$\begin{aligned} F_Y(x) &= \\ &= P(Y \leq x) \\ &= P(cX \leq x) \\ &= P\left(X \leq \frac{x}{c}\right) \\ &= F_X\left(\frac{x}{c}\right) \\ &= 1 - e^{-\frac{\lambda}{c} x} \end{aligned}$$

□

## 5.10 Modello gaussiano

Una variabile  $X$  di modello gaussiano (o normale), è una variabile aleatoria continua definita da due parametri:

$$X \sim N(\mu, \sigma) \quad \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$$

### 5.10.1 Funzione di densità di probabilità

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

Studiando la funzione di densità si verifica algebricamente la famosa forma “a campana”:

- $\lim_{x \rightarrow \pm\infty} f_X(x) = 0$
- $f'_x(x) = \frac{1}{\sqrt{2\pi} \sigma^3} e^{-\frac{(x-\mu)^2}{2\sigma^2}} (\mu - x)$
- $f'_x(x) \geq 0 \Leftrightarrow x \leq \mu$
- $f''_x(x) = \left(\frac{x-\mu}{\sigma}\right)^2 - 1$
- $f''_x(x) \geq 0 \Leftrightarrow x \geq \mu + \sigma \vee x \leq \mu - \sigma$

Modificare  $\mu$  significa ovviamente traslare la curva parallelamente all'asse delle  $x$ . Aumentare il valore di  $\sigma$  significa diminuire l'ordinata del massimo e, conseguentemente, “allargare la campana” (in quanto l'area totale sottesa deve rimanere invariata). Vale ovviamente il viceversa per una diminuzione.

Si può dimostrare che l'area sottesa alla funzione di densità è 1:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$

### 5.10.2 Funzione di ripartizione

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy$$

### 5.10.3 Valore atteso

$$\mathbb{E}[X] = \mu$$

### 5.10.4 Varianza

$$\text{var}(X) = \sigma^2$$

### 5.10.5 Distribuzione normale standard

A partire da una variabile gaussiana  $X$  si può costruire la variabile  $Z$  come standardizzazione (normalizzazione) di  $X$ :

$$Z = \frac{x - \mu}{\sigma}$$



Si verificano i seguenti risultati

$$\begin{aligned}\mathbb{E}[Z] &= \mathbb{E}\left[\frac{1}{\sigma}\mathbb{E}[X - \mu]\right] \\ &= \frac{1}{\sigma}(\mathbb{E}[X] - \mu) = 0\end{aligned}$$

$$\begin{aligned}\text{var}(Z) &= \frac{1}{\sigma^2} \text{var}(X - \mu) \\ &= \frac{1}{\sigma^2} \text{var}(X) = 1\end{aligned}$$

Ergo

$$X \sim N(\mu, \sigma) \Rightarrow Z \sim N(0, 1)$$

Le variabili normali standard si indicano solitamente con  $Z$ , mentre le relative funzioni di densità e di ripartizione di indicano rispettivamente con  $\phi(z)$  e  $\Phi(z)$ .

### 5.10.6 Risultati notevoli

**Trasformazioni lineari** Trasformando linearmente la variabile aleatoria  $X \sim N(\mu, \sigma)$ , si ottiene una variabile aleatoria gaussiana  $Y$ :

$$X \sim N(\mu, \sigma) \Rightarrow Y \sim N(a\mu + b, a\sigma)$$

**Riproducibilità** Date variabili  $X_1, \dots, X_n$  gaussiane indipendenti tali che  $\forall i X_i \sim N(\mu_i, \sigma_i)$ :

$$Y \sim N\left(\sum_{i=1}^n \mu_i, \sqrt{\sum_{i=1}^n \sigma_i^2}\right)$$

Anche il modello binomiale e l'ipergeometrico, ad esempio, godono della proprietà di riproducibilità.

**Funzione di ripartizione normale e standard** è possibile ricavare la funzione di ripartizione di una variabile aleatoria gaussiana qualsiasi conoscendo la funzione di ripartizione di una variabile standard:

$$\begin{aligned}F_X(x) &= P(X \leq x) \\ &= P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{x - \mu}{\sigma}\right) \\ &= F_Z\left(\frac{x - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{x - \mu}{\sigma}\right)\end{aligned}$$

### 5.11 Risultati notevoli sui modelli

#### 5.12 Indici di variabili aleatorie

**Definizione 5.1.** Data una variabile aleatoria  $X$ , la mediana di  $X$  è un numero  $m \in \mathbb{R}$  tale che  $P(X \leq m) = P(X > m) = 1/2$ .

**Definizione 5.2.** Data una variabile aleatoria  $X$ , la moda di  $X$  è la specificazione di densità (o massa di probabilità) massima.

**Definizione 5.3.** Data una variabile aleatoria  $X$ , il quantile di livello  $q \in [0, 1]$  di  $X$  è la specificazione  $x_q \in \mathbb{R}$  tale che  $P(X \leq x_q) = q$ .

#### 5.13 Teorema centrale del limite

**Teorema 5.1.** Siano  $X_1, \dots, X_n$  variabili aleatorie indipendenti e identicamente distribuite, ossia tali che  $\forall i \quad \mathbb{E}[X_i] = \mu \wedge \text{var}(X_i) = \sigma^2$ . Allora per  $n$  grandi le variabili sono distribuite in modo approssimativamente<sup>3</sup> normale:

$$\sum_{i=1}^n X_i \sim N(n\mu, \sqrt{n}\sigma)$$

O, standardizzando

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \sim N(0, 1)$$

Ovverosia:

$$\lim_{n \rightarrow +\infty} P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x\right) = \Phi(x)$$

##### 5.13.1 Funzione cumulativa empirica

La funzione cumulativa empirica dà una misura del numero di osservazioni che superano un dato input:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(x_i)$$

Fatta una selezione di osservazioni sul campione, a patto che tale selezione sia coerente con la funzione di densità/massa, la funzione cumulativa empirica è un'approssimazione tanto più buona della funzione di ripartizione della selezione quanto grande è la selezione sul campione.

Per il teorema centrale del limite, variabili aleatorie bernoulliane di parametri alti possono essere approssimate con il modello normale:

$$\begin{aligned} X &\sim B(n, p) \\ X = \sum_{i=1}^n X_i &\sim N(np, \sqrt{np(1-p)}) \\ \frac{x - np}{\sqrt{np(1-p)}} &\sim N(0, 1) \end{aligned}$$

<sup>3</sup>Il simbolo  $\sim$  indica l'appartenenza approssimativa a un modello.

## 6 Statistica inferenziale

La statistica inferenziale tenta di applicare un'inferenza attraverso l'induzione. La statistica inferenziale riprende concetti della statistica descrittiva ma li reinterpreta in senso probabilistico applicando l'induzione.

### 6.1 Definizioni

Gli attori fondamentali della statistica inferenziale sono la popolazione, il campione la statistica o stimatore<sup>4</sup>, e la stima.

- La popolazione è descritta come una variabile aleatoria  $X$ . La branca della statistica inferenziale in cui la distribuzione di  $X$  è nota a meno di uno o più parametri  $\theta$  si dice parametrica:  $X \sim F(\theta)$ . Se anche il modello della distribuzione è ignoto, si tratta di statistica inferenziale non parametrica. Nel resto di questo testo si studierà la statistica inferenziale parametrica;
- Poiché non sempre si vuole ricavare o approssimare il parametro di distribuzione ignoto  $\theta$ , ma talvolta un valore che ne deriva, si introduce la funzione  $\tau(\theta)$  a indicare tale valore;
- L'informazione conosciuta sulla popolazione è data da un campione, talvolta espresso come una serie di variabili aleatorie  $X_1, \dots, X_n$  indipendenti e identicamente distribuite, talvolta come una tupla di loro specificazioni  $x_1, \dots, x_n$ ;
- L'operazione di stimare  $\tau(\theta)$  consiste nella statistica o stimatore, che è una funzione  $t$  che associa a una tupla possibili specificazioni del campione a un numero reale che stima il valore cercato:  $t : D_X^n \rightarrow \mathbb{R}$ . Essendo la statistica una variabile aleatoria, la si rappresenta talvolta (quando si vuole evidenziare tale lettura) con  $T$ ;
- Una stima  $\hat{\tau}$  è un'immagine della statistica e approssima  $\tau(\theta)$ :  $\hat{\tau} = t(x_1, \dots, x_n)$  con istanze  $X_1 = x_1, \dots, X_n = x_n$ .

#### 6.1.1 Stimatori non devianti

**Definizione 6.1** (stimatore non deviato). Uno stimatore  $t$  è non deviato per una quantità  $\tau(\theta)$  quando il suo valore atteso è uguale a  $\tau(\theta)$ :

$$\mathbb{E}[t(X_1, \dots, X_n)] = \tau(\theta)$$

*Esempio 6.1.1.* Uno stimatore non deviato per stimare il valore atteso, indipendentemente dalla distribuzione, è quello che fa corrispondere alle specificazioni date la loro media:

$$t(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i \quad \tau(\theta) = \mathbb{E}[X]$$

<sup>4</sup>in alcuni contesti, i termini statistica e stimatore differiscono leggermente di significato. Tuttavia in questo testo verranno usati equivalentemente.

Infatti:

$$\begin{aligned}\mathbb{E}[t] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \\ &= \frac{1}{n} \cdot n \mathbb{E}[X_i] \\ &= \mathbb{E}[X]\end{aligned}$$

Inoltre, calcolando la varianza dello stimatore ci si accorge che essa tende a 0 per campioni molto grandi:

$$\text{var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{n}{n^2} \text{var}(X) = \frac{\text{var}(X)}{n}$$

## 6.2 Errore quadratico medio

Il errore quadratico medio (Mean Square Error o MSE) è un modo di valutare uno stimatore di una quantità ignota:

**Definizione 6.2** (errore quadratico medio).

$$\text{MSE}_{\tau(\theta)}(T) = \mathbb{E}[(T - \tau(\theta))^2]$$

con  $T = t(X_1, \dots, X_n)$ .

**Definizione 6.3.** Il bias di uno stimatore  $T$  su  $\tau(\theta)$  è la differenza tra il suo valore atteso e  $\tau(\theta)$ :

$$b_{T(\theta)}(T) = \mathbb{E}[T] - \tau(\theta)$$

In virtù delle definizioni di MSE e bias si può trarre un interessante risultato:

$$\begin{aligned}\text{MSE}_{\tau(\theta)}(T) &= \mathbb{E}[(T - \tau(\theta))^2] \\ &= \mathbb{E}[(T - \mathbb{E}[T] + \mathbb{E}[T] - \tau(\theta))^2] \\ &= \mathbb{E}[(T - \mathbb{E}[T])^2 + 2(T - \mathbb{E}[T])(\mathbb{E}[T] - \tau(\theta)) + (\mathbb{E}[T] - \tau(\theta))^2] \\ &= \mathbb{E}[(T - \mathbb{E}[T])^2] + 2(\mathbb{E}[T] - \tau(\theta)) \underbrace{\mathbb{E}[T - \mathbb{E}[T]]}_0 + (\mathbb{E}[T] - \tau(\theta))^2 \\ &= \text{var}(T) + (\mathbb{E}[T] - \tau(\theta))^2 \\ &= \text{var}(T) + (b_{T(\theta)}(T))^2\end{aligned}$$

Per definizione, se  $T$  è uno stimatore non deviato per  $\tau(\theta)$  allora  $b_{\tau(\theta)}(T) = 0$  e quindi l'errore quadratico medio di  $T$  è uguale alla sua varianza.

## 6.3 Stimatori consistenti

**Definizione 6.4** (Stimatore consistente in media quadratica). Uno stimatore  $T_n$ <sup>5</sup> è consistente in media quadratica per  $\tau(\theta)$  se

$$\lim_{n \rightarrow +\infty} \text{MSE}_{\tau(\theta)}(T_n) = 0$$

<sup>5</sup>Più precisamente, si parla di una famiglia di stimatori  $T_n$  i cui membri differiscono per dimensione del campione

**Definizione 6.5** (Stimatore debolmente consistente). Uno stimatore  $T_n = t(X_1, \dots, X_n)$  è debolmente consistente rispetto a una quantità ignota  $\tau(\theta)$  se e solo se

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow +\infty} P(\tau(\theta) - \varepsilon \leq T_n \leq \tau(\theta) + \varepsilon) = 1$$

**Teorema 6.1.** *Se uno stimatore  $T$  è consistente in media quadratica per  $\tau(\theta)$  allora  $T$  è debolmente consistente per  $\tau(\theta)$ .*

*Dimostrazione.*

$$\begin{aligned} P(-\varepsilon \leq T_n - \tau(\theta) \leq \varepsilon) &= P(|T_n - \tau(\theta)| \leq \varepsilon) \\ &= P((T_n - \tau(\theta))^2 \leq \varepsilon^2) \\ &= 1 - P((T_n - \tau(\theta))^2 > \varepsilon^2) \\ &> 1 - \frac{\mathbb{E}[(T_n - \tau(\theta))^2]}{\varepsilon^2} \quad \text{disuguaglianza 4.1 di Markov} \\ &= 1 - \frac{\text{MSE}_{\tau(\theta)}(T_n)}{\varepsilon^2} \rightarrow 1 \end{aligned}$$

Essendo le probabilità limitate superiormente da 1, per il teorema del confronto la probabilità iniziale tende a 1.  $\square$

## 6.4 Legge dei grandi numeri

**Teorema 6.2** (legge dei grandi numeri forte).

$$\lim_{n \rightarrow +\infty} P(\bar{X}_n = \mu) = 1$$

**Teorema 6.3** (legge dei grandi numeri debole).

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow +\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$$

## 6.5 Problema dello scarto

Un problema tipo che fa uso del teorema centrale del limite nella statistica inferenziale è il seguente.

Sia  $\bar{X}_n$  la media campionaria dipendente da  $n$  elementi e sia  $\mu$  il suo valore atteso. Si vuole che lo scarto  $|\bar{X}_n - \mu|$  tra i due valori sia minore di una soglia  $r$  con probabilità superiore a  $1 - \delta$  (con  $\delta$  piccolo a piacere):

$$P(|\bar{X}_n - \mu| \leq r) \geq 1 - \delta$$

Normalizzando:

$$\begin{aligned} P\left(\frac{|\bar{X}_n - \mu|}{\sigma/\sqrt{n}} \leq \frac{r}{\sigma/\sqrt{n}}\right) &= P\left(\left|\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right| \leq \frac{r\sqrt{n}}{\sigma}\right) \\ &\approx P\left(|Z| \leq \frac{r}{\sigma}\sqrt{n}\right) \\ &= P\left(-\frac{r}{\sigma}\sqrt{n} \leq Z \leq \frac{r}{\sigma}\sqrt{n}\right) \\ &= \Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - \Phi\left(-\frac{r}{\sigma}\sqrt{n}\right) \\ &= \Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - \left(1 - \Phi\left(\frac{r}{\sigma}\sqrt{n}\right)\right) \end{aligned}$$

Ergo:

$$\begin{aligned} 2\Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - 1 &\geq 1 - \delta \\ \Phi\left(\frac{r}{\sigma}\sqrt{n}\right) &\geq 1 - \frac{\delta}{2} \\ \frac{r}{\sigma}\sqrt{n} &\geq \Phi^{-1}\left(1 - \frac{\delta}{2}\right) \\ n &\geq \left(\frac{\sigma}{r}\Phi^{-1}\left(1 - \frac{\delta}{2}\right)\right)^2 \end{aligned}$$

Allo stesso modo si possono ricavare gli altri parametri del problema:

$$\begin{aligned} r &= \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\delta}{2}\right) \\ \delta &\geq 2\left(1 - \Phi\left(\frac{r}{\sigma}\sqrt{n}\right)\right) \end{aligned}$$

Mentre  $\sigma$  tipicamente è stimato come  $\sqrt{s^2}$  (dove  $s$  è la deviazione standard campionaria).

Un'alternativa a questo tipo di stima si può fare applicando la disuguaglianza di Tchebishev:

$$\begin{aligned} P(|X - \mu| \geq r) &\leq \frac{\sigma^2}{r^2} \\ P(|X - \mu| < r) &= 1 - P(|X - \mu| \geq r) \geq 1 - \frac{\sigma^2}{r^2} \\ P(|\bar{X} - \mu| < \varepsilon) &\geq 1 - \frac{\sigma^2}{n\varepsilon^2} \geq 1 - \delta \\ \frac{\sigma^2}{n\varepsilon^2} &\leq \delta \end{aligned}$$

$$\begin{aligned} n &\geq \frac{\sigma^2}{\delta\varepsilon^2} \Rightarrow P(|\bar{X} - \mu| < \varepsilon) \geq 1 - \delta \\ \delta &\geq \frac{\sigma^2}{n\varepsilon^2} \\ \varepsilon &\geq \sqrt{\frac{\sigma^2}{n\delta}} \end{aligned}$$

## 6.6 Processo di Poisson

Il processo di Poisson dimostra che sotto certe ipotesi il problema del numero di eventi che avvengono in un certo intervallo di tempo è descritto da un modello di Poisson.

**Teorema 6.4.** *Sia  $t$  una variabile temporale e  $N(t)$  la variabile aleatoria che esprime il numero di eventi che avvengono nell'intervallo di tempo  $[0, t)$ . Verificate le seguenti ipotesi:*

$$1. N(0) = 0;$$

2. istanze di  $N$  sono indipendenti per intervalli disgiunti;

$$3. \lim_{h \rightarrow 0} \frac{P(N(h) = 1)}{h} = \lambda;$$

$$4. \lim_{h \rightarrow 0} \frac{P(N(h) \geq 2)}{h} = 0.$$

allora  $N(t)$  è distribuita secondo il modello di Poisson con parametro  $\lambda t$ :

$$N(t) \sim P(\lambda t)$$

*Dimostrazione.* Volendo calcolare la massa di probabilità  $P(N(t) = k)$ , si sceglie  $n \in \mathbb{N}$  e si divide l'intervallo  $[0, t]$  in  $n$  parti uguali  $[\frac{m}{n}t, \frac{m+1}{n}t)$ . Potendo scegliere  $n$  grande a piacere si può scegliere in particolare  $n > k$ .

L'evento  $N(t) = k$ , di cui si vuole calcolare la probabilità, è esprimibile come l'unione disgiunta di due eventi:

- $A$ : ognuno dei  $k$  eventi avviene in un intervallo diverso, ossia: ogni intervallo contiene al più un evento, per un totale di  $k$  eventi;
- $B$ : esiste almeno un intervallo che contiene due eventi, per un totale di  $k$  eventi.

Ricordando che  $n$  è grande a piacere, per l'ipotesi 4 la probabilità che uno specifico intervallo contenga due eventi tende a 0. Inoltre, per l'ipotesi 2 la probabilità dell'evento  $B$  è calcolabile come il prodotto delle singole. La probabilità complessiva è quindi il prodotto di fattori tendenti a 0 ed è pertanto tendente a 0 per  $n \rightarrow +\infty$ . Si ha quindi:

$$P(N(t) = k) = P(A)$$

Siano  $B_i$  le variabili aleatorie bernoulliane che descrivono l'esistenza o meno di un evento nell' $i$ -esimo intervallo. Per l'ipotesi 3 si ha, per  $n \rightarrow +\infty$ :

$$B_i \sim B\left(\lambda \frac{t}{n}\right)$$

Si noti inoltre che l'evento  $A$  corrisponde a una particolare specificazione della variabile aleatoria binomiale  $B'$  che descrive il numero di eventi complessivi in  $n$  intervalli, ossia la somma dei  $B_i$ :

$$B' \sim B\left(n, \lambda \frac{t}{n}\right)$$

Ma allora:

$$P(N(t) = k) = P(A) = P(B' = k)$$

$$N(t) = B'$$

Poiché il prodotto dei parametri della binomiale trovata è costante, ed essendo  $n \rightarrow +\infty$ , per quanto visto al paragrafo 5.5.4 la distribuzione è ben descritta da un modello di Poisson di parametro  $\lambda t$ :

$$N(t) \sim P(\lambda t)$$

□