

Probabilità semplice (per davvero)

Rovesti Gabriel

Attenzione



Il file non ha alcuna pretesa di correttezza; di fatto, è una riscrittura attenta di appunti, slide, materiale sparso in rete, approfondimenti personali dettagliati al meglio delle mie capacità. Credo comunque che, per scopo didattico e di piacere di imparare (sì, io studio per quello e non solo per l'esame) questo file possa essere utile. Semplice si pone, per davvero ci prova.

Thank me sometimes, it won't kill you that much.

Gabriel

Sommario

Insiemi	3
Calcolo combinatorio.....	4
Vettori.....	6
Matrici	12
Spazi di probabilità	24
Probabilità condizionali e indipendenza.....	38
Indipendenza di due o più eventi	44
Variabili aleatorie	52
Variabili aleatorie discrete e valor medio.....	60
Disuguaglianza di Markov-Chebyshev e approssimazione di Poisson.....	77
Vettori aleatori discreti.....	82
Indipendenza, covarianza e correlazione	88
Funzioni di ripartizioni per variabili aleatorie reali.....	94
Variabili aleatorie assolutamente continue	99
Legge dei grandi numeri e teorema limite centrale	108
Statistica descrittiva	119
Lettura delle tavole.....	123
Lettura tavola Normale Standard	123
Lettura Tavole con Poisson.....	124
Media, moda, mediana.....	128
Quantili e vari tipi	131
Indici di dispersione.....	134
Stime.....	135
Distribuzione degli stimatori	143
Intervalli di confidenza: classici, unilaterali e bilaterali.....	145



Unione: $A \cup B$

Intersezione: $A \cap B$

Differenza: $A \setminus B$

Le **LEGGI DI DE MORGAN** sono due: iniziamo col vedere la prima.

La **PRIMA LEGGE DI DE MORGAN** afferma che il *complementare dell'intersezione di due insiemi è uguale all'unione del complementare del primo insieme col complementare del secondo insieme.*

In altri termini:

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

La **SECONDA LEGGE DI DE MORGAN** afferma che il *complementare dell'unione di due insiemi è uguale all'intersezione del complementare del primo insieme col complementare del secondo insieme.*

In altri termini:

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

L'**insieme \mathbb{N}** è l'insieme dei numeri naturali. I suoi elementi sono tutti i numeri interi non negativi.

$$\mathbb{N} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, \dots\}$$

L'**insieme \mathbb{Z}** è l'insieme dei numeri interi relativi. Gli elementi di questo insieme sono tutti i numeri interi caratterizzati da un segno, che può essere positivo (+), negativo (-) o nullo: in particolare, l'unico elemento con segno nullo è lo zero.

$$\mathbb{Z} = \{\dots, -7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, \dots\}$$

L'**insieme \mathbb{Q}** è l'insieme dei numeri razionali relativi, cioè di quei numeri che si esprimono attraverso una frazione e sono preceduti da segno positivo (+), negativo (-) o nullo.

Gli elementi dell'insieme \mathbb{Q} si presentano nella forma $\frac{a}{b}$, dove a e b sono due qualsiasi numeri interi positivi, con b diverso da zero.

$$c \in \mathbb{Q} \Leftrightarrow c = \frac{a}{b} \text{ con } a, b \in \mathbb{Z} \text{ e } b \neq 0$$

L'**insieme \mathbb{I}** è l'insieme dei numeri irrazionali, cioè di quei numeri che non possono essere espressi attraverso una frazione. Gli elementi di questo insieme sono tutti i numeri decimali illimitati non periodici, cioè quei numeri per cui non esiste una frazione generatrice: ad esempio $\pi, \sqrt{2}, \sqrt{3}, -\sqrt[3]{5}$.

L'**insieme \mathbb{R}** è l'insieme dei numeri reali ed definito come l'unione tra l'insieme dei numeri razionali e l'insieme dei numeri irrazionali.

$$\mathbb{R} = \mathbb{Q} \cup \mathbb{I}$$

Pertanto i suoi elementi sono quei numeri che possono essere espressi attraverso una rappresentazione decimale, sia limitata che illimitata, sia periodica che non periodica: cioè, sono tutti quei numeri positivi e negativi (zero incluso) che ci vengono in mente e con cui abbiamo a che fare nella vita di tutti i giorni.

Calcolo combinatorio

- Disposizioni semplici (senza ripetizione)

- Figurano « k » oggetti senza ripetizione ed è gruppo ordinato (conta l'ordine)

$$\# \text{possibili scelte ordinate} = n(n-1) \dots (n-k+1)$$

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

Esempio: In quanti modi diversi 5 alunni si possono sedere su 3 sedie numerate?

- Disposizioni con ripetizione

- Tutti i possibili raggruppamenti di oggetti tali che in ciascuno appaiono k oggetti un massimo di k volte

- n^k

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

- Permutazioni semplici (senza ripetizione)

- Prendo tutti gli n oggetti e formo un gruppo con tutti questi (conta l'ordine)

- $n!$

- Quanti anagrammi anche senza senso si possono formare con la parola LIBRO?

- Permutazioni con ripetizione

- Tutti i possibili raggruppamenti di oggetti tali che in ciascuno appaiono k oggetti un massimo di k volte (conta l'ordine) e non tutti gli elementi sono diversi

- $\frac{n!}{r_1!r_2!\dots r_k!}$

- Quanti anagrammi anche senza senso si possono formare con la parola MAMMA?



- **Combinazioni semplici (senza ripetizione)**
 - Prendo tra gli oggetti alcuni tali che ciascun gruppo differisca dai restanti almeno per uno degli elementi in esso contenuti (non conta l'ordine)
 - $\frac{n!}{k!(n-k)!} \rightarrow$ Coefficiente binomiale
- **Combinazioni con ripetizione**
 - Prendo tra gli oggetti alcuni tali che ciascun gruppo differisca dai restanti almeno per uno degli elementi in esso contenuti un massimo di k volte (non conta l'ordine)
 - $\frac{(n+k-1)!}{k!(n-1)!}$

$n =$ numero di oggetti $k =$ numero di posti	senza ripetizione di oggetti	con ripetizione r di oggetti
Permutazioni <ul style="list-style-type: none"> • $n = k$ • conta l'ordine 	$P_n = n!$	$P_n^r = \frac{n!}{r_1! r_2! \dots r_k!}$
Disposizioni <ul style="list-style-type: none"> • $n \neq k$ • conta l'ordine 	$D_{n,k} = \frac{n!}{(n-k)!} \quad n > k$	$D_{n,k}^r = n^k$
Combinazioni <ul style="list-style-type: none"> • $n \neq k$ • non conta l'ordine 	$C_{n,k} = \frac{n!}{k!(n-k)!} \quad n > k$	$C_{n,k}^r = \frac{(n+k-1)!}{k!(n-1)!}$

Vettori

Un vettore è un insieme di numeri che, in qualche modo, sono fra loro associati. Ad es. un vettore potrebbe rappresentare una variabile misurata su più soggetti o un soggetto con tutte le “misurazioni” effettuate su di lui. Questi sono due esempi numerici di vettori:

$$\begin{bmatrix} 5 \\ 12 \\ 4 \\ 9 \end{bmatrix} \quad [5 \quad 12 \quad 4 \quad 9]$$

Il primo dei due vettori potrebbe rappresentare una variabile ed è chiamato un *vettore colonna*, mentre il secondo potrebbe rappresentare un soggetto ed è chiamato un *vettore riga*.

Quando ci si vuole riferire ad un vettore generico, si usa una specifica notazione: una lettera minuscola (in grassetto) indica il vettore nel suo insieme, mentre la stessa lettera in corsivo seguita da un indice indica un singolo valore del vettore, chiamato “elemento”. Il vettore colonna precedente può essere formalizzato come:

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}$$

dove a_1 indica il primo elemento del vettore (cioè 5), a_2 il secondo (12) e così via.

In modo ancora più generico, possiamo scrivere:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix}$$

dove n è il numero di elementi che compongono il vettore ed è anche chiamato la *dimensione* o l'*ordine* del vettore. Per motivi che diverranno chiari più avanti, la dimensione di un vettore colonna è definibile come $n \times 1$ (n righe e 1 colonna) e quella di un vettore riga come $1 \times n$ (1 riga e n colonne).

In genere, con il solo termine “vettore” si intende un vettore colonna, mentre i vettori riga vengono indicati con un apice dopo la lettera:

$$\mathbf{v}' = [v_1 \quad v_2 \quad \dots \quad v_n]$$

$$\mathbf{a} = [11 \ 12 \ 13 \ 14] \quad \mathbf{b} = [1 \ 2 \ 3 \ 4]$$

Uguaglianza. Due vettori \mathbf{a} e \mathbf{b} si dicono uguali se ogni elemento a_i di \mathbf{a} è uguale al corrispondente elemento b_i di \mathbf{b} . Si scrive $\mathbf{a}=\mathbf{b}$. Se non è vero che ogni elemento di \mathbf{a} corrisponde al corrispondente elemento di \mathbf{b} , si scrive $\mathbf{a} \neq \mathbf{b}$. Esempio:

$$\mathbf{a} = [1 \ 0 \ 3 \ 2 \ 4 \ 1] \quad \mathbf{b} = [1 \ 0 \ 3 \ 2 \ 4 \ 1]$$

Vettore zero. Il **vettore zero** è un vettore in cui tutti gli elementi che lo compongono sono uguali a 0 e viene indicato con uno zero in grassetto.

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{0}' = [0 \ 0 \ 0 \ 0]$$

Vettore unità. Il **vettore unità** è un vettore in cui tutti gli elementi che lo compongono sono uguali a 1 e viene indicato con un 1 in grassetto (oppure, in testi che si rifacciano alla notazione anglosassone, con una \mathbf{u} in grassetto).

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{1}' = [1 \ 1 \ 1 \ 1]$$

3.2.1 Operazioni con gli scalari

Addizione. Aggiungere uno scalare ad un vettore significa sommare lo scalare a ciascuno degli elementi del vettore:

$$x + \mathbf{v} = \begin{bmatrix} x + v_1 \\ x + v_2 \\ \dots \\ x + v_n \end{bmatrix}$$

L'ordine con cui viene effettuata l'operazione non cambia il risultato: $x + \mathbf{v} = \mathbf{v} + x$. Esempi:

$$3 + \begin{bmatrix} 5 \\ 12 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 3 + 5 \\ 3 + 12 \\ 3 + 4 \\ 3 + 9 \end{bmatrix} = \begin{bmatrix} 8 \\ 15 \\ 7 \\ 12 \end{bmatrix}$$

Sottrazione. La sottrazione si può pensare come l'addizione di un numero negativo ad un vettore o l'addizione di uno scalare ad un vettore opposto:

$$x + (-\mathbf{v}) = x - \mathbf{v} \quad \mathbf{v} + (-x) = \mathbf{v} - x$$

Ai fini pratici, sottrarre un vettore da uno scalare significa sottrarre allo scalare ciascuno degli elementi del vettore:

$$x + (-\mathbf{v}) = x - \mathbf{v} = \begin{bmatrix} x + (-v_1) \\ x + (-v_2) \\ \dots \\ x + (-v_n) \end{bmatrix} = \begin{bmatrix} x - v_1 \\ x - v_2 \\ \dots \\ x - v_n \end{bmatrix}$$

Ai fini pratici, sottrarre uno scalare ad un vettore significa sottrarre lo scalare a ciascuno degli elementi del vettore. Nel caso della sottrazione l'ordine con cui viene effettuata l'operazione cambia il risultato: $x - \mathbf{v} \neq \mathbf{v} - x$.

$$\mathbf{v} + (-x) = \mathbf{v} - x = \begin{bmatrix} v_1 - x \\ v_2 - x \\ \dots \\ v_n - x \end{bmatrix}$$

Esempi:

$$3 - \begin{bmatrix} 5 \\ 12 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 3 - 5 \\ 3 - 12 \\ 3 - 4 \\ 3 - 9 \end{bmatrix} = \begin{bmatrix} -2 \\ -9 \\ -1 \\ -6 \end{bmatrix}$$

$$\begin{bmatrix} 5 \\ 12 \\ 4 \\ 9 \end{bmatrix} - 3 = \begin{bmatrix} 5 - 3 \\ 12 - 3 \\ 4 - 3 \\ 9 - 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 9 \\ 1 \\ 6 \end{bmatrix}$$

Moltiplicazione. La moltiplicazione fra un vettore e uno scalare produce un vettore i cui elementi sono ciascuno moltiplicati per lo scalare:

$$x \times \mathbf{v} = \begin{bmatrix} x \times v_1 \\ x \times v_2 \\ \dots \\ x \times v_n \end{bmatrix}$$

L'ordine con cui viene effettuata l'operazione non cambia il risultato: $x \times \mathbf{v} = \mathbf{v} \times x$. Esempi:

$$3 \times \begin{bmatrix} 5 \\ 12 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 3 \times 5 \\ 3 \times 12 \\ 3 \times 4 \\ 3 \times 9 \end{bmatrix} = \begin{bmatrix} 15 \\ 36 \\ 12 \\ 27 \end{bmatrix}$$

$$\begin{bmatrix} 5 \\ 12 \\ 4 \\ 9 \end{bmatrix} \times 3 = \begin{bmatrix} 5 \times 3 \\ 12 \times 3 \\ 4 \times 3 \\ 9 \times 3 \end{bmatrix} = \begin{bmatrix} 15 \\ 36 \\ 12 \\ 27 \end{bmatrix}$$

Valgono le seguenti proprietà:

- commutativa: $k\mathbf{a} = \mathbf{a}k$
- associativa: $k(x\mathbf{a}) = (kx)\mathbf{a}$
- distributiva: $k(\mathbf{a} + \mathbf{b}) = k\mathbf{a} + k\mathbf{b}$ e $(k + x)\mathbf{a} = k\mathbf{a} + x\mathbf{a}$
- $1\mathbf{a} = \mathbf{a}$: un qualunque vettore moltiplicato per 1, non cambia
- $0\mathbf{a} = \mathbf{0}$: un qualunque vettore moltiplicato per 0 si annulla
- $-1\mathbf{a} = -\mathbf{a}$: moltiplicando un vettore per -1, si ottiene il vettore opposto

Divisione. In algebra matriciale non esiste una vera e propria divisione fra uno scalare e un vettore, ma è sufficiente lavorare con il reciproco dello scalare. Vale a dire che:

$$\mathbf{v} \div x = \frac{1}{x}\mathbf{v} = \mathbf{v}\frac{1}{x}$$

Esempio:

$$\frac{1}{3} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 4/3 \\ 3/3 \end{bmatrix} = \begin{bmatrix} 1.33 \\ 1 \end{bmatrix}$$

3.2.2 Operazioni fra vettori

Addizione. Si possono addizionare fra loro solo vettori che hanno la stessa dimensione (2 vettori colonna o due vettori riga) e l'operazione consiste nell'addizionare gli elementi corrispondenti dei due vettori. Non si possono addizionare fra loro un vettore riga e un vettore colonna.

$$\mathbf{a} + \mathbf{b} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \dots \\ a_n + b_n \end{bmatrix}$$

$$\mathbf{a}' + \mathbf{b}' = [a_1 \ a_2 \ \dots \ a_n] + [b_1 \ b_2 \ \dots \ b_n] = [a_1 + b_1 \ a_2 + b_2 \ \dots \ a_n + b_n]$$

L'ordine dell'operazione non cambia il risultato: $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$

Esempi:

$$\begin{bmatrix} 4 \\ 3 \\ 5 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 4+2 \\ 3+4 \\ 5+5 \\ 2+3 \end{bmatrix} = \begin{bmatrix} 6 \\ 7 \\ 10 \\ 5 \end{bmatrix}$$

$$[4 \ 3 \ 5 \ 2] + [2 \ 4 \ 5 \ 3] = [4+2 \ 3+4 \ 5+5 \ 2+3] = [6 \ 7 \ 10 \ 5]$$

L'addizione fra vettori ha alcune proprietà:

- proprietà commutativa: $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$
- proprietà associativa: $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$
- qualunque vettore riga (o colonna) se viene sommato al corrispondente vettore zero, non cambia: $\mathbf{a} + \mathbf{0} = \mathbf{0} + \mathbf{a} = \mathbf{a}$
- la somma di un vettore (\mathbf{a}) con il suo opposto ($-\mathbf{a}$) produce il vettore zero: $\mathbf{a} + (-\mathbf{a}) = (-\mathbf{a}) + \mathbf{a} = \mathbf{0}$

Sottrazione. Anche in questo caso, la sottrazione è in realtà un'addizione con un vettore opposto $\mathbf{a} + (-\mathbf{b})$. Di conseguenza, si possono sottrarre fra loro solo vettori che hanno la stessa dimensione (e lo stesso orientamento, cioè due vettori riga o due vettori colonna) e l'operazione consiste nel sottrarre gli elementi corrispondenti dei due vettori. Essendo una sottrazione, l'ordine degli operandi è importante: $\mathbf{a} - \mathbf{b} \neq \mathbf{b} - \mathbf{a}$

$$\mathbf{a} - \mathbf{b} = \mathbf{a} + (-1)\mathbf{b} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix} - \begin{bmatrix} b_1 \\ a_2 \\ \dots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 - b_1 \\ a_2 - b_2 \\ \dots \\ a_n - b_n \end{bmatrix}$$

$$\mathbf{b} - \mathbf{a} = \mathbf{b} + (-1)\mathbf{a} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix} - \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix} = \begin{bmatrix} b_1 - a_1 \\ b_2 - a_2 \\ \dots \\ b_n - a_n \end{bmatrix}$$

$$\mathbf{a}' - \mathbf{b}' = [a_1 \ a_2 \ \dots \ a_n] - [b_1 \ b_2 \ \dots \ b_n] = [a_1 - b_1 \ a_2 - b_2 \ \dots \ a_n - b_n]$$

Esempi:

$$\begin{bmatrix} 4 \\ 3 \\ 5 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 4-2 \\ 3-4 \\ 5-5 \\ 2-3 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} 2 \\ 4 \\ 5 \\ 3 \end{bmatrix} - \begin{bmatrix} 4 \\ 3 \\ 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 2-4 \\ 4-3 \\ 5-5 \\ 3-2 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Moltiplicazione. La moltiplicazione fra due vettori è possibile solo fra vettori riga e vettori colonna (o viceversa) purché della stessa dimensione.

Il risultato dell'operazione fra un vettore riga e un vettore colonna è uno scalare. Ogni elemento del primo vettore viene moltiplicato per il corrispondente elemento del secondo vettore ed, infine, i prodotti vengono sommati fra loro.

$$[b_1 \quad b_2 \quad \dots \quad b_n] \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix} = a_1(b_1) + a_2(b_2) + \dots + a_n(b_n)$$

Esempio:

$$[1 \quad 2 \quad -1] \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = 1(2) + 2(1) + -1(3) = 2 + 2 + (-3) = 1$$

Non è possibile moltiplicare fra loro due vettori riga o due vettori colonna. Tuttavia, quando si parla di **prodotto scalare** (o *prodotto interno*) fra due vettori, si intende che la *trasposta* del primo vettore viene moltiplicata per il secondo vettore, ovvero che il *prodotto scalare* di **a** e **b** è **a'b**. In questo modo il prodotto scalare di un vettore per se stesso, permette di calcolare la somma dei quadrati degli elementi $\sum x^2$, in quanto ogni elemento viene moltiplicato per se stesso e quindi sommato agli altri:

$$\mathbf{x}'\mathbf{x} = [2 \quad 1 \quad 3] \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = 2(2) + 1(1) + 3(3) = 4 + 1 + 9 = 14$$

E il prodotto scalare del vettore unità con un vettore **a**, permette di calcolare la somma degli elementi $\sum x$, infatti ogni elemento viene moltiplicato per 1 e sommato agli altri:

$$\mathbf{1}'\mathbf{x} = [1 \quad 1 \quad 1] \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = 1(2) + 1(1) + 1(3) = 2 + 1 + 3 = 6$$

Un particolare tipo di moltiplicazione fra vettori è quello fra un vettore colonna e un vettore riga (*prodotto esterno*), ma il risultato di questa operazione è una matrice. Per cui affronteremo questa operazione più avanti (p. 18).

A volte, ai termini che indicano la moltiplicazione, si fanno precedere dai prefissi *pre-* e *post-*. Per la moltiplicazione **a'a**, si può dire che il vettore **a'** pre-moltiplica **a** oppure che il vettore **a** viene post-moltiplicato ad **a'**.

Matrici

4.1 Definizione di una matrice

Una matrice è un insieme ordinato di numeri disposti in righe e colonne, come una tabella. Ecco alcuni esempi numerici di matrici:

$$\begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix} \quad \begin{bmatrix} 3 & 4 & 5 & -1 \\ 7 & 3 & 2 & 6 \\ 6 & 1 & 6 & 3 \end{bmatrix} \quad \begin{bmatrix} 2 \\ 6 \\ 8 \end{bmatrix}$$

Potete pensare a una matrice come un insieme di vettori riga o di vettori colonna oppure ai vettori come a delle matrici composte da una sola riga o da una sola colonna.

Più in generale una matrice viene indicata formalmente come:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}$$

In questo caso, si tratta di un insieme ordinato di 3 righe e 4 colonne e viene chiamata una matrice rettangolare. E ancor più in generale, una matrice può essere indicata come:

$$\mathbf{A} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

In questo caso, m e n indicano rispettivamente il numero di righe e di colonne di cui è composta la matrice. La matrice viene anche indicata con una lettera maiuscola in grassetto. Le dimensioni della matrice (numero di righe e di colonne) sono indicate come $m \times n$ in cui m (il primo indice) indica il numero di righe e n (il secondo indice) il numero di colonne. Le dimensioni delle tre matrici numeriche di esempio sono, rispettivamente, 2×2 , 3×4 , 3×1 . Per indicare una matrice di una certa dimensione si può usare $\mathbf{A} : n \times m$ oppure $\mathbf{A}_{n \times m}$.

Le parentesi quadre indicano usualmente la matrice, ma è possibile trovare matrici indicate con parentesi tonde o doppie righe verticali.

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \left\| \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right\|$$

Poiché le matrici sono insiemi ordinati di numeri posti su righe e colonne (tabelle), si utilizza un sistema duplice di indici per identificare gli elementi della matrice: il primo indice corrisponde alla riga e il secondo alla colonna. L'elemento a_{23} è l'elemento che si trova all'incrocio fra la riga 2 e la colonna 3; l'elemento a_{ij} è l'elemento che si trova all'incrocio fra la riga i e la colonna j , dove i e j indicano genericamente una qualunque riga o una qualunque colonna. Per questo motivo è possibile indicare una matrice anche in forma (super-)abbreviata:

$$\mathbf{A} = [a_{ij}]$$

Un vettore può essere quindi considerato come una matrice che ha una delle due dimensioni uguali a 1. Un vettore riga è una matrice $1 \times n$, mentre un vettore colonna è una matrice $m \times 1$.

Uno scalare può essere considerato come una matrice di dimensione 1×1 .

In ambito statistico, una tabella di dati può essere considerata come una matrice e una matrice può essere usata per lavorare con una tabella dati.

4.3 Matrice trasposta

La trasposta di una matrice \mathbf{A} è un'altra matrice \mathbf{A}' , derivata scambiando le righe con le colonne, in modo che la *riga* i diventi la *colonna* i della trasposta. In altre parole, ogni vettore riga di una matrice diventa un vettore colonna della trasposta.

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 4 & 1 \end{bmatrix} \quad \mathbf{A}' = \begin{bmatrix} 1 & 2 \\ 0 & 4 \\ 3 & 1 \end{bmatrix}$$

Se la dimensione di una matrice è $n \times m$, la sua trasposta sarà $m \times n$. La trasposta della trasposta corrisponde alla matrice di partenza:

$$(\mathbf{A}')' = \mathbf{A}$$

4.4 Operazioni possibili con le matrici

Sulle matrici si possono eseguire le operazioni dell'addizione, sottrazione, moltiplicazione e divisione per uno scalare, con vettori o con altre matrici.

Elementi corrispondenti. Gli elementi di due matrici diverse che hanno gli stessi indici, si chiamano corrispondenti perché occupano la stessa posizione. Ad esempio:

$$\mathbf{A} = \begin{bmatrix} 11 & 12 \\ 13 & 14 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

L'elemento $a_{21} = 13$ della matrice \mathbf{A} è corrispondente all'elemento $b_{21} = 3$ della matrice \mathbf{B} , perché hanno la stessa posizione, indicata dagli stessi indici di riga e colonna (21).

Uguaglianza. Due matrici \mathbf{A} e \mathbf{B} (di uguali dimensioni) si dicono uguali se ogni elemento a_{ij} di \mathbf{A} è uguale ad ogni corrispondente elemento b_{ij} di \mathbf{B} . Si scrive $\mathbf{A}=\mathbf{B}$. Se non è vero che ogni elemento di \mathbf{A} corrisponde al corrispondente elemento di \mathbf{B} , si scrive $\mathbf{A} \neq \mathbf{B}$. Esempio:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 4 & 1 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 4 & 1 \end{bmatrix}$$

Matrice opposta o negativa. La matrice negativa di \mathbf{A} è la matrice $-\mathbf{A}$ i cui elementi corrispondenti sono uguali a quelli di \mathbf{A} ma di segno opposto.

$$\mathbf{A} = \begin{bmatrix} 4 & -1 & 3 \\ 2 & 4 & 1 \end{bmatrix} \quad -\mathbf{A} = \begin{bmatrix} -4 & 1 & -3 \\ -2 & -4 & -1 \end{bmatrix}$$

4.4.1 Operazioni con gli scalari

Le operazioni di addizione e moltiplicazione per uno scalare godono della proprietà commutativa (ovvero $x + \mathbf{A} = \mathbf{A} + x$ e $x\mathbf{A} = \mathbf{A}x$), mentre sottrazione e divisione, no.

Addizione. L'addizione fra uno scalare e una matrice, produce una matrice i cui elementi sono tutti addizionati allo scalare:

$$x + \mathbf{A} = x + [a_{ij}] = \begin{bmatrix} x + a_{11} & \cdots & x + a_{1n} \\ \cdots & \cdots & \cdots \\ x + a_{m1} & \cdots & x + a_{mn} \end{bmatrix}$$

Esempio:

$$3 + \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} = \begin{bmatrix} 3+1 & 3+2 & 3+3 & 3+4 \\ 3+5 & 3+6 & 3+7 & 3+8 \end{bmatrix} = \begin{bmatrix} 4 & 5 & 6 & 7 \\ 8 & 9 & 10 & 11 \end{bmatrix}$$

Sottrazione Anche in questo caso, la sottrazione fra uno scalare e una matrice può essere pensata come l'addizione con una matrice opposta. Produce una matrice i cui elementi sono il risultati della sottrazione fra lo scalare e l'elemento stesso. La sottrazione non gode della proprietà commutativa, per cui l'ordine con cui viene effettuata l'operazione produce risultati diversi ($x - \mathbf{A} \neq \mathbf{A} - x$), in quanto è l'addizione con due operandi diversi.

$$x + (-\mathbf{A}) = x - [a_{ij}] = \begin{bmatrix} x - a_{11} & \cdots & x - a_{1n} \\ \cdots & \cdots & \cdots \\ x - a_{m1} & \cdots & x - a_{mn} \end{bmatrix}$$

$$\mathbf{A} + (-x) = [a_{ij}] - x = \begin{bmatrix} a_{11} - x & \cdots & a_{1n} - x \\ \cdots & \cdots & \cdots \\ a_{m1} - x & \cdots & a_{mn} - x \end{bmatrix}$$

Esempi:

$$3 - \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} = \begin{bmatrix} 3-1 & 3-2 & 3-3 & 3-4 \\ 3-5 & 3-6 & 3-7 & 3-8 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 & -1 \\ -2 & -3 & -4 & -5 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} - 3 = \begin{bmatrix} 1-3 & 2-3 & 3-3 & 4-3 \\ 5-3 & 6-3 & 7-3 & 8-3 \end{bmatrix} = \begin{bmatrix} -2 & -1 & 0 & 1 \\ 2 & 3 & 4 & 5 \end{bmatrix}$$

Notate i risultati delle due operazioni; gli elementi corrispondenti hanno segno opposto.

Moltiplicazione La moltiplicazione fra uno scalare e una matrice produce una matrice i cui elementi sono il risultato della moltiplicazione fra lo scalare e l'elemento stesso.

$$x * \mathbf{A} = x * [a_{ij}] = \begin{bmatrix} x * a_{11} & \dots & x * a_{1n} \\ \dots & \dots & \dots \\ x * a_{m1} & \dots & x * a_{mn} \end{bmatrix}$$

Esempio:

$$3 * \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} = \begin{bmatrix} 3*1 & 3*2 & 3*3 & 3*4 \\ 3*5 & 3*6 & 3*7 & 3*8 \end{bmatrix} = \begin{bmatrix} 3 & 6 & 9 & 12 \\ 15 & 18 & 21 & 24 \end{bmatrix}$$

Valgono le stesse proprietà valide fra uno scalare e un vettore (v. pag. 8):

- associativa: $k(x\mathbf{A}) = (kx)\mathbf{A}$
- distributiva: $k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B} = (\mathbf{A} + \mathbf{B})k$ e $(k + x)\mathbf{A} = k\mathbf{A} + x\mathbf{A}$
- $1\mathbf{A} = \mathbf{A}$: una qualunque matrice moltiplicata per 1, non cambia
- $0\mathbf{A} = \mathbf{0}$: una matrice moltiplicata per 0 si annulla
- $-1 \times \mathbf{A} = -\mathbf{A}$: moltiplicando una matrice per -1, si ottiene la matrice opposta o negativa, in quanto ogni elemento cambia di segno.

Divisione Anche in questo caso, la divisione fra uno scalare e una matrice si ottiene tramite la moltiplicazione con l'inverso dello scalare.

$$\mathbf{A} \div x = \frac{1}{x} [a_{ij}] = \begin{bmatrix} \frac{1}{x} a_{11} & \dots & \frac{1}{x} a_{1n} \\ \dots & \dots & \dots \\ \frac{1}{x} a_{m1} & \dots & \frac{1}{x} a_{mn} \end{bmatrix}$$

Esempio:

$$\frac{1}{3} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/3 & 3/3 & 4/3 \\ 5/3 & 6/3 & 7/3 & 8/3 \end{bmatrix} = \begin{bmatrix} 0.333 & 0.667 & 1 & 1.333 \\ 1.667 & 2 & 2.333 & 2.667 \end{bmatrix}$$

4.4.2 Operazioni fra vettori e matrici

Abbiamo visto che un vettore può essere considerato una matrice in cui una delle due dimensioni è uguale a 1. In tal caso (e diventerà chiaro nella prossima sezione), l'unica operazione possibile è la moltiplicazione fra un vettore e una matrice (o viceversa) purché siano valide le condizioni di compatibilità. Poiché questo tipo di operazione è solo un caso particolare della moltiplicazione fra matrici, la affronteremo più oltre.

4.4.3 Operazioni fra matrici

Anche fra matrici si possono eseguire operazioni, purché le dimensioni delle matrici siano fra loro compatibili.

Addizione. La somma di due matrici **A** e **B** aventi lo stesso numero di righe e di colonne (cioè la stessa dimensione) è una matrice **C** i cui elementi sono la somma dei corrispondenti elementi di **A** e di **B**.

$$\begin{matrix} \mathbf{A} & + & \mathbf{B} & = & \mathbf{C} \\ n \times m & & n \times m & & n \times m \end{matrix}$$

Si può scrivere, per qualunque elemento di **C**:

$$\mathbf{C} = [c_{ij}] = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}]$$

Se le due matrici non sono dello stesso ordine (ovvero non hanno le stesse dimensioni), l'operazione non si può eseguire e si dice che le matrici non sono *compatibili* o *conformabili*.

L'ordine con cui vengono sommate le matrici non è importante:

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \cdots & \cdots & \cdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}$$

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$$

Esempio:

$$\begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix} + \begin{bmatrix} 2 & 3 & 6 \\ 4 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1+2 & 2+3 & 4+5 \\ 1+4 & 3+2 & 5+1 \end{bmatrix} = \begin{bmatrix} 3 & 5 & 9 \\ 5 & 5 & 6 \end{bmatrix}$$

Le proprietà dell'addizione sono:

- commutativa: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
- associativa: $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$
- $(\mathbf{A} + \mathbf{B} + \mathbf{C})' = \mathbf{A}' + \mathbf{B}' + \mathbf{C}'$: la trasposta di una somma è uguale alla somma delle trasposte

Sottrazione. La sottrazione è analoga all'addizione, in quanto possiamo pensare alla sottrazione come all'addizione di una matrice **A** con l'opposta di **B** (cioè **-B**), per cui per qualunque elemento di **C**:

$$\mathbf{C} = [c_{ij}] = \mathbf{A} + (-1 \times \mathbf{B}) = \mathbf{A} - \mathbf{B} = [a_{ij}] - [b_{ij}]$$

Anche in questo caso, se le due matrici non sono dello stesso ordine, l'operazione non si può eseguire.

L'ordine con cui vengono sottratte le matrici è importante, in quanto

$$\mathbf{A} - \mathbf{B} \neq \mathbf{B} - \mathbf{A}$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \cdots & \cdots & \cdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}$$

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} a_{11} - b_{11} & \cdots & a_{1n} - b_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} - b_{m1} & \cdots & a_{mn} - b_{mn} \end{bmatrix}$$

$$\mathbf{B} - \mathbf{A} = \begin{bmatrix} b_{11} - a_{11} & \cdots & b_{1n} - a_{1n} \\ \cdots & \cdots & \cdots \\ b_{m1} - a_{m1} & \cdots & b_{mn} - a_{mn} \end{bmatrix}$$

Esempi:

$$\begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix} - \begin{bmatrix} 2 & 3 & 6 \\ 4 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1-2 & 2-3 & 4-5 \\ 1-4 & 3-2 & 5-1 \end{bmatrix} = \begin{bmatrix} -1 & -1 & -1 \\ -3 & 1 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 3 & 6 \\ 4 & 2 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 2-1 & 3-2 & 6-4 \\ 4-1 & 2-3 & 1-5 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 3 & -1 & -4 \end{bmatrix}$$

Da quanto detto sopra, risulta che se $\mathbf{A} + \mathbf{B} = \mathbf{0}$ allora dev'essere $\mathbf{B} = -\mathbf{A}$ ovvero che \mathbf{B} è l'opposta di \mathbf{A} .

Moltiplicazione. E' possibile moltiplicare due matrici fra loro se, e solamente se, il numero di colonne della prima matrice è uguale al numero di righe della seconda, ovvero se la dimensione di \mathbf{A} è $n \times m$ e quella di \mathbf{B} è $m \times p$. La matrice risultante avrà dimensioni corrispondenti alle righe della prima e alle colonne della seconda, ovvero $n \times p$:

$$\begin{matrix} \mathbf{A} & * & \mathbf{B} & = & \mathbf{C} \\ n \times m & & m \times p & & n \times p \end{matrix}$$

La matrice risultato \mathbf{C} si ottiene moltiplicando ogni vettore riga di \mathbf{A} per ogni vettore colonna di \mathbf{B} . Se

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ \cdots & \cdots & \cdots & \cdots \\ b_{m1} & b_{m2} & \cdots & b_{mp} \end{bmatrix}$$

La matrice \mathbf{C} avrà n righe e p colonne:

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{np} \end{bmatrix}$$

Il primo elemento, il c_{11} , si ottiene con la sommatoria dei prodotti degli elementi corrispondenti del primo vettore riga di \mathbf{A} con il primo vettore colonna di \mathbf{B} :

$$c_{11} = [a_{11} \quad a_{12} \quad \cdots \quad a_{1m}] \begin{bmatrix} b_{11} \\ b_{21} \\ \cdots \\ b_{m1} \end{bmatrix} =$$

$$c_{11} = a_{11} * b_{11} + a_{12} * b_{21} + a_{13} * b_{31} + \dots + a_{1m} * b_{m1}$$

E l'ultimo elemento della matrice **C**, il c_{np} , si ottiene dalla moltiplicazione dell'ultimo vettore riga di **A** con l'ultimo vettore colonna di **B**:

$$c_{np} = [a_{n1} \quad a_{n2} \quad \dots \quad a_{nm}] \begin{bmatrix} b_{1p} \\ b_{2p} \\ \dots \\ b_{mp} \end{bmatrix} =$$

$$= a_{n1} * b_{1p} + a_{n2} * b_{2p} + a_{n3} * b_{3p} + \dots + a_{nm} * b_{mp}$$

Notate che, ogni volta, gli elementi che vengono moltiplicati fra loro hanno lo stesso indice interno ($a_{n2} * b_{2p}$, in questo caso il 2).

Un esempio numerico può essere più esplicativo:

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 2 & 1 \\ -1 & 2 & 3 & 5 \\ 3 & 4 & 1 & 2 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ 3 & 4 \\ -2 & 1 \end{bmatrix}$$

Poiché **A** è una matrice 3×4 e **B** 4×2 , la matrice **C** avrà ordine 3×2 . I singoli elementi si ottengono come segue:

$$\begin{aligned} c_{11} &= 1 \times 1 + 3 \times 2 + 2 \times 3 + 1 \times (-2) &&= 1 + 6 + 6 - 2 \\ c_{12} &= 1 \times 3 + 3 \times 1 + 2 \times 4 + 1 \times 1 &&= 3 + 3 + 8 + 1 \\ c_{21} &= -1 \times 1 + 2 \times 2 + 3 \times 3 + 5 \times (-2) &&= -1 + 4 + 9 - 10 \\ c_{22} &= -1 \times 3 + 2 \times 1 + 3 \times 4 + 5 \times 1 &&= -3 + 2 + 12 + 5 \\ c_{31} &= 3 \times 1 + 4 \times 2 + 1 \times 3 + 2 \times (-2) &&= 3 + 8 + 3 - 4 \\ c_{32} &= 3 \times 3 + 4 \times 1 + 1 \times 4 + 2 \times 1 &&= 9 + 4 + 4 + 2 \end{aligned}$$

La matrice **C** sarà quindi:

$$\begin{bmatrix} 11 & 15 \\ 2 & 16 \\ 10 & 19 \end{bmatrix}$$

Un'altro modo per indicare la moltiplicazione fra matrici è:

$$\mathbf{AB} = \mathbf{C} = [c_{ij}] = \sum a_{ik} b_{kj}$$

Si può capire perché la moltiplicazione di due matrici è possibile soltanto quando il numero di righe della matrice da moltiplicare è uguale al numero di colonne della matrice moltiplicativa. Ne segue che **AB** non è uguale a **BA** se le matrici sono rettangolari. In tal caso **BA** potrebbe non essere neppure calcolabile. Se la matrice **A** è di ordine 2×3 ($\mathbf{A}_{2 \times 3}$) e $\mathbf{B}_{3 \times 2}$, allora **BA** è possibile e sarà di dimensione 3×3 . Mentre se $\mathbf{A}_{2 \times 3}$ e $\mathbf{B}_{3 \times 4}$, allora **BA** non è conformabile. Se invece le matrici sono quadrate e dello stesso ordine (ossia $n \times n$), **BA** è possibile, ma **AB** \neq **BA**.

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 3 & 4 \\ -2 & 1 \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} 1(3) + 3(-2) & 1(4) + 3(1) \\ 2(3) + 1(-2) & 2(4) + 1(1) \end{bmatrix} = \begin{bmatrix} 3 - 6 & 4 + 3 \\ 6 - 2 & 8 + 1 \end{bmatrix} = \begin{bmatrix} -3 & 7 \\ 4 & 9 \end{bmatrix}$$

$$\mathbf{BA} = \begin{bmatrix} 3(1) + 4(2) & 3(3) + 4(1) \\ -2(1) + 1(2) & -2(3) + 1(1) \end{bmatrix} = \begin{bmatrix} 3 + 8 & 9 + 4 \\ -2 + 2 & -6 + 1 \end{bmatrix} = \begin{bmatrix} 11 & 13 \\ 0 & -5 \end{bmatrix}$$

Nella sezione sui vettori abbiamo presentato la moltiplicazione fra vettori riga per vettori colonna, il cui risultato è uno scalare. Dovrebbe essere chiaro ora perché non è possibile moltiplicare fra loro due vettori riga (o due vettori colonna). E' invece possibile moltiplicare un vettore colonna per un vettore riga, ma il risultato è una matrice.

$$\begin{matrix} \mathbf{a}' & * & \mathbf{b} & = & c & \text{mentre} & \mathbf{b} & * & \mathbf{a}' & = & \mathbf{C} \\ 1 \times m & & m \times 1 & & 1 \times 1 & & m \times 1 & & 1 \times m & & m \times m \end{matrix}$$

L'operazione di moltiplicazione è la stessa: ogni riga del vettore moltiplicatore viene moltiplicata per ogni colonna del secondo vettore. Solo che in questo caso è uno scalare moltiplicato per un'altro scalare.

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} [b_1 \ b_2 \ b_3 \ b_4] = \begin{bmatrix} a_1 b_1 & a_1 b_2 & a_1 b_3 & a_1 b_4 \\ a_2 b_1 & a_2 b_2 & a_2 b_3 & a_2 b_4 \\ a_3 b_1 & a_3 b_2 & a_3 b_3 & a_3 b_4 \\ a_4 b_1 & a_4 b_2 & a_4 b_3 & a_4 b_4 \end{bmatrix}$$

Esempio:

$$\begin{bmatrix} 2 \\ 3 \\ 4 \\ 5 \end{bmatrix} [-1 \ 3 \ 0 \ 2] = \begin{bmatrix} 2(-1) & 2(3) & 2(0) & 2(2) \\ 3(-1) & 3(3) & 3(0) & 3(2) \\ 4(-1) & 4(3) & 4(0) & 4(2) \\ 5(-1) & 5(3) & 5(0) & 5(2) \end{bmatrix} = \begin{bmatrix} -2 & 6 & 0 & 4 \\ -3 & 9 & 0 & 6 \\ -4 & 12 & 0 & 8 \\ -5 & 15 & 0 & 10 \end{bmatrix}$$

Proprietà. Si può osservare che, per quanto riguarda la moltiplicazione, le seguenti proprietà sono soddisfatte (le prime due derivano dall'algebra dei numeri):

- a) associativa, $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) = \mathbf{ABC}$: l'ordine con cui si moltiplicano fra loro le prime due matrici non influisce sul risultato finale (è implicito che le matrici devono avere le dimensioni giuste per poter essere moltiplicate fra loro);
- b) distributiva, $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$: la moltiplicazione di una matrice \mathbf{A} per la somma di due matrici \mathbf{B} e \mathbf{C} , equivale a sommare il prodotto di ciascuna delle matrici \mathbf{B} e \mathbf{C} per la matrice \mathbf{A} ;
- c) $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$: la trasposta di un prodotto è uguale al prodotto delle trasposte (ma in ordine inverso).

Valgono anche le normali regole delle potenze, ovviamente per matrici conformabili:

- a) $\mathbf{AA} = \mathbf{A}^2$
- b) $\mathbf{A}^m \mathbf{A}^n = \mathbf{A}^{m+n}$
- c) $(\mathbf{A}^m)^n = (\mathbf{A}^n)^m = \mathbf{A}^{mn}$

4.4.4 Combinazione lineare

Se \mathbf{b} è un vettore (ad es. di ordine 2) e \mathbf{A} una matrice (ad es. di ordine 3×2), il prodotto \mathbf{Ab} può essere sviluppato come:

$$b_1 A_{i1} + b_2 A_{i2} + b_3 A_{i3}$$

in quanto

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} b_1 a_{11} + b_2 a_{12} \\ b_1 a_{21} + b_2 a_{22} \\ b_1 a_{31} + b_2 a_{32} \end{bmatrix}$$

Il prodotto \mathbf{Ab} è chiamato combinazione lineare del vettore \mathbf{b} con la matrice \mathbf{A} .

Più in generale, se \mathbf{x} , \mathbf{y} , \mathbf{w} e \mathbf{z} sono dei vettori e i , j , k ed l sono degli scalari, la somma dei prodotti fra gli scalari e i vettori è chiamata *combinazione lineare*:

$$ix + jy + kw + lz$$

ed è considerata come la somma pesata dei vettori. Per cui, una retta di regressione multipla del tipo:

$$Y_i = b_1 X_{i1} + b_2 X_{i2} + \dots + b_m X_{im}$$

è una combinazione lineare fra gli scalari b racchiusi nel vettore \mathbf{b} e i vettori \mathbf{X}_k raccolti nella matrice \mathbf{X} .

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1m} \\ X_{21} & X_{22} & & X_{2m} \\ \vdots & & & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nm} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{bmatrix}$$

4.5 Matrice quadrata

Si definisce come matrice quadrata una matrice che ha lo stesso numero di righe e di colonne. La dimensione di una matrice quadrata è indicata con un solo numero e si dice di ordine n .

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 3 & 5 & -1 \\ 7 & 2 & 6 \\ 6 & -6 & 3 \end{bmatrix}$$

2×2 3×3

In una matrice quadrata ci sono 2 due diagonali e la diagonale che va dall'elemento a_{11} fino all'elemento a_{nn} (ovvero che hanno indice uguale) si chiama **diagonale principale**. La somma degli elementi sulla diagonale principale si chiama **traccia** della matrice:

$$tr(\mathbf{A}) = \sum_i a_{ii}$$

La traccia delle due matrici precedenti è:

$$tr(\mathbf{A}) = 1 + 6 = 7$$

$$tr(\mathbf{B}) = 3 + 2 + 3 = 8$$

4.5.1 Matrice simmetrica

E' una *matrice quadrata* che, se viene trasposta, non cambia ovvero $\mathbf{A}' = \mathbf{A}$. In pratica, rispetto alla diagonale principale, le due metà sono fra loro speculari. Ecco un esempio di matrice simmetrica:

$$\mathbf{S} = \begin{bmatrix} 1 & 2 & 5 & 6 \\ 2 & 4 & 3 & 3 \\ 5 & 3 & 5 & 2 \\ 6 & 3 & 2 & 3 \end{bmatrix}$$

In questa matrice gli elementi con indici speculari sono uguali fra loro ($a_{12} = a_{21}$, $a_{13} = a_{31}$, $a_{23} = a_{32}$, ...).

Un esempio di matrice simmetrica è la matrice delle correlazioni. Ipotizziamo di avere 4 variabili e di calcolare la correlazione lineare di Pearson fra tutte le variabili. Inseriamo ora le 6 correlazioni in una struttura matriciale: le righe e le colonne indicheranno le variabili mentre i singoli elementi corrisponderanno alle correlazioni. Ovviamente, poiché la correlazione fra x e y è uguale a quella fra y e x , la matrice sarà simmetrica.

$$\mathbf{R} = \begin{bmatrix} 1.00 & .70 & .19 & .55 \\ .70 & 1.00 & .36 & .50 \\ .19 & .36 & 1.00 & .16 \\ .55 & .50 & .16 & 1.00 \end{bmatrix}$$

4.5.2 Matrice diagonale

È una *matrice quadrata simmetrica*, in cui vengono utilizzati solo i valori lungo la diagonale principale; tutti gli altri elementi sono nulli. La diagonale può contenere qualunque scalare, anche 0 e 1.

$$\mathbf{D} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix}$$

Moltiplicare una qualunque matrice per una matrice diagonale (o viceversa) equivale a moltiplicare un elemento della matrice per quello della diagonale, in quanto viene usato solo l'elemento diverso da zero.

$$\mathbf{DA} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 3 & 5 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 4 \times 2 & 4 \times 4 \\ 2 \times 3 & 2 \times 5 \\ 3 \times 3 & 3 \times 2 \end{bmatrix} = \begin{bmatrix} 8 & 16 \\ 6 & 10 \\ 9 & 6 \end{bmatrix}$$

$$\mathbf{AD} = \begin{bmatrix} 2 & 4 & 3 \\ 3 & 5 & 2 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 2 \times 4 & 4 \times 2 & 3 \times 3 \\ 3 \times 4 & 5 \times 2 & 2 \times 3 \end{bmatrix} = \begin{bmatrix} 8 & 8 & 9 \\ 12 & 10 & 6 \end{bmatrix}$$

A volte, gli elementi 0 (zero) della parte superiore destra e inferiore sinistra possono essere interamente omissi (facilitando la lettura). Scrivendo la matrice a mano, si può usare un solo 0 sottolineato (0) anziché riempire tutta l'area.

4.5.3 Matrice scalare

È una *matrice quadrata simmetrica diagonale* in cui tutti gli elementi della diagonale principale sono uguali fra loro:

$$\mathbf{F} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$$

4.5.4 Matrice identica

La *matrice identica* o *identità* è una *matrice quadrata simmetrica diagonale scalare* i cui elementi sono tutti 0 con l'eccezione della diagonale i cui elementi sono invece uguali a 1.

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Si indica con la lettera **I** oppure con \mathbf{I}_n , in cui n indica l'ordine della matrice, ovvero il numero di righe e di colonne.

La traccia di una matrice identica è N (ovvero corrisponde all'ordine della matrice).

$$tr(I_4) = 1 + 1 + 1 + 1 = 4$$

La sua funzione in algebra matriciale è analogo a quello del valore scalare 1 nell'algebra normale. In algebra, il numero 1 è quello che moltiplicato per un qualunque altro numero, non lo modifica: $1 \times x = x$. Allo stesso modo la matrice *identità* moltiplicata per una qualunque altra matrice non la modifica.

$$\mathbf{IA} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix} = \begin{bmatrix} 1(1) + 0(4) & 1(3) + 0(6) \\ 0(1) + 1(4) & 0(3) + 1(6) \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix}$$

Avendo una matrice $\mathbf{A}_{n \times m}$ è possibile definire il prodotto $\mathbf{A}_{n \times m} \mathbf{I}_m$ e anche $\mathbf{I}_n \mathbf{A}_{n \times m}$, ma non è invece possibile, in generale, individuare il prodotto $\mathbf{I}_m \mathbf{A}_{n \times m}$. Si noti anche che $\mathbf{A}_{n \times m} \mathbf{I}_m = \mathbf{I}_n \mathbf{A}_{n \times m} = \mathbf{A}$.

Moltiplicando **I** per uno scalare si ottiene una matrice scalare in cui i valori della diagonale vengono sostituiti dallo scalare stesso, ovvero *una matrice scalare*.

$$3 \times \mathbf{I} = 3 \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3(1) & 3(0) \\ 3(0) & 3(1) \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

4.5.5 Matrice triangolare

E' una *matrice quadrata* in cui una delle due metà è nulla, in genere quella superiore. Ecco un esempio di matrice triangolare:

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 5 & 3 & 5 & 0 \\ 6 & 3 & 2 & 3 \end{bmatrix}$$

4.5.6 Matrice ortogonale*

Una matrice quadrata si dice ortogonale se è vero che

$$\mathbf{AA}' = \mathbf{A}'\mathbf{A} = \mathbf{I}$$

cioè se la moltiplicazione fra la matrice e la sua trasposta (o viceversa) origina la matrice identica.

Inoltre, il prodotto di due matrici ortogonali **A** e **B** è ancora una matrice ortogonale, infatti:

$$(\mathbf{AB})(\mathbf{AB})' = \mathbf{ABB}'\mathbf{A}' = \mathbf{AIA}' = \mathbf{AA}' = \mathbf{I}$$

4.5.7 Matrici congruenti*

Si chiamano congruenti due matrici quadrate **A** e **B** dello stesso ordine, per cui esista una matrice **C** che rende vera l'uguaglianza:

$$\mathbf{A} = \mathbf{C}'\mathbf{BC}$$

4.6 Matrice nulla o matrice zero

E' una matrice in cui tutti gli elementi sono uguali a 0:

$$\emptyset_{4 \times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Una matrice quadrata nulla è anche simmetrica e diagonale.

La matrice nulla svolge le stesse funzioni dello 0 (zero) in algebra scalare:

- a) $\mathbf{A} + \emptyset = \emptyset + \mathbf{A} = \mathbf{A}$; sommando (o sottraendo) la matrice nulla a una qualunque altra matrice, quest'ultima non cambia;
- b) $\mathbf{A} - \mathbf{A} = \mathbf{A} + (-\mathbf{A}) = (-\mathbf{A}) + \mathbf{A} = \emptyset$; una matrice sottratta a se stessa, origina la matrice nulla;
- c) $\mathbf{A}\emptyset = \emptyset$; la matrice nulla moltiplicata per una qualunque matrice, produce la matrice nulla.

Si noti tuttavia che $\mathbf{AB} = \emptyset$ non implica che \mathbf{A} o \mathbf{B} siano necessariamente matrici nulle.

Spazi di probabilità

La teoria della probabilità si occupa di studiare modelli matematici per *esperimenti aleatori*, cioè osservazioni riguardo ad un qualunque fenomeno il cui esito non è determinabile con certezza a priori.

Ad esempio:

- Lancio di un dado/una moneta
- Risultato di un test clinico
- Verificarsi di un terremoto o temporale
- Misurazione di una qualunque grandezza fisica

Per poter descrivere gli esperimenti aleatori, abbiamo bisogno di tre componenti:

- L'insieme di tutti gli esiti possibili dell'esperimento, all'interno di uno spazio campionario, in un insieme non vuoto Ω .

Esempio: lancio di un dado da gioco a sei facce;
osservazione riguarda il numero segnato
 \leadsto spazio campionario "minimale": $\Omega = \{1, \dots, 6\}$.

- L'insieme delle affermazioni ammissibili sull'esito dell'esperimento, definito come *sistema degli eventi*, cioè il sottoinsieme \mathcal{F} delle parti di Ω .

Esempio del dado: possiamo scegliere $\mathcal{F} = \mathcal{P}(\Omega)$.

Interpretazione di un sottoinsieme $A \in \mathcal{P}(\Omega)$, quindi $A \in \mathcal{F}$, come affermazione sull'esito dell'esperimento:

$A \iff$ "si è verificato un $\omega \in A$ "

Nell'esempio: • $A = \{1\} \iff$ "Il dado lanciato segna 1."
• $A = \{1, 2\} \iff$ "Il dado lanciato segna 1 o 2."
• $A = \{1, 3, 5\} \iff$ "Il dado lanciato segna o un numero dispari"
(oppure: "È uscito un numero dispari.")

- Assegnazione di un "grado di fiducia" ad ogni affermazione ammissibile, intesa come *misura di probabilità*, cioè una mappa/funzione compresa tra 0 (minimo gr. di fiducia) e 1 (massimo gr. di fiducia)

Di base, abbiamo che:

Def.: Un tripletta (Ω, \mathcal{F}, P) si dice spazio di probabilità se

- (i) Ω è un insieme non-vuoto, lo spazio campionario,
- (ii) \mathcal{F} è una σ -algebra in Ω , il sistema degli eventi,
- (iii) P è una misura di probabilità su \mathcal{F} .

Definiamo la sigma-algebra:

Def.: \mathcal{F} si dice σ -algebra in Ω , $\Omega \neq \emptyset$, se

- (a) $\emptyset \in \mathcal{F}$;
- (b) se $A \in \mathcal{F}$ allora $A^c \in \mathcal{F}$;
- (c) se $(A_i)_{i \in \mathbb{N}} \subset \mathcal{F}$ allora $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$.

Nota: Se \mathcal{F} σ -algebra in Ω , allora:

- $\emptyset, \Omega \in \mathcal{F}$
- $(A_i)_{i \in \mathbb{N}} \subset \mathcal{F}$ implica $\bigcap_{i \in \mathbb{N}} A_i \in \mathcal{F}$.

Dato un insieme Ω , si definisce σ -algebra su Ω una famiglia \mathfrak{F} di sottoinsiemi di Ω tale che:^[2]

- L'insieme Ω appartiene a \mathfrak{F} .
- Se un insieme A è in \mathfrak{F} , allora il suo **complementare** è in \mathfrak{F} .
- Se gli elementi A_i di una famiglia **numerabile** di insiemi $\{A_i\}_{i \in \mathbb{N}}$ sono in \mathfrak{F} , allora la loro **unione**:

$$A = \bigcup_{i=1}^{\infty} A_i$$

appartiene a \mathfrak{F} .

Se \mathfrak{F} è una σ -algebra su Ω , allora Ω si dice **spazio misurabile** e gli elementi di \mathfrak{F} sono detti **insiemi misurabili** in Ω .^[2]

Una σ -algebra, in particolare, è un'**algebra di insiemi**, poiché la terza condizione sopraindicata implica la stabilità per unione *finita* richiesta nella definizione di struttura di algebra. In tal caso si richiede la stabilità anche per unioni numerabili, da cui l'identificativo σ , un'abbreviazione per **successione**.

In matematica e in particolare in teoria delle probabilità, una **sigma-algebra** è un insieme di eventi (sottoinsiemi) che soddisfano determinate proprietà.

Un insieme di eventi forma una sigma-algebra se:

- Il completo vuoto e l'insieme totale appartengono alla sigma-algebra.
- Ogni evento nella sigma-algebra è un insieme chiuso rispetto alla sua complementare.
- L'unione di eventi appartenenti alla sigma-algebra appartiene anche alla sigma-algebra.

L'utilità della sigma-algebra in teoria delle probabilità è che descrive l'universo di eventi in cui si verifica la probabilità. Inoltre, la sigma-algebra definisce il concetto di evento misurabile, ovvero un evento per cui si può assegnare una probabilità.

Ad esempio, in una sigma-algebra di eventi legati al lancio di una moneta, l'insieme totale potrebbe essere tutte le possibili combinazioni di testa o croce, mentre gli eventi appartenenti alla sigma-algebra potrebbero essere eventi come "lancio di testa" o "lancio di croce".

In teoria delle probabilità, una **misura di probabilità** è una funzione che assegna un valore reale compreso tra 0 e 1 ad ogni evento in un universo di eventi descritto da una sigma-algebra. Questo valore rappresenta la probabilità che l'evento si verifichi.

Una misura di probabilità deve soddisfare tre proprietà fondamentali:

- **Non negatività:** la probabilità di ogni evento deve essere maggiore o uguale a zero.
- **Normalizzazione:** la probabilità dell'insieme totale deve essere uguale a 1.
- **Additività:** se gli eventi sono mutuamente esclusivi, allora la loro probabilità totale è la somma delle probabilità individuali (detta anche sigma-additività quando riferito ad una sigma-algebra)

Ciò implica:

- L'additività finita di eventi
- L'evento omega pari ad 1

La più semplice distribuzione di interesse generale è quella in cui si assegna lo stesso grado di fiducia a tutte le possibili realizzazioni. Essa è chiamata distribuzione uniforme discreta.

Esempio: Sia $\Omega \neq \emptyset$ insieme finito.

Poniamo $\mathcal{F} = \mathcal{P}(\Omega)$, e definiamo P tramite

$$P(A) = \frac{|A|}{|\Omega|}, \quad A \subseteq \Omega.$$

$\leadsto P$ è la distribuzione uniforme discreta su Ω .

In questa, gli eventi sono disgiunti a due a due e descrive una situazione in cui ogni possibile valore all'interno di un insieme finito di valori ha la stessa probabilità di verificarsi.

Applicazione al lancio di un dado a sei facce:

$$\Omega = \{1, \dots, 6\}. \quad \mathcal{F} = \mathcal{P}(\Omega).$$

Sia P la distribuzione uniforme discreta.

$$\leadsto P(\{\omega\}) = \frac{1}{6} \quad \text{per ogni } \omega \in \Omega.$$

\leadsto dado regolare ed equilibrato.

Con queste scelte: Probabilità che esca un numero pari?

$$\leadsto P(\{2, 4, 6\}) = \frac{|\{2, 4, 6\}|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}.$$

Nell'esempio del dado, come potremmo modellizzare un dado non equilibrato?

Ad esempio, un dado che segna sempre sei:

Poniamo $\tilde{P}(\{6\}) \doteq 1$

$\leadsto \tilde{P}(\{1, \dots, 5\}) = 0$

$1 = \tilde{P}(\{6\} \cup \{1, \dots, 5\}) \stackrel{\text{add.}}{=} \underbrace{\tilde{P}(\{6\})}_{\doteq 1} + \tilde{P}(\{1, \dots, 5\})$

Vale in generale: $P(A^c) = 1 - P(A)$

Un dado che dia uguale peso ai numeri 2, 5 e non faccia uscire gli altri numeri:

$\hat{P}(\{2\}) \doteq \frac{1}{2}, \quad \hat{P}(\{5\}) \doteq \frac{1}{2}$

$\leadsto \hat{P}(\{1, 3, 4, 6\}) = 0$

$\leadsto \hat{P}(\{1\}) = \hat{P}(\{3\}) = \hat{P}(\{4\}) = \hat{P}(\{6\}) = 0$

Definiamo ora un concetto importante, cioè la densità discreta. Essa è una funzione di probabilità che descrive una distribuzione di probabilità discreta. Esprime la probabilità che una variabile casuale discreta assuma un determinato valore all'interno di un insieme finito di valori possibili. La somma delle probabilità per tutti i valori possibili deve essere uguale a 1, poiché la probabilità totale di un evento è definita come 1.

Def.: Una funzione $p: \Omega \rightarrow [0, 1]$ si dice densità discreta su Ω se

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

Sia p una densità discreta su Ω . Allora

$$P(A) \doteq \sum_{\omega \in A} p(\omega), \quad A \subseteq \Omega,$$

definisce una misura di probabilità su $\mathcal{P}(\Omega)$.

Esempio: Lancio di due monete non necessariamente equilibrate

Osservazione: facce (teste o croce) delle due monete

Spazio campionario: $\mathcal{N} = \{T, C\} \times \{T, C\}$
 $\leadsto \mathcal{N} = \{(T, T), (T, C), (C, T), (C, C)\}$.

(a) Supponiamo che le due monete siano equilibrate.

Densità discreta su \mathcal{N} : \nearrow uniforme discreta

$$p((T, T)) = p((T, C)) = p((C, T)) = p((C, C)) = \frac{1}{4}.$$

(b) Supponiamo ora che la prima moneta sia equilibrata, la seconda truccata con risultato sempre teste:

$$p((T, T)) = \frac{1}{2}, \quad p((C, T)) = \frac{1}{2}, \quad p((C, C)) = p((T, C)) = 0$$

(c) Supponiamo invece che la prima moneta sia equilibrata e la seconda dia teste con probab. $\frac{1}{3}$:

Una soluzione: $p((T, T)) = \frac{1}{6}, \quad p((C, T)) = \frac{1}{6}$

$$p((C, C)) = \frac{1}{3}, \quad p((T, C)) = \frac{1}{3}.$$

$$P(\{(T, T), (T, C)\}) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}, \quad P(\{(T, T), (C, T)\}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

Ma esistono altre soluzioni (\nearrow foglio II)!

Dato un insieme di campioni osservati, possiamo inoltre definire la densità campionaria. Essa è una stima della densità di probabilità di una distribuzione sotto-ospite. La densità campionaria viene calcolata come la distribuzione normalizzata delle osservazioni, dove la normalizzazione viene effettuata dividendo il numero di osservazioni in un intervallo specifico per la lunghezza di questo intervallo. La densità campionaria viene utilizzata frequentemente in statistica e in analisi dei dati per stimare la forma della distribuzione sotto-ospite in modo non parametrico.

Sia $(x_i)_{i \in \{1, \dots, n\}}$ un campione di numerosità n di dati e valori in \mathcal{X} (ad. es., $\mathcal{X} = \mathbb{R}$).

Definiamo $p: \mathcal{X} \rightarrow [0, 1]$ tramite

$$p(x) = \frac{\#\{i \in \{1, \dots, n\} : x_i = x\}}{n}, \quad x \in \mathcal{X}.$$

Allora p è una densità discreta su \mathcal{X} ,
 la densità campionaria di $(x_i)_{i \in \{1, \dots, n\}}$

$p(x)$ è la frequenza relativa del valore $x \in \mathcal{X}$

$\leadsto p$ induce una misura di probab. su $\mathcal{P}(\mathcal{X})$:

$$P(A) = \sum_{x \in A} p(x), \quad A \subseteq \mathcal{X}$$

$\leadsto P$ misure di probab. campionaria (o empirica).

Listiamo delle proprietà fondamentali delle misure di probabilità.

- L'evento complementare/Almeno uno

	1) Se $A \subseteq B$, allora $P(B \setminus A) = P(B) - P(A)$.
"almeno uno"	In particolare, $P(A^c) = 1 - P(A)$, $P(\emptyset) = 0$.

- Inclusione/Esclusione

inclusione/ esclusione"	2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
	In particolare, se $A \cap B = \emptyset$, allora $P(A \cup B) = P(A) + P(B)$.

La sigma-additività è una proprietà fondamentale delle misure di probabilità che descrive come la probabilità di un insieme di eventi possa essere calcolata come la somma delle probabilità dei singoli eventi che compongono l'insieme. In altre parole, se A e B sono eventi disgiunti (ovvero, non hanno alcun evento in comune), allora la probabilità dell'insieme unione $A \cup B$ è data dalla somma delle probabilità degli eventi A e B: $P(A \cup B) = P(A) + P(B)$.

Questa proprietà è nota come sigma-additività perché utilizza la notazione sigma (Σ) per descrivere la somma delle probabilità. Da questa discende la successiva:

3) Se $(A_n)_{n \in \mathbb{N}} \in \mathcal{F}$, allora

" σ -sub-additività"

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} P(A_n).$$

In particolare (\rightarrow assioma dell'additività σ),

se $A_i \cap A_j = \emptyset$ per $i \neq j$, allora $P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n)$.

La formula delle probabilità totali afferma che la probabilità di un evento A è data dalla somma delle probabilità degli eventi che compongono A, soggette alle condizioni date. Può essere espressa come:

$$P(A) = \sum P(A | B) * P(B)$$

dove B è un insieme di eventi mutuamente esclusivi che coprono l'universo e $P(A | B)$ è la probabilità condizionale di A soggetta alle condizioni date da B. In altre parole, questa formula calcola la probabilità di un evento A come la somma delle probabilità dei singoli eventi che compongono A soggette a diverse condizioni B.

4) Sia $(A_n)_{n \in \mathbb{N}} \in \mathcal{F}$ tale che $A_i \cap A_j = \emptyset$ per $i \neq j$

"probab. totali"

e $\bigcup_{n \in \mathbb{N}} A_n = \Omega$.

Allora $P(B) = \sum_{n \in \mathbb{N}} P(B \cap A_n)$

In particolare, $P(B) = P(B \cap A) + P(B \cap A^c)$.

Similmente, se la successione di eventi è crescente/decrescente lo è di conseguenza la loro convergenza in probabilità:

5) Sia $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{F}$.

a) Se (A_n) è crescente, cioè $A_n \subseteq A_{n+1} \forall n \in \mathbb{N}$, allora

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

b) Se (A_n) è decrescente, cioè $A_n \supseteq A_{n+1} \forall n \in \mathbb{N}$, allora

$$P\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Partendo dalla uniforme discreta:

\leadsto P è la misura di probab. definita su $\mathcal{P}(N)$ data da

$$P(A) = \frac{|A|}{|N|}, \quad A \subseteq N.$$

\leadsto calcolare la probab. di un evento A si riduce al conteggio del numero di elementi di A
 \rightarrow calcolo combinatorio

Applicazione: estrazioni di palline da un'urna

Esperimento: estrazione di n palline da un'urna con N palline di cui M rosse e $N-M$ verdi.

Osservazione di interesse: numero di palline rosse tra le n palline estratte.

Gli schemi con reinserimento e senza reinserimento sono due schemi utilizzati per generare campioni casuali da una distribuzione di probabilità.

Nello schema senza reinserimento, un elemento viene estratto dalla distribuzione di probabilità e viene rimosso dalla distribuzione per il resto del processo di campionamento. Questo significa che ogni elemento viene selezionato una sola volta. Questo schema è utile quando è necessario generare un insieme di eventi distinti.

Nello schema con reinserimento, un elemento viene estratto dalla distribuzione di probabilità e viene reinserito nella distribuzione prima di ogni successiva estrazione. Questo significa che un elemento può essere selezionato più di una volta. Questo schema è utile quando è necessario generare un insieme di eventi non distinti.

Partendo da una distribuzione uniforme discreta, ovvero una distribuzione in cui ogni elemento ha la stessa probabilità di essere estratto, lo schema senza reinserimento genera una distribuzione uniforme discreta ridotta, ovvero una distribuzione in cui ogni elemento ha una probabilità inferiore a quella originale a causa della rimozione di alcuni elementi dalla distribuzione durante il processo di campionamento.

Lo schema con reinserimento genera una distribuzione uniforme discreta modificata, ovvero una distribuzione in cui alcuni elementi hanno una probabilità maggiore rispetto alla distribuzione originale a causa della loro selezione più frequente durante il processo di campionamento.

Due schemi:
a) estrazioni con reinserimento
b) " " senza reinserimento.

Schemi \Rightarrow (con reinserimento):

Esempio pratico: conteggio di veicoli (due tipi diversi)

Modello: $\Omega \equiv \{1, \dots, N\}^n$

palline numerate da 1 a N , prime M palline rosse.

$P \equiv$ distribuzione uniforme

Eventi d'interesse: $A_K \equiv$ "esattamente K delle n palline estratte sono rosse", $K \in \{0, \dots, n\}$.

$$\begin{aligned} \leadsto A_K &= \left\{ \omega \in \Omega : \sum_{i=1}^n \mathbb{1}_{\{1, \dots, M\}}(\omega_i) = K \right\} \\ &= \left\{ \omega \in \Omega : \#\{i \in \{1, \dots, n\} : \omega_i \in \{1, \dots, M\}\} = K \right\}. \end{aligned}$$

$P(A_K) = ?$

$$\begin{matrix} P_{\text{uniforme}} \\ \leadsto \end{matrix} \quad P(A_K) = \frac{|A_K|}{|\Omega|}$$

Nota: $|\Omega| = |\{1, \dots, N\}^n| = N^n$.

Ad esempio: $N=7$ palline di cui $M=3$ rosse,
 $n=4$ estrazioni

$$\leadsto \Omega = \left\{ \underset{\cdot \cdot \cdot \cdot}{(1,1,1,1)}, \underset{\cdot \cdot \cdot \cdot}{(1,1,1,2)}, \dots, \underset{\cdot \cdot \cdot \cdot}{(7,7,7,7)} \right\}$$

La distribuzione binomiale descrive la probabilità di ottenere un certo numero di successi in una sequenza di n prove indipendenti, ognuna con la stessa probabilità p di successo.

Nel caso dell'estrazione di palline da un'urna nello schema con reinserimento, ogni estrazione rappresenta una prova indipendente e la distribuzione binomiale descrive la probabilità di ottenere un certo numero di palline di un certo colore in n estrazioni. La probabilità di successo in questo caso è pari alla frazione di palline del colore desiderato nell'urna.

La distribuzione binomiale può essere descritta come segue:

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

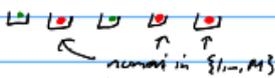
dove X è il numero di successi ottenuti in n prove, k è il numero specifico di successi, p è la probabilità di successo in ogni prova e $\binom{n}{k}$ è il coefficiente binomiale, che descrive il numero di modi in cui k successi possono essere ottenuti in n prove.

In questo esempio, la distribuzione binomiale descrive la probabilità di ottenere un certo numero di palline di un certo colore in n estrazioni da un'urna con reinserimento ed è una densità discreta ottenuta con questi passi:

a) scelta delle posizioni delle k palline rosse (quindi anche delle verdi): $\binom{n}{k}$ alternative

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

b) scelta delle identità delle palline rosse: $\rightarrow M^k$ alternative



c) scelta delle identità delle palline verdi: $\rightarrow (N-M)^{n-k}$ alternative

principio fondamentale

\rightsquigarrow del calcolo combinatorio $|A_k| = \binom{n}{k} \cdot M^k \cdot (N-M)^{n-k}$

$$|U| = N^n$$

$$\rightsquigarrow P(A_k) = \frac{|A_k|}{|U|} = \binom{n}{k} \cdot \left(\frac{M}{N}\right)^k \cdot \left(\frac{N-M}{N}\right)^{n-k}, \quad k \in \{0, 1, \dots, n\}$$

Nota: $\frac{M}{N}$ proporzione delle palline rosse nell'urna

$$\frac{N-M}{N} = 1 - \frac{M}{N} \text{ proporzione delle palline verdi.}$$

$$P(k) \equiv \binom{n}{k} \cdot \left(\frac{M}{N}\right)^k \cdot \left(\frac{N-M}{N}\right)^{n-k}, \quad k \in \{0, 1, \dots, n\}$$

definisce una densità discreta su $\{0, 1, \dots, n\}$ (o su \mathbb{R}).

(\rightarrow distribuzione binomiale).

Schemi b) ; Estrazioni senza reinserimento

Esempio : controllo di qualità

Spazio campionario $\mathcal{N} = \{1, \dots, N\}^n$ come prima,

$P =$ distribuzione uniforme su \mathcal{N} .

Poniamo $\tilde{\mathcal{N}} = \{ \omega \in \mathcal{N} : \omega : \{1, \dots, n\} \rightarrow \{1, \dots, N\} \text{ è iniettiva} \}$

$= \{ \omega \in \mathcal{N} : \omega_i \neq \omega_j \text{ per } i \neq j \}$

$\omega = (\omega_1, \dots, \omega_n)$

Definiamo una nuova misura di probab. \tilde{P} tramite

$$\tilde{P}(A) = \frac{P(A \cap \tilde{\mathcal{N}})}{P(\tilde{\mathcal{N}})}, \quad A \subseteq \mathcal{N}$$

$$\rightsquigarrow \tilde{P}(A) = \frac{|A \cap \tilde{\mathcal{N}}|}{|\tilde{\mathcal{N}}|}, \quad A \subseteq \mathcal{N}$$

Domanda come prima : $\tilde{P}(A_K) = ?$

\uparrow esattamente K palline rosse

Note: $A_K \cap \tilde{\mathcal{N}} = \{ \omega \in \tilde{\mathcal{N}} : \sum_{i=1}^n \mathbb{1}_{\{1, \dots, M\}}(\omega_i) = K \}$

La distribuzione ipergeometrica descrive la probabilità di ottenere un certo numero di successi in una sequenza di n prove indipendenti senza reinserimento, ognuna con la stessa probabilità p di successo.

Nel caso dell'estrazione di palline da un'urna nello schema senza reinserimento, ogni estrazione rappresenta una prova indipendente e la distribuzione ipergeometrica descrive la probabilità di ottenere un certo numero di palline di un certo colore in n estrazioni, con la condizione che una volta estratta, una pallina non venga reinserita nell'urna. La probabilità di successo in questo caso è pari alla frazione di palline del colore desiderato nell'urna.

La distribuzione ipergeometrica può essere descritta come segue:

$$P(X = k) = \binom{m}{k} \cdot \binom{N-m}{n-k} / \binom{N}{n}$$

dove X è il numero di successi ottenuti in n prove, k è il numero specifico di successi, m è il numero di successi possibili nella popolazione, N è il numero totale di elementi nella popolazione, $\binom{m}{k}$ è il coefficiente binomiale che descrive il numero di modi in cui k successi possono essere ottenuti da m elementi e $\binom{N-m}{n-k}$ è il coefficiente binomiale che descrive il numero di modi in cui $n-k$ fallimenti possono essere ottenuti da $N-m$ elementi.

In questo esempio, la distribuzione ipergeometrica descrive la probabilità di ottenere un certo numero di palline di un certo colore in n estrazioni da un'urna senza reinserimento.

Da calcolare: $|\tilde{\Omega}|$, $|A_k \cap \tilde{\Omega}|$

$$|\tilde{\Omega}| = N \cdot (N-1) \cdot \dots \cdot (N-n+1) = \frac{N!}{(N-n)!}$$

Gli elementi $A_k \cap \tilde{\Omega}$ sono determinati attraverso tre scelte successive:

a) scelta delle posizioni delle palline rosse (quindi anche di quelle verdi):

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

b) scelta delle identità delle k palline rosse:

$$\frac{M!}{(M-k)!} \quad \text{2 alternative}$$

c) scelta delle identità delle $n-k$ palline verdi:

$$\frac{(N-M)!}{(N-M-(n-k))!}$$

$$\leadsto |A_k \cap \tilde{\mathcal{A}}| = \binom{n}{k} \cdot \frac{M!}{(M-k)!} \cdot \frac{(N-M)!}{(N-M-(n-k))!}$$

$$\begin{aligned} \leadsto \tilde{P}(A_k) &= \frac{(N-n)! \cdot n! \cdot M! \cdot (N-M)!}{N! \cdot k! \cdot (n-k)! \cdot (M-k)! \cdot (N-M-(n-k))!} \\ &= \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}, \quad k \in \{0, 1, \dots, n, M\} \end{aligned}$$

↗ distribuzione ipergeometrica

Probabilità condizionali e indipendenza

Parliamo di probabilità condizionata, sapendo che in teoria della probabilità la probabilità condizionata di un evento A rispetto a un evento B è la probabilità che si verifichi A sapendo che B è verificato.

Sia (Ω, \mathcal{F}, P) uno spazio di probabilità.

Def.: Sia $B \in \mathcal{F}$ un evento con $P(B) > 0$.

Per $A \in \mathcal{F}$, la quantità

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

si dice probabilità condizionale di A dato B.

Esempio: Nella situazione di estrazioni di palline da un'urna **senza reinserimento**:
 (→ Lezione 5, pp. 10-11)

$$\mathcal{N} = \{1, \dots, N\}^n, \quad w = (w_1, \dots, w_n),$$

P distribuzione uniforme su \mathcal{N}

$$\tilde{\mathcal{N}} = \{w \in \mathcal{N} : w_i \neq w_j \text{ per } i \neq j\}$$

evento di estrarre n palline distinte.

Abbiamo posto

$$\tilde{P}(A) = \frac{P(A \cap \tilde{\mathcal{N}})}{P(\tilde{\mathcal{N}})}, \quad A \subseteq \mathcal{N}.$$

→ $\tilde{P}(A) = P(A | \tilde{\mathcal{N}})$ è la probabilità condizionale di A dato $\tilde{\mathcal{N}}$.

Le proprietà fondamentali delle probabilità condizionali sono:

- Positività: La probabilità condizionale di un evento A dato che l'evento B si è verificato deve sempre essere una quantità positiva o zero: $0 \leq P(A | B) \leq 1$.
- Normalizzazione: La probabilità condizionale di un evento dato che un altro evento si è verificato deve sempre essere normalizzata rispetto all'evento dato: $P(B) \neq 0 \Rightarrow P(A | B) = P(A \text{ and } B) / P(B)$.
- Definizione della probabilità totale: La probabilità totale di un evento A è data dalla somma delle probabilità condizionali di A dato che gli eventi B_1, B_2, \dots, B_n si sono verificati: $P(A) = \sum P(A | B_i) * P(B_i)$, dove B_i sono gli eventi esclusivi che coprono l'universo.
- l'intersezione degli eventi corrisponde a prodotto, per l'indipendenza. Utile nel caso di calcolo concreto. L'unica proprietà utile negli esercizi del nostro è la seguente, espressione sotto forma di produttoria che è una *famiglia*/insieme di eventi è indipendente se appunto esprimibile come produttoria. Questa è la *regola della catena*.
- Proprietà della somma: La probabilità della unione di due eventi A e B è data dalla somma delle probabilità condizionali dell'evento A dato che l'evento B si è verificato e dell'evento B dato che l'evento A si è verificato: $P(A \text{ or } B) = P(A | B) * P(B) + P(B | A) * P(A)$.

Queste proprietà sono importanti per l'analisi e la modellizzazione di situazioni complesse in probabilità e statistica, dove è necessario tenere conto dei fatti noti per calcolare la probabilità di eventi futuri.

1) Sia $B \in \mathcal{F}$ con $P(B) > 0$. Allora la mappa

$$\mathcal{F} \ni A \mapsto P(A|B) \in [0,1]$$

definisce una misura di probabilità su \mathcal{F}

$\leadsto (\Omega, \mathcal{F}, P(\cdot|B))$ è uno spazio di probabilità.

In particolare: $P(A^c|B) = 1 - P(A|B)$.

"regola della catena" / "regola di moltiplicazione"
 2) Siano $A_1, A_2, \dots, A_n \in \mathcal{F}$ eventi con $P(\bigcap_{i=1}^n A_i) > 0$. Allora

$$P(\bigcap_{i=1}^n A_i) = P(A_1) \cdot \prod_{i=2}^n P(A_i | A_1 \cap \dots \cap A_{i-1}).$$

In particolare, se $P(A_1 \cap A_2) > 0$, allora

!

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 | A_1).$$

"formula delle probabilità totali" (versione condizionata)
 3) Sia $\{B_i\}_{i \in I}$ una **partizione** di Ω più numerabile di Ω , cioè I è più numerabile, $B_i \cap B_j = \emptyset$ se $i \neq j$, e $\bigcup_{i \in I} B_i = \Omega$, tale che $B_i \in \mathcal{F}$ con $P(B_i) > 0$ per ogni $i \in I$.

Allora, per ogni $A \in \mathcal{F}$,

$$P(A) = \sum_{i \in I} P(A|B_i) \cdot P(B_i)$$

In particolare, se $B \in \mathcal{F}$ tale che $P(B) \in (0,1)$, allora

!

$$P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot (1 - P(B))$$

per ogni $A \in \mathcal{F}$.

La regola della catena della probabilità condizionale, anche nota come formula di Bayes, stabilisce una relazione tra la probabilità condizionale di un evento e la probabilità condizionale di un altro evento, dato che entrambi gli eventi sono verificati. La formula è:

$$P(A | B) = P(B | A) * P(A) / P(B)$$

dove $P(A | B)$ è la probabilità condizionale di A dato che B è verificato, $P(B | A)$ è la probabilità condizionale di B dato che A è verificato, $P(A)$ è la probabilità di A e $P(B)$ è la probabilità di B.

La regola della catena della probabilità condizionale è molto utile per la risoluzione di problemi in probabilità e statistica, in quanto consente di passare da una probabilità condizionale a un'altra utilizzando informazioni sulle probabilità a priori e sulle probabilità condizionali. Ad esempio, può essere utilizzato per calcolare la probabilità di una diagnosi medica dato un sintomo, tenendo conto della probabilità a priori della diagnosi e della probabilità condizionale del sintomo dato che la diagnosi sia vera.

Teorema: *formula di Bayes*

Siano $A, B \in \mathcal{F}$ eventi in (Ω, \mathcal{F}, P) con $P(A) > 0, P(B) > 0$. Allora

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

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

Esempio: Test clinico

Test per determinare la presenza di un virus negli individui di una popolazione

Dati sulla qualità del test:

a) sensibilità del test: probabilità che il test dia risultato positivo se applicato a un individuo portatore del virus
 $\hat{=} P_V$ "veri positivi"

b) specificità del test: probabilità che il test dia risultato positivo se il virus è assente nell'individuo
 $\hat{=} P_F$ "falsi positivi"

Risposta senza specificare lo spazio di probabilità:

Eventi di interesse:

$A =$ "l'individuo è portatore del virus"

$B =$ "il test dà risultato positivo"

Domanda corrisponde a calcolare $P(A|B)$

Dati sulla qualità del test:

$$P_V = P(B|A), \quad P_F = P(B|A^c)$$

Utilizziamo la formula di Bayes (e il corollario):

$$\begin{aligned}
 P(A|B) &= \frac{P(A)}{P(B)} \cdot P(B|A) \\
 &= \frac{P(A) \cdot \overset{=P_V}{P(B|A)}}{\underbrace{P(B|A) \cdot P(A)}_{=P_V} + \underbrace{P(B|A^c) \cdot (1-P(A))}_{=P_F}} \\
 \text{qui} &= \frac{P(A) \cdot P_V}{P(A) \cdot P_V + (1-P(A)) \cdot P_F}
 \end{aligned}$$

Dato mancante: $P(A)$ è l'incidenza del virus nella popolazione

Supponiamo che $P(A)$ sia la probabilità a priori che un individuo sia portatore del virus, $P(B | A)$ sia la sensibilità del test, ovvero la probabilità che il test dia risultato positivo dato che l'individuo è effettivamente portatore del virus, e $P(B | \sim A)$ sia la falsa positività del test, ovvero la probabilità che il test dia risultato positivo dato che l'individuo non è portatore del virus.

Allora, possiamo utilizzare la regola della catena della probabilità condizionale per calcolare $P(A | B)$:

$$P(A | B) = P(B | A) * P(A) / (P(B | A) * P(A) + P(B | \sim A) * P(\sim A))$$

dove $P(\sim A)$ è la probabilità a priori che un individuo non sia portatore del virus, ovvero $1 - P(A)$.

In altre parole, $P(A | B)$ è la probabilità che un individuo sia portatore del virus dato che il test ha dato risultato positivo, tenendo conto sia della probabilità a priori che l'individuo sia portatore del virus, sia della sensibilità e della falsa positività del test.

Poniamo $q \doteq P(A)$ incidenza del virus

$$\leadsto P(A|B) = \frac{q \cdot P_V}{q \cdot P_V + (1-q) \cdot P_F}$$

Se q è piccolo rispetto a 1 e anche a P_F

$$\leadsto P(A|B) \approx \frac{q \cdot P_V}{P_F}$$

Ad esempio: $P_V = 99\%$, $P_F = 0,5\%$

$$q = 4 \cdot 10^{-4} \quad \leadsto q \ll 1 \text{ e } q \ll P_F$$

$$\leadsto P(A|B) \approx \frac{4 \cdot 10^{-4} \cdot \frac{99}{100}}{\frac{1}{200}} = 8 \cdot 10^{-2} = 8\%$$

probabilità condizionata bassissima!

Indipendenza di due o più eventi

Due eventi sono detti indipendenti quando la probabilità di verificarsi di un evento non è influenzata dalla verifica o meno dell'altro evento. In altre parole, se conosciamo che un evento è accaduto, la probabilità che l'altro evento accada non cambia.

Def.: Siano A, B eventi in uno spazio di probabilità (Ω, \mathcal{F}, P) .

Allora

A, B si dicono indipendenti se

$$P(A \cap B) = P(A) \cdot P(B).$$

Osservazioni:

1) In generale, l'indipendenza di eventi dipende dalla misura di probab.

2) Se $P(A) \in \{0, 1\}$ (o $P(B) \in \{0, 1\}$) allora A, B sono indipendenti.

Infatti: Ad esempio: $P(A) = 0$. Allora $P(A \cap B) = 0$, ma anche $\underbrace{P(A)}_{=0} \cdot P(B) = 0$.

In particolare, se $A \in \{\emptyset, \Omega\}$, allora A, B indipendenti per ogni scelta di B .

3) Se $P(A) > 0$ e $P(B) > 0$ e $A \cap B = \emptyset$, allora A e B non indipendenti.

4) Se A, B sono indipendenti, allora lo sono anche A e B^c , A^c e B , e anche A^c e B^c .

Infatti: $P(A) = P(A \cap B) + P(A \cap B^c)$

Se A e B indipendenti, allora $P(A \cap B) = P(A) \cdot P(B)$

$$\leadsto P(A) = P(A) \cdot P(B) + P(A \cap B^c)$$

$$\leadsto P(A \cap B^c) = P(A) \cdot (1 - P(B)) = P(A) \cdot P(B^c)$$

$\leadsto A$ e B^c indipendenti.

5) Se $P(B) > 0$, allora:

A, B indipendenti se e solo se

$$P(A|B) = P(A).$$

In particolare, se A, B sono indipendenti e $P(B) > 0$, allora sapere se si è verificato o meno B non cambia la valutazione di A .

Questo è il senso in cui due eventi sono indipendenti.

Esempio: lancio di tre dadi regolari a sei facce

Osservazione di interesse: numeri segnati dai dadi

Modello: $\Omega = \{1, \dots, 6\}^3 \leadsto |\Omega| = 6^3 = 216$

$\mathcal{F} = \mathcal{P}(\Omega)$, $P =$ distribuzione uniforme su Ω

Consideriamo gli eventi:

$E_1 =$ "il primo dado segna 3"

$E_2 =$ "il secondo dado segna 2"

$E_3 =$ "il terzo dado segna 4"

$$\leadsto E_1 = \{\omega \in \Omega : \omega_1 = 3\}$$

$$E_2 = \{\omega \in \Omega : \omega_2 = 2\}$$

$$E_3 = \{\omega \in \Omega : \omega_3 = 4\}$$

Nota: $|E_1| = 6^2 = 36 = |E_2| = |E_3|$

P uniforme $\leadsto P(E_1) = \frac{|E_1|}{|\Omega|} = \frac{1}{6} = P(E_2) = P(E_3)$

Dall'altra parte, $P(E_1 \cap E_2) = \frac{|E_1 \cap E_2|}{|\Omega|} = \frac{1}{36}$

poiché $|E_1 \cap E_2| = 6$

Def.: Sia (Ω, \mathcal{F}, P) uno spazio di probab.

Siano $A_1, \dots, A_n \in \mathcal{F}$. Allora

A_1, \dots, A_n si dicono *indipendenti* come famiglie se per ogni scelta di $\emptyset \neq J \subseteq \{1, \dots, n\}$ si ha

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

Anzitutto si definisce l'indipendenza per famiglie arbitrarie:

Sia $(A_i)_{i \in I} \subseteq \mathcal{F}$ una famiglia di eventi.

$I \neq \emptyset$
arbitrario

Allora $(A_i)_{i \in I}$ si dice *indipendente* come famiglie se per ogni scelta di $J \subseteq I$ *finito* si ha che gli eventi $A_j, j \in J$, sono indipendenti come famiglie nel senso di prima.

Applicazione: prove ripetute e indipendenti

Supponiamo di poter ripetere una prova in "condizioni analoghe e indipendenti"; singola prova ha solo due esiti: o "successo" o "fallimento"; probabilità di successo rimane costante.

Esempio: lancio ripetuto di una moneta non necess. equilibrata

→ esito di una singola prova: "testa" o "croce"

Modello probabilistico per n prove ripetute e indipendenti

Def.: Sia $q \in [0,1]$, e sia $n \in \mathbb{N}$.

Uno spazio di probabilità (Ω, \mathcal{F}, P) con eventi C_1, \dots, C_n si dice un modello per n prove ripetute e indipendenti con probabilità di successo q se

C_1, \dots, C_n sono indipendenti come famiglie
e $P(C_i) = q$ per ogni $i \in \{1, \dots, n\}$.

Interpretazione: C_i è l'evento "i-esima prova ha successo."

Calcolo di alcune probabilità fondamentali nelle prove ripetute e indipendenti:

Siano $q \in [0,1]$, $n \in \mathbb{N}$. Sia (Ω, \mathcal{F}, P) con eventi C_1, \dots, C_n un modello per n prove ripetute e indipendenti con probabilità di successo q

1) Probab. che nessuna delle n prove abbia successo:

$$P\left(\bigcup_{i=1}^n C_i\right)^c = P\left(\bigcap_{i=1}^n C_i^c\right) = \prod_{i=1}^n P(C_i^c) = (1-q)^n.$$

2) Probab. che almeno una delle n prove abbia successo:

$$P\left(\bigcup_{i=1}^n C_i\right) = 1 - P\left(\left(\bigcup_{i=1}^n C_i\right)^c\right) \stackrel{1)}{=} 1 - (1-q)^n.$$

3) Probab. che il primo successo avvenga alla prova l -esima: $l \in \{1, \dots, n\}$

$$\begin{aligned} P(C_1^c \cap \dots \cap C_{l-1}^c \cap C_l) &= P(C_1^c) \cdot \dots \cdot P(C_{l-1}^c) \cdot P(C_l) \\ &= (1-q)^{l-1} \cdot q. \end{aligned}$$

4) Probab. che esattamente K delle n prove abbiano successo:

Riprendendo lo stesso schema fissato in un esempio sopra con le nuove conoscenze, possiamo caratterizzare la distribuzione binomiale. In essa le prove sono indipendenti e misura il numero di successi in un caso con reinserimento con prove con successo e senza successo. La binomiale approssima la ipergeometrica campionando un piccolo numero di valori.

$$\rightarrow p(K) \doteq \binom{n}{K} \cdot q^K \cdot (1-q)^{n-K}, \quad K \in \{0, \dots, n\},$$

definisce una densità discreta su $\{0, \dots, n\}$

↗ distribuzione binomiale

Variabili aleatorie

Le variabili aleatorie sono delle quantità matematiche che assegnano un valore numerico a ogni esito di un esperimento aleatorio. In altre parole, una variabile aleatoria è una funzione che mappa gli esiti di un esperimento aleatorio a dei valori numerici.

Le variabili aleatorie sono utili per descrivere e quantificare incertezza e randomicità di un esperimento. La distribuzione di probabilità di una variabile aleatoria descrive la probabilità che la variabile assuma un determinato valore o un intervallo di valori.

Ad esempio: lancio di tre dadi da gioco

X_1 numero segnato del primo dado

X_2 " " " " secondo "

X_3 " " " " terzo "

Descrizione spesso conveniente dell'esperimento.

Nell'esempio: $X_1 + X_2$ somme dei punteggi

$X_1 + X_2 + X_3$

→ operazioni aritmetiche

Definizione "pratica" di variabile aleatoria (v.z.)

Sia (Ω, \mathcal{F}, P) uno spazio di probabilità, e sia

E un insieme non-vuoto.

Una variabile aleatoria (su (Ω, \mathcal{F}, P)) a valori in E

è una funzione $\Omega \rightarrow E$.

Caso $E = \mathbb{R}$: Una v.z. a valori in \mathbb{R} si dice

v.z. reale o, semplicemente, variabile aleatoria.

Non tutte le funzioni $\Omega \rightarrow E$ sono anche delle variabili aleatorie.

Similmente vale il legame seguente.

Sia $B \in E$. L'anti-immagine di B sotto X è il sottoinsieme di \mathcal{N} dato da

$$X^{-1}(B) \doteq \{\omega \in \mathcal{N} : X(\omega) \in B\}.$$

Notazione: $\underbrace{\{X \in B\}}_{\text{sottoinsieme di } \mathcal{N}} \doteq X^{-1}(B) = \{\omega \in \mathcal{N} : X(\omega) \in B\}$

Torniamo al caso generale:

X v.z. su $(\mathcal{N}, \mathcal{F}, P)$ a valori in E .

Fatto: Se \mathcal{E} è una σ -algebra in E , allora

$$\sigma(X) \doteq \sigma_{\mathcal{E}}(X) \doteq \{X^{-1}(B) : B \in \mathcal{E}\}$$

è una σ -algebra in \mathcal{N} ,

è la σ -algebra generata da X (rispetto a \mathcal{E})

Nell'esempio dei tre dadi: $E = \mathbb{R}$, $\mathcal{E} = \mathcal{O}(\mathbb{R})$

$$\sigma(X_1) = \{A \times \{1, \dots, 6\}^2 : A \in \mathcal{O}(\mathbb{R})\}$$

\uparrow
 σ -algebra generata da X_1 (numero segnato dal primo dado)

Ricorda: $\mathcal{N} = \{1, \dots, 6\}^3$

Se $B \in \mathcal{R}$ con $B \cap \{1, \dots, 6\} = \emptyset$, allora $X_1^{-1}(B) = \emptyset$.

Se $B = \{1\}$, allora $X_1^{-1}(B) = \{\omega \in \mathcal{N} : \omega_1 = 1\}$
 $= \{1\} \times \{1, \dots, 6\}^2$

Se $B = \{2, 4, 6, 8, \dots\}$, allora $X_1^{-1}(B) = \{\omega \in \mathcal{N} : \omega_1 \in \{2, 4, 6\}\}$
 $= \{2, 4, 6\} \times \{1, \dots, 6\}^2$.

Le variabili aleatorie discrete sono delle variabili aleatorie che possono assumere solo un numero finito o contabile di valori. Questi valori sono distinti e limitati.

Ad esempio, la variabile aleatoria che descrive il risultato di un lancio di un dado è una variabile aleatoria discreta, poiché può assumere solo 6 valori distinti (1, 2, 3, 4, 5 o 6). Anche la variabile aleatoria che descrive il numero di successi in un certo numero di esperimenti indipendenti e identicamente distribuiti è una variabile aleatoria discreta.

La distribuzione di probabilità di una variabile aleatoria discreta è descritta dalla funzione di probabilità massima, che assegna una probabilità a ogni valore possibile della variabile. Questa funzione soddisfa la proprietà fondamentale della probabilità che la somma delle probabilità di tutti i valori possibili della variabile deve essere uguale a 1.

Sia X una v.z. su (Ω, \mathcal{F}, P) a valori in E .

L'immagine di X è il sottoinsieme di E dato da

$$\text{Im}(X) \equiv \{X(\omega) : \omega \in \Omega\} \subseteq E.$$

\rightarrow $\text{Im}(X)$ è l'insieme dei valori assunti da X .

Def.: Una v.z. X a valori in E si dice discreta se $\text{Im}(X)$ è al più numerabile.

In particolare, una v.z. discreta è una funzione $X: \Omega \rightarrow \mathbb{R}$ tale che $\text{Im}(X)$ è al più numerabile.

Esistono vari tipi di variabili aleatorie discrete:

- 1) Variabili aleatorie costanti: Sia $c \in \mathbb{R}$. Allora $X(\omega) \equiv c$ per ogni $\omega \in \Omega$ definisce una v.z. discreta con $\text{Im}(X) = \{c\}$.

2) Variabili aleatorie indicatorie:

Sia $A \in \mathcal{F}$ un evento. Allora

$$X(\omega) \doteq \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{se } \omega \in A, \\ 0 & \text{altrimenti,} \end{cases} \quad \omega \in \Omega,$$

definisce una v.z. discreta con $\text{Im}(X) = \{0, 1\}$
(e meno che A non sia \emptyset o Ω); in ogni
caso $\text{Im}(X) \subseteq \{0, 1\}$.

3) Prove ripetute ed indipendenti:

Siano C_1, \dots, C_n indipendenti ed equiprobabili (e $C_i \neq \emptyset$,
 $C_i \neq \Omega$)

Poniamo $X_i(\omega) \doteq \mathbb{1}_{C_i}(\omega)$,

$$S_n(\omega) \doteq \sum_{i=1}^n X_i(\omega), \quad \omega \in \Omega.$$

$$T_n(\omega) \doteq \inf \{ i \in \{1, \dots, n\} : X_i(\omega) = 1 \},$$

$\leadsto X_1, \dots, X_n, S_n, T_n$ v.z. discrete con

$$\text{Im}(X_i) = \{0, 1\}, \quad \text{Im}(S_n) = \{0, \dots, n\},$$

$$\text{Im}(T_n) = \{1, \dots, n\} \cup \{\infty\} \quad [\text{Convenzione: } \inf \emptyset = \infty]$$

Interpretazione: S_n numero di successi in n prove,

T_n indice ("tempo") del primo successo.

La densità di probabilità di una variabile aleatoria discreta è una funzione che descrive la probabilità che la variabile aleatoria assuma un determinato valore. La funzione di densità di probabilità discreta è una variante della funzione di probabilità massima, con la differenza che la funzione di densità è definita su un intervallo di valori anziché su un insieme di valori singoli. Questa funzione soddisfa la proprietà fondamentale della probabilità che la somma delle probabilità di tutti i valori possibili della variabile deve essere uguale a 1.

Def.: Sia X una v.e. discreta su (Ω, \mathcal{F}, P)
e valori in E . La funzione

$$p_X(x) \equiv P(X=x), \quad x \in E,$$

si dice densità discreta di X .

La distribuzione/legge di una variabile aleatoria descrive la probabilità che la variabile aleatoria assuma un certo valore. La distribuzione di una variabile aleatoria è rappresentata da una funzione di probabilità che può essere discreta o continua.

Se la variabile aleatoria è discreta, la distribuzione è descritta dalla sua funzione di densità di probabilità discreta.

La densità discreta determina la legge:

$$P_X(B) \equiv P(X \in B), \quad B \subseteq E$$

è la legge di X (o distribuzione di X)

Applicando gli esempi di cui sopra:

1) Variabili aleatorie costanti: Sia $c \in \mathbb{R}$

$$X(\omega) \equiv c, \quad \omega \in \Omega.$$

\leadsto densità discreta di X è data da

$$p_X(x) = \begin{cases} 1 & \text{se } x=c, \\ 0 & \text{altrimenti.} \end{cases}$$

\equiv
 $P(X=x)$

2) Variabili aleatorie indicatorie: Sia $A \in \mathcal{F}$.

$$X(\omega) \doteq \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{se } \omega \in A \\ 0 & \text{altrimenti} \end{cases}$$

$$\rightarrow P_X(x) = \begin{cases} P(A) & \text{se } x=1 \\ 1-P(A) & \text{se } x=0 \\ 0 & \text{altrimenti} \end{cases}$$

$$X^{-1}(\{1\}) = A, \quad X^{-1}(\{0\}) = A^c$$

Sul modello delle prove ripetute ed indipendenti, si può caratterizzare una variabile aleatoria di Bernoulli. Essa è una variabile aleatoria discreta che assume solo due valori, 0 o 1. Essa viene utilizzata per descrivere la probabilità di successo o fallimento in un singolo esperimento che può essere classificato come successo o fallimento. La probabilità di successo è denotata come p e la probabilità di fallimento come $(1 - p)$.

Ad esempio, se si lancia una moneta, la variabile aleatoria di Bernoulli potrebbe rappresentare l'esito del lancio, con successo rappresentato da "testa" e fallimento rappresentato da "croce". La probabilità di ottenere testa sarebbe $p = 0.5$, mentre la probabilità di ottenere croce sarebbe $(1 - p) = 0.5$.

Def.: Una v.z. Y con densità data da

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

Si dice v.z. di Bernoulli di parametro q .
In questo caso, si dice che Y ha distribuzione di Bernoulli di parametro q , in simboli

$$Y \sim \text{Ber}(q).$$

La sua densità discreta è data dal numero di successi e fallimenti in modo indipendente, quindi caratterizzata da una variabile aleatoria binomiale.

Densità discreta di S_n :

S_n è a valori in $\{0, \dots, n\}$ e dà il numero di successi in n prove ripetute e indipendenti

$$\leadsto P(S_n = k) = \binom{n}{k} \cdot q^k \cdot (1-q)^{n-k}, \quad k \in \{0, \dots, n\}$$

$$\leadsto P_{S_n}(x) = \begin{cases} \binom{n}{x} \cdot q^x \cdot (1-q)^{n-x} & \text{se } x \in \{0, \dots, n\}, \\ 0 & \text{altrimenti} \end{cases}$$

Def.: Una v.z. Y con densità data da

$$P_Y(x) = \begin{cases} \binom{n}{x} \cdot q^x \cdot (1-q)^{n-x} & \text{se } x \in \{0, \dots, n\}, \\ 0 & \text{altrimenti} \end{cases}$$

si dice v.z. binomiale di parametri $n \in \mathbb{N}$ e $q \in [0, 1]$.

In questo caso si dice che Y ha distribuzione binomiale di parametri n, q , in simboli

$$Y \sim \text{Bin}(n, q).$$

Considerando T_n l'indice del primo successo in n prove ripetute, avremo una caratterizzazione che considera successi/insuccessi ottenuti in un certo tempo.

Densità discreta di T_n :

T_n dà l'indice ("tempo"/"istante") del primo successo in n prove ripetute

$$\leadsto P(T_n = k) = (1-q)^{k-1} \cdot q \quad \text{se } k \in \{1, \dots, n\}$$

$$P(T_n = \infty) = (1-q)^n$$

$$\leadsto p_{T_n}(x) = \begin{cases} q \cdot (1-q)^{x-1} & \text{se } x \in \{1, \dots, n\}, \\ (1-q)^n & \text{se } x = \infty, \\ 0 & \text{altrimenti.} \end{cases}$$

Da qui caratterizziamo la variabile aleatoria geometrica, che descrive il numero di tentativi necessari per ottenere un successo in una sequenza di esperimenti indipendenti che hanno lo stesso risultato di successo o fallimento con la stessa probabilità. Ogni tentativo viene rappresentato come una variabile aleatoria di Bernoulli con una probabilità di successo p . La distribuzione geometrica modella il numero di tentativi necessari prima di ottenere il primo successo.

La distribuzione di T_n si avvicina, per $n \rightarrow \infty$ (numero di prove $\rightarrow \infty$), alla distribuzione geometrica:

Def.: Una v.z. Y con densità data da

$$p_Y(x) = \begin{cases} q \cdot (1-q)^{x-1} & \text{se } x \in \mathbb{N}, \\ 0 & \text{altrimenti,} \end{cases}$$

si dice v.z. geometrica di parametro $q \in (0, 1]$.
In questo caso, si dice che Y ha distribuzione geometrica di parametro q , in simboli

$$Y \sim \text{Geo}(q).$$

Variabili aleatorie discrete e valor medio

Ora possiamo descrivere l'insieme delle distribuzioni notevoli di variabili aleatorie discrete. Partiamo dalla definizione generale:

Sia X una v.z. discreta su (Ω, \mathcal{F}, P) con densità discreta p_X .

Ricorda: • $X: \Omega \rightarrow \mathbb{R}$ con $Im(X) \geq 1$ più numerabile (X : v.z. discreta),

• $p_X: \mathbb{R} \rightarrow [0,1]$ con $p_X(z) = P(X=z)$, $z \in \mathbb{R}$

Distribuzione di Bernoulli: La distribuzione di Bernoulli descrive un esperimento che ha solo due possibili risultati, successo o fallimento. Ad esempio, il lancio di una moneta. La probabilità di successo, p , è fissata e la probabilità di fallimento è $(1 - p)$. La distribuzione di Bernoulli può essere descritta da una funzione di probabilità $P(X = x) = p^x * (1-p)^{(1-x)}$, dove X è una variabile aleatoria che assume il valore 1 per successo e 0 per fallimento.

1) X ha distribuzione di Bernoulli di parametro $q \in [0,1]$
 se

$$p_X(z) = \begin{cases} q & \text{se } z=1, \\ 1-q & \text{se } z=0, \\ 0 & \text{altrimenti;} \end{cases}$$

in simboli, $X \sim \text{Ber}(q)$.

[$\rightarrow X$ dà l'esito di una singola prova]

Distribuzione di Rademacher: La distribuzione di Rademacher descrive una variabile aleatoria che assume solo i valori -1 o 1 con la stessa probabilità di 1/2. È spesso utilizzata come funzione di base in molte tecniche di modellizzazione della distribuzione di probabilità.

2) X ha distribuzione di Rademacher di parametro $q \in [0,1]$
se

$$P_X(z) = \begin{cases} q & \text{se } z=1, \\ 1-q & \text{se } z=-1, \\ 0 & \text{altrimenti.} \end{cases}$$

In simboli, $X \sim \text{Rad}(q)$.

Distribuzione binomiale: La distribuzione binomiale descrive il numero di successi in n esperimenti indipendenti identici con la stessa probabilità di successo p . Ad esempio, il numero di volte che si ottiene un risultato positivo in n test clinici. La distribuzione binomiale può essere descritta da una funzione di probabilità $P(X=x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$, dove X è una variabile aleatoria che descrive il numero di successi, $\binom{n}{x}$ è il coefficiente binomiale e p è la probabilità di successo.

3) X ha distribuzione binomiale di parametri $n \in \mathbb{N}$, $q \in [0,1]$
se

$$P_X(z) = \begin{cases} \binom{n}{z} \cdot q^z \cdot (1-q)^{n-z} & \text{se } z \in \{0, \dots, n\}, \\ 0 & \text{altrimenti.} \end{cases}$$

In simboli, $X \sim \text{Bin}(n, q)$.

[↗ numero di successi in n prove ripetute e indipendenti]

Distribuzione ipergeometrica: La distribuzione ipergeometrica descrive il numero di successi in n estrazioni a caso senza reinserimento da un insieme di N elementi, di cui K sono successi. Ad esempio, il numero di palline giuste estratte da un'urna contenente sia palline giuste che sbagliate. La distribuzione ipergeometrica può essere descritta da una funzione di probabilità $P(X = x) = \frac{\binom{K}{x} \cdot \binom{N-K}{n-x}}{\binom{N}{n}}$, dove X è una variabile aleatoria che descrive il numero di successi, $\binom{K}{x}$ è il coefficiente binomiale e $\binom{N}{n}$ è il numero di modi per scegliere n elementi da N .

4) X ha distribuzione ipergeometrica di parametri
 $N \in \mathbb{N}$, $M \in \{0, \dots, N\}$, $n \in \{1, \dots, N\}$ se

$$P_X(z) = \begin{cases} \frac{\binom{M}{z} \cdot \binom{N-M}{n-z}}{\binom{N}{n}} & \text{se } z \in \{0, \dots, n \wedge M\} \\ 0 & \text{altrimenti.} \end{cases}$$

minimo
↓

In simboli, $X \sim \text{Iper}(N, M, n)$.

[↗ numero di palline rosse in n estrazioni *senza reinserimento* da un'urna contenente N palline di cui esattamente M rosse]

Distribuzione geometrica: La distribuzione geometrica descrive il numero di tentativi necessari per ottenere il primo successo in una sequenza di esperimenti indipendenti con la stessa probabilità di successo p . Ad esempio, il numero di volte che si deve lanciare una moneta prima di ottenere il primo risultato positivo. La distribuzione geometrica può essere descritta da una funzione di probabilità $P(X = x) = p \cdot (1-p)^{x-1}$, dove X è una variabile aleatoria che descrive il numero di tentativi e p è la probabilità di successo.

5) X ha distribuzione geometrica di parametro $q \in [0, 1]$
 se

$$P_X(z) = \begin{cases} q \cdot (1-q)^{z-1} & \text{se } z \in \mathbb{N}, \\ 0 & \text{altrimenti.} \end{cases}$$

In simboli, $X \sim \text{Geo}(q)$.

[↗ prova di primo successo in infinite prove ripetute e indipendenti]

Distribuzione di Poisson: La distribuzione di Poisson descrive il numero di eventi che accadono in un intervallo di tempo o in una regione specifica con una media di successo λ , in particolare modo sapendo che avvengono in un determinato intervallo di tempo o di spazio. Ad esempio, il numero di chiamate ricevute in un call center in un determinato intervallo di tempo. La distribuzione di Poisson può essere descritta da una funzione di probabilità $P(X = x) = (e^{-\lambda} * \lambda^x) / x!$, dove X è una variabile aleatoria che descrive il numero di successi.

6) X ha distribuzione di Poisson di parametro $\lambda > 0$
se

$$P_X(z) = \begin{cases} e^{-\lambda} \cdot \frac{\lambda^z}{z!} & \text{se } z \in \mathbb{N}_0, \\ 0 & \text{altrimenti.} \end{cases}$$

In simboli, $X \sim \text{Poiss}(\lambda)$.

[\nearrow numero di "arrivi" (di richieste, clienti, ...) in un determinato intervallo temporale; λ intensità]

Distribuzione uniforme discreta: La distribuzione uniforme discreta descrive la distribuzione di una variabile aleatoria che assume valori interi compresi tra un valore minimo e un valore massimo, con una probabilità uguale per ogni valore possibile. La distribuzione uniforme discreta è spesso usata per descrivere la distribuzione del risultato di un lancio di un dado o di un lancio di una moneta. La funzione di probabilità di una variabile aleatoria uniforme discreta è data da: $P(x) = 1 / (b - a + 1)$ se x è compreso tra a e b , altrimenti è 0.

7) X ha distribuzione uniforme discreta su $A \subset \mathbb{R}$ finito
se

$$P_X(z) = \begin{cases} \frac{1}{|A|} & \text{se } z \in A, \\ 0 & \text{altrimenti} \end{cases}$$

In simboli, $X \sim \text{Unif}(A)$.

Parliamo ora del valor medio per variabili aleatorie discrete. Esso è la somma di un valore moltiplicato per la probabilità che appaia quel valore stesso. Il tutto sotto forma di serie. Si calcola come il prodotto del valore di X e la probabilità che X assuma quel valore, sommati per tutti i valori possibili che X può assumere. Il valore medio è una misura di posizione che descrive la "tendenza centrale" di una distribuzione di probabilità.

Def.: Sia X una v.z. reale discreta su (Ω, \mathcal{F}, P) con densità discreta p_X . Si dice che X ammette valor medio finito (o valor atteso finito) se

$$\sum_{z \in \mathbb{R}} |z| \cdot p_X(z) < \infty.$$

In questo caso, il valor medio (valor atteso) di X è dato da

$$E[X] = \sum_{z \in \mathbb{R}} z \cdot p_X(z).$$

Il valore medio di una variabile aleatoria dipende solo dalla distribuzione delle variabili aleatorie; ciò significa che la media non dipende dalla realizzazione specifica della variabile aleatoria, ma dipende solo dalla forma della sua distribuzione. La distribuzione descrive la probabilità di ogni valore che la variabile aleatoria può assumere e, di conseguenza, descrive anche la media attesa del valore.

Ad esempio, se abbiamo una variabile aleatoria X con una distribuzione uniforme, sappiamo che la media è uguale al centro dell'intervallo uniforme, indipendentemente da quale valore specifico X assume. La media dipende solo dalla forma della distribuzione e non dalla realizzazione specifica della variabile aleatoria.

Esempi di calcolo:

1) X una v.z. quasi certamente costante, cioè esiste $c \in \mathbb{R}$ tale che $P(X=c) = 1$.

$$\leadsto p_X(z) = \begin{cases} 1 & \text{se } z=c \\ 0 & \text{altrimenti} \end{cases}$$

$$\leadsto E[X] = \sum_{z \in \mathbb{R}} z \cdot p_X(z) = c \cdot 1 = c$$

! In particolare, $E[c] = c$ per ogni $c \in \mathbb{R}$.

2) Sia $X = \mathbb{1}_A$ per un evento $A \in \mathcal{F}$.

$$\leadsto P_X(z) = \begin{cases} P(A) & \text{se } z=1 \\ 1-P(A) & \text{se } z=0 \\ 0 & \text{altrimenti} \end{cases}$$

$$[\text{Qui: } \{X=1\} = \{\omega \in \Omega: \mathbb{1}_A(\omega)=1\} = A]$$

$$\leadsto E[X] = 1 \cdot P(A) + 0 \cdot (1-P(A)) = P(A)$$

! $\leadsto E[\mathbb{1}_A] = P(A)$.

3) Sia $X \sim \text{Ber}(q)$, $q \in [0,1]$.

$$\leadsto P_X(z) = \begin{cases} q & \text{se } z=1, \\ 1-q & \text{se } z=0, \\ 0 & \text{altrimenti} \end{cases}$$

$$\leadsto E[X] = 1 \cdot q + 0 \cdot (1-q) = q$$

Nota: Se $X = \mathbb{1}_A$ per un evento $A \in \mathcal{F}$,
allora

$$X \sim \text{Ber}(P(A)).$$

4) Sia $X \sim \text{Unif}(\{1, \dots, 6\})$.

$$\leadsto P_X(z) = \begin{cases} \frac{1}{6} & \text{se } z \in \{1, \dots, 6\}, \\ 0 & \text{altrimenti} \end{cases}$$

$$\begin{aligned} \leadsto E[X] &= \sum_{i=1}^6 i \cdot \frac{1}{6} = \frac{1}{6} \left(\sum_{i=1}^6 i \right) \\ &= \frac{21}{6} = \frac{7}{2} = 3,5 \end{aligned}$$

Il valor medio ha una serie di proprietà:

- Linearità: Se X e Y sono variabili aleatorie e a e b sono costanti, allora $E(aX + bY) = aE(X) + bE(Y)$.
- Esistenza: Se X è una variabile aleatoria definita su un insieme finito o numerabile di possibili risultati, allora esiste un valore medio.
- Invarianza sotto trasformazione: Se X è una variabile aleatoria e $g(X)$ è una funzione continua e invertibile, allora $E(g(X)) = g(E(X))$.
- Additività: Se X e Y sono variabili aleatorie indipendenti, allora $E(X + Y) = E(X) + E(Y)$.

Siano X, Y v.z. reali discrete su (Ω, \mathcal{F}, P) tali che X, Y ammettono valor medio finito. Allora:

1) Monotoniz: Se $X \leq Y$ P-q.c. (cioè $P(X \leq Y) = 1$), allora $E[X] \leq E[Y]$.

2) Stima fondamentale: $|E[X]| \leq E[|X|]$.

3) Linearità: Per ogni $\alpha, \beta \in \mathbb{R}$:

$$E[\alpha \cdot X + \beta \cdot Y] = \alpha \cdot E[X] + \beta \cdot E[Y].$$

Esempio:

Si è $X = \sum_{i=1}^n \mathbb{1}_{A_i}$ con $A_1, \dots, A_n \in \mathcal{F}$ eventi

Allora

$$\begin{aligned} E[X] &= E\left[\sum_{i=1}^n \mathbb{1}_{A_i}\right] && | \text{linearità} \\ &= \sum_{i=1}^n E[\mathbb{1}_{A_i}] && | \nearrow 2) \text{ sopra} \\ &= \sum_{i=1}^n P(A_i) \end{aligned}$$

Se gli eventi A_1, \dots, A_n sono equiprobabili con $P(A_i) = q$, allora

$$\sum_{i=1}^n P(A_i) = n \cdot q.$$

Se gli eventi A_1, \dots, A_n sono indipendenti ed equiprobabili, allora

$$X = \sum_{i=1}^n \mathbb{1}_{A_i} \sim \text{Bin}(n, q)$$

Siccome il valor medio di una v.z. dipende solo dalla sua distribuzione, trovato il valor medio di una v.z. binomiale:

Se $X \sim \text{Bin}(n, q)$, allora $E[X] = n \cdot q$.

La formula di trasformazione del valor medio nel caso discreto per una variabile aleatoria X e una funzione f è data da $E(f(X)) = \sum f(x) \cdot P(X = x)$, dove x è un valore possibile che la variabile aleatoria X può assumere e $P(X = x)$ è la probabilità che X assuma quel valore. Questa formula è utilizzata per calcolare il valore atteso o il valor medio di una funzione di una variabile aleatoria.

Formula di trasformazione (caso discreto)

Sia X una v.z. reale discreta su $(\Omega, \mathcal{F}, \mathbb{P})$, e sia $g: \mathbb{R} \rightarrow \mathbb{R}$ una funzione.

Poniamo $Z = g(X)$, cioè $Z(\omega) = g(X(\omega))$, $\omega \in \Omega$.

Allora Z è una v.z. discreta con densità discreta data da

$$p_Z(z) = \sum_{\substack{x \in \mathbb{R}: \\ g(x)=z}} p_X(x), \quad z \in \mathbb{R},$$

dove p_X è la densità discreta di X .

Inoltre, Z ammette valor medio finito se e solo se

$$\sum_{x \in \mathbb{R}} |g(x)| \cdot p_X(x) < \infty.$$

In questo caso,

$$E[Z] = \sum_{x \in \mathbb{R}} g(x) \cdot p_X(x),$$

∇₀ quindi $E[g(X)] = \sum_{x \in \mathbb{R}} g(x) \cdot p_X(x) = \sum_{x \in \text{Im}(X)} g(x) \cdot p_X(x).$

formule di trasformazione

La varianza nel caso discreto di una variabile aleatoria X è una misura della dispersione dei valori attorno al valor medio. Può essere definita come la somma delle probabilità ponderate delle deviazioni delle variabili aleatorie dal loro valore medio, ovvero: $\text{Var}(X) = E((X - E(X))^2)$. In altre parole, la varianza descrive la distanza quadratica media tra ogni valore possibile che la variabile aleatoria X può assumere e il suo valor medio.

Altre quantità riassuntive importante, varianza

Def.1 Sia X una v.z. reale discreta su (Ω, \mathcal{F}, P) con densità discreta p_X .

La varianza di X (se esiste) è data da

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

Osservazioni:

1) La varianza di X esiste finita se e solo se X^2 ammette valor medio finito

2) Confronto con la varianza campionaria:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{con } \bar{x} \text{ media campionaria}$$

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

$$= E[X^2 - 2X \cdot E[X] + E[X]^2]$$

$$= E[X^2] - 2E[X \cdot E[X]] + E[E[X]^2]$$

$$= E[X^2] - 2E[X]^2 + E[X]^2$$

! $\leadsto \text{var}(X) = E[X^2] - E[X]^2$.

Grazie alle formule di trasformazione:

$$\text{var}(X) = E[(X - E[X])^2] = \sum_{x \in \mathbb{R}} (x - E[X])^2 \cdot p_X(x),$$

dove $E[X] = \sum_{x \in \mathbb{R}} x \cdot p_X(x).$

In alternativa:

$$\text{var}(X) = E[X^2] - E[X]^2 = \sum_{x \in \mathbb{R}} x^2 \cdot p_X(x) - \left(\sum_{x \in \mathbb{R}} x \cdot p_X(x) \right)^2.$$

Esempio: varianze di una v.v. di Bernoulli:

Sia $X \sim \text{Ber}(q)$, $q \in [0, 1]$.

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

$$\leadsto E[X] = 1 \cdot q + 0 \cdot (1-q) = q$$

$$E[X^2] = 1^2 \cdot q + 0^2 \cdot (1-q) = q$$

$$\leadsto \text{var}(X) = q - q^2 = q \cdot (1-q).$$

Nota: $[0, 1] \ni q \mapsto q \cdot (1-q)$ assume il suo massimo in $q = \frac{1}{2}$.

Due variabili aleatorie reali X e Y sono dette indipendenti se la probabilità di verificare un insieme di eventi associato a X non dipende dall'evento associato a Y e viceversa. Matematicamente, questo può essere espresso come $P(X \cap Y) = P(X) \cdot P(Y)$, dove \cap rappresenta l'intersezione.

In altre parole, l'indipendenza significa che l'occorrenza di un evento associato a X non influisce sulla probabilità di un evento associato a Y e viceversa. Questo è un concetto importante in probabilità e statistica, in quanto molte proprietà e teoremi si basano sulla presunzione di indipendenza delle variabili aleatorie, in senso di dominio sui reali.

Indipendenza per v.z. reali.

Def:

Siano X, Y v.z. reali su (Ω, \mathcal{F}, P) .

Allora X e Y si dicono indipendenti se per ogni scelta di $x, y \in \mathbb{R}$ si ha

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y).$$

Note: X e Y sono indipendenti se e solo se

gli eventi $\{X \leq x\}, \{Y \leq y\}$ sono indipendenti per ogni scelta di $x, y \in \mathbb{R}$.

Ciò vale equivalentemente per le variabili aleatorie discrete, sapendo che, in particolar modo la seconda conseguenza ha una caratterizzazione definita come legge/distribuzione congiunta, cioè la distribuzione delle probabilità che descrive la relazione tra due o più variabili aleatorie. Esso descrive la probabilità congiunta di un insieme di valori possibili per ciascuna delle variabili aleatorie.

Caratterizzazione equivalente per v.z. discrete:

Siano X, Y v.z. reali discrete su (Ω, \mathcal{F}, P)

con densità discrete p_X, p_Y .

Allora sono equivalenti le seguenti condizioni:

(i) X, Y indipendenti;

Ricorda:

$$p_X(x) = P(X=x)$$

(ii) $P(X=x, Y=y) = p_X(x) \cdot p_Y(y)$ per ogni $x, y \in \mathbb{R}$;

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

per ogni scelte di $A, B \in \mathcal{R}$. //

Supponiamo di avere due variabili aleatorie discrete X e Y, con distribuzioni marginali date da:

$$P(X=1) = 0.3$$

$$P(X=2) = 0.4$$

$$P(X=3) = 0.3$$

e

$$P(Y=1) = 0.2$$

$$P(Y=2) = 0.5$$

$$P(Y=3) = 0.3$$

La distribuzione congiunta di X e Y è data da $P(X=x, Y=y)$ per ogni coppia di valori (x, y) . Ad esempio, per calcolare $P(X=2, Y=3)$, possiamo usare la distribuzione congiunta:

$$P(X=2, Y=3) = 0.1$$

Questo significa che la probabilità che $X=2$ e $Y=3$ contemporaneamente è 0.1.

Nel calcolo della distribuzione congiunta ho effettuato una moltiplicazione tra le probabilità condizionali, che descrivono le probabilità della comparsa di due eventi A e B in relazione tra loro, e le probabilità marginali dei singoli eventi. Questo mi ha permesso di trovare la probabilità congiunta, ovvero la probabilità che gli eventi A e B accadano insieme.

Dalla congiunta si può anche esprimere la distribuzione marginale, d'altra parte, descrive la distribuzione di probabilità per una sola variabile aleatoria, ignorando le relazioni tra le variabili. Ciò viene ottenuto sommando o integrando la distribuzione congiunta su tutti i valori possibili delle altre variabili. In altre parole, la distribuzione marginale descrive la distribuzione delle probabilità per una variabile aleatoria, considerando tutte le possibili combinazioni di valori per le altre variabili.

Supponiamo di avere due variabili aleatorie discrete X e Y e di conoscere la loro distribuzione congiunta $P(X=x, Y=y)$. Il calcolo della distribuzione marginale di X consiste nel sommare la distribuzione congiunta su tutti i possibili valori di Y:

$$P(X=x) = \sum P(X=x, Y=y)$$

Il calcolo della distribuzione marginale di Y è analogo.

Un esempio pratico potrebbe essere quello di considerare X e Y come le variabili che descrivono i risultati dei lanci di due dadi diversi. La distribuzione congiunta $P(X=x, Y=y)$ descrive la probabilità che il risultato del primo dado sia x e quello del secondo dado sia y, mentre la distribuzione marginale di X descrive la probabilità che il risultato del primo dado sia x, indipendentemente dal risultato del secondo dado. Allo stesso modo, la distribuzione marginale di Y descrive la probabilità che il risultato del secondo dado sia y, indipendentemente dal risultato del primo dado.

Qualche volta si può parlare di momento; fondamentalmente si tratta di calcolare il valore atteso per una serie di potenze (momento secondo è il valor atteso per la potenza di ordine 2, il momento terzo è il valore atteso per la potenza di ordine 3 e così via). Nel discreto e continuo si caratterizza così:

- **Momento (ordinario) r -esimo** di una variabile aleatoria unidimensionale X

$$\mu_r = E(X^r)$$

– **Caso discreto:**

$$\mu_r = \sum_{t \in R_X} t^r p_X(t)$$

– **Caso continuo:**

$$\mu_r = \int_{\mathbb{R}} t^r f_X(t) dt$$

Valor medio del prodotto di v.z. discrete indipendenti

Siano X, Y v.z. reali discrete su (Ω, \mathcal{F}, P) con valor medio finito.

Se X, Y sono indipendenti, allora

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

Attenzione: Non vale in generale quando X, Y non indipendenti!

Applicazione: varianza di una v.z. binomiale

$$\text{Sia } X \doteq \sum_{i=1}^n I_{A_i}, \text{ dove}$$

A_1, \dots, A_n sono eventi indipendenti ed equiprobabili
con $P(A_i) = q$.

$$\leadsto X \sim \text{Bin}(n, q).$$

$$\text{Nota: } I_{A_i} \sim \text{Ber}(q) \quad \leadsto E[I_{A_i}] = q,$$

$$\text{var}(I_{A_i}) = q \cdot (1 - q).$$

Ora per linearità,

$$E[X] = E\left[\sum_{i=1}^n I_{A_i}\right] = \sum_{i=1}^n E[I_{A_i}] = n \cdot q.$$

Per la varianza di X :

$$E[X] = n \cdot q$$

$$\text{var}(X) = E\left[\left(\sum_{i=1}^n I_{A_i} - n \cdot q\right)^2\right]$$

$$= E\left[\left(\sum_{i=1}^n (I_{A_i} - q)\right)^2\right]$$

$$= E\left[\sum_{i=1}^n (I_{A_i} - q)^2 + \sum_{i \neq j} (I_{A_i} - q) \cdot (I_{A_j} - q)\right]$$

linearità

$$= \sum_{i=1}^n E[(I_{A_i} - q)^2] + \sum_{i \neq j} E[(I_{A_i} - q) \cdot (I_{A_j} - q)]$$

$$= \text{var}(I_{A_i})$$

indipendenza

$$= E[I_{A_i} - q] \cdot E[I_{A_j} - q]$$

$$| E[I_{A_i} - q] = E[I_{A_i}] - q = 0$$

$$\leadsto \text{var}(X) = \sum_{i=1}^n \text{var}(I_{A_i})$$

$$= \sum_{i=1}^n q \cdot (1 - q)$$

$$\leadsto \text{var}(X) = n \cdot q \cdot (1 - q) \quad \text{se } X \sim \text{Bin}(n, q).$$

In alternativa: $\text{var}(X) = E[X^2] - E[X]^2$.

Abbiamo $E[X] = n \cdot q$

$$X = \sum_{i=1}^n I_{A_i}$$

Ora $E[X^2] = E\left[\left(\sum_{i=1}^n I_{A_i}\right)^2\right]$

$$= E\left[\sum_{i=1}^n I_{A_i}^2\right] + E\left[\sum_{i \neq j} I_{A_i} \cdot I_{A_j}\right]$$

$$= \underbrace{\sum_{i=1}^n E[I_{A_i}]}_{n \cdot q} + \sum_{i \neq j} \underbrace{E[I_{A_i} \cdot I_{A_j}]}_{\substack{\text{indir} \\ = E[I_{A_i}] \cdot E[I_{A_j}]} \\ = q^2}$$

| A_1, \dots, A_n independ.

$$\leadsto E[X^2] = n \cdot q + n \cdot (n-1) \cdot q^2$$

$$\leadsto \text{var}(X) = n \cdot q + n \cdot (n-1) \cdot q^2 - n^2 \cdot q^2$$

$$= n \cdot q - n \cdot q^2$$

$$= n \cdot q \cdot (1 - q) \quad \checkmark$$

Disuguaglianza di Markov-Chebyshev e approssimazione di Poisson

Partiamo dalla definizione della disuguaglianza di Markov, che afferma che per due variabili aleatorie X e Y , la varianza di X è sempre inferiore o uguale alla varianza di $X + Y$. In altre parole, la varianza di una somma di variabili aleatorie non può essere inferiore alla varianza di ciascuna variabile aleatoria che la compone. Questa disuguaglianza ha molte applicazioni, ad esempio in ottimizzazione, elaborazione di segnali, e modellizzazione di sistemi dinamici.

Teorema: disuguaglianza di Markov - Chebyshev

Sia X una v.z. reale su (Ω, \mathcal{F}, P) [vale tutto per v.z. reali generali]

1) Markov: Se $X \geq 0$, allora $\forall \varepsilon > 0$:

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

2) Markov generalizzato: Se $X \geq 0$ e $f: [0, \infty) \rightarrow [0, \infty)$ è crescente ($f(x) \leq f(y)$ se $x \leq y$) con $f(x) > 0$ se $x > 0$, allora

$$P(X \geq \varepsilon) \leq \frac{E[f(X)]}{f(\varepsilon)}$$

Avendo X una variabile non negativa, la disuguaglianza ci dice che finché X è positiva, allora la probabilità che X sia il doppio più grande del suo valore atteso è al più $\frac{1}{2}$ e, più in generale, la probabilità che una variabile aleatoria ottenga valori molto più grandi del suo valore atteso è piccola.

Vediamo poi la disuguaglianza di Chebyshev.

Essa afferma che una variabile aleatoria con varianza finita sia concentrata attorno al valore atteso. Più piccola la varianza, più grande la concentrazione. Fornisce un limite ben più forte di Markov.

Entrambe le disuguaglianze servono ad affermare che, nella maggior parte delle volte, le variabili aleatorie non ottengano valori "inattesi". La disuguaglianza di Chebyshev è un teorema matematico che fornisce un limite inferiore per la varianza di una variabile aleatoria rispetto alla distanza del suo valore medio.

3) Chebyshev: Se X ammette varianza, allora $\forall \varepsilon > 0$:

$$P(|X - E[X]| \geq \varepsilon) \leq \frac{\text{var}(X)}{\varepsilon^2}$$

La disuguaglianza di Chebyshev permette di stimare la probabilità di deviazioni dal valor medio in termini della deviazione standard:

Sia X una v.a. reale su (Ω, \mathcal{F}, P) con varianza $\text{var}(X) \in (0, \infty)$.

Poniamo $\sigma^2 \doteq \text{var}(X) \Rightarrow \sigma$ è la deviazione standard di X

Grazie alla disuguaglianza di Chebyshev, per $c > 0$,

$$P(|X - E[X]| \geq c \cdot \sigma) \leq \frac{\text{var}(X)}{c^2 \cdot \sigma^2} = \frac{1}{c^2}.$$

In particolare:

$c=1$ • Deviazioni dal valor medio fino $\pm 1 \cdot \sigma$ (una dev. std.) sono "tipiche" / da aspettarsi

• $c=2$: $P(|X - E[X]| \geq 2\sigma) \leq \frac{1}{4} = 25\%$

• $c=3$: $P(|X - E[X]| \geq 3\sigma) \leq \frac{1}{9} \approx 11\%$

• $c=5$: $P(|X - E[X]| \geq 5\sigma) \leq \frac{1}{25} = 4\%$

$\leadsto P(|X - E[X]| < 5\sigma) \geq \frac{24}{25} = 96\%$

Abbiamo una nozione importante, ossia la legge dei piccoli numeri, che va al di là del concetto di equiprobabilità e considera la dimensione del campione rispetto ai possibili eventi e conseguenti esiti. In particolare, a seguito di esperimenti ripetuti considerando un campione più piccolo, è molto più semplice allontanarsi dal valore atteso, banalmente perché avendo meno valori da considerare vi è più probabilità che essa si approssimi ad un certo valore, sottostimando il numero di campioni per stime accurate. Essa fu teorizzata da Kahneman.

Considerando una distribuzione binomiale, nel corso del tempo, possiamo approssimare asintoticamente, tramite un grande numero di osservazioni e con una probabilità di occorrenza estremamente piccola, il parallelismo con una Poisson.

Ricorda: Una v.z. binomiale di parametri n, q dà il numero di successi in n prove ripetute ed indipendenti con probab. di successo di una singola prova uguale a q .

Supponiamo di contare il numero di arrivi (di clienti / messaggi, pacchetti di dati, particelle, ...)
in un intervallo di tempo $[0, 1]$ (1 minuto, 1 ore, ...)

Dividiamo l'intervallo in n intervallini di uguale lunghezza:



con n grande

Sia $q_{n,k}$ la probab. di vedere un arrivo nell'intervallo k -esimo, cioè in $[\frac{k-1}{n}, \frac{k}{n}]$.

Se $S_n \sim \text{Bin}(n, q_n)$ e $q_n \approx \frac{1}{n}$, n grande,

allora S_n conta all'incirca il numero di arrivi in $[0, 1]$.

Cosa succede per $n \rightarrow \infty$?

Nota: Per ipotesi, $n \cdot q_n \xrightarrow{n \rightarrow \infty} \lambda$.

Teorema (legge dei piccoli numeri):

Sia $(q_n)_{n \in \mathbb{N}} \subset [0,1]$ tale che $n \cdot q_n \xrightarrow{n \rightarrow \infty} \lambda$
 per una costante $\lambda > 0$.

Sia p_n la densità discreta della $\text{Bin}(n, q_n)$,
 e sia \hat{p}_λ " " " " della $\text{Pois}(\lambda)$.

Ricorda:

$$p_n(k) = \begin{cases} \binom{n}{k} \cdot q_n^k \cdot (1-q_n)^{n-k} & \text{se } k \in \{0, \dots, n\}, \\ 0 & \text{altrimenti,} \end{cases}$$

$$\hat{p}_\lambda(k) = \begin{cases} e^{-\lambda} \cdot \frac{\lambda^k}{k!} & \text{se } k \in \mathbb{N}_0, \\ 0 & \text{altrimenti.} \end{cases}$$

Allora $\sum_{k=0}^{\infty} |p_n(k) - \hat{p}_\lambda(k)| \xrightarrow{n \rightarrow \infty} 0$

Applicazione: approssimazione di Poisson

Sia $S \sim \text{Bin}(n, q)$ con n "grande" e q "piccolo"

Allora S ha distribuzione vicina alla distribuzione di Poisson di parametro $\lambda = n \cdot q$.

Di conseguenza, per ogni $B \in \mathbb{R}$,

$$\begin{aligned} P(S \in B) &= \sum_{x \in B} p_S(x) && | \text{ } p_S \text{ densità discreta} \\ &\approx \sum_{x \in B} \hat{p}_\lambda(x) && \text{della } \text{Bin}(n, q) \\ &= P(Y \in B), && \left| \begin{array}{l} \lambda = n \cdot q, \hat{p}_\lambda \\ \text{densità discreta della Poiss}(\lambda) \end{array} \right. \end{aligned}$$

dove $Y \sim \text{Poiss}(\lambda)$, $\lambda = n \cdot q$.

In particolare ($B = (-\infty, x]$): $\forall x \in \mathbb{R}$:

$$P(S \leq x) \approx P(Y \leq x), \text{ dove } S \sim \text{Bin}(n, q), Y \sim \text{Poiss}(n \cdot q).$$

Vantaggi dell'approssimazione di Poisson:

- probab. in termini della distribuzione di Poisson più facili da calcolare risp. alla binomiale (n grande, q piccolo)
- ! • La distribuzione di Poisson determinata da un solo parametro, quella binomiale invece ne richiede due.

Vettori aleatori discreti

Definiamo un vettore aleatorio come una funzione definita sui reali fino a D dimensioni. Le componenti di questo vettore sono, chiaramente, variabili aleatorie reali. Occorre porre attenzione all'ordine delle componenti.

Def. "pratica":

Sia (Ω, \mathcal{F}, P) uno spazio di probabilità.

Un vettore aleatorio d -dimensionale è una funzione $\Omega \rightarrow \mathbb{R}^d$.

Sia X un vettore aleatorio d -dim. Allora

$X = (X_1, \dots, X_d)$ dove X_1, \dots, X_d sono funzioni $\Omega \rightarrow \mathbb{R}$

\leadsto Le componenti di X sono delle v.z. reali

Viceversa, se X_1, \dots, X_d sono v.z. reali, allora

$X(\omega) = (X_1(\omega), \dots, X_d(\omega))$, $\omega \in \Omega$,

definisce un vettore aleatorio d -dim.

Se l'insieme è al più numerabile, abbiamo un insieme "limitato" di numeri, quindi discreto. Esso è tale se e solo se è composto da variabili aleatorie discrete.

Def.: Sia X un vettore aleatorio d -dim. su (Ω, \mathcal{F}, P) .

Allora X si dice **discreto** se

$\text{Im}(X) = \{X(\omega) : \omega \in \Omega\}$ è al più numerabile.

La distribuzione di un vettore aleatorio discreto è determinata dalla sua densità discreta:

Sia X un vettore aleatorio discreto d -dim. su (Ω, \mathcal{F}, P) .
La densità discreta di X è la funzione $p_X: \mathbb{R}^d \rightarrow [0,1]$ data da

$$p_X(z) \doteq \underbrace{P(X_1=z_1, \dots, X_d=z_d)}_{= P(X=z)}, \quad z \in \mathbb{R}^d.$$

Qui approfondiamo meglio, attraverso questa generalizzazione, la densità discreta congiunta e le densità marginali.

Densità discrete congiunte e densità marginali

Def.: Siano X_1, \dots, X_d v.z. reali discrete su (Ω, \mathcal{F}, P) .
Poniamo $X \doteq (X_1, \dots, X_d)$.

La densità discreta del vettore aleatorio X si dice **densità discreta congiunta** di X_1, \dots, X_d (ordine!).

Le densità discrete delle componenti X_1, \dots, X_d si dicono **densità marginali** del vettore aleatorio X .

La densità discreta congiunta determina le densità discrete marginali:

$$X = (X_1, \dots, X_d)$$

$$p_{X_i}(x) = \sum_{\substack{z \in \mathbb{R}^d \\ z_i = x}} p_X(z) \quad x \in \mathbb{R}$$

↑
densità discreta della componente X_i

$$= \sum_{\substack{z_j \in \mathbb{R}, \\ j \neq i}} p_X(z_1, \dots, z_{i-1}, x, z_{i+1}, \dots, z_d)$$

Nota: basta sommare su $z_j \in \text{Im}(X_j)$, $j \neq i$.

In particolare: densità discreta di X_1
(prima densità marginale):

$$P_{X_1}(x) = \sum_{z_2 \in \text{Im}(X_2)} \dots \sum_{z_d \in \text{Im}(X_d)} P_X(x, z_2, \dots, z_d), \quad x \in \mathbb{R}$$

Esempio: Sia $X = (X_1, X_2)$ un vettore aleatorio bi-dim.
a valori in $\{1, \dots, 6\} \times \{1, \dots, 8\}$

$$\leadsto |\text{Im}(X)| = 48$$

Supponiamo che X abbia distribuzione uniforme su $\text{Im}(X)$:

$$P_X(z) = \begin{cases} \frac{1}{48} & \text{se } z \in \{1, \dots, 6\} \times \{1, \dots, 8\}, \\ 0 & \text{altrimenti,} \end{cases} \quad z \in \mathbb{R}^2.$$

Densità marginali:

$$\begin{aligned} P_{X_1}(x) &= \sum_{z_2 \in \mathbb{R}} P_X(x, z_2) \stackrel{\text{qui}}{=} \sum_{z_2 \in \{1, \dots, 8\}} \underbrace{P_X(x, z_2)}_{= \begin{cases} \frac{1}{48} & \text{se } x \in \{1, \dots, 6\}, \\ 0 & \text{altrimenti} \end{cases}} \\ &= \begin{cases} \frac{8}{48} & \text{se } x \in \{1, \dots, 6\} \\ 0 & \text{altrimenti} \end{cases} \end{aligned}$$

$\leadsto X_1$ ha distribuzione uniforme su $\{1, \dots, 6\}$.

Anzitutto, X_2 ha distribuzione uniforme su $\{1, \dots, 8\}$.

Attenzione: In generale, le densità marginali
non determinano la densità congiunta.

Sia $X = (X_1, X_2)$ un vettore aleatorio bi-dim. con densità discrete

$$P_X(z) = \begin{cases} \frac{1}{36} & \text{se } z \in \{1, \dots, 6\}^2 \\ 0 & \text{altrimenti} \end{cases}$$

come sopra

\leadsto densità (marginali) di X_1, X_2 :

$$X_1 \sim \text{Unif}(\{1, \dots, 6\}), \text{ ma anche } X_2 \sim \text{Unif}(\{1, \dots, 6\})$$

Poniamo $\hat{X} = (X_1, X_1)$ componenti uguali

$$\leadsto \hat{X}_1 = X_1 \sim \text{Unif}(\{1, \dots, 6\}), \hat{X}_2 = X_1 \sim \text{Unif}(\{1, \dots, 6\})$$

$\leadsto X, \hat{X}$ hanno le stesse distribuzioni marginali

Ma: densità di \hat{X} data da

$$\begin{aligned} P_{\hat{X}}(z) &= P(X_1 = z_1, X_1 = z_2) \\ &= \begin{cases} \frac{1}{6} & \text{se } z_1 = z_2 \in \{1, \dots, 6\}, \\ 0 & \text{altrimenti} \end{cases} \end{aligned}$$

$$\leadsto \hat{X} \sim \text{Unif}(\{(x, x) : x \in \{1, \dots, 6\}\}), \text{ mentre } X \sim \text{Unif}(\{1, \dots, 6\}^2)$$

$\leadsto \hat{X}$ e X non hanno la stessa densità (congiunta).

Le densità marginali descrivono la distribuzione di una singola variabile all'interno di un sistema a due o più variabili, ma non forniscono informazioni complete sulla distribuzione congiunta di tutte le variabili. Per ottenere la distribuzione congiunta è necessario considerare la relazione tra tutte le variabili.

Se si considerano solo le distribuzioni marginali di componenti X_1, \dots, X_n , si perde l'informazione sulle dipendenze fra le v.z.

Se le v.z. (componenti) sono indipendenti, allora la distribuzione congiunta è determinata dalle distribuzioni marginali.

Def.: Siano X_1, \dots, X_n v.z. reali su (Ω, \mathcal{F}, P) .

Allora X_1, \dots, X_n si dicono **indipendenti** (come famiglie) se per ogni scelta di $x_1, \dots, x_n \in \mathbb{R}$ si ha

$$(*) \quad P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i).$$

Per una famiglia di variabili aleatorie reali X e Y , si dice che X e Y sono indipendenti se e solo se la distribuzione congiunta di X e Y è uguale al prodotto delle loro distribuzioni marginali: $P(X, Y) = P(X) * P(Y)$. Questo significa che la probabilità di osservare un certo insieme di valori per X e un certo insieme di valori per Y non dipende dalla presenza dell'altra variabile. In altre parole, conoscendo il valore di X non influirà sulle probabilità di ottenere un certo valore di Y e viceversa.

In questo modo, possiamo dire che una famiglia di eventi indipendenti è una produttrice tra tutte le occorrenze. Se le variabili della successione sono discrete, allora il loro prodotto corrisponde alle densità discrete.

Se X_1, \dots, X_n sono v.z. discrete e p_X è la loro densità congiunta, allora

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = \sum_{\substack{z \in \mathbb{R}^n: \\ z_i \leq x_i, 1 \leq i \leq n}} p_X(z).$$

Osservazione importante:

Se X_1, \dots, X_n sono indipendenti (come famiglie),
 allora $X_j, j \in \mathcal{I}$, sono indipendenti per ogni scelta
 di $\mathcal{I} \subseteq \{1, \dots, n\}$ con $|\mathcal{I}| \geq 2$.

In particolare: X_1, \dots, X_n indipendenti
 implicano X_1, \dots, X_n indipendenti a due a due
 (cioè, X_i e X_j indipendenti per $i \neq j$).

Caratterizzazione equivalente per v.v. discrete.

Siano X_1, \dots, X_n v.v. reali discrete su (Ω, \mathcal{F}, P) .

Allora sono equivalenti:

(i) X_1, \dots, X_n indipendenti;

$$\begin{aligned} \text{(ii)} \quad P_{(X_1, \dots, X_n)}(z) &= \prod_{i=1}^n P_{X_i}(z_i) \quad \text{per ogni } z \in \mathbb{R}^n \\ &\stackrel{\#}{=} P(X_1=z_1, \dots, X_n=z_n) \quad = P(X_i=z_i) \\ &\text{densità congiunte} \qquad \qquad \text{densità marginali} \end{aligned}$$

$$\text{(iii)} \quad P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n P(X_i \in B_i)$$

per ogni scelte di $B_1, \dots, B_n \subseteq \mathbb{R}$.

Indipendenza, covarianza e correlazione

La covarianza tra due variabili aleatorie X e Y misura la loro relazione lineare. La covarianza è un valore che indica se le due variabili tendono a crescere o diminuire insieme (covarianza positiva) o se tendono a crescere in direzioni opposte (covarianza negativa). Il valore assoluto della covarianza non fornisce informazioni sul grado di correlazione tra le due variabili, ma solo sulla direzione del loro cambiamento congiunto.

Se X e Y sono indipendenti, la loro covarianza è pari a zero, poiché le informazioni su una variabile non influiscono sul valore dell'altra variabile. In altre parole, non esiste alcuna relazione lineare tra le due variabili e quindi la loro covarianza è nulla.

In pratica, la covarianza è importante nell'analisi dei dati perché fornisce informazioni sulle relazioni tra le variabili e può essere utilizzata nella modellizzazione per prevedere il comportamento delle variabili in futuro. Inoltre, la covarianza è un ingrediente fondamentale nella definizione di un altro importante concetto, la correlazione, che misura il grado di associazione tra due variabili.

Misura importante per quantificare la "dipendenza"
tra due v.z. reali: covarianza

Def.: Siano X, Y v.z. reali con momento secondo finito
(cioè $E[X^2] < \infty$, $E[Y^2] < \infty$).

La covarianza di X e Y è la quantità data da

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

Osservazioni:

La covarianza prende valori in \mathbb{R} ;

covarianza strettamente positiva indica che le due v.z. tendenzialmente superano insieme oppure stanno insieme sotto i rispettivi valori medi; al contrario per covarianza negativa

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

La covarianza dipende dalla distribuzione congiunta di X e Y .

In particolare, se X, Y sono v.z. reali discrete, allora

$$\text{cov}(X, Y) = \sum_{(x,y) \in \mathbb{R}^2} (x - E[X]) \cdot (y - E[Y]) \cdot P_{(X,Y)}(x, y),$$

dove $P_{(X,Y)}$ è la densità discreta congiunta di X, Y ,

mentre $E[X] = \sum_{z \in \mathbb{R}} z \cdot P_X(z)$ $E[Y] = \sum_{z \in \mathbb{R}} z \cdot P_Y(z)$

↑ densità discreta marginale ↑

Esempio (lancio di due dadi)

Siano Z_1, Z_2 v.z. indipendenti e identicamente distribuite con $Z_i \sim \text{Unif}(\{1, \dots, 6\})$. Poniamo

$$X = Z_1, \quad Y = Z_1 + Z_2.$$

$\text{cov}(X, Y) = ?$ Ci aspettiamo $\text{cov}(X, Y) > 0$.

Calcolare la densità discreta congiunta di X e Y :

	2	3	4	5	6	7	8	9	10	11	12
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	0	0	0	0
2	0	$\frac{1}{36}$	$\frac{1}{36}$	0	0
3	0	0	$\frac{1}{36}$	$\frac{1}{36}$	0	..	0
4	0	..	0	$\frac{1}{36}$	$\frac{1}{36}$	0	0
5	0	0	$\frac{1}{36}$	$\frac{1}{36}$	0
6	0	0	0	0	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

Esempio (cont.)

Ricorda: Se $Z_i \sim \text{Unif}(\{1, \dots, 6\})$ allora $E[Z_i] = \frac{7}{2}$

$$\leadsto E[X] = \frac{7}{2}, \quad E[Y] = E[Z_1] + E[Z_2] = 7$$

$$\leadsto \text{cov}(X, Y) = \sum_{x \in \{1, \dots, 6\}} \sum_{y \in \{2, \dots, 12\}} (x - \frac{7}{2}) \cdot (y - 7) \cdot P_{(X, Y)}(x, y)$$

$$= \frac{1}{36} \cdot \left(\begin{aligned} &(-\frac{5}{2}) \cdot (-5 + -4 + -3 + -2 + -1 + 0) \\ &+ (-\frac{3}{2}) \cdot (-4 + -3 + -2 + -1 + 0 + 1) \\ &+ (-\frac{1}{2}) \cdot (-3 + -2 + -1 + 0 + 1 + 2) \\ &+ (\frac{1}{2}) \cdot (-2 + -1 + 0 + 1 + 2 + 3) \\ &+ (\frac{3}{2}) \cdot (-1 + 0 + 1 + 2 + 3 + 4) \\ &+ (\frac{5}{2}) \cdot (0 + 1 + 2 + 3 + 4 + 5) \end{aligned} \right)$$

$$= \frac{35}{12}$$

La covarianza ha una serie di proprietà. Si noti che se le variabili sono indipendenti, la covarianza come detto sopra è pari a zero, dato che le informazioni su una variabile non influiscono sull'altra.

Proprietà della covarianza:

Siano X, Y, Z v.v. reali con momento secondo finito.

1) Simmetria: $\text{cov}(X, Y) = \text{cov}(Y, X)$

2) Bi-linearità: $\forall \alpha, \beta \in \mathbb{R}$:

$$\begin{aligned} \text{cov}(\alpha X + \beta Y, Z) &= \alpha \cdot \text{cov}(X, Z) + \beta \cdot \text{cov}(Y, Z) \\ &= \text{cov}(Z, \alpha X + \beta Y) \end{aligned}$$

3) $\text{var}(c \cdot X) = \text{cov}(c \cdot X, c \cdot X) = c^2 \cdot \text{var}(X)$ per ogni $c \in \mathbb{R}$.

! 4) Indipendenza: Se X e Y sono indipendenti, allora $\text{cov}(X, Y) = 0$.

Ad 4): Siano X, Y indipendenti. Allora $(X - E[X])$ e $(Y - E[Y])$ indipendenti

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

$$\stackrel{\text{indip.}}{=} \underbrace{E[X - E[X]]}_{=0} \cdot \underbrace{E[Y - E[Y]]}_{=0} = 0. //$$

Due v.z. X e Y con $\text{cov}(X, Y) = 0$ si dicono *incorrelate*.

! Attenzione: Se X e Y sono indipendenti, allora sono anche incorrelate.

L'implicazione inversa *non* vale in generale!

(Contro-) Esempio: Sia $Z \sim \text{Unif}(\{-1, 0, 1\})$

Postiamo

$$X = \begin{cases} 1 & \text{se } Z=0, \\ 0 & \text{altrimenti,} \end{cases} \quad Y = Z.$$

$$\leadsto E[X] = 1 \cdot \frac{1}{3} = \frac{1}{3}, \quad E[Y] = \frac{1}{3} \cdot (-1 + 0 + 1) = 0.$$

$$\leadsto \text{cov}(X, Y) = E[(X - \frac{1}{3}) \cdot Y]$$

$$= E[X \cdot Y] - \frac{1}{3} \cdot \underbrace{E[Y]}_{=0} = E[X \cdot Y]$$

$$= E[\mathbb{1}_{\{0\}}(Z) \cdot Z] = \frac{1}{3} \cdot (1 \cdot 0 + 0 \cdot (-1) + 0 \cdot 1) = 0.$$

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

M₂ X e Y **non** sono indipendenti! Infatti:

$$P(X=1, Y=1) = 0, \quad \text{mentre}$$

$$P(X=1) = \frac{1}{3} > 0, \quad P(Y=1) = \frac{1}{3} > 0$$

$$\leadsto P(X=1, Y=1) \neq P(X=1) \cdot P(Y=1).$$

Applicazione: passeggiate aleatorie semplici e simmetria

Siano ξ_1, ξ_2, \dots indipendenti e identicamente distribuite
con $\xi_i \sim \text{Rad}(\frac{1}{2})$.

Poniamo $S_0 \doteq 0$, $S_n \doteq \sum_{i=1}^n \xi_i$. $\xi_i \sim \text{Rad}(\frac{1}{2})$
 $\leadsto E[\xi_i] = 0$
 $\text{var}(\xi_i) = 1$

Interpretazioni:

a) S_n dà la posizione dopo n passi di una particella che parte in zero e ad ogni passo si sposta di una unità a destra o sinistra in maniera casuale (con probab. $\frac{1}{2}, \frac{1}{2}$).

b) S_n dà il guadagno dopo n giocate di un gioco "leale" in cui la posta in ogni giocata è di un'unità (1€, 1\$...).

Nota: $E[S_n] = \sum_{i=1}^n E[\xi_i] = 0$

Grazie all'indipendenza di ξ_1, \dots, ξ_n :

$$\text{var}(S_n) = \sum_{i=1}^n \text{var}(\xi_i) = n \cdot 1 = n$$

Dalla disuguaglianza di Markov-Chebyshev: $\forall \varepsilon > 0$

$$P(|S_n| \geq \varepsilon) \leq \frac{\text{var}(S_n)}{\varepsilon^2} = \frac{n}{\varepsilon^2}$$

\leadsto Valori di S_n in $[-\sqrt{n}, \sqrt{n}]$ "tipici".

Il coefficiente di correlazione è un valore che misura la relazione lineare tra due variabili aleatorie. Questo valore varia tra -1 e 1 e indica il grado di associazione tra le variabili:

- Un coefficiente di correlazione pari a 1 indica una correlazione perfettamente positiva, cioè le due variabili crescono o diminuiscono insieme in modo proporzionale.
- Un coefficiente di correlazione pari a -1 indica una correlazione perfettamente negativa, cioè le due variabili crescono o diminuiscono in direzioni opposte in modo proporzionale.
- Un coefficiente di correlazione pari a 0 indica che non esiste alcuna relazione lineare tra le due variabili.

Il coefficiente di correlazione è calcolato come il rapporto tra la covarianza delle due variabili e il prodotto delle loro deviazioni standard. Questo normalizza la covarianza, rendendola indipendente dalla scala di misura delle variabili, e la rende più facile da interpretare. In pratica, il coefficiente di correlazione è utilizzato in molte applicazioni, tra cui la previsione delle variabili, la modellizzazione dei dati, la verifica dell'ipotesi e la scoperta di relazioni tra le variabili.

Come per i campioni, possiamo definire il coefficiente di correlazione tra v.z. reali:

Def.: Siano X, Y v.z. reali con $\text{var}(X), \text{var}(Y) \in (0, \infty)$
 Il coefficiente di correlazione tra X e Y è dato da

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}}$$

↑
rho

Osservazioni:

1) $\rho(X, Y)$ non dipende dall'unità di misura per X o Y .
 In particolare,

$$\rho(a \cdot X + b, Y) = \rho(X, Y) \quad \forall a, b \in \mathbb{R}$$

2) Fatto: $\rho(X, Y) \in [-1, 1]$

• $\rho(X, Y) = 1$ se e solo esistono $a > 0, b \in \mathbb{R}$ tali che
 $Y = a \cdot X + b$ P-q.c.

• $\rho(X, Y) = -1$ se e solo esistono $a < 0, b \in \mathbb{R}$ tali che
 $Y = a \cdot X + b$ P-q.c.

• $\rho(X, Y) = 0$ se e solo se X e Y sono incorrelate.

Funzioni di ripartizioni per variabili aleatorie reali

La funzione di ripartizione (o funzione di distribuzione cumulativa) per una variabile aleatoria X descrive la probabilità che X assuma un valore inferiore o uguale a un dato valore x . La funzione di ripartizione, $F(x)$, è definita come:

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

In altre parole, la funzione di ripartizione fornisce la probabilità che una variabile aleatoria assuma un valore uguale o inferiore a un dato valore x .

La funzione di ripartizione è una proprietà fondamentale delle variabili aleatorie e fornisce un modo semplice per descrivere la distribuzione di una variabile. Essa è continua e crescente da 0 a 1 e ha come limite inferiore 0 e limite superiore 1

Sia X una v.z. reale su (Ω, \mathcal{F}, P) .

Definiamo una funzione $F_X: \mathbb{R} \rightarrow [0, 1]$ tramite

$$F_X(x) \doteq P(X \leq x), \quad x \in \mathbb{R}.$$

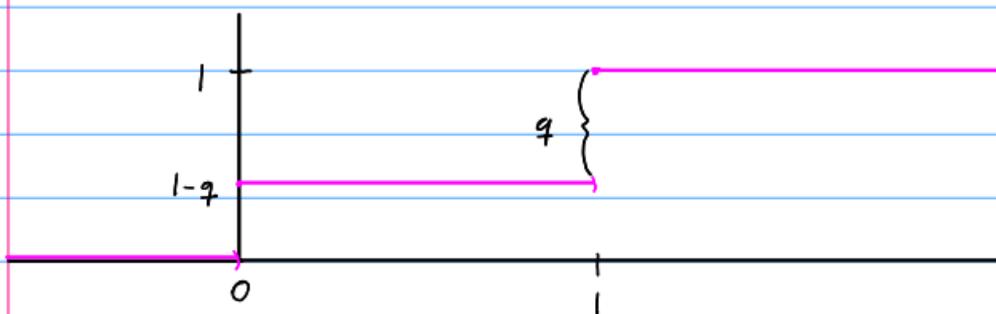
$\leadsto F_X(x)$ dà probabilità che X sia in $(-\infty, x]$.

La funzione F_X si dice funzione di ripartizione di X .

Visivamente si vede come funzione definita a tratti, come si intuisce qui:

Esempio: Sia X una v.z. con $X \sim \text{Ber}(q)$

$$\leadsto P(X=1) = q, \quad P(X=0) = 1-q$$



La funzione di ripartizione determina univocamente la distribuzione di una variabile aleatoria reale. La funzione di ripartizione $F(x)$ descrive la probabilità che una variabile aleatoria X assuma un valore uguale o inferiore a un dato valore x . Se conosciamo la funzione di ripartizione di una variabile aleatoria, possiamo calcolare la probabilità che la variabile assuma qualsiasi valore in un intervallo determinato.

Inoltre, se la funzione di ripartizione è continua e crescente da 0 a 1, è possibile calcolare la densità di probabilità della variabile aleatoria come derivata della funzione di ripartizione. La densità di probabilità descrive la distribuzione continua della variabile aleatoria.

Si può vedere che ricalca le proprietà fondamentali del calcolo integrale, sottraendo quindi la funzione calcolata negli estremi superiore ed inferiore.

Sia X una v.z. reale, F_X la sua funzione di ripartizione.
 Per definizione di F_X ,

$$P(X \in (-\infty, b]) = P(X \leq b) = F_X(b) \quad \text{per ogni } b \in \mathbb{R}$$

Ora $P(X \in (a, b]) = P(X \leq b) - P(X \leq a)$ se $a < b$

! $\rightarrow P(X \in (a, b]) = F_X(b) - F_X(a)$ per ogni $a, b \in \mathbb{R}$
 con $a < b$.

$a \uparrow b$
 $\rightarrow P(X = x) = F_X(x) - \lim_{a \rightarrow 0^+} F_X(x-a)$ per ogni $x \in \mathbb{R}$.

Enunciamo in modo evidente le proprietà:

Proprietà della F_X :

- 1) F_X è crescente: $F_X(x) \leq F_X(y)$ se $x \leq y$.
- 2) F_X è continua a destra: $\lim_{a \rightarrow 0^+} F_X(x+a) = F_X(x)$ per ogni $x \in \mathbb{R}$.

Inoltre, F_X ha limiti a sinistra:
 $\lim_{a \rightarrow 0^+} F_X(x-a)$ esiste in \mathbb{R} per ogni $x \in \mathbb{R}$.

- 3) $\lim_{x \rightarrow \infty} F_X(x) = 1$, $\lim_{x \rightarrow -\infty} F_X(x) = 0$.

Esempio: distribuzioni uniformi discrete su $\{\frac{k}{N} : k \in \{1, \dots, N\}\}$

Per $N \in \mathbb{N}$, sia P_N la densità discreta della distribuzione uniforme su $\{\frac{k}{N} : k \in \{1, \dots, N\}\}$:

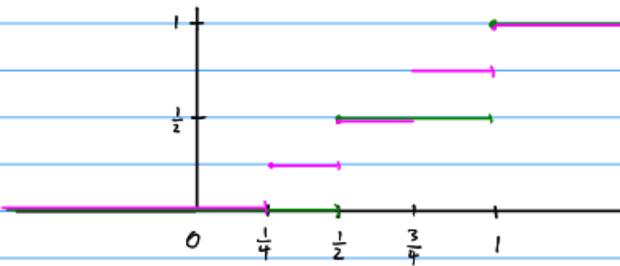
$$P_N(x) = \begin{cases} \frac{1}{N} & \text{se } x \in \{\frac{k}{N} : k \in \{1, \dots, N\}\}, \\ 0 & \text{altrimenti} \end{cases}$$

Sia F_N la corrispondente funzione di ripartizione:

$$F_N(x) = \sum_{y \leq x} P_N(y), \quad x \in \mathbb{R}.$$

$$\leadsto F_N(x) = \begin{cases} 0 & \text{se } x < \frac{1}{N} \\ \frac{k}{N} & \text{se } x \in [\frac{k}{N}, \frac{k+1}{N}) \text{ per un } k \in \{1, \dots, N-1\}, \\ 1 & \text{se } x \geq 1 \end{cases}$$

Gratuito per $N=2$ e $N=4$:



F_N è una funzione costante a tratti.

Per $N \rightarrow \infty$, $F_N(x)$ tende a $F(x)$ con

$$F(x) = \begin{cases} 0 & \text{se } x < 0, \\ x & \text{se } x \in [0, 1) \\ 1 & \text{se } x \geq 1 \end{cases}$$

Nota: F è continua!

Date tutte queste proprietà, si possono riassumere come segue. Il fatto che sia continua preserva la convergenza in probabilità su un intervallo rispetto ad una variabile aleatoria.

Def.: Una funzione $F: \mathbb{R} \rightarrow [0, 1]$ si dice funzione di ripartizione se

- (i) F è crescente;
- (ii) F è continua e destra;
- (iii) $\lim_{x \rightarrow \infty} F(x) = 1, \quad \lim_{x \rightarrow -\infty} F(x) = 0.$

! Fatto:

Sia F una funzione di ripartizione (nel senso delle def.)

Allora esistono uno spazio di probabilità (Ω, \mathcal{F}, P)

e una v.z. reale X su (Ω, \mathcal{F}, P) tali che

F è la funzione di ripartizione di X nel senso che

$$F(x) = P(X \leq x) \quad \text{per ogni } x \in \mathbb{R}.$$

La distribuzione uniforme continua è una distribuzione di probabilità continua in cui ogni valore nell'intervallo specificato ha la stessa probabilità di verificarsi.

Esempio: distribuzione uniforme continua

Siano $a, b \in \mathbb{R}$ con $a < b$. Definiamo

$F_{\text{Unif}(a,b)}: \mathbb{R} \rightarrow [0, 1]$ tramite

$$F_{\text{Unif}(a,b)}(x) = \begin{cases} 0 & \text{se } x < a, \\ \frac{x-a}{b-a} & \text{se } x \in [a, b), \\ 1 & \text{se } x \geq b. \end{cases}$$

Allora $F_{\text{Unif}(a,b)}$ è una funzione di ripartizione; è

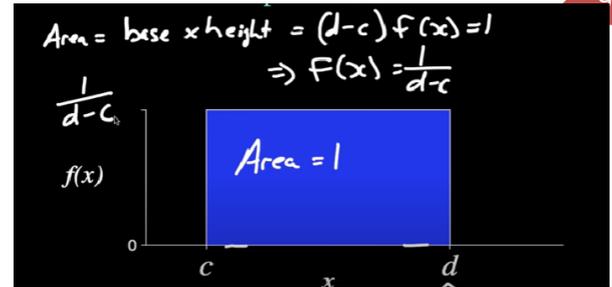
la funzione di ripartizione della distribuzione uniforme continua su (a, b) .

Def.: Si dice che una v.z. reale X ha distribuzione uniforme continua su (a, b) , in simboli

$$X \sim \text{Unif}(a, b),$$

se $F_X = F_{\text{Unif}(a,b)}$, cioè $P(X \leq x) = F_{\text{Unif}(a,b)}(x)$ per ogni $x \in \mathbb{R}$.

La uniforme continua usa $f(x)$ come costante su tutti i valori di x . A lato si vede che, considerando la stessa altezza, la probabilità uniforme si calcola sottraendo gli estremi. Grazie a questi, riusciamo uniformemente a "spezzare" a metà la distribuzione nel calcolo di media/varianza (per quello uguali), data infatti l'uniformità.



La distribuzione esponenziale è una distribuzione di probabilità continua che descrive il tempo trascorso tra eventi avvenuti in un certo intervallo di tempo con una certa media. Essa è utile per descrivere il tempo trascorso tra eventi come la vita di un componente elettronico, il tempo trascorso tra fallimenti di un sistema o il tempo trascorso tra eventi in un processo Poisson. La distribuzione esponenziale è utilizzata in molte applicazioni, come la previsione dei fallimenti, la modellizzazione della durata della vita e la previsione del tempo tra eventi.

Esempio: distribuzione esponenziale

Sia $\lambda > 0$. Definiamo $F_{\text{Exp}(\lambda)}: \mathbb{R} \rightarrow [0,1]$ tramite

$$F_{\text{Exp}(\lambda)}(x) \doteq \begin{cases} 0 & \text{se } x < 0, \\ 1 - e^{-\lambda x} & \text{se } x \geq 0. \end{cases}$$

Allora $F_{\text{Exp}(\lambda)}$ è una funzione di ripartizione; è la funzione di ripartizione della distribuzione esponenziale di parametro λ .

Def.: Si dice che una v.z. reale X ha distribuzione esponenziale di parametro $\lambda > 0$, in simboli

$$X \sim \text{Exp}(\lambda),$$

se $F_X = F_{\text{Exp}(\lambda)}$, cioè $P(X \leq x) = F_{\text{Exp}(\lambda)}(x)$ per ogni $x \in \mathbb{R}$.

Variabili aleatorie assolutamente continue

Una funzione assolutamente continua è una funzione per la quale vale il Teorema fondamentale del calcolo integrale, sostanzialmente. Infatti si dimostra che la definizione di assoluta continuità non è altro che la richiesta di avere una specie di "derivata" che faccia tornare il Teorema di integrazione per parti.

Def.: Sia $F: \mathbb{R} \rightarrow [0,1]$ una funzione di ripartizione.
Allora F si dice **assolutamente continua** se esiste una funzione $f: \mathbb{R} \rightarrow [0, \infty)$ "integrabile" tale che

$$F(x) = \int_{-\infty}^x f(y) dy \quad \text{per ogni } x \in \mathbb{R}.$$

In questo caso, f si dice **densità** di F .

In particolare, una funzione, per essere assolutamente continua, deve essere definita dopo il calcolo dell'integrale su ogni punto e non avere nessuna discontinuità.

Note: Se f è una densità, allora $\int_{-\infty}^{\infty} f(y) dy = 1$.

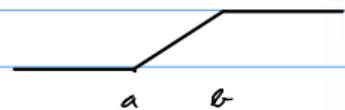
Def.: Sia X una v.z. reale. Allora X si dice **assolutamente continua** se la sua funzione di ripartizione F_X è assolutamente continua.

Diamo esempi di calcolo sulle varie distribuzioni date le nozioni appena descritte:

Esempio: **Uniforme continua**

Sia $X \sim \text{Unif}(a,b)$ con $a < b$. Allora

$$F_X(x) = \begin{cases} 0 & \text{se } x < a, \\ \frac{x-a}{b-a} & \text{se } x \in [a,b), \\ 1 & \text{se } x \geq b. \end{cases}$$



$\leadsto \bar{F}_X$ è differenziabile su $\mathbb{R} \setminus \{a, b\}$ con derivate

$$\bar{F}_X'(x) = \frac{1}{b-a} \cdot \mathbb{1}_{(a,b)}(x), \quad x \in \mathbb{R} \setminus \{a, b\}.$$

Poniamo $f_X(x) \equiv \frac{1}{b-a} \cdot \mathbb{1}_{(a,b)}(x)$, $x \in \mathbb{R}$.

$\leadsto f_X$ non-negativa e definita su \mathbb{R} .

Note: f_X è continua su $\mathbb{R} \setminus \{a, b\}$, mentre è discontinua nei punti a, b .

Inoltre,

$$\int_{-\infty}^x f_X(y) dy = \int_{-\infty}^x \frac{1}{b-a} \cdot \mathbb{1}_{(a,b)}(y) dy$$

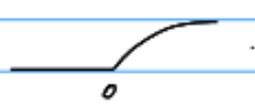
$$= \begin{cases} 0 & \text{se } x \leq a \\ \int_a^x \frac{1}{b-a} dy = \frac{x - \overset{\text{minimo}}{a}}{b-a} & \text{se } x > a. \end{cases}$$

$\leadsto \bar{F}_X(x) = \int_{-\infty}^x f_X(y) dy$ per ogni $x \in \mathbb{R}$.

$\leadsto \frac{1}{b-a} \cdot \mathbb{1}_{(a,b)}(\cdot)$ è la densità della distribuzione uniforme continua su (a, b) .

Esempio: distribuzione esponenziale $\lambda > 0$

Sia $X \sim \text{Exp}(\lambda)$. Allora

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0, \\ 1 - e^{-\lambda x} & \text{se } x \geq 0. \end{cases}$$


Note: F_X è differenziabile su $\mathbb{R} \setminus \{0\}$ con derivate

$$F_X'(x) = \lambda \cdot e^{-\lambda x} \cdot \mathbb{1}_{(0, \infty)}(x), \quad x \in \mathbb{R} \setminus \{0\}$$

Poniamo $f_X(x) = \lambda \cdot e^{-\lambda x} \cdot \mathbb{1}_{(0, \infty)}(x), \quad x \in \mathbb{R}$
 \uparrow discontinua in 0

$\leadsto f_X$ funzione non-negativa definita su \mathbb{R}

In più,

$$\int_{-\infty}^x f_X(y) dy = \int_{-\infty}^x \lambda \cdot e^{-\lambda y} \cdot \mathbb{1}_{(0, \infty)}(y) dy$$

$$= \begin{cases} 0 & \text{se } x \leq 0 \\ \int_0^x \lambda \cdot e^{-\lambda y} dy = 1 - e^{-\lambda x} & \text{se } x > 0. \end{cases}$$

$$\leadsto F_X(x) = \int_{-\infty}^x f_X(y) dy \quad \text{per ogni } x \in \mathbb{R}$$

$\leadsto x \mapsto \lambda \cdot e^{-\lambda x} \cdot \mathbb{1}_{(0, \infty)}(x)$ è la densità della distribuzione esponenziale di parametro λ .

Diamo inoltre alcune condizioni:

! Condizione **sufficiente** per la continuità assoluta:

Sia F una funzione di ripartizione.

Se F è continua ed esiste un insieme $\gamma \subset \mathbb{R}$ di punti isolati (in particolare, γ finito) tale che F è di classe C^1 su $\mathbb{R} \setminus \gamma$, allora F è assolutamente continua con densità

$$f(x) = F'(x) \cdot \mathbb{1}_{\mathbb{R} \setminus \gamma}(x), \quad x \in \mathbb{R}.$$

Esempi:
• distribuzione uniforme continua
• " " esponenziale

Nota: Una funzione di ripartizione assolutamente continua è **necessariamente** continua, ma esistono funzioni di ripartizione continue (su \mathbb{R}) che non sono assolutamente continue; ad esempio, la funzione di Cantor ("scalotta del diavolo").

In questo contesto si inserisce il *metodo Monte Carlo*: la scelta di valori casuali, tali da avere una scelta imparziale (unbiased) se permettiamo alla simulazione di crescere casualmente; grazie alla legge dei grandi numeri siamo sicuri che, prima o poi, ci avvicineremo al valore medio. Più misure ci sono, più accurato è il risultato.

- Una simulazione Monte Carlo è un modello utilizzato per prevedere la probabilità di una serie di risultati in presenza di variabili casuali, specie quando continue
- Le simulazioni Monte Carlo aiutano a spiegare l'impatto del rischio e dell'incertezza nei modelli di previsione e predizione.
- Una simulazione Monte Carlo richiede l'assegnazione di più valori a una variabile incerta per ottenere più risultati e poi fare la media dei risultati per ottenere una stima.

A livello di limite viene definita la convergenza del "p-quasi certamente", che significa che a livello discreto/continuo la distribuzione ha una regolare convergenza ad 1:

Una successione di variabili casuali $(X_n)_{n \in \mathbb{N}}$ si dice **convergere quasi certamente** (o "quasi ovunque" se non anche "**P quasi certamente**" intendendo con P la probabilità, abbreviabile come "P q.c.") alla variabile casuale X , in simboli $X_n \xrightarrow{q.c.} X$ o $X_n \xrightarrow{q.o.} X$, se

$$P(\lim_{n \rightarrow \infty} X_n = X) = 1.$$

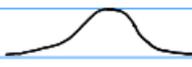
Viene poi definita la distribuzione normale standard, intuitivamente definita come la distribuzione più utile:

Esempio: distribuzione normale standard

La funzione $f_{N(0,1)}$ data da

$$f_{N(0,1)}(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R},$$

è una densità.



Fatto: $\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}$

La funzione

$$\Phi(x) = \int_{-\infty}^x f_{N(0,1)}(y) dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy, \quad x \in \mathbb{R},$$

è la funzione di ripartizione della distribuzione normale standard.

Essa viene definita come l'approssimazione uniforme verso il valor medio. Normalmente è un tipo di distribuzione "a campana" tipo così:



Si può vedere quindi che si approssima attorno al valor medio, descrivendo qual è la "normalità" della distribuzione di dati; se per esempio usassimo la distribuzione per rappresentare l'altezza, vedremmo che gli estremi rappresenterebbero i bassi e gli alti e ci approssimiamo intorno ad un'altezza media. Inoltre, le curve hanno una variazione misurata dalla deviazione standard (chiamata così proprio perché devia rispetto alla normale standard).

Più in generale, definiamo la distribuzione normale di media μ e varianza σ^2 :

Siano $\mu \in \mathbb{R}$, $\sigma > 0$. Poniamo

$$f_{N(\mu, \sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

Allora $f_{N(\mu, \sigma^2)}$ è la densità della

distribuzione normale (o gaussiana) di media μ e varianza σ^2 .

Def.: Si dice che una v.e. reale X ha distribuzione normale di media μ e varianza σ^2 in simboli $X \sim N(\mu, \sigma^2)$,

se la sua funzione di ripartizione è assolutamente continua con densità $f_{N(\mu, \sigma^2)}$.

Note: Sia $F_{N(\mu, \sigma^2)}$ la funzione di ripartizione di una normale di media μ e varianza σ^2 .

Allora

$$\begin{aligned} F_{N(\mu, \sigma^2)}(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \quad | \quad \tilde{y} = \frac{y-\mu}{\sigma} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{\tilde{y}^2}{2}} d\tilde{y} \end{aligned}$$

! \leadsto $F_{N(\mu, \sigma^2)}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ funzione di ripartizione della normale standard

Possiamo parlare anche qui di valor medio e varianza per quanto riguarda le variabili continue. Per calcolarle, utilizziamo necessariamente le variabili assolutamente continue.

Teorema: **Formule di trasformazione**

Sia X una v.v. assolutamente continua con densità f_X .

Sia $g: \mathbb{R} \rightarrow \mathbb{R}$. Allora

$$E[g(X)] = \int_{-\infty}^{\infty} g(y) \cdot f_X(y) dy$$

è patto che l'integrale a destra sia ben definito.

In particolare, se X è assol. continua con densità f_X :

$$\cdot E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx \quad (g = \text{id}_{\mathbb{R}})$$

$$\cdot E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f_X(x) dx \quad (g(x) = x^2)$$

$$\cdot \text{var}(X) = \int_{-\infty}^{\infty} (x - E[X])^2 \cdot f_X(x) dx \quad (g(x) = (x - E[X])^2)$$

$$= E[X^2] - E[X]^2$$

Ciò permette di determinare i valori "generali/notevoli" delle distribuzioni, applicando le nozioni date precedentemente.

Esempio: Sia $X \sim \text{Unif}(a, b)$.

Valor medio: $E[X] = \int_{-\infty}^{\infty} y \cdot \frac{1}{b-a} \cdot \mathbb{1}_{(a,b)}(y) dy$
 $g = \text{id}$

$$\leadsto E[X] = \frac{1}{b-a} \cdot \int_a^b y dy = \frac{1}{b-a} \cdot \left[\frac{y^2}{2} \right]_a^b$$

$$= \frac{1}{b-a} \cdot \left(\frac{b^2}{2} - \frac{a^2}{2} \right) \quad | \quad b^2 - a^2 = (b+a) \cdot (b-a)$$

$$= \frac{a+b}{2}$$

Momento secondo: $E[X^2] = \int_{-\infty}^{\infty} y^2 \cdot \frac{1}{b-a} \cdot \mathbb{1}_{(a,b)}(y) dy$
 $g(y) = y^2$

$$= \frac{1}{b-a} \cdot \int_a^b y^2 dy = \frac{1}{b-a} \cdot \left[\frac{y^3}{3} \right]_a^b$$

$$= \frac{1}{3(b-a)} (b^3 - a^3) \quad \left| \begin{array}{l} (b-a) \cdot (b^2 + a^2) \\ = b^3 + ba^2 - ab^2 - a^3 \\ = b^3 - a^3 + ab(a-b) \end{array} \right.$$

$$= \frac{1}{3} (b^2 + a^2 + ab)$$

$$\begin{aligned} \leadsto \text{var}(X) &= E[X^2] - E[X]^2 \\ &\stackrel{\text{qui}}{=} \frac{1}{3} (b^2 + a^2 + ab) - \frac{(a+b)^2}{4} \\ &= \frac{1}{12} (4b^2 + 4a^2 + 4ab - 3a^2 - 6ab - 3b^2) \\ &= \frac{1}{12} (b^2 + a^2 - 2ab) = \frac{(b-a)^2}{12} \end{aligned}$$

Esempio: $X \sim \text{Exp}(\lambda)$

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \int_0^{\infty} x \cdot \lambda \cdot e^{-\lambda x} dx \\ &\stackrel{\text{integrazione per parti}}{=} \underbrace{[x \cdot (-e^{-\lambda x})]_0^{\infty}}_{=0} + \int_0^{\infty} e^{-\lambda x} dx \\ &= 0 + \left[-\frac{1}{\lambda} \cdot e^{-\lambda x} \right]_0^{\infty} = \frac{1}{\lambda} \end{aligned}$$

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 \cdot f_X(x) dx = \int_0^{\infty} x^2 \cdot \lambda \cdot e^{-\lambda x} dx \\ &\stackrel{\text{integrazione per parti}}{=} \underbrace{[x^2 \cdot (-e^{-\lambda x})]_0^{\infty}}_{=0} + 2 \int_0^{\infty} x \cdot e^{-\lambda x} dx \\ &= \frac{2}{\lambda^2} \text{ da sopra} \end{aligned}$$

$$\leadsto E[X^2] = \frac{2}{\lambda^2}$$

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

Poi, il calcolo con i parametri della normale standard, da cui derivano le condizioni standard:

Una v.z. normale di media μ e varianza σ^2
 ha quindi davvero valor medio uguale a μ
 e varianza uguale a σ^2 .

Esempio: Sia $X \sim N(\mu, \sigma^2)$

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{-\infty}^{\infty} x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu) \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad | \quad x = x-\mu + \mu \\ &\quad + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \mu \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \mu \quad \text{poiché} \quad \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 \end{aligned}$$

$$\int_{-\infty}^{\infty} (x-\mu) \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \left[-\sigma^2 \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right]_{-\infty}^{\infty} = 0$$

$$\leadsto E[X] = \mu.$$

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 \cdot f_X(x) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{-\infty}^{\infty} x^2 \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{-\infty}^{\infty} x \cdot (x-\mu) \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &\quad + \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{-\infty}^{\infty} x \cdot \mu \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \mu^2 \quad \text{da sopra} \end{aligned}$$

$$\int_{-\infty}^{\infty} x \cdot (x-\mu) \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

integra
per parti

$$\underbrace{\left[x \cdot (-\sigma^2) \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right]_{-\infty}^{\infty}}_{=0} + \sigma^2 \cdot \underbrace{\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx}_{=\sqrt{2\pi\sigma^2}}$$

$$X \sim N(\mu, \sigma^2)$$

$$\leadsto E[X^2] = \underline{\underline{\sigma^2 + \mu^2}}$$

$$\leadsto \text{var}(X) = E[X^2] - E[X]^2 = \underline{\underline{\sigma^2}}$$

Legge dei grandi numeri e teorema limite centrale

Citiamo anche i *teoremi limite* parlando della legge dei grandi numeri, dove in pratica si ha il concetto di *media empirica*, descrivendo una equiprobabilità negli eventi delle distribuzioni. Questa dice che, tendendo ad infinito, il valore atteso tenderà per n al valor medio rilevato dai calcoli, tali che non conta quante prove facciamo, convergerà sempre. Questo per descrivere il comportamento asintotico della *media empirica* di una successione di variabili aleatorie reali.

Obiettivo: descrivere il comportamento asintotico delle medie empiriche di una successione $(X_n)_{n \in \mathbb{N}}$ di v.z. reali:

comportamento di $\bar{S}_n \doteq \frac{1}{n} \sum_{i=1}^n X_i$ per $n \rightarrow \infty$.

Si parla inoltre di una serie di variabili *i.i.d* (*indipendenti ed identicamente distribuite*) qualora le variabili hanno tutte la stessa distribuzione di probabilità; le variabili sono tutte statisticamente indipendenti. L'abbreviazione *i.i.d*. (spesso anche *iid*, a volte *IID*) è particolarmente comune in statistica, dove le osservazioni di un campione sono presupposte (più o meno) *i.i.d* per l'inferenza statistica. Il presupposto (o requisito) che le osservazioni siano *i.i.d* tende a facilitare la matematica di molti metodi statistici. Tuttavia, nelle applicazioni pratiche, questo può non essere sempre realistico.

Della legge dei grandi numeri esistono due versioni:

- 1) la legge forte dei grandi numeri

Se, data una successione di variabili casuali $X_1, X_2, \dots, X_n, \dots$ indipendenti e identicamente distribuite con media μ , si considera la *media campionaria*

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

la legge (forte) dei grandi numeri afferma che

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

ossia lo stimatore media campionaria converge quasi certamente al valore atteso comune delle X_i .

2) la legge debole dei grandi numeri

Se, data una successione di variabili casuali $X_1, X_2, \dots, X_n, \dots$ aventi la stessa media μ , la stessa varianza finita e indipendenti, si considera la media campionaria

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

la legge (debole) dei grandi numeri afferma che per ogni $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1.$$

ossia la media campionaria converge in probabilità al valore atteso comune alle X_i .

Generalmente, è comunque definita livello di limite da:

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

$$\lim_{n \rightarrow +\infty} P(|\bar{X}_n - \mu| \leq \varepsilon) = 1 \quad \forall \varepsilon > 0$$

Similmente, il concetto di legge dei grandi numeri si estende allo spazio quadratico, descrivendo che “la distanza della media empirica delle misurazioni tenda a 0”:

Teorema: Legge dei grandi numeri (versione L^2)

Siano X_1, X_2, \dots v.z. reali i.i.d. con varianze finite. Allora

$$\bar{S}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{n \rightarrow \infty} E[X_1] \text{ in } L^2$$

nel senso che $E[|\bar{S}_n - E[X_1]|^2] \xrightarrow{n \rightarrow \infty} 0$.

Per far valere tale affermazioni, abbiamo varie proprietà (utili a livello logico):

- $\text{cov}(X_i, X_j) = 0$ per $i \neq j$, cioè le v.z. sono **incorrelate e due a due**;
- $E[X_1] = E[X_2] = \dots$ stesso valor medio
- $\text{var}(X_1) = \text{var}(X_2) = \dots$ stessa varianza finite.

! \leadsto Legge dei grandi numeri nella versione L^2
 • vale per $(X_n)_{n \in \mathbb{N}}$ successione v.z. reali **incorrelate e due a due** con **stesso valor medio** e **stessa varianza finite**.

\nearrow
 Sufficiente che le varianze siano uniformemente limitate.

Di fatto, grazie a Chebyshev, possiamo utilizzare quanto descritto sopra per dare un limite, a parità di distanza/varianza, al valore atteso e all'errore:

Inoltre, grazie alla disuguaglianza di Chebyshev, per ogni $\varepsilon > 0$,

$$P(|\bar{S}_n - \mu| \geq \varepsilon) \leq \frac{\text{var}(\bar{S}_n)}{\varepsilon^2}$$

ma dalla dimostrazione: $\text{var}(\bar{S}_n) = \frac{\text{var}(X_1)}{n}$

$$\leadsto P(|\bar{S}_n - \mu| \geq \varepsilon) \leq \frac{\text{var}(X_1)}{n \cdot \varepsilon^2} .$$

Stima dell'errore!

Approfondiamo ora meglio l'integrazione Monte Carlo, che di fatto è letteralmente un teorema della media integrale:

Applicazione: Integrazione Monte Carlo

Sia $g: [a, b] \rightarrow \mathbb{R}$ una funzione "misurabile" e limitata,

cioè $\|g\|_\infty = \sup_{x \in [a, b]} |g(x)| < \infty$.

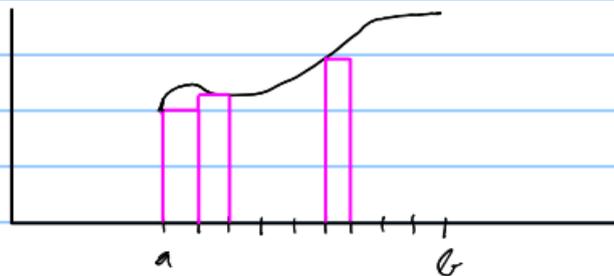
Obiettivo: calcolare numericamente l'integrale

$$\int_a^b g(x) dx.$$

Procedure deterministiche (se g è continua):

somme di Riemann per approssimare l'integrale:

$$\sum_{i=1}^N g(x_i^N) \cdot \frac{b-a}{N}, \quad \text{dove } x_i^N \in \left[a + \frac{b-a}{N} \cdot (i-1), a + \frac{b-a}{N} \cdot i \right]$$



qui approssimazione dal basso

Metodo Monte Carlo: metodo stocastico

Osservazione di partenza: Sia $\xi \sim \text{Unif}(a, b)$.

Allora

$$E[g(\xi)] \stackrel{\text{formule di trasformazione}}{=} \int_{\mathbb{R}} g(x) \cdot f_\xi(x) dx \stackrel{\xi \sim \text{Unif}(a, b)}{=} \frac{1}{b-a} \int_a^b g(x) dx.$$

Siano $\xi_1, \xi_2, \xi_3, \dots$ i.i.d. con comune distribuzione $Unif(a, b)$.

Postiamo $X_n \doteq g(\xi_n)$, $n \in \mathbb{N}$

$\leadsto X_1, X_2, \dots$ sono v.z. i.i.d. con varianze finite poiché g limitata

Grazie alla legge dei grandi numeri,

$$\bar{S}_n \doteq \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{n \rightarrow \infty} E[X_1] \text{ in } L^2,$$

ma $E[X_1] = E[g(\xi_1)] = \frac{1}{b-a} \int_a^b g(x) dx,$

mentre $\bar{S}_n = \frac{1}{n} \sum_{i=1}^n g(\xi_i)$ $\stackrel{!}{=} \mu$

implementabile! distr. uniforme
↓
continue

Si scelgono n numeri pseudo-casuali secondo una $Unif(a, b)$

\leadsto corrisponde a prendere $w \in \Omega \leadsto \xi_1(w), \dots, \xi_n(w)$

$\leadsto \bar{S}_n(w) = \frac{1}{n} \sum_{i=1}^n g(\xi_i(w))$ media campionaria di $g(\xi_1(w)), \dots, g(\xi_n(w))$.

Stima dell'errore: per ogni $\varepsilon > 0$,

$$P(|\bar{S}_n - \mu| \geq \varepsilon) \leq \frac{\text{var}(X_1)}{n \cdot \varepsilon^2}.$$

Si considera poi l'approssimazione del limite centrale, che considera la varianza e la successione di tutti gli eventi come segue (utile perché si usa nelle gaussiane).

Il teorema limite centrale afferma che la distribuzione della somma di un numero elevato di variabili casuali indipendenti e identicamente distribuite tende a distribuirsi "normalmente" (gaussiana), indipendentemente dalla distribuzione delle singole variabili.

Per questo motivo siamo inclini a pensare che, usando la media campionaria, essa converge alla normale standard, in quanto, in condizioni "standard", il valor medio vale 0 e la varianza vale 1.

Teorema del limite centrale:

Sia $(X_i)_{i \in \mathbb{N}}$ una successione di v.z. reali *i.i.d* con variazze finite non-zero. Poniamo

$$\mu = E[X_1], \quad \sigma^2 = \text{var}(X_1),$$

$$Z_n = \frac{1}{\sigma \cdot \sqrt{n}} \cdot \left(\sum_{i=1}^n (X_i - \mu) \right).$$

Allora Z_n converge in distribuzione, per $n \rightarrow \infty$, a una normale standard nel senso che, per ogni $x \in \mathbb{R}$,

$$P(Z_n \leq x) \xrightarrow{n \rightarrow \infty} \Phi(x)$$

↑ funzione di ripartizione della normale standard

Quanto sopra si dimostra esattamente qui:

Nota: $E[Z_n] = 0, \quad \text{var}(Z_n) = 1,$

$$Z_n = \frac{\sqrt{n}}{\sigma} \cdot (\bar{S}_n - \mu)$$

Infatti:

- $E[Z_n] \stackrel{\text{linearità}}{=} \frac{1}{\sigma \cdot \sqrt{n}} \cdot \sum_{i=1}^n \overbrace{E[X_i - \mu]}^{=0 \text{ (v.z. ident. distribuite)}} = 0.$

- $\text{var}(Z_n) = \frac{1}{\sigma^2 \cdot n} \cdot \text{var}\left(\sum_{i=1}^n (X_i - \mu)\right) \quad | \text{ v.z. indipendenti}$
 $= \frac{1}{\sigma^2 \cdot n} \cdot \sum_{i=1}^n \underbrace{\text{var}(X_i - \mu)}_{= \text{var}(X_i) = \sigma^2 \text{ (v.z. ident. distn)}} = 1.$

Esempio: controllo di qualità

Azienda produce viti di un certo tipo. Effettua un controllo di qualità testando n viti scelte a caso per determinare la proporzione di viti difettose.

Modello: X_1, \dots, X_n v. z. a valori in $\{0, 1\}$
 con $X_i = 1$ se i -esima vite testata è difettosa,
 $X_i = 0$ altrimenti

$\Rightarrow X_1, \dots, X_n$ i.i.d. con comune distribuzione di Bernoulli di parametro $q \in [0, 1]$,
 dove q è la probabilità che una vite scelta a caso sia difettosa.

Parametro q è incognito!

Possiamo usare la media empirica $\bar{S}_n = \frac{1}{n} \sum_{i=1}^n X_i$ come stimatore di q .

Ricorda: $E[\bar{S}_n] \stackrel{\text{i.i.d.}}{=} E[X_1] \stackrel{X_1 \sim \text{Ber}(q)}{=} q$.

Esempio (cont.)

Obiettivo: Scegliere $n \in \mathbb{N}$ e $d_n > 0$ tali che

$$P(q \in (\bar{S}_n - d_n, \bar{S}_n + d_n)) \geq 0,95.$$

↑ intervallo di confidenza (qui al 95%)

Siano $n \in \mathbb{N}$, $d_n > 0$. Allora

$$P(q \in (\bar{S}_n - d_n, \bar{S}_n + d_n)) = P(\bar{S}_n - d_n < q < \bar{S}_n + d_n)$$

$$= P(-d_n < q - \bar{S}_n < d_n)$$

$$= P(-d_n < \bar{S}_n - q < d_n)$$

$$= P(|\bar{S}_n - q| < d_n).$$

Vogliamo avere $P(|\bar{S}_n - q| < d_n) \geq 0,95$.

1° metodo: disuguaglianza di Chebyshev:

$$P(|\bar{S}_n - q| < d_n) = 1 - P(|\bar{S}_n - q| \geq d_n)$$

Orz, grazie a Chebyshev,

$$P(|\bar{S}_n - q| \geq d_n) \leq \frac{\text{var}(\bar{S}_n)}{d_n^2}$$

$$\leadsto P(|\bar{S}_n - q| \geq d_n) \leq \frac{q \cdot (1-q)}{n \cdot d_n^2}$$

var. i.i.d.

$$\downarrow \text{var}(X_i) = \frac{\text{var}(X_i)}{n}$$

$$= \frac{q \cdot (1-q)}{n}$$

$$\uparrow X_i \sim \text{Ber}(q)$$

$$\leadsto P(q \in (\bar{S}_n - d_n, \bar{S}_n + d_n)) \geq 1 - \frac{q \cdot (1-q)}{n \cdot d_n^2}$$

Vogliamo quindi che $1 - \frac{q \cdot (1-q)}{n \cdot d_n^2} \geq 0,95$

$$\leadsto n \cdot d_n^2 \geq \frac{1}{0,05} \cdot \underbrace{q \cdot (1-q)}_{\leq \frac{1}{4}}, \quad \text{dove } q \in [0,1]$$

$$\leadsto n \cdot d_n^2 \geq \frac{20}{4} = 5.$$

Ad esempio:

$$a) \quad n = 45 \quad \leadsto \quad d_n^2 \geq \frac{1}{9} \quad \leadsto \quad d_n \geq \frac{1}{3}$$

$$b) \quad d_n = \frac{1}{10} \quad \leadsto \quad n \geq 500.$$

Confronto con l'approssimazione normale:

Dal Teorema del limite centrale abbiamo che

$\frac{\sqrt{n}}{\sqrt{\text{var}(X)}} \cdot (\bar{S}_n - q)$ è vicina in distribuzione a una normale standard se n è abbastanza grande.

$$\leadsto P(|\bar{S}_n - q| < d_n) \quad | \quad \sqrt{\text{var}(X)} = \sqrt{q \cdot (1-q)}$$

$$= P\left(|\frac{\sqrt{n}}{\sqrt{q \cdot (1-q)}} \cdot (\bar{S}_n - q)| < \frac{d_n \cdot \sqrt{n}}{\sqrt{q \cdot (1-q)}}\right) \quad | \quad Z \sim N(0,1)$$

$$\approx P(|Z| < \frac{d_n \cdot \sqrt{n}}{\sqrt{q \cdot (1-q)}})$$

$$= P\left(Z \leq \frac{d_n \cdot \sqrt{n}}{\sqrt{q \cdot (1-q)}}\right) - P\left(Z \leq -\frac{d_n \cdot \sqrt{n}}{\sqrt{q \cdot (1-q)}}\right)$$

$$\leadsto P(|\bar{S}_n - q| < d_n)$$



$$\approx \Phi\left(\frac{d_n \cdot \sqrt{n}}{\sqrt{q \cdot (1-q)}}\right) - \Phi\left(-\frac{d_n \cdot \sqrt{n}}{\sqrt{q \cdot (1-q)}}\right) \quad | \quad \Phi(-x) = 1 - \Phi(x)$$

$$= 2\Phi\left(\frac{d_n \cdot \sqrt{n}}{\sqrt{q \cdot (1-q)}}\right) - 1$$

Cerchiamo $y \in \mathbb{R}$ minimo tale che

$$2\Phi(y) - 1 \geq 0,95$$

$$\Leftrightarrow \Phi(y) \geq \frac{1,95}{2} = 0,975$$

tavola

$$\leadsto y \geq 1,96$$

$$\leadsto \frac{d_n \cdot \sqrt{n}}{\sqrt{q \cdot (1-q)}} \geq 1,96 \quad | \quad q \cdot (1-q) \leq \frac{1}{4} \text{ per } q \in [0,1]$$

$$\leadsto d_n \cdot \sqrt{n} \geq \frac{1,96}{2}$$

Nell'esempio:

$$a) \quad n=45 \quad \leadsto \quad d_n \geq 0,98 \cdot \frac{1}{\sqrt{45}}$$

$$\leadsto \quad d_n \geq \frac{1}{7}$$

$$b) \quad d_n = \frac{1}{10} \quad \leadsto \quad \sqrt{n} \geq 10 \cdot \frac{1,96}{2} = 9,8$$

$$\leadsto \quad n \geq 9,8^2 \approx 100$$

Stime migliori risp. a Chebyshev

L' approssimazione normale può essere applicata anche a somme di v.z. i.i.d. piuttosto che alle loro medie empiriche:

Siano X_1, \dots, X_n v.z. i.i.d. $\text{Ber}(q)$

$$S_n = \sum_{i=1}^n X_i$$

v.z. i.i.d. \rightarrow $E[S_n] = n \cdot E[X_i] = n \cdot q,$

$$\text{var}(S_n) = n \cdot \text{var}(X_i) = n \cdot q \cdot (1-q)$$

Grazie al teorema del limite centrale, per n abbastanza grande,

$\frac{(S_n - n \cdot q)}{\sqrt{n \cdot q \cdot (1-q)}}$ è vicina in distribuzione a una normale standard.

In particolare, se vogliamo approssimare probabilità della forma

$P(S_n \leq x)$ per n grande, possiamo scrivere: $\hat{=}$ somme di v.z. X_1, \dots, X_n i.i.d.

$$P(S_n \leq x) = P(S_n - E[S_n] \leq x - E[S_n])$$

$$= P\left(\frac{S_n - E[S_n]}{\sqrt{\text{var}(S_n)}} \leq \frac{x - E[S_n]}{\sqrt{\text{var}(S_n)}}\right)$$

teorema del
 \approx
limite centrale

$$\Phi\left(\frac{x - E[S_n]}{\sqrt{\text{var}(S_n)}}\right),$$

dove $E[S_n] = n \cdot E[X_i],$
 $\sqrt{\text{var}(S_n)} = \sqrt{n \cdot \text{var}(X_i)}$
poiché X_1, \dots, X_n i.i.d.

Statistica descrittiva

Definiamo poi la statistica descrittiva come la sintesi di dati attraverso funzione sui dati detti *statistiche*. In particolare, matematicamente, consideriamo una successione finita di numeri reali in un insieme ordinato sui reali, detto *campione* di numerosità n .

A tal proposito, possiamo definire la media campionaria, come una misura di tendenza centrale che indica la posizione media di un insieme di dati. La media campionaria è calcolata come la somma dei valori dei dati divisa per il numero di dati. I valori con frequenza massima sono i *valori modali*.

Esempio:

Supponiamo di avere i voti ottenuti da un gruppo di 10 studenti in un esame: 90, 95, 88, 92, 83, 87, 85, 86, 94, 91. La media campionaria può essere calcolata come segue:

$$(90 + 95 + 88 + 92 + 83 + 87 + 85 + 86 + 94 + 91) / 10 = 90$$

In questo esempio, la media campionaria è 90, il che significa che la maggior parte dei voti è intorno a 90.

Dato quindi un insieme di valori, abbiamo che la *frequenza assoluta* è il numero di volte che un determinato valore o categoria si verifica in un insieme di dati.

Esempio:

Supponiamo di avere i risultati di un sondaggio sulla marca di gelato preferita da 100 persone.

Le marche disponibili sono: Gelato Algida, Ben & Jerry's e Häagen-Dazs. La tabella seguente mostra le frequenze assolute per ciascuna marca:

Marca	Frequenza Assoluta
Algida	40
Ben & Jerry's	30
Häagen-Dazs	30

In questo esempio, la frequenza assoluta per Gelato Algida è 40, il che significa che 40 persone hanno scelto Gelato Algida come la loro marca preferita di gelato.

Quanto appena descritto è la *statistica sul centro dei dati*, ma ne esistono diverse altre:

- d'ordine, intendendo minimo e massimo
- sulla dispersione dei dati, definendo
 - o La *varianza campionaria*, misura di dispersione che quantifica la deviazione media dei valori di un insieme di dati rispetto alla media campionaria. La varianza campionaria viene calcolata come la somma del quadrato delle deviazioni di ciascun valore rispetto alla media campionaria, divisa per il numero di dati meno uno.

Def.: Per $n \geq 2$, la *varianza campionaria* di $(x_i)_{i \in \{1, \dots, n\}}$ è data da

$$s^2 := \frac{1}{n-1} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right),$$

dove \bar{x} è la *media campionaria*.

La quantità $s := \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)}$ si dice *deviazione standard campionaria*.

La *disuguaglianza di Chebyshev* in statistica descrittiva afferma che per qualsiasi insieme di dati e qualsiasi livello di tolleranza k , almeno il $(1 - 1/k^2)\%$ dei dati si trova all'interno di k volte la deviazione standard dalla media.

Questa disuguaglianza è utile perché fornisce una stima inferiore rigorosa del numero di dati che si trovano in una determinata regione intorno alla media. Ad esempio, se $k = 2$, la disuguaglianza di Chebyshev afferma che almeno l'75% dei dati si trova entro 2 deviazioni standard dalla media.

Un esempio di applicazione della disuguaglianza di Chebyshev in statistica descrittiva è la verifica di eventuali valori anomali in un insieme di dati. Se un valore si trova al di fuori di k volte la deviazione standard dalla media, allora può essere considerato un valore anomalo. In questo modo, la disuguaglianza di Chebyshev può essere utilizzata per identificare eventuali valori anomali che potrebbero influire sulla rappresentazione dei dati.

In matematica, data la media campionaria ed il campione, diamo una stima rigorosa della variazione dei dati intorno alla media, mettendo tutto insieme:

Disuguaglianze di Chebyshev (versioni campionarie)

Siano \bar{x} la media campionaria e s la deviazione standard campionaria del campione $(x_i)_{i \in \{1, \dots, n\}} \subset \mathbb{R}$.

Se $s > 0$, allora per ogni $\alpha > 0$:

- (i) $\frac{\#\{i \in \{1, \dots, n\} : |x_i - \bar{x}| < \alpha \cdot s\}}{n} \geq 1 - \frac{(n-1)}{n \cdot \alpha^2} > 1 - \frac{1}{\alpha^2}$
- (ii) $\frac{\#\{i \in \{1, \dots, n\} : |x_i - \bar{x}| \leq \alpha \cdot s\}}{n} > 1 - \frac{(n-1)}{n \cdot \alpha^2} > 1 - \frac{1}{\alpha^2}$
- (iii) $\frac{\#\{i \in \{1, \dots, n\} : |x_i - \bar{x}| \geq \alpha \cdot s\}}{n} < \frac{1}{\alpha^2}$
- Unilaterale (iv) $\frac{\#\{i \in \{1, \dots, n\} : x_i - \bar{x} \geq \alpha \cdot s\}}{n} < \frac{1}{1 + \alpha^2}$

Avendo una permutazione crescente dei valori (distinguendo un insieme di 100 valori, con k compreso qui), il percentile campionario k -esimo in statistica descrittiva è un valore che separa l'insieme di dati in k percentuali uguali. Ad esempio, il 50° percentile, noto anche come mediana campionaria, separa l'insieme di dati in due parti uguali, con la metà dei dati al di sotto e la metà al di sopra del valore del 50° percentile.

Il valore del k -esimo percentile può essere calcolato ordinando i dati in modo crescente e individuando il valore che corrisponde alla posizione $(n * k/100)$, dove n è il numero di dati. Se la posizione non è un numero intero, il valore del percentile viene interpolato tra i valori più vicini.

Il percentile campionario è una misura robusta della posizione che non viene influenzata da valori estremi come la media e la deviazione standard, ed è utile per descrivere la distribuzione dei dati in modo più dettagliato. Ad esempio, i percentili possono essere utilizzati per descrivere la distribuzione dei salari o dei tempi di completamento di un compito in un insieme di dati.

In matematiche notiamo quanto sotto:

Def.: Il percentile campionario k -esimo, con $k \in \{0, \dots, 100\}$, del campione è dato da

$$\bar{P}_k = \begin{cases} X_{\sigma(\frac{n \cdot k}{100})} & \text{se } \frac{n \cdot k}{100} \notin \mathbb{N}_0, \\ \frac{1}{2} (X_{\sigma(\frac{n \cdot k}{100})} + X_{\sigma(\frac{n \cdot k}{100} + 1)}) & \text{altrimenti.} \end{cases}$$

La covarianza campionaria in statistica descrittiva è una misura della relazione tra due variabili aleatorie. La covarianza campionaria misura il grado di variazione congiunta delle due variabili e il loro senso di relazione.

Se x e y sono le due variabili aleatorie, la covarianza campionaria (denotata come $cov(x, y)$) può essere calcolata come la media dei prodotti delle deviazioni di x e y dalle loro medie:

$$cov_{x,y} = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}),$$

dove \bar{x} media campionaria di $(x_i)_{i \in \{1, \dots, n\}}$,

\bar{y} media campionaria di $(y_i)_{i \in \{1, \dots, n\}}$.

Se la covarianza campionaria è positiva, significa che x e y aumentano e diminuiscono insieme. Se la covarianza campionaria è negativa, significa che x e y si muovono in direzioni opposte. Se la covarianza campionaria è zero, significa che x e y non hanno alcuna relazione lineare.

Per avere una quantità che non dipenda dalla scelta delle unità di misura, si normalizza la covarianza usando le deviazioni standard. Usiamo la correlazione campionaria, che esprime la forza e la direzione della relazione lineare tra le due variabili.

Def.: La correlazione campionaria tra $(x_i)_{i \in \{1, \dots, n\}}$ e $(y_i)_{i \in \{1, \dots, n\}}$ è data da

$$\text{corr}_{x,y} = \frac{\text{cov}_{x,y}}{s_x \cdot s_y},$$

dove s_x deviazione standard di $(x_i)_{i \in \{1, \dots, n\}}$,

s_y " " " " di $(y_i)_{i \in \{1, \dots, n\}}$.

Essendo definita tra -1 ed 1, il fattore di correlazione varia rispettivamente da 1 e -1 qualora $a > 0$ oppure < 0 .

$$\text{corr}_{x,y} \in [-1, 1].$$

$\text{corr}_{x,y} = 1$ se e solo se $\exists a, b \in \mathbb{R}$:
 $a > 0$ e
 $x_i = a \cdot y_i + b \quad \forall i \in \{1, \dots, n\}$.

$\text{corr}_{x,y} = -1$ se e solo se $\exists a, b \in \mathbb{R}$:
 $a < 0$ e
 $x_i = a \cdot y_i + b \quad \forall i \in \{1, \dots, n\}$.

Letture delle tavole

Letture tavola Normale Standard

$$\Phi(y) \geq 0,98$$

tavola \rightarrow $y \geq 2,06$

Per calcolarla si cerca dentro le colonne delle coordinate di riga e colonna che diano un numero ≥ 0.98 . In questo caso il numero è 2,06 e dall'immagine si mostra perché:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Letture Tavole con Poisson

Si calcola il parametro come al solito:

Nota: $S \sim \text{Bin}(1000, \frac{1}{500})$

\leadsto distribuzione di S è vicina alla distribuzione di Poisson di parametro

$$\lambda = 1000 \cdot \frac{1}{500} = 2.$$

$$\leadsto P(S \leq k) \approx F_{\text{Pois}(2)}(k)$$

Scegliere $k \in \mathbb{N}$ minimo tale che

$$F_{\text{Pois}(2)}(k) \geq 0,98.$$

tavole
 $\leadsto k \geq 5$

\leadsto stima per N_x : $N_x = 5.$

Anche qui serve cercare un valore maggiore usando la colonna di sinistra del parametro lambda. In questo caso il parametro è uguale a 2 e quindi si trova nella stessa riga il primo numero ≥ 0.98 in questo caso è 5 e si mostra a pagina dopo il perché:

λ	$x=0$	1	2	3	4	5	6	7	8	9
0.02	0.980	1.000								
0.04	0.961	0.999	1.000							
0.06	0.942	0.998	1.000							
0.08	0.923	0.997	1.000							
0.10	0.905	0.995	1.000							
0.15	0.861	0.990	1.000	1.000						
0.20	0.819	0.983	0.999	1.000						
0.25	0.779	0.974	0.998	1.000						
0.30	0.741	0.963	0.996	1.000						
0.35	0.705	0.951	0.995	1.000						
0.40	0.670	0.938	0.992	0.999	1.000					
0.45	0.638	0.925	0.989	0.999	1.000					
0.50	0.607	0.910	0.986	0.998	1.000					
0.55	0.577	0.894	0.982	0.998	1.000					
0.60	0.549	0.878	0.977	0.997	1.000					
0.65	0.522	0.861	0.972	0.996	0.999	1.000				
0.70	0.497	0.844	0.966	0.994	0.999	1.000				
0.75	0.472	0.827	0.960	0.993	0.999	1.000				
0.80	0.449	0.809	0.953	0.991	0.999	1.000				
0.85	0.427	0.791	0.945	0.989	0.998	1.000				
0.90	0.407	0.772	0.937	0.987	0.998	1.000				
0.95	0.387	0.754	0.929	0.984	0.997	1.000				
1.00	0.368	0.736	0.920	0.981	0.996	0.999	1.000			
1.1	0.333	0.699	0.900	0.974	0.995	0.999	1.000			
1.2	0.301	0.663	0.879	0.966	0.992	0.999	1.000			
1.3	0.273	0.627	0.857	0.957	0.989	0.998	1.000			
1.4	0.247	0.592	0.834	0.946	0.986	0.997	0.999	1.000		
1.5	0.223	0.558	0.809	0.934	0.981	0.996	0.999	1.000		
1.6	0.202	0.525	0.783	0.921	0.976	0.994	0.999	1.000		
1.7	0.183	0.493	0.757	0.907	0.970	0.992	0.998	1.000		
1.8	0.165	0.463	0.731	0.891	0.964	0.990	0.997	0.999	1.000	
1.9	0.150	0.434	0.704	0.875	0.956	0.987	0.997	0.999	1.000	
2.0	0.135	0.406	0.677	0.857	0.947	0.983	0.996	0.999	1.000	

Quello è il parametro che abbiamo noi

Primo parametro > 0.98 e quindi il numero è 5.

Cumulative probability, Poisson distribution

λ	$x=0$	1	2	3	4	5	6	7	8	9
0.02	0.980	1.000								
0.04	0.961	0.999	1.000							
0.06	0.942	0.998	1.000							
0.08	0.923	0.997	1.000							
0.10	0.905	0.995	1.000							
0.15	0.861	0.990	1.000	1.000						
0.20	0.819	0.983	0.999	1.000						
0.25	0.779	0.974	0.998	1.000						
0.30	0.741	0.963	0.996	1.000						
0.35	0.705	0.951	0.995	1.000						
0.40	0.670	0.938	0.992	0.999	1.000					
0.45	0.638	0.925	0.989	0.999	1.000					
0.50	0.607	0.910	0.986	0.998	1.000					
0.55	0.577	0.894	0.982	0.998	1.000					
0.60	0.549	0.878	0.977	0.997	1.000					
0.65	0.522	0.861	0.972	0.996	0.999	1.000				
0.70	0.497	0.844	0.966	0.994	0.999	1.000				
0.75	0.472	0.827	0.960	0.993	0.999	1.000				
0.80	0.449	0.809	0.953	0.991	0.999	1.000				
0.85	0.427	0.791	0.945	0.989	0.998	1.000				
0.90	0.407	0.772	0.937	0.987	0.998	1.000				
0.95	0.387	0.754	0.929	0.984	0.997	1.000				
1.00	0.368	0.736	0.920	0.981	0.996	0.999	1.000			
1.1	0.333	0.699	0.900	0.974	0.995	0.999	1.000			
1.2	0.301	0.663	0.879	0.966	0.992	0.999	1.000			
1.3	0.273	0.627	0.857	0.957	0.989	0.998	1.000			
1.4	0.247	0.592	0.834	0.946	0.986	0.997	0.999	1.000		
1.5	0.223	0.558	0.809	0.934	0.981	0.996	0.999	1.000		
1.6	0.202	0.525	0.783	0.921	0.976	0.994	0.999	1.000		
1.7	0.183	0.493	0.757	0.907	0.970	0.992	0.998	1.000		
1.8	0.165	0.463	0.731	0.891	0.964	0.990	0.997	0.999	1.000	
1.9	0.150	0.434	0.704	0.875	0.956	0.987	0.997	0.999	1.000	
2.0	0.135	0.406	0.677	0.857	0.947	0.983	0.996	0.999	1.000	
2.2	0.111	0.355	0.623	0.819	0.927	0.975	0.993	0.998	1.000	
2.4	0.091	0.308	0.570	0.779	0.904	0.964	0.988	0.997	0.999	1.000
2.6	0.074	0.267	0.518	0.736	0.877	0.951	0.983	0.995	0.999	1.000
2.8	0.061	0.231	0.469	0.692	0.848	0.935	0.976	0.992	0.998	0.999
3.0	0.050	0.199	0.423	0.647	0.815	0.916	0.967	0.988	0.996	0.999
3.2	0.041	0.171	0.380	0.603	0.781	0.895	0.955	0.983	0.994	0.998
3.4	0.033	0.147	0.340	0.558	0.744	0.871	0.942	0.977	0.992	0.997
3.6	0.027	0.126	0.303	0.515	0.706	0.844	0.927	0.969	0.988	0.996
3.8	0.022	0.107	0.269	0.473	0.668	0.816	0.909	0.960	0.984	0.994

continued on next page

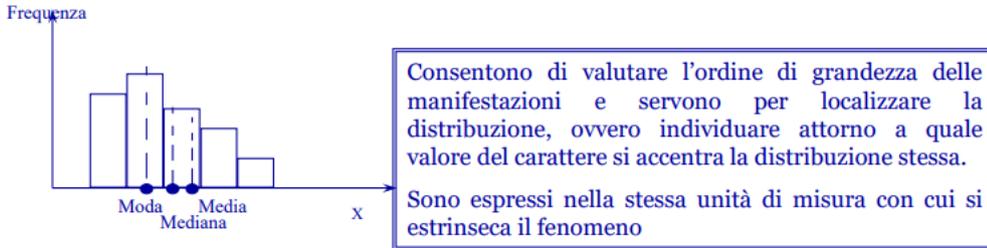
Parti utili normale standard:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Media, moda, mediana

Questi tre valori sono caratteristici di un insieme dati statistico. In particolare:

- Media è il rapporto tra la somma dei dati numerici e il numero dei dati
- Moda è il valore che si presenta con maggiore frequenza
- Mediana è il valore centrale tra i dati numerici



MODA

E' definita come il valore che ha la frequenza più alta.

MEDIANA

E' quel valore al di sotto del quale cadono la metà dei valori campionari.

MEDIA Aritmetica

E' quel valore che corrisponde alla somma di tutti i valori diviso il numero dei valori stessi.

Rappresenta il valore che sostituito a ciascun x_i lascia invariata la intensità totale (somma)

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

dove:

X_i = esito i-ma misura - n = numero dei dati (taglia del campione)

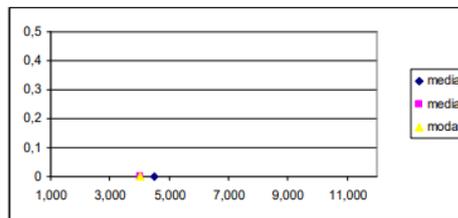
Moda e media sono indici di posizione, poiché la loro variazione sposta appunto la posizione della curva (verso destra o verso sinistra) in funzione del segno della variazione.

ESEMPIO 1

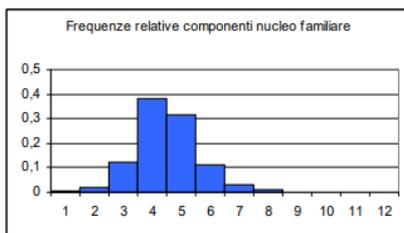
Numero di componenti familiari

studenti.xls utilizzo strumento di analisi: statistica descrittiva

COMPON	
Mean	4,513
Median	4
Mode	4



Media = 4.5 > 4 = mediana = moda



Ci sono più osservazioni alla sinistra della media

La distribuzione è più concentrata alla sinistra della media

MA

la coda è più lunga a destra

ESEMPIO 2

studenti.xls utilizzo strumento di analisi: statistica descrittiva

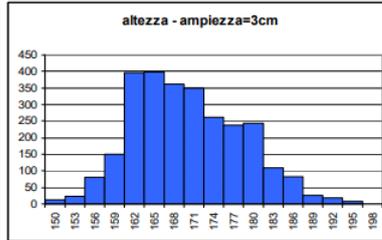
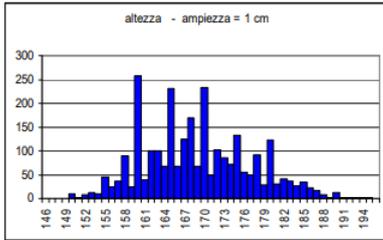
altezza

ALTEZZA	
Mean	169,011961
Median	168
Mode	160

media = 169 > 168 = mediana > Moda=160

Ci sono più osservazioni alla sinistra della media

La distribuzione è più concentrata alla sinistra della media
e
la coda è più lunga a destra



MODA

La moda di una distribuzione di frequenze è il punto centrale della classe di misure più frequente.

Distribuzione zeromodale: nessun valore ha una frequenza più elevata degli altri.

Distribuzione unimodale: c'è un solo valore con una frequenza più elevata degli altri. Es. [2, 4, 1, 3, 7, 3, 5, 3]

Distribuzione bimodale: ci sono due valori con una frequenza più elevata degli altri. Es. [7, 4, 7, 3, 7, 3, 5, 3]

3

MEDIANA

La mediana è il valore che occupa la posizione centrale quando le osservazioni di un campione sono **ordinate** in base al loro valore.

Quando **n** è **dispari**, la mediana corrisponde al punteggio dell'individuo numero $(n + 1)/2$.

Quando n è pari, la mediana corrisponde al valore intermedio tra il punteggio dell'individuo numero $n/2$ e il punteggio dell'individuo $(n/2) + 1$.

6, 6.7, 3.8, 7, 5.8, 9.975

I valori ordinati sono:

3.8, 5.8, 6, 6.7, 7, 9.975

| |

2 middle values

$$\text{Median} = \frac{6 + 6.7}{2} = 6.35$$

Quantili e vari tipi

Oltre alla mediana, nella statistica descrittiva vengono usati altri indici di posizione detti quantili, i quali dividono le distribuzioni in determinate percentuali. Questi sono detti indici di posizione non centrale o di non centralità. A seconda delle percentuali in cui la distribuzione viene suddivisa, si classificano diversi tipi di quantili:

- Mediana (quantile di ordine $1/2$);
- Quartili (quantili di ordine $1/4$, $1/2$, $3/4$);
- Decili (quantili di ordine $1/10$);
- Centili o percentili (quantili di ordine $1/100$).

Quartili

I **quartili**, come dice la parola stessa, si ottengono dividendo l'insieme di dati ordinati in 4 parti uguali ed esattamente:

- il **primo quartile** Q_1 è il valore che lascia alla sua sinistra il 25% degli elementi della distribuzione; Q_1 è anche detto **25-esimo percentile** ($P_{0.25}$).
- Il **secondo quartile** Q_2 coincide con la mediana dato che è quello che lascia alla sua sinistra il 50% dei dati della distribuzione; Q_2 è anche detto **50-esimo percentile** ($P_{0.5}$).
- Il **terzo quartile** Q_3 è il valore che lascia il 75% degli elementi a sinistra e il 25% a destra; Q_3 è anche detto **75-esimo percentile** ($P_{0.75}$).

Percentili

I **percentili**, invece, dividono la distribuzione in 100 parti, ad esempio:

- il 1° percentile $P_{0.01}$ lascia alla sua sinistra un centesimo degli elementi della distribuzione, ossia l'1%;
- il 10° percentile $P_{0.1}$ lascia alla sua sinistra il 10% degli elementi della distribuzione;
- il 50° percentile (che coincide con la mediana e con il secondo quartile) lascia alla sua sinistra il 50% della distribuzione.

Decili

I **decili**, invece, dividono la distribuzione in 10 parti, ad esempio:

- il primo decile $P_{0.1}$ lascia alla sua sinistra un decimo degli elementi della distribuzione, ossia il 10%;
- il terzo decile $P_{0.3}$ lascia alla sua sinistra il 30% degli elementi della distribuzione;
- il quinto decile (che coincide con la mediana e con il secondo quartile) lascia alla sua sinistra il 50% della distribuzione.

Regola pratica per il calcolo dei quartili e dei percentili

Per calcolare i quartili (o anche i percentili) di una distribuzione, seguiamo i passi di seguito indicati:

1. Si ordinano gli n dati della distribuzione in ordine crescente;
2. Indicato con p il percentile in decimale ($p = 0.25$ per il 25° percentile o 1° quartile, $p = 0.37$ per il 37° percentile), si calcola il prodotto $k = np$;
3. se k è un intero, il quartile (percentile) si ottiene facendo la media del k -esimo e del $(k + 1)$ -esimo valore dei dati ordinati;
4. se k non è un intero, si arrotonda k per eccesso al primo intero successivo e si sceglie come quartile (percentile) il corrispondente valore dei dati ordinati

Esempio calcolo quartili

Calcolare il primo e il terzo quartile dell'insieme di dati

32.2 32 30.4 31 31.2 31.3 30.3 29.6 30.5
30.7

Ordiniamo i dati:

29.6 30.3 30.4 30.5 30.7 31 31.2
31.3 32 32.2

Per il primo quartile abbiamo:

$$n = 10, \quad p = 0.25, \quad k = np = 2.5$$

Poichè k non è intero, si arrotonda per eccesso ($k = 3$): il primo quartile è il terzo dei dati ordinati, ovvero,

$$Q_1 = 30.4$$

Per il secondo quartile, ovvero la mediana, abbiamo:

$$n = 10, \quad p = 0.5, \quad k = np = 5$$

Poichè k è intero, si fa la media tra il quinto e il sesto dato ottenendo

$$Q_2 = \frac{30.7 + 31}{2} = 30.85$$

Per il terzo quartile abbiamo:

$$n = 10, \quad p = 0.75, \quad k = np = 7.5$$

Poichè k non è intero, si arrotonda per eccesso ($k = 8$): il terzo quartile è l'ottavo dei dati ordinati, ovvero,

$$Q_3 = 31.3$$

Esempio calcolo percentili

Calcolare il 95-esimo percentile per i dati (già ordinati) della seguente tabella

6.2	7.7	8.3	9.0	9.4	9.8	10.5	10.7	11.0	11.2
11.8	12.3	12.8	13.2	13.3	13.5	13.9	14.4	14.5	14.7
15.2	15.5	15.8	15.9	16.2	16.7	16.9	17.0	17.3	17.5
17.6	17.9	18.0	18.0	18.1	18.1	18.4	18.5	18.7	19.0
19.1	19.2	19.3	19.4	19.4	20.0	20.1	20.1	20.4	20.5
20.8	20.9	21.4	21.6	21.9	22.3	22.5	22.7	22.7	22.9
23.0	23.5	23.7	23.9	24.1	24.3	24.6	24.6	24.8	25.7
25.9	26.1	26.4	26.6	26.8	27.5	28.5	28.6	29.6	31.8

Seguendo lo stesso procedimento adottato per calcolare i quartili otteniamo:

$$n = 80, \quad p = 0.95, \quad k = np = 76$$

Essendo k intero, si fa la media tra il 76-esimo e il 77-esimo dato. Risulta:

$$P_{0.95} = \frac{27.5 + 28.5}{2} = 28$$

Il 95-esimo percentile fornisce un'importante informazione: soltanto il 5% dei dati sono maggiori di 28.

Calcolo quartili e percentili nel caso di dati raggruppati

La formula già vista qui per il calcolo della mediana nel caso di dati raggruppati in classi, si può adattare per il calcolo di qualsiasi quartile o percentile. Ad esempio, supponiamo di voler calcolare il primo quartile e il 79-esimo percentile; una volta determinata la classe di appartenenza (vedi procedura nel caso della mediana), le formule da utilizzare sono rispettivamente:

$$Q_1 = x_{i-1} + (x_i - x_{i-1}) \frac{0.25 - f_c^{i-1}}{f_c^i - f_c^{i-1}}$$

e

$$P_{0.79} = x_{i-1} + (x_i - x_{i-1}) \frac{0.79 - f_c^{i-1}}{f_c^i - f_c^{i-1}}$$

Indici di dispersione

I valori degli indici di posizione sono importanti per la descrizione sintetica di un fenomeno statistico, ma non sono sufficienti per la comprensione complessiva del fenomeno in quanto non sono in grado di fornire alcuna informazione sulla dispersione dei dati (omogeneità o disomogeneità).

Per cui si debbono analizzare altri indici statistici, detti appunto di misura della dispersione, quali i seguenti:

MISURE DI DISPERSIONE

$x_{max} - x_{min}$	RANGE (Campo di variazione)
$\frac{1}{n} \sum_{i=1}^n X_i - \bar{X} $	SCARTO MEDIO ASSOLUTO
$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$	VARIANZA (Media degli Scarti al Quadrato)
$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$	VARIANZA CAMPIONARIA (Calcolata da Excel)
$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$	DEVIAZIONE STANDARD CAMPIONARIA
$CV = \frac{\sigma}{ \bar{X} }$	COEFFICIENTE DI VARIAZIONE

Stime

Oggi studiamo le proprietà della stima che ricaviamo da un campione. Si chiama teoria della stima. La stima statistica consiste nel trarre delle conclusioni su alcune proprietà statistiche della popolazione mediante informazioni su campioni. La stima può essere:

- Puntuale, si risolve in un valore assunto a rappresentare una proprietà statistica (un parametro) della popolazione
- Intervallare, si risolve nel fissare due valori, tra cui si presume sia compreso un parametro della popolazione

Variabile casuale campionaria detta anche **Statistica campionaria** è una funzione che assume valore nell'universo dei campioni, definita quindi rispetto alla n -pla di variabili casuali (X_1, X_2, \dots, X_n)

Basandoci sull'universo dei campioni possiamo valutare le caratteristiche di una particolare statistica campionaria (es. la media), analizzandone il comportamento su tutti i potenziali campioni estraibili dalla popolazione

Una statistica campionaria ci permette di «stimare» il valore della caratteristica incognita della popolazione. Questa statistica si chiama «**Stimatore**»

Lo stimatore potrà avere vari valori al variare del campione.

Il valore realizzato sul campione effettivamente osservato si chiama «**stima**».

Stimatore e stima

P : popolazione

X : variabile oggetto di studio con una sua distribuzione

θ : Parametro incognito (media, varianza,...) da stimare
→ determinare un valore

$C: (x_1, x_2, \dots, x_n) \rightarrow$ campione casuale bernoulliano di n elementi da P ,

$(X_1, X_2, \dots, X_n) \rightarrow$ variabili casuali indipendenti e identicamente distribuite

Stimatore: $T=f(X_1, X_2, \dots, X_n)$ per un caratteristica/parametro di X

Stima puntuale: $t=f(x_1, x_2, \dots, x_n)$

Differenza tra stimatore e stima

Lo stimatore è una variabile aleatoria nell'universo dei campioni:

$T = f(X_1, X_2, \dots, X_n)$ stimatore puntuale

è **una variabile aleatoria** funzione di variabili aleatorie.

La stima non è una variabile ma un valore:

$t = f(x_1, x_2, \dots, x_n)$ stima puntuale

è **una costante del campione**, una determinazione empirica di T , ossia di uno stato di grandezza di tale variabile

Quando si conosce la famiglia di appartenenza delle v.a. che formano il c.c.s., non si conoscono i parametri caratterizzanti la distribuzione stessa, ma una sua funzione.

Tale funzione, $T_n = T(X_1, \dots, X_n)$, si chiama *stimatore (puntuale)* del parametro θ della distribuzione.

La quantità $T(x_1, \dots, x_n) = \hat{\theta}$ si chiama *stima* (o *statistica*) del parametro.

È bene osservare che, a seconda del campione estratto, si avrà una stima diversa, dunque $T(X_1, \dots, X_n) = T_n$ avrà una propria distribuzione, detta *distribuzione campionaria*.

Proprietà di uno stimatore

- Lo stimatore è uno strumento teorico che permette di dare dei giudizi sulla bontà della stima.
- È necessario individuare lo stimatore più adeguato per stimare i parametri della popolazione.
- **Proprietà finite**: valgono per qualsiasi numerosità del campione
- **Proprietà asintotiche**: valgono solo per campioni di grande numerosità

Proprietà finite di uno stimatore

- Correttezza: riferita al valore atteso
- Efficienza: riferita alla variabilità

Definizione di Correttezza: uno stimatore T è corretto se il suo valore atteso coincide col parametro θ che si vuole stimare

$$E(T) = \theta$$

- Fra tutti i campioni ce ne sono alcuni che forniscono sotto-stime e altri sovra-stime del parametro, altri ancora che danno valori molto lontani, altri molto vicini o anche uguali, Stimatore è corretto se sovra-stime e sottostime si compensano, e in media lo stimatore coincide con il valore vero incognito del parametro.
- Uno stimatore non corretto si dice **distorto** e indichiamo con B la distorsione

$$B(T) = E(T) - \theta$$

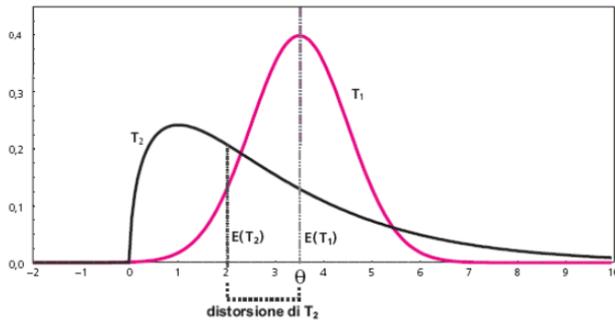
Correttezza. Lo stimatore T_n si dice *corretto* o *non distorto* se il suo valore atteso coincide con il valore teorico del parametro θ da stimare:

$$E(T_n) = \theta .$$

Se lo stimatore non è corretto, si dice anche essere *distorto*. La quantità $b(T_n) = E(T_n) - \theta \neq 0$ si chiama *distorsione* o *bias* dello stimatore.

Esempio

Si considerino i due stimatori T_1 e T_2



T_1 è corretto mentre T_2 no è quindi **distorto**

Media Aritmetica campionaria

Esempio: La media aritmetica campionaria è uno stimatore corretto per la media della X nella popolazione.

$$E(\bar{X}) = \mu?$$

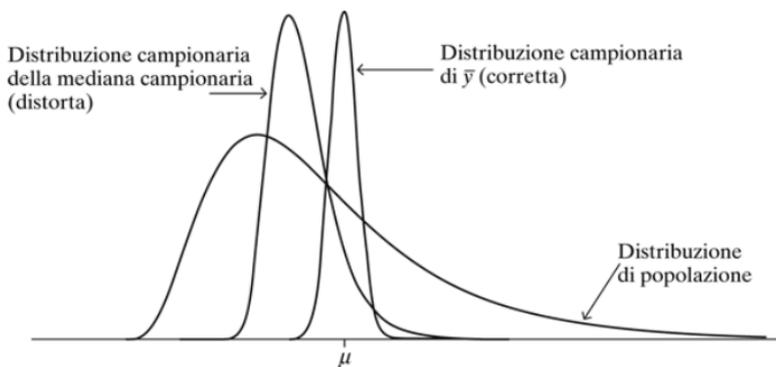
•Lo abbiamo visto empiricamente nell'esempio sul voto di maturità, vediamo ora teoricamente

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\sum_{i=1}^n E(X_i)}{n} = \frac{n\mu}{n} = \mu$$

La dimostrazione si basa:

- sulla proprietà per cui il valore atteso di una somma è uguale alla somma dei valori attesi
- Tutte le X_i hanno la stessa media μ in virtù del piano di campionamento bernoulliano

Correttezza



Varianza campionaria

Esempio: La varianza campionaria è uno stimatore corretto per la varianza della X nella popolazione?

$$E(S) = \sigma^2?$$

•Lo stimatore S è definito come:

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

Si dimostra che

$$E(S) \neq \sigma^2$$

Pertanto lo stimatore è **distorto**

Uno stimatore corretto si può ottenere in questo modo

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

Efficienza di uno stimatore

- Un'altra proprietà desiderabile per uno stimatore è quella di essere poco variabile, quindi di determinare in media stime del parametro più vicine al valore vero incognito
- La variabilità solitamente è misurata dalla deviazione standard che nel caso di uno stimatore è anche detta **errore standard**.
- Efficienza è una proprietà relativa e riguarda la variabilità di uno stimatore.
- Uno stimatore è detto più **efficiente** di un altro se determina stime del parametro più vicine al vero valore, in media, rispetto ad altri stimatori.
- Si parla di efficienza di uno stimatore in termini di confronto con quella di un altro stimatore

Efficienza. Lo stimatore T_n si dice *efficiente* se rende minimo il grado di dispersione della sua distribuzione, rispetto al parametro θ oggetto di stima.

Dati due stimatori $T_n^{(1)}$ e $T_n^{(2)}$, si dice che $T_n^{(1)}$ è *più efficiente* di $T_n^{(2)}$ se

$$EQM(T_n^{(1)}) \leq EQM(T_n^{(2)}) .$$

Se i due stimatori sono corretti, per quanto detto finora, $EQM(T_n) = Var(T_n)$, dunque $T_n^{(1)}$ è più efficiente di $T_n^{(2)}$ se

$$Var(T_n^{(1)}) \leq Var(T_n^{(2)}) .$$

Una statistica $T_n = T_n(X_1, \dots, X_n)$ risulta il miglior stimatore del parametro θ se è il più efficiente tra gli stimatori corretti e consistenti.

Nella pratica, si usa il metodo *BLUE* (Best Linear Unbiased Estimator), che consiste nello scegliere lo stimatore nella classe degli stimatori *lineari, corretti, con varianza minima*. Tale risultato discende dal teorema di Cramer-Rao, che fornisce un valore che minimizza la varianza, dato da

$$\frac{1}{n \cdot E \left[\frac{\partial^2}{\partial \theta^2} (\ln f(x, \theta)) \right]}$$

Un metodo per determinare la stima puntuale dei parametri va sotto il nome di *metodo della massima verosimiglianza*. Siano X_1, \dots, X_n n v.a.i.i.d., con densità $f(x_i, \theta)$, $i = 1, \dots, n$.

Consideriamo la quantità (nota come *funzione di verosimiglianza*)

$$L(X_1, \dots, X_n; \theta) := \prod_{i=1}^n f(x_i; \theta)$$

Cerchiamo il valore da assegnare al parametro θ affinché la funzione di verosimiglianza sia massima (i.e., assuma il valore più grande possibile.) Per far questo, se $L(X, \theta)$ è derivabile rispetto a θ , allora esistono le derivate di L . Si procede ponendo uguale a 0 tali derivate e si risolve il sistema così ottenuto. Per alleggerire i calcoli, invece di considerare la funzione di verosimiglianza L , si preferisce studiare la funzione di *log-verosimiglianza*, $\ln(L)$.

Efficienza

Per valutare la variabilità di T intorno a θ possiamo usare la varianza (o anche l'errore standard) ma se lo stimatore è distorto è più opportuno usare l'**errore quadratico medio** dato dal valore atteso della differenza al quadrato tra lo stimatore e il valore incognito che si vuole stimare:

$$MSE(T) = E[(T - \theta)]^2$$

Si dimostra che:

$$MSE(T) = E[(T - \theta)]^2 = Var(T) + B(T)^2$$

$$Var(T) = E[T - E(T)]^2$$

dove

Diciamo che lo stimatore T_1 è più efficiente di T_2 se

$$MSE(T_1) < MSE(T_2)$$

Per tutti i possibili valori di θ

Con MSE \rightarrow Mean Squared Error

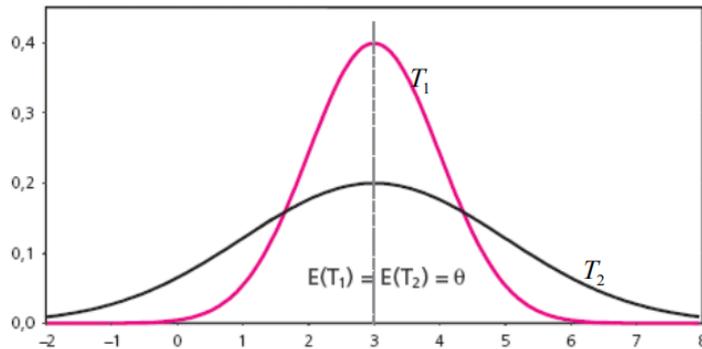
Efficienza

Se lo stimatore è corretto e quindi è nulla la distorsione si ha:

$$MSE(T) = Var(T)$$

Dati due stimatori corretti T_1 e T_2 . Si dirà che T_1 è più efficiente di T_2 se

$$Var(T_1) < Var(T_2)$$



Proprietà asintotiche

Consistenza

Lo stimatore T_n di un parametro θ , dove l'indice indica la dipendenza dello stimatore dalla numerosità campionaria, è uno stimatore **consistente in media quadratica** se

$$\lim_{n \rightarrow \infty} MSE(T_n) = \lim_{n \rightarrow \infty} E(T_n - \theta)^2 = 0$$

Quindi

$$\lim_{n \rightarrow \infty} MSE(T_n) = 0 \quad \text{se e solo se} \quad \lim_{n \rightarrow \infty} Var(T_n) = 0 \quad \lim_{n \rightarrow \infty} B(T_n) = 0$$

Correttezza asintotica

Uno stimatore T_n di un parametro θ è uno **asintoticamente corretto** se:

$$\lim_{n \rightarrow \infty} E(T_n) = \theta$$

per ogni possibile valore di θ

Consistenza. Lo stimatore T_n si dice *consistente (in media quadratica)* se, al crescere della taglia n del c.c.s., la distribuzione campionaria di T_n si concentra sempre di più attorno a θ , cioè

$$\lim_{n \rightarrow \infty} E[(T_n - \theta)^2] = 0 .$$

La quantità $E[(T_n - \theta)^2]$ si chiama *errore quadratico medio (EQM)* e misura di quanto, in media, le realizzazioni di T_n , su tutti i c.c.s. di taglia n , distano da θ .

Una formula operativa per il calcolo di EQM è la seguente:

$$EQM(T_n) = Var(T_n) + b^2(T_n) .$$

Infatti:

$$\begin{aligned} EQM(T_n) &= E[(T_n - \theta)^2] = E[T_n^2 + \theta^2 - 2\theta T_n] \\ &= E[T_n^2] + \theta^2 - 2\theta E[T_n] \\ &= E[T_n^2] - E[T_n]^2 + E[T_n]^2 + \theta^2 - 2\theta E[T_n] \\ &= Var(T_n) + b^2(T_n) . \end{aligned}$$

Distribuzione di probabilità di alcuni stimatori

- Se la variabile X nella popolazione si distribuisce normalmente anche la media aritmetica campionaria si distribuisce normalmente
- Se la X non ha distribuzione normale (ad esempio bernoulliana) possiamo fare ricorso al

Teorema del Limite Centrale: la somma (o la media) di un numero elevato di variabili casuali indipendenti con la stessa distribuzione è approssimativamente una normale.

- In altre parole per campioni casuali di elevata ampiezza n , la distribuzione di una qualsiasi variabile casuale campionaria che presenta queste caratteristiche ha approssimativamente una distribuzione normale.
- Quindi nel caso di variabile X bernoulliana nella popolazione, lo stimatore Media aritmetica campionaria ha per grandi campioni distribuzione approssimativamente normale
- Queste considerazioni servono per costruire intervalli di confidenza e verificare ipotesi

Distribuzione degli stimatori

Media campionaria. Sia $\{X_1, \dots, X_n\}$ un c.c.s. estratto dalla popolazione. Si definisce *media campionaria* la v.a.

$$\bar{X} := \frac{X_1 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Valore atteso di \bar{X} : se $E(X_i) = \mu$, per ogni $i = 1, \dots, n$, allora

$$E(\bar{X}) = E\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{E(X_1) + \dots + E(X_n)}{n} = \mu.$$

Quindi, la distribuzione della media campionaria è centrata sulla media delle singole v.a. che formano il campione. Questo garantisce che la media campionaria è uno stimatore corretto della media di popolazione.

Varianza di \bar{X} : se $Var(X_i) = \sigma^2$, per ogni $i = 1, \dots, n$, allora

$$Var(\bar{X}) = Var\left(\frac{X_1 + \dots + X_n}{n}\right) = Var\left(\frac{X_1}{n}\right) + \dots + Var\left(\frac{X_n}{n}\right) = \frac{\sigma^2}{n}.$$

Quindi, la varianza della media campionaria è inversamente proporzionale alla taglia del c.c.s. Inoltre,

$$EQM(\bar{X}) = Var(\bar{X}) \rightarrow 0, \text{ se } n \rightarrow \infty,$$

ossia, la media campionaria è anche uno stimatore consistente della media di popolazione.

Se $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $i = 1, \dots, n$, allora $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$. (senza dimostrazione).

Se, invece, il c.c.s. $\{X_1, \dots, X_n\}$ non proviene da una popolazione gaussiana, ma la taglia del campione è abbastanza grande, e supponendo $m = E(X_i)$, $v^2 = Var(X_i)$, $i = 1, \dots, n$ si ha $\bar{X} \sim \mathcal{N}(m, v^2/n)$ (la dimostrazione segue immediatamente dall'applicazione del Teorema del limite centrale).

Varianza campionaria. Sia $\{X_1, \dots, X_n\}$ un c.c.s. estratto dalla popolazione. Si definisce *varianza campionaria* la v.a.

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

Valore atteso di S^2 : innanzitutto, osserviamo che

$$\begin{aligned} S^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n (X_i^2 + \bar{X}^2 - 2\bar{X}X_i) \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 + \frac{1}{n} \sum_{i=1}^n \bar{X}^2 - 2\frac{1}{n} \sum_{i=1}^n \bar{X}X_i \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 + \bar{X}^2 - 2\bar{X} \underbrace{\frac{1}{n} \sum_{i=1}^n X_i}_{n\bar{X}} \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 + \bar{X}^2 - 2\bar{X}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \end{aligned}$$

Dunque, si ha:

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

D'altro canto, posto $m = E(X_i)$ e $v^2 := \text{Var}(X_i)$, $i = 1, \dots, n$ e ricordando la formula operativa $E(X_i^2) = \text{Var}(X_i) + E(X_i)^2 = v^2 + m^2$ della varianza di v.a., avremo quindi

$$\begin{aligned} E(S^2) &= \frac{1}{n} \sum_{i=1}^n (\text{Var}(X_i) + E(X_i)^2) - (\text{Var}(\bar{X}) + E(\bar{X})^2) \\ &= \frac{1}{n} \sum_{i=1}^n (v^2 + m^2) - \left(\frac{v^2}{n} + m^2\right) \\ &= v^2 + m^2 - \frac{v^2}{n} - m^2 = (n-1) \frac{v^2}{n}. \end{aligned}$$

Questo ci dice che la varianza campionaria non è uno stimatore corretto, in quanto $E(S^2) \neq \frac{v^2}{n}$.

Si può correggere la varianza campionaria nel seguente modo:

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

\bar{S}^2 si chiama *varianza campionaria corretta*. Inoltre, si può dimostrare (ma noi non lo faremo) che

$$EQM(\bar{S}^2) = Var(\bar{S}^2) \rightarrow 0, \text{ se } n \rightarrow \infty,$$

ossia, che la varianza campionaria corretta è anche uno stimatore consistente della varianza di popolazione.

Se il c.c.s. proviene da una popolazione gaussiana, \bar{S}^2 è uno stimatore efficiente (senza dimostrazione).

Si dimostra che

$$S^2 \sim \frac{\sigma^2}{n-1} \chi_{n-1}^2,$$

da cui

$$\bar{S}^2 \sim \chi_{n-1}^2.$$

Intervalli di confidenza: classici, unilaterali e bilaterali

Stima per intervalli

Se il carattere in una popolazione ha una distribuzione normale con media μ e varianza σ^2 , anche lo stimatore della media $Th = \bar{X}$ si distribuisce come una normale con media μ e varianza σ^2/n , dove n è la numerosità del campione

Possiamo usare tale risultato per chiederci:

Quali sono gli estremi dell'intervallo all'interno del quale il nostro stimatore per una prefissata percentuale di possibili risultati campionari?

Tale percentuale si chiama **livello di confidenza** ed è usualmente indicata come **(1-a)**, con a compreso tra 0 e 1.

Con la stima puntuale, si ottiene un valore empirico *approssimato* del parametro da stimare. Quindi, se si conoscesse la distribuzione campionaria della statistica usata per la stima, si potrebbe valutare il grado di errore commesso e determinare, con una probabilità molto prossima a 1, l'intervallo in cui si trova il valore vero del parametro che stiamo stimando.

Si chiama *intervallo di confidenza* l'intervallo $[\theta_{min}, \theta_{max}]$ tale che

$$P(\theta_{min} \leq \theta \leq \theta_{max}) > \alpha,$$

dove α rappresenta il *livello di significatività* misurato in percentuale, della stima effettuata.

Stima intervallare: l'intervallo di confidenza

- Per avere un'informazione più completa sul parametro incognito che vogliamo stimare, è utile dare indicazioni su quanto precisa sia la stima rispetto al vero valore del parametro.
- Le indicazioni sulla precisione della stima puntuale sono basate sull'ampiezza della **stima intervallare** di un parametro.
- Poiché le stime intervallari contengono il parametro con un certo livello di fiducia, essi vengono indicati come **intervalli di confidenza**.
- Un **intervallo di confidenza** per un parametro è un intervallo di valori entro cui si ritiene ricada il valore di un parametro.
- La probabilità associata al fatto che l'intervallo contenga il parametro è denominata **livello di confidenza**. Questo è un numero prossimo a 1, tipicamente 0,95 o 0,99 indicato come

$$1 - \alpha$$

Intervallo di confidenza

- La costruzione di un intervallo di confidenza è legata alla distribuzione campionaria dello stimatore puntuale.
- In termini non rigorosi per costruire un intervallo di confidenza, si aggiunge e si sottrae dalla stima puntuale un multiplo del suo errore standard.
- Questo multiplo dell'errore standard è il **marginale di errore**.
- Un intervallo di confidenza è quindi dato da:

Stima puntuale \pm Margine di errore

Definizione 6.0.1 (Intervallo di confidenza). Sia X_1, X_2, \dots, X_n un campione statistico e sia θ un parametro (ignoto) che caratterizza la distribuzione del campione.

Siano $L_i = l_i(X_1, X_2, \dots, X_n)$ e $L_s = l_s(X_1, X_2, \dots, X_n)$ due statistiche del campione e sia $\alpha \in (0, 1)$. Dico che l'intervallo (L_i, L_s) è un *intervallo di confidenza* (o di fiducia) di livello $1 - \alpha$ se $\mathbb{P}(\theta \in (L_i, L_s)) \geq 1 - \alpha$, ovvero che (L_i, L_s) è un intervallo di confidenza (o di fiducia) di errore α se $\mathbb{P}(\theta \notin (L_i, L_s)) \leq \alpha$.

Dico che la semiretta $(L_i, +\infty)$ è un *intervallo di confidenza unilaterale superiore* di livello $1 - \alpha$ se $\mathbb{P}(\theta > L_i) \geq 1 - \alpha$

Dico che la semiretta $(-\infty, L_s)$ è un *intervallo di confidenza unilaterale inferiore* di livello $1 - \alpha$ se $\mathbb{P}(\theta < L_s) \geq 1 - \alpha$

Osservazione 6.0.3. 1. La scelta dei nomi delle due statistiche non è casuale: L_i sta per limitazione inferiore mentre L_s sta per limitazione superiore.

2. Di solito si è interessati a *piccoli* valori di α , più precisamente a $\alpha \in (10^{-2}, 10^{-1})$.
3. La disuguaglianza di Chebyshev ci ha fornito un intervallo di confidenza per la media μ del campione nel caso in cui la varianza σ^2 sia nota

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

ovvero

$$\mathbb{P}(|\bar{X} - \mu| < t) \geq 1 - \frac{\sigma^2}{t^2} \quad \forall t > 0$$

cioè

$$\mathbb{P}(\bar{X} - t < \mu < \bar{X} + t) \geq 1 - \frac{\sigma^2}{t^2} \quad \forall t > 0.$$

Fissato $\alpha \in (0, 1)$ scelgo $t = \frac{\sigma}{\sqrt{\alpha}}$. La disuguaglianza di Chebyshev si legge allora

$$\mathbb{P}\left(\bar{X} - \frac{\sigma}{\sqrt{\alpha}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{\alpha}}\right) \geq 1 - \alpha \quad \forall \alpha \in (0, 1).$$

6.1.1. Campione gaussiano di cui è nota la varianza

Intervallo bilaterale

Sia X_1, X_2, \dots, X_n un campione gaussiano di media μ incognita e varianza σ^2 nota.

Sia Z una v.a. gaussiana standard e sia $\alpha \in (0, 1)$. Calcolo $\mathbb{P}\left(|Z| \leq z_{1-\frac{\alpha}{2}}\right)$:

$$\begin{aligned} \mathbb{P}\left(|Z| \leq z_{1-\frac{\alpha}{2}}\right) &= \mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}\right) \\ &= \mathbb{P}\left(Z \leq z_{1-\frac{\alpha}{2}}\right) - \mathbb{P}\left(Z \leq -z_{1-\frac{\alpha}{2}}\right) = \mathbb{P}\left(Z \leq z_{1-\frac{\alpha}{2}}\right) - \mathbb{P}\left(Z \leq z_{\frac{\alpha}{2}}\right) \quad (6.1) \\ &= \Phi\left(z_{1-\frac{\alpha}{2}}\right) - \Phi\left(z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha. \end{aligned}$$

Sappiamo che $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ e che dunque $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$. Applichiamo quindi la

disuguaglianza (6.1) a $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$. Si ha:

$$\begin{aligned} 1 - \alpha &= \mathbb{P}\left(\left|\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right| \leq z_{1-\frac{\alpha}{2}}\right) = \mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\mu - \bar{X}}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\frac{\alpha}{2}}\right) \\ &= \mathbb{P}\left(\frac{-\sigma z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leq \mu - \bar{X} \leq \frac{\sigma z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) \quad (6.2) \\ &= \mathbb{P}\left(\bar{X} - \frac{\sigma z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{\sigma z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) \end{aligned}$$

L'intervallo

$$\left(\bar{X} - \frac{\sigma z_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \bar{X} + \frac{\sigma z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right)$$

è dunque un intervallo di confidenza di livello $1 - \alpha$ per la media μ del campione.

Intervallo unilaterale superiore

Sia $Z \sim \mathcal{N}(0, 1)$. Sappiamo che

$$\mathbb{P}(Z \leq t) = 1 - \alpha \quad \text{se e solo se} \quad t = z_{1-\alpha}.$$

Abbiamo dunque

$$1 - \alpha = \mathbb{P}\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\alpha}\right) = \mathbb{P}\left(\bar{X} - \mu \leq \frac{\sigma z_{1-\alpha}}{\sqrt{n}}\right) = \mathbb{P}\left(\mu \geq \bar{X} - \frac{\sigma z_{1-\alpha}}{\sqrt{n}}\right).$$

Quindi la semiretta

$$\left(\bar{X} - \frac{\sigma z_{1-\alpha}}{\sqrt{n}}, +\infty\right)$$

è un intervallo di confidenza unilaterale superiore di livello $1 - \alpha$.

Intervallo unilaterale inferiore

Sia $Z \sim \mathcal{N}(0, 1)$. Sappiamo che

$$\mathbb{P}(Z \geq t) = 1 - \alpha \quad \text{se e solo se} \quad \mathbb{P}(Z \leq t) = \alpha \quad \text{se e solo se} \quad t = z_\alpha.$$

Abbiamo dunque

$$1 - \alpha = \mathbb{P}\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq z_\alpha\right) = \mathbb{P}\left(\bar{X} - \mu \geq \frac{\sigma z_\alpha}{\sqrt{n}}\right) = \mathbb{P}\left(\mu \leq \bar{X} - \frac{\sigma z_\alpha}{\sqrt{n}}\right).$$

Quindi la semiretta

$$\left(-\infty, \bar{X} - \frac{\sigma z_\alpha}{\sqrt{n}}\right) = \left(-\infty, \bar{X} + \frac{\sigma z_{1-\alpha}}{\sqrt{n}}\right)$$

è un intervallo di confidenza unilaterale inferiore di livello $1 - \alpha$.

Costruzione di un intervallo di confidenza.

1. Si sceglie il livello di confidenza α (oppure $1 - \alpha$)
2. Si identifica uno stimatore T di θ
3. Si cerca una funzione di T e di θ , detta *quantità pivotale* e indicata con $Q(T, \theta)$, la cui distribuzione di probabilità sia completamente nota
4. Indichiamo con $q_{\alpha/2}$ e $q_{1-\alpha/2}$ i quantili della quantità pivotale di ordine $\alpha/2$ e $1 - \alpha/2$, rispettivamente, e si scrive l'intervallo di confidenza come segue:

$$P(q_{\alpha/2} \leq \theta \leq q_{1-\alpha/2}) > \alpha$$

5. Si ricava l'intervallo di confidenza, risolvendo la disequazione rispetto al parametro θ

Maggiori al riferimento: <https://www.corsi.univr.it/documenti/OccorrenzaIns/matdid/matdid247804.pdf>

Intervallo di confidenza per la media

- Consideriamo una variabile X distribuita normalmente

$$X \approx N(\mu, \sigma^2)$$

con media incognita e varianza che ipotizziamo nota.

- Si sa che lo stimatore media aritmetica campionaria Allora

$$\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Standardizziamo la variabile

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

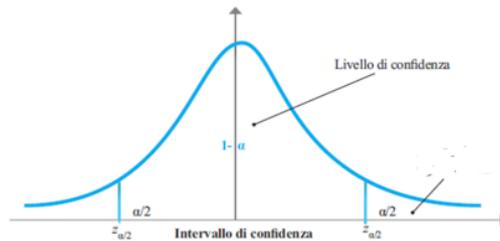
Intervallo di confidenza per la media

- Si può determinare la probabilità di un valore compreso tra due valori $+z$ e $-z$
- Scegliamo come valore di confidenza $1 - \alpha = 0,95$
- allora dalle tavole appositamente calcolate per la normale standardizzata sappiamo che $+z_{\alpha/2} = 1,96$ e $-z_{\alpha/2} = -1,96$

$$p(-z_{\alpha/2} < z < +z_{\alpha/2}) = 1 - \alpha$$

$$p\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < +z_{\alpha/2}\right) = 1 - \alpha$$

$$p\left(\bar{X} - z_{\alpha/2}\sigma / \sqrt{n} < \mu < \bar{X} + z_{\alpha/2}\sigma / \sqrt{n}\right) = 1 - \alpha$$



Intervallo di confidenza per la media

Si ha quindi

$$p\left(\bar{X} - \frac{1,96\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{1,96\sigma}{\sqrt{n}}\right) = 1 - 0,05 = 0,95$$

La quantità

$$\bar{X} \pm 1,96 \times \frac{\sigma}{\sqrt{n}}$$

È aleatoria fino a quando non estraiamo un campione. Possiamo solo dire che il 95% degli intervalli così costruiti contiene il valore incognito μ

Estraiamo quindi il campione e allo stimatore \bar{X} sostuiamo la stima nel campione \bar{x} allora determiniamo uno dei possibili campioni

$$\bar{x} \pm 1,96 \times \frac{\sigma}{\sqrt{n}}$$

Che contiene il valore μ con un livello di confidenza del 0,95

Esempio

La temperatura massima X a Palermo nel mese di aprile si distribuisce come una normale di media incognita μ e di varianza pari a 50.

La media delle temperature massima registrate negli ultimi 25 giorni dello scorso aprile è risultata 27. Si costruisca l'intervallo di confidenza al livello 0,95.

La varianza della temperatura è 50, e quindi la varianza della media campionaria è $\sigma^2/n = 50/25 = 2$

$$\bar{x} \pm 1,96 \times \sqrt{2}$$

$$\bar{x} \pm 1,96 \times 1,41$$

$$\bar{x} \pm 2,8$$

Estremo inferiore $27 - 2,8 = 24,2$ Estremo superiore $27 + 2,8 = 29,8$

$$[24,2; 29,8]$$

Intervallo di confidenza per la media Grandi campioni

Qualora la varianza σ^2 non sia nota allora si deve stimarla. Si usa in questo caso la varianza campionaria corretta

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

Che nel campione estratto diventa

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Questa quantità si sostituisce nella formula $\bar{x} \pm 1,96 \times \frac{\sigma}{\sqrt{n}}$

Ottenendo

$$\bar{x} \pm 1,96 \times \frac{s}{\sqrt{n}}$$

Questo risultato vale se la numerosità del campione è elevata. Se il campione è piccolo allora bisogna procedere in modo diverso

Intervallo di confidenza per la media Piccoli campioni

La variabile $\frac{\bar{X} - \mu}{s/\sqrt{n}}$

Non ha più distribuzione normale standardizzata ma distribuzione di probabilità detta *t* di Student.

- La distribuzione *t* è campanulare e simmetrica intorno alla media 0.
- La sua deviazione standard è leggermente più grande di 1 e il valore esatto dipende da quelli che vengono chiamati gradi di libertà indicati con *gdl che sono pari a n-1*
- La distribuzione *t* presenta un'ampiezza leggermente diversa per ciascun differente valore dei *gdl*.
- Quanto più elevato è il valore dei *gdl* tanto più la distribuzione tenderà a rassomigliare a una normale standardizzata.

Intervallo di confidenza per la media Piccoli campioni-Esempio

Torniamo all'esempio sulla temperatura a Palermo. Immaginiamo quindi di non conoscere la varianza. Sappiamo però che nel campione la varianza campionaria corretta è risultata pari a 75

La varianza della media campionaria è $\sigma^2/n = 75/25 = 3$

$$\bar{x} \pm t_{\alpha/2; n-1} \times \sqrt{3}$$

Dalle tavole della t per

$$1 - \alpha = 0,95 \quad \alpha/2 = 0,025$$

$$n - 1 = 24 \quad t_{0,025; 24} = 2,0639$$

$$\bar{x} \pm t_{\alpha/2; n-1} \times \sqrt{3} \quad 27 \pm 2,0639 \times 1,73 \quad 27 \pm 3,57$$

Estremo inferiore $27 - 3,57 = 23,43$ Estremo superiore $27 + 3,57 = 30,57$

L'intervallo è più ampio di quello del caso di varianza nota