



UNIVERSITÀ DI PISA

Dipartimento di Informatica
Corso di Laurea Triennale in Informatica

Corso 2° anno - 6 CFU

Statistica

Professore:
Prof. Francesco Grotto

Autore:
Filippo Ghirardini

Anno Accademico 2023/2024

Contents

1	Statistica descrittiva	4
1.0.1	Campioni statistici	4
1.1	Rappresentazione grafica	4
1.1.1	Diagramma a barre	5
1.1.2	Diagramma a torta	5
1.1.3	Istogramma	5
1.2	Indici statistici	6
1.2.1	Indici di centralità	6
1.2.2	Indici di dispersione	7
1.2.3	Quantili	8
1.3	Dati multi-variati	9
2	Probabilità e indipendenza	12
2.1	Spazi di probabilità	12
2.2	Probabilità	12
2.3	Modello uniforme	13
2.3.1	Calcolo combinatorio	13
2.4	Probabilità condizionata	14
2.5	Indipendenza	14
2.6	Entropia di Shannon	15
2.7	Densità di probabilità	15
3	Variabili aleatorie	16
3.1	Legge di una variabile aleatoria	16
3.2	Tipi di variabili aleatorie	16
3.2.1	Variabili discrete	16
3.2.2	Variabili continue	16
3.3	Funzione di ripartizione	16
3.3.1	Funzioni di variabili discrete	17
3.3.2	Funzioni di variabili continue	17
3.4	β -quantile	17
3.5	Variabili discrete notevoli	18
3.5.1	Binomiali	18
3.5.2	Geometriche	18
3.5.3	Ipergeometriche	18
3.5.4	Poisson	19
3.6	Variabili con densità notevoli	19
3.6.1	Uniformi su intervalli	19
3.6.2	Esponenziali	19
3.6.3	Gamma	19
3.6.4	Pareto	20
3.6.5	Gaussiane standard	20
3.6.6	Chi-Quadro	21
3.6.7	Student	21
3.6.8	Gaussiane non standard	21
3.7	Trasformazioni di variabili con densità	22
3.8	Valore atteso	22
3.8.1	Valore atteso di trasformazioni	23
3.8.2	Momenti	23
3.8.3	Varianza di una variabile aleatoria	23
3.8.4	Momenti notevoli	24
3.9	Variabili doppie	25
3.9.1	Distribuzioni marginali	25
3.9.2	Variabili doppie discrete	26

3.9.3	Variabili doppie con densità	26
3.10	Indipendenza di variabili aleatorie	26
3.10.1	Indipendenza di variabili doppie	26
3.10.2	Indipendenza di funzioni di variabili indipendenti	27
3.11	Correlazione	27
3.12	Covarianza	27
3.13	Teoremi limite	28
3.13.1	Legge debole dei grandi numeri	28
3.13.2	Teorema centrale del limite	28
4	Stima parametrica	30
4.1	Campioni	30
4.2	Stimatori	30
4.2.1	Scelta di uno stimatore	31
5	Intervalli di fiducia	32
5.1	Campione Gaussiano	32
5.1.1	Bilateri	32
5.1.2	Unilateri	33
5.1.3	Intervalli per la varianza	33
5.2	Campione di Bernoulli	33
6	Test statistici	34
6.1	Formulazione	34
6.1.1	Ipotesi	34
6.1.2	Regione critica	34
6.1.3	Valutazione del test	34
6.2	Tipologie di test	35
6.2.1	Z-Test	35

Statistica

Realizzato da: Filippo Ghirardini

A.A. 2023-2024

1 Statistica descrittiva

La statistica si occupa dello studio dei dati, ovvero della sua **raccolta**, **analisi** ed **interpretazione**. Le risposte dipendono dai dati e dalla conoscenza pregressa del problema, quindi da eventuali ipotesi ed assunzioni.

- Statistica **descrittiva**: quando i dati vengono analizzati senza fare assunzioni esterne per evidenziarne la struttura e rappresentarli in modo efficace
- **Inferenza statistica**: studia i dati utilizzando un modello probabilistico, ovvero supponendo che i dati siano valori assunti da *variabili aleatorie* con una certa *distribuzione di probabilità* dipendente da parametri non noti che devono essere stimati. Il modello potrà poi fare previsioni.

1.0.1 Campioni statistici

Definizione 1.0.1 (Popolazione o Universo). *Insieme di oggetti o fenomeni che si vuole studiare. Può essere **ideale**, ovvero tutti i possibili esiti, o **reale**.*

Definizione 1.0.2 (Carattere). *Caratteristica degli individui della popolazione, ottenuta con la stessa procedura. Può essere:*

- **Quantitativo** se gli esiti sono numeri paragonabili tra loro
- **Qualitativo** altrimenti

Definizione 1.0.3 (Modalità). *Uno dei possibili valori o attributi che può assumere il carattere. Se il carattere è quantitativo, si usa il termine **valore**.*

Definizione 1.0.4 (Campione statistico). *Un sottoinsieme della popolazione scelto per rappresentarla.*

Definizione 1.0.5 (Dati). *Gli esiti delle misure dei caratteri considerati sugli elementi del campione.*

Osservazione 1.0.1 (Campione rappresentativo). Il problema di quando un campione sia o meno rappresentativo di tutta la popolazione è complesso e, in un certo senso, dà origine alla seguente distinzione:

- Statistica **descrittiva** o **deduttiva**: analizza il gruppo senza trarre alcuna conclusione su quello più grande
- Statistica **induttiva** o **referenziale**: dato un campione cerca di trarre conclusioni circa la popolazione. Queste sono valide se e solo se il campione è veramente significativo. Non ci sono certezze e le conclusioni sono rappresentate nel linguaggio della **probabilità**.

Definizione 1.0.6 (Frequenza di una modalità). *Può essere:*

- **Assoluta**: il numero di volte in cui la modalità compare nei dati
- **Relativa**: frazione di volte in cui la modalità compare sul totale dei dati

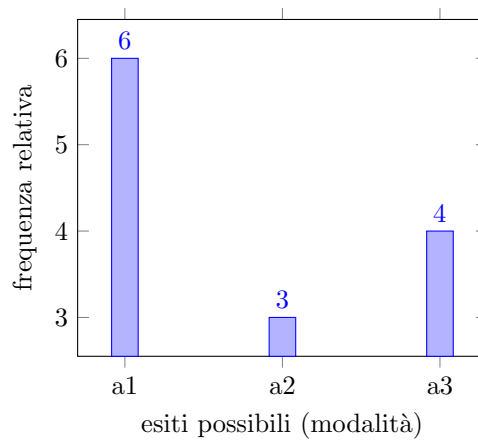
1.1 Rappresentazione grafica

La rappresentazione grafica dei dati di un campione dipende dalla natura del carattere:

- **Discreto**: se può assumere una quantità finita e relativamente piccola di modalità o valori
- **Continuo**: se può assumere un grande numero di dati

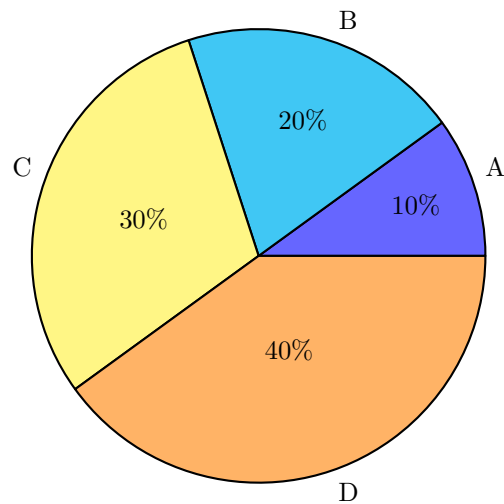
1.1.1 Diagramma a barre

Permette di rappresentare caratteri discreti.



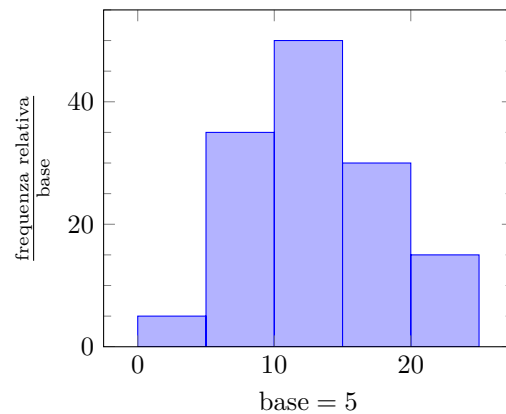
1.1.2 Diagramma a torta

Sempre per rappresentare caratteri discreti, specialmente se le modalità sono poche.



1.1.3 Istogramma

Consiste in una serie di colonne ognuna delle quali ha per base un intervallo numerico e per area la frequenza relativa dei dati contenuti nell'intervallo.



$$\text{Area} = \frac{\text{frequenza assoluta}}{\text{totale}} = b \cdot h$$

Osservazione 1.1.1. In alcuni libri di testo si prende come *altezza*, invece che come area, la frequenza relativa dei dati nell'intervallo considerato. Se la base è uguale per tutti i rettangoli, ovvero tutti gli intervalli hanno la stessa ampiezza, la rappresentazione va comunque bene. Il problema nasce quando vogliamo utilizzare intervalli con ampiezze diverse.

Osservazione 1.1.2. La scelta delle ampiezze degli intervalli di base è cruciale. L'idea è di cercare un compromesso che ci permetta di non perdere troppe informazioni (ampiezza troppo grande) senza però rendere l'istogramma troppo dipendente da piccole variazioni (troppi intervalli).

1.1.3.1 Forme

Un istogramma può avere varie forme:

- **Normale** se ha la forma di una *campana simmetrica*
- **Unimodale** se si concentra su una colonna più alta o **bimodale** se su due. Può essere asimmetrica a *destra* o a *sinistra* in base alla concentrazione dei dati in base al picco

1.2 Indici statistici

Dato un vettore $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ di dati numerici (quantitativi) di un campione di taglia n , gli indici statistici sono quantità numeriche che riassumono alcune proprietà significative della distribuzione dei dati x_1, \dots, x_n . Possono essere:

- Centralità o di posizione, e.g. media, mediana, moda e quantile
- Dispersione o di variabilità, e.g. varianza e deviazione standard
- Concentrazione
- Diversità
- Correlazione, e.g. covarianza
- Forma, e.g. indice di simmetria e curtosi

1.2.1 Indici di centralità

Gli indici di centralità indicano dove si trova il centro di una distribuzione di dati.

Definizione 1.2.1 (Media campionaria). *La media aritmetica dei dati:*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

Definizione 1.2.2 (Media campionaria date le frequenze assolute). *Supponiamo di avere dei dati x_1, \dots, x_k che capitano rispettivamente f_1, \dots, f_n volte, ovvero con f_i frequenza assoluta del valore x_i nel campione. Allora:*

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_k \cdot f_k}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{i=1}^k x_i \cdot f_i}{\sum_{i=1}^k f_i} = \frac{\sum_{i=1}^k x_i \cdot f_i}{N} \quad (2)$$

Definizione 1.2.3 (Media campionaria date le frequenze relative). *Ricordando che la frequenza relativa è $\frac{\text{frequenza assoluta}}{N}$, partendo dalla formula precedente otteniamo:*

$$\bar{x} = \frac{\sum_{i=1}^k x_i \cdot f_i}{N} = \sum_{i=1}^k x_i \cdot \frac{f_i}{N} \quad (3)$$

Definizione 1.2.4 (Media sfrondata). *È la media campionaria effettuata sui dati che rimangono dopo aver tolto una certa frazione di dati più grandi e una di quelli più piccoli.*

Definizione 1.2.5 (Mediana). *Il dato x_i tale che la metà degli altri valori è minore o uguale ad esso e l'altra metà maggiore o uguale. Operativamente:*

1. Ordino i dati x_i in modo crescente

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

2. Se n è **dispari** la mediana è il valore centrale dei dati ordinati, ovvero

$$x_{\left(\frac{n+1}{2}\right)}$$

3. Se n è **pari** la mediana è la media campionaria dei due valori centrali, over

$$\frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

Osservazione 1.2.1. La **mediana** è utile nel caso di dati molto **asimmetrici** ed è robusta rispetto alle code delle distribuzioni. Al contrario la **media campionaria** viene facilmente spostata da dati molto piccoli o grandi.

Definizione 1.2.6 (Moda). *Il valore che si presenta con la più alta frequenza, ovvero quello più comune. Nel caso di simmetria unimodale sarà il valore centrale, mentre nel caso di simmetria bimodale avremo due mode, una per campana, e la media corrisponderà alla mediana.*

1.2.2 Indici di dispersione

Questo tipo di indici misura la dispersione dei dati attorno a valori tipici individuati da quelli di posizione.

Definizione 1.2.7 (Variabilità). *Attitudine delle osservazioni ad essere una diversa dall'altra.*

Definizione 1.2.8 (Varianza campionaria). *Si usa per misurare la concentrazione dei dati attorno alla media campionaria.*

$$var(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (4)$$

È nulla se i dati sono tutti uguali. Possiamo mappare x diversamente:

- $x \mapsto x^2$ misura la media dei punti della media campionaria
- $x \mapsto x^3$ misura la **sample skewness**, ovvero l'asimmetria della distribuzione

$$b = \frac{1}{\sigma} \cdot \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \quad (5)$$

In particolare b sarà positivo quando avremo i valori più grandi a destra della media campionaria e negativo per il caso opposto.

- $x \mapsto x^4$ misura la piattezza della distribuzione dei dati, ovvero la **curtosi**

Definizione 1.2.9 (Varianza empirica). *Rappresenta la media degli scarti quadratici.*

$$var_e(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (6)$$

Dimostrazione 1.2.1.

$$var_e(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (7)$$

$$= \frac{1}{n} \cdot \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \quad (8)$$

$$= \frac{1}{n} \cdot \sum_{i=1}^n x_i^2 - 2\bar{x} \cdot \frac{1}{n} \sum_{i=1}^n x_i + \frac{1}{n} \cdot n\bar{x}^2 \quad (9)$$

$$= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \quad (10)$$

Osservazione 1.2.2. Vediamo la relazione tra la varianza campionaria e quella empirica:

$$(n-1) \cdot var_e(x) = \sum_{i=1}^n (x_i - \bar{x})^2 = n \cdot var_e(x) \Rightarrow var_e(x) = \frac{n}{n-1} \cdot var_e(x)$$

Definizione 1.2.10 (Varianza date le frequenze relative). *Supponiamo di conoscere per ogni x_i la frequenza assoluta f_i e quindi la frequenza relativa $\frac{f_i}{n}$.*

$$var_e(x) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \quad (11)$$

$$= \frac{1}{n} \underbrace{(x_1^2 + \dots + x_1^2)}_{f_1} + \dots + \underbrace{(x_n^2 + \dots + x_n^2)}_{f_n} - \bar{x}^2 \quad (12)$$

$$= \frac{f_1}{n} \cdot x_1^2 + \frac{f_2}{n} \cdot x_2^2 + \dots + \frac{f_n}{n} \cdot x_n^2 - \bar{x}^2 \quad (13)$$

$$= \sum_{i=1}^n x_i^2 \cdot \frac{f_i}{n} - \bar{x}^2 \quad (14)$$

Definizione 1.2.11 (Scarto quadratico medio o deviazione standard). *Essendo che la varianza è misurata in un'unità di misura diversa da quella dei dati (è elevata al quadrato), per ottenere un indice più facilmente interpretabile ne consideriamo la radice:*

$$\sigma(x) = \sqrt{var_e(x)} \quad \sigma_e(x) = \sqrt{var_e(x)} \quad (15)$$

Osservazione 1.2.3.

$$\sigma(x) = 0 \Leftrightarrow var_e(x) = 0 \Leftrightarrow \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 0 \Rightarrow \bar{e} = 0 \Leftrightarrow x_i = \bar{x}, \forall i$$

1.2.3 Quantili

Definizione 1.2.12 (Funzione di ripartizione empirica). *Dato $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ un vettore di dati quantitativi, la funzione empirica è la funzione $F_e : \mathbb{R} \rightarrow \mathbb{R}$ definita come:*

$$F_e(t) = \frac{\#\{i | x_i \leq t\}}{n} \quad (16)$$

*Per ogni $t \in \mathbb{R}$ restituisce la frequenza relativa dei dati minori o uguali a t . È sempre **non decrescente** e $F_e(-\infty) = 0$, $F_e(+\infty) = 1$.*

Definizione 1.2.13 (β -quantile). Il dato $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ vettore di dati quantitativi e sia k un numero naturale con $0 < k < 100$ e $\beta = \frac{k}{100}$. Definiamo il **k -esimo percentile** o **β -quantile** il dato x_i tale che:

- almeno il $k\%$ dei dati siano $\leq x_i$
- almeno $(1 - k)\%$ dei dati siano $\geq x_i$

Se due dati soddisfano le condizioni si prende la media aritmetica.

Operativamente parlando, dato il vettore x e $\beta = \frac{k}{100}$ eseguiamo:

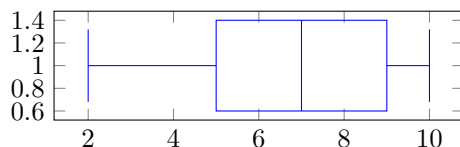
1. Ordiniamo i dati in senso crescente
2. Dato β , $\beta \cdot n$ può essere:
 - Non intero vale $x_{(\lceil \beta n \rceil)}$
 - Intero è la media aritmetica tra $x_{(\beta n)}$ e $x_{(\beta n + 1)}$

Osservazione 1.2.4. La mediana rappresenta il 50-mo percentile o $\frac{1}{2}$ -quantile o II-quartile. Con $k = 25$ abbiamo il 25-mo percentile o I-quartile e con $k = 75$ abbiamo il 75-mo percentile o III-quartile.

1.2.3.1 Boxplot

Il boxplot dei dati è un grafico che fornisce simultaneamente informazioni sulla posizione, sulla variabilità e sulla forma di una distribuzione di dati. È costituito da tre elementi:

- Una **linea** in corrispondenza della mediana, per rappresentare il centro della distribuzione
- Un **rettangolo** (box) che va dal I° al III° quartile, la cui altezza indica qual è la variabilità della metà centrale dei dati (**scarto interquartile**)
- Due **segmenti** (baffi) disegnati a partire dal rettangolo le cui lunghezze sono determinate dagli estremi dei dati x_i e x_n



Definizione 1.2.14 (Outlier). Un outlier è un valore che differisce in modo significativo dalla grande maggioranza dei dati. Possono essere derivati da:

- Errori di misurazione
- Individui molto rari

Una possibile regola per identificarli (non certa) è verificare se ci sono dati al di fuori del seguente intervallo:

$$[Q_1 - 1.5 \cdot L, Q_3 + 1.5 \cdot L] \quad L = Q_3 - Q_1$$

1.3 Dati multi-variati

A volte può capitare di avere più dati misurati su uno stesso campione, che diventano quindi k -uple di vettori del tipo $(x, y, z, \dots) \in \mathbb{R}^{k \times N}$. Nello specifico, con $k = 2$ avremo dati **bivariati** della forma:

$$(x, y) \in \mathbb{R}^{2 \cdot n}$$

Definizione 1.3.1 (Covarianza campionaria ed empirica). *Misura di quanto gli x_i e gli y_i si discostano dalle rispettive medie e se le fluttuazioni sono concordi ($cov > 0$) o discordi ($cov < 0$).*

$$cov(x, y) = \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1} \quad cov_e(x, y) = \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{n} \quad (17)$$

Osservazione 1.3.1. Notiamo due uguaglianze:

$$cov(x, y) = cov(y, x) \quad (18)$$

$$cov(x, y) \cdot (n-1) = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = n \cdot cov_e(x, y) \implies cov(x, y) = \frac{n}{n-1} \cdot cov_e(x, y) \quad (19)$$

Osservazione 1.3.2. Per eseguire i calcoli è comodo esprimere la covarianza come:

$$cov_e(x, y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} \quad (20)$$

Dimostrazione 1.3.1.

$$\begin{aligned} cov_e(x, y) &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{n} \sum_{i=1}^n [x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y}] \\ &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n \frac{x_i}{n} - \bar{x} \sum_{i=1}^n \frac{y_i}{n} + \frac{1}{n} \cdot n \bar{x} \bar{y} \\ &= \frac{1}{n} \cdot \sum_{i=1}^n x_i y_i - \bar{y} \bar{x} - \bar{x} \bar{y} + \bar{x} \bar{y} \\ &= \frac{1}{n} \cdot \sum_{i=1}^n x_i y_i - \bar{y} \bar{x} \end{aligned}$$

Definizione 1.3.2 (Coefficiente di correlazione). *Serve per determinare la forza e la direzione di una relazione lineare tra due variabili. Dati $\sigma(x) \neq 0$ e $\sigma(y) \neq 0$:*

$$r(x, y) = \frac{cov(x, y)}{\sigma(x)\sigma(y)} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (21)$$

Osservazione 1.3.3. Il coefficiente di correlazione non varia se sostituiamo alla covarianza e deviazione standard con quelle empiriche.

$$\frac{cov}{\sigma \cdot \sigma} = \frac{\frac{n}{n-1} \cdot cov_e}{\sigma_e \cdot \sigma_e \cdot \sqrt{\frac{n}{n-1}} \cdot \sqrt{\frac{n}{n-1}}} = \frac{cov_e}{\sigma_e \sigma_e}$$

Proposizione 1.3.1 (Disuguaglianza di Cauchy-Schwarz).

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \leq \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (22)$$

e quindi $r(x, y) \in [-1, 1]$, ovvero:

$$|r(x, y)| \leq 1 \quad (23)$$

Più r è vicino a 0 e più è debole la relazione, mentre più è vicino a 1 e più è forte in senso positivo (e in negativo per -1).

La **retta di regressione** è un'approssimazione dei dati con y_i con una combinazione lineare affine a $q + mx_i$, ottenuta cercando il minimo della distanza dai dati da questa retta con i quadrati degli scarti. L'obiettivo è quindi di cercare i parametri q e m calcolando

$$\min_{q, m \in \mathbb{R}} \sum_{i=1}^n (y_i - q - mx_i)^2 \quad (24)$$

Teorema 1.3.1 (Retta di regressione). Se $\sigma(x) \neq 0$ e $\sigma(y) \neq 0$, esiste un unico minimo al variare di $q, m \in \mathbb{R}$ della quantità 24, dato da:

$$m^* = \frac{\text{cov}(x, y)}{\text{var}(x)} \quad q^* = \bar{y} - m^* \bar{x} \quad (25)$$

Inoltre vale:

$$\min_{q, m \in \mathbb{R}} \sum_{i=1}^n (y_i - q - mx_i)^2 = (1 - r(x, y)^2) \sum_{i=1}^n (y_i - \bar{y})^2 \quad (26)$$

Quanto più $r(x, y)$ è vicino a ± 1 , tanto più i valori tendono ad allinearsi con la retta. Se vale ± 1 vuol dire che i punti sono tutti sulla retta. Il segno di $r(x, y)$ corrisponde al segno del coefficiente angolare. Se è prossimo a 0 allora non è una buona approssimazione, ma non esclude che non ci sia un altro tipo di relazione.

2 Probabilità e indipendenza

Molte decisioni sono prese in condizioni di incertezza. Questo perché non è possibile conoscere con certezza il futuro in quanto qualsiasi affermazione possiamo fare potrebbe avverarsi come no. La teoria della probabilità nasce per quantificare questa incertezza, misurando la fiducia che un evento possa accadere.

2.1 Spazi di probabilità

Definizione 2.1.1 (Esperimento aleatorio). *Un esperimento di cui non si conosce con certezza il risultato.*

Definizione 2.1.2 (Esito). *Un ipotetico risultato di un esperimento aleatorio.*

Definizione 2.1.3 (Spazio campionario). *Lo spazio di probabilità Ω è l'insieme di tutti gli esiti possibili ω dell'esperimento.*

Definizione 2.1.4 (Evento). *Un sottoinsieme dello spazio campionario che rappresenta un'affermazione che possiamo fare sull'esito dell'esperimento. Quando ha un solo elemento si dice **elementare**. Ω è un **evento certo** in quanto sottoinsieme improprio dello spazio campionario mentre \emptyset è un **evento impossibile**.*

Consideriamo come esperimento il lancio di un dado:

$$A = \{\omega_2, \omega_4, \omega_6\} \quad B = \{\omega_4, \omega_5, \omega_6\}$$

vediamo la corrispondenza tra le operazioni insiemistiche e logiche:

Operazione insiemistica	Operazione logica	Esempio
$A \cup B$	Accade A oppure B	$A \cup B = \{\omega_2, \omega_4, \omega_5, \omega_6\}$
$A \cap B$	Accadono A e B	$A \cap B = \{\omega_4, \omega_6\}$
A^c	Non accade A	$A^c = \{\omega_1, \omega_3, \omega_5\}$
$A \setminus B$	Accade A ma non B	$A \setminus B = A \cap B^c = \{\omega_2\}$
Relazione insiemistica	Relazione logica	Esempio
$A \subseteq B$	Se accade A allora accade B	$\{\omega_5\} \subset \{\omega_5, \omega_6\}$
$A \cap B = \emptyset$	Eventi incompatibili	C esce un numero < 2 e B

2.2 Probabilità

Osservazione 2.2.1. Se due eventi sono incompatibili la probabilità che si realizzi uno qualsiasi dei due è la somma delle probabilità dei singoli eventi.

Definizione 2.2.1 (Probabilità classica). *Consideriamo uno spazio campionario Ω relativo ad un esperimento e $A \subseteq \Sigma$ un evento. La probabilità di A è un valore $P(A) \in [0, 1]$ che misura il grado di fiducia nell'evento.*

$$P(A) = \frac{\# \text{casi favorevoli ad } A}{\# \text{casi possibili}} = \frac{\#A}{\#\Omega} \quad (27)$$

Questa definizione è adatta solo se assumiamo che gli eventi elementari sono ugualmente probabili quindi solo per esperimenti su popolazione reale.

Definizione 2.2.2 (Probabilità frequentista). *Consideriamo un evento articolato in n prove, nel corso del quale si verificano k eventi elementari $\omega_1, \dots, \omega_k$ incompatibili ma non equiprobabili. Supponiamo che ω_i si sia manifestato n_i volte (frequenza assoluta), allora:*

$$P(\omega_i) = \lim_{n \rightarrow \infty} \frac{n_i}{n} \quad (28)$$

Questa definizione è adatta ai fenomeni fisici e biologici ma non sempre è possibile effettuare davvero tante prove (anche se oggi si simula). Inoltre non è detto che esista sempre il limite.

Definizione 2.2.3 (Probabilità associativa). *Dato uno spazio campionario Ω la probabilità è una funzione $P : \mathbb{P} \rightarrow \mathbb{R}$ tale che:*

- $0 \leq P(A) \leq 1 \quad \forall A \subseteq \Omega$
- $\mathbb{P}(\Omega) = 1$
- se $(A_n)_{n=1,2,\dots}$ è una successione di eventi incompatibili vale l'**addittività**

$$P\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} P(A_n) \quad (29)$$

e se la successione è infinita

$$P\left(\bigcup_{n=1}^N A_n\right) = \sum_{n=1}^N P(A_n) \quad (30)$$

Note 2.2.1. Si dice **trascurabile** un evento A tale che $\mathbb{P}(A) = 0$ e **quasi certo** un evento A tale che $\mathbb{P}(A) = 1$.

Definizione 2.2.4 (Spazio di probabilità). *La coppia Ω, P si dice spazio di probabilità.*

Proposizione 2.2.1. *Proprietà della probabilità:*

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ e di conseguenza $\mathbb{P}(\emptyset) = 0$
- $B \subseteq A \implies \mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Proposizione 2.2.2 (Limite di una successione di eventi). *Data una successione di eventi A_1, \dots, A_n, \dots , questa può essere:*

- **Crescente:** $A_n \subseteq A_{n+1}$ e quindi $A = \bigcup_{n=1}^{+\infty} A_n = \lim_{n \rightarrow \infty} A_n$
- **Decrescente:** $A_n \supseteq A_{n+1}$ e quindi $A = \bigcap_{n=1}^{+\infty} A_n = \lim_{n \rightarrow \infty} A_n$

In entrambi i casi vale:

$$\mathbb{P}(A) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) \quad (31)$$

2.3 Modello uniforme

Definizione 2.3.1 (Modello uniforme). *Definiamo il modello uniforme come lo spazio di probabilità (Ω, P) tale che Ω è finito e ogni esito $\omega \in \Omega$ è equiprobabile, ovvero $P(\{\omega\})$ è la stessa $\forall \omega \in \Omega$.*

2.3.1 Calcolo combinatorio

Alcune formule notevoli:

- **Sequenze ordinate con ripetizione** di k numeri da 1 a n : n^k
- **Ordinamenti possibili** di $\{1, \dots, n\}$: $n!$
- **Sequenze ordinate senza ripetizione** di k numeri di $1, \dots, n$

$$\frac{n!}{(n-k)!} \quad 0 \leq k \leq n$$

- **Sottoinsiemi** di $\{1, \dots, n\}$ formati da k elementi

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad 0 \leq k \leq n$$

2.4 Probabilità condizionata

Quando si è a conoscenza della realizzazione di un evento, cambia la valutazione di probabilità di ogni altro evento.

Definizione 2.4.1 (Probabilità condizionata). *Dati due eventi A, B con B non trascurabile ($P(B) \neq 0$), la probabilità condizionata di A rispetto a B è*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (32)$$

Osservazione 2.4.1. Fissato B non trascurabile la funzione $A \mapsto P(A|B)$ è una probabilità. Non vale il contrario.

Proposizione 2.4.1 (Regola dei prodotti). *Se l'intersezione di eventi $A_1 \cap \dots \cap A_{n-1}$ non è trascurabile vale*

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \dots \cdot \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}) \quad (33)$$

Dimostrazione 2.4.1. Vediamo la dimostrazione della regola dei prodotti nel caso con A e B . Abbiamo:

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

Però, sapendo che $P(A \cap B) = P(B \cap A)$ possiamo dire che

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

Definizione 2.4.2 (Partizione). *Una partizione di Ω è una collezione di n eventi B_1, \dots, B_n a due a due disgiunti tali che*

$$B_1 \cup \dots \cup B_n = \Omega \quad (34)$$

Definizione 2.4.3 (Sistema di alternative). *È una partizione di Ω in eventi non trascurabili.*

Teorema 2.4.1 (Formula della probabilità o della fattorizzazione). Dato B_1, \dots, B_n un sistema di alternative, per un qualunque evento A vale

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|B_i) \mathbb{P}(B_i) \quad (35)$$

Definizione 2.4.4 (Formula di Bayes). *Dati A e B due eventi non trascurabili vale*

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B) \mathbb{P}(B)}{\mathbb{P}(A)} \quad (36)$$

Definizione 2.4.5 (Formula di Bayes - Alternative). *Dati A un evento e B_1, \dots, B_n un sistema di alternative vale*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \mathbb{P}(B_i)}{\sum_{j=1}^n \mathbb{P}(A|B_j) \mathbb{P}(B_j)} \quad (37)$$

2.5 Indipendenza

L'idea è che la conoscenza che si è realizzato un certo evento non modifica la valutazione di probabilità di un altro evento.

Definizione 2.5.1. *Dati n eventi A_1, \dots, A_n , questi sono indipendenti se per ogni k con $2 \leq k \leq n$ e per ogni scelta di interi $1 \leq i_1 < i_2 < \dots < i_k \leq n$ vale*

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \dots \cdot \mathbb{P}(A_{i_k}) \quad (38)$$

Osservazione 2.5.1 (Complessità). Il numero di uguaglianze da verificare per n eventi è

$$2^n - n - 1$$

Proposizione 2.5.1 (Spazi prodotto). *Si consideri*

$$\Omega = \{a = (a_1, \dots, a_n) | a_i = 0, 1\} = \{0, 1\}^n$$

su cui definiamo per ogni a la probabilità

$$\mathbb{P}(\{a\}) = p^{\#\{i:a_i=1\}}(a-p)^{\#\{i:a_i=0\}} = p^{\sum_{i=1}^n a_i}(a-p)^{n-\sum_{i=1}^n a_i}$$

E gli eventi

$$A_i = \{a \in \Omega : a_i = 1\} \quad i = 1, \dots, n$$

sono indipendenti tra di loro, così come i complementari A_i^c .

Osservazione 2.5.2. Due eventi possono essere indipendenti anche in presenza di una relazione causale. Viceversa due eventi possono essere dipendenti anche in assenza di una relazione causale.

2.6 Entropia di Shannon

Una misura di probabilità può essere uno strumento per quantificare l'informazione.

Definizione 2.6.1 (Entropia). *Data una misura di probabilità discreta \mathbb{P} su $\Omega = \{x_1, \dots, x_n\}$, con $p_i = \mathbb{P}(\{x_i\})$, la sua entropia è data dalla funzione*

$$H^{(n)}(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log(p_i) \quad (39)$$

Proposizione 2.6.1. *Valgono:*

1. *La funzione dell'entropia è **simmetrica**: scambiando p_i e p_j non cambia*
2. *$H^{(n)}(1, 0, \dots, 0) = 0$*
3. *È coerente tra n diversi: $H^{(n)}(p_1 = 0, p_2, \dots, p_n) = H^{(n-1)}(p_2, \dots, p_n)$*
4. *$h^{(n)}(p_1, \dots, p_n) \leq H^{(n)}(\frac{1}{n}, \dots, \frac{1}{n})$, ovvero la massima entropia è data dalla distribuzione uniforme di probabilità*
5. *Data una probabilità su $n \times m$ oggetti $\Omega = \{x_{11}, \dots, x_{ij}, \dots, x_{nm}\}$ con $\mathbb{P}(x_{ij}) = q_{ij}$, considerando gli eventi $A_i = \{x_{i1}, \dots, x_{im}\}$ con $\mathbb{P}(A_i) = p_i$ vale*

$$H^{nm}(q_{11}, \dots, q_{ij}, \dots, q_{nm}) = H^{(n)}(p_1, \dots, p_n) + \sum_{i=1}^n p_i H^{(m)}\left(\frac{q_{i1}}{p_i}, \dots, \frac{q_{im}}{p_i}\right)$$

ovvero l'entropia è data da quella relative al sistema di alternative A_i più la media pesata delle entropie relative nei blocchi A_i .

Teorema 2.6.1 (Shannon). Una funzione che soddisfa le 5 proprietà ha la forma

$$cH^{(n)} \quad c > 0 \quad (40)$$

2.7 Densità di probabilità

Definizione 2.7.1 (Densità di probabilità). *Una funzione non negativa $f : \mathbb{R} \rightarrow [0, +\infty]$, integrabile e tale che*

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

La sua probabilità è

$$\mathbb{P}(A) = \int_A f(x) dx \quad A \subseteq \Omega \quad (41)$$

Osservazione 2.7.1. La probabilità di ogni singolo punto è nulla

$$\mathbb{P}(\{t\}) = \int_{\{t\}} f(x) dx = 0 \quad (42)$$

e in generale

$$\mathbb{P}(A) = 0 \quad \forall A \subset \mathbb{R} \quad (43)$$

3 Variabili aleatorie

Le variabili aleatorie sono funzioni dello spazio di probabilità. Permettono di scrivere osservazioni diverse fatte su uno stesso spazio Ω .

Definizione 3.0.1 (Variabile aleatoria). È una funzione

$$X : \Omega \rightarrow \mathbb{R} \quad (44)$$

definita su uno spazio di probabilità.

3.1 Legge di una variabile aleatoria

Ad una variabile aleatoria sono associati eventi del tipo "X prende valori in un insieme $A \subseteq \mathbb{R}$:"

$$\{X \in A\} = X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$$

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$$

Definizione 3.1.1 (Legge di probabilità di una v.a.). La funzione \mathbb{P}_X è una probabilità su \mathbb{R} ed è detta **legge di probabilità** di X.

Note 3.1.1. Quando due variabili aleatorie hanno la stessa legge di probabilità sono dette **equi distribuite**.

3.2 Tipi di variabili aleatorie

3.2.1 Variabili discrete

Definizione 3.2.1 (Variabile aleatoria discreta). Una variabile aleatoria è discreta se la sua immagine $X(\Omega) \subset \mathbb{R}$ è un sottoinsieme al più numerabile di \mathbb{R} o se la sua legge di probabilità è discreta. Se $A \subseteq \mathbb{R}$ vale

$$p_X(A) = \mathbb{P}(x \in A) = \sum_{x_i \in A} p_X(x_i)$$

3.2.2 Variabili continue

Definizione 3.2.2 (Variabile aleatoria continua). Una variabile aleatoria è detta con densità o continua se la sua legge di probabilità è definita da una densità f , ovvero se esiste una f tale che

$$\mathbb{P}_X(A) = \mathbb{P}\{X \in A\} = \int_A f(x)dx \quad (45)$$

Se $A = [a, b]$ è un segmento, vale

$$\mathbb{P}\{X \in A\} = \mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx \quad (46)$$

3.3 Funzione di ripartizione

Per studiare una legge di probabilità di una variabile aleatoria è conveniente usare una funzione su \mathbb{R} .

Definizione 3.3.1 (Funzione di ripartizione). La funzione di ripartizione (c.d.f.) su X è

$$F_X : \mathbb{R} \rightarrow [0, 1] \quad F_X(x) = \mathbb{P}\{X \leq x\} \quad (47)$$

Proposizione 3.3.1. Data $F = F_X$ la funzione di ripartizione di una variabile aleatoria X, valgono:

- F è **non decrescente**

$$x < y \implies F(x) \leq F(y) \quad (48)$$

- $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow +\infty} F(x) = 1$

- F è **continua a destra**

$$\forall x \in \mathbb{R} \quad F(x_n) \rightarrow F(x) \quad (49)$$

per ogni successione $x_n \rightarrow x \quad x_n \geq x$

Proposizione 3.3.2. La probabilità che X cada in un dato intervallo $[a, b]$ per $a < b$ è

$$\mathbb{P}\{a < X \leq b\} = F(b) - F(a) \quad (50)$$

3.3.1 Funzioni di variabili discrete

Data una variabile aleatoria discreta X , la sua c.d.f. che assume valori x_1, x_2, \dots è

$$F_X(t) = \sum_{x_i \leq t} p(x_i) \quad (51)$$

Questa è una funzione a **gradini** che esegue un salto in ogni punto x tale che $\mathbb{P}(X = x) > 0$ di ampiezza pari alla probabilità di quel punto. Vale quindi

$$\mathbb{P}\{X = x\} = F(x) - F_-(x) \quad (52)$$

3.3.2 Funzioni di variabili continue

Quando la variabile ha densità f la sua funzione di ripartizione (**continua**) è

$$F(x) = \int_{-\infty}^x dt \quad (53)$$

o nel caso in cui è **continua a tratti** si ottiene:

$$f(x) = \frac{dF(x)}{dx} \quad (54)$$

3.4 β -quantile

Definizione 3.4.1 (β -quantile). Data una variabile aleatoria X ed un numero $0 < \beta < 1$ il β -quantile è:

$$r_\beta = \inf\{r \in \mathbb{R} : F(r) \geq \beta\} \quad \beta \in (0, 1) \quad (55)$$

Definizione 3.4.2 (Inversa generalizzata). L'inversa generalizzata di F è

$$F^\leftarrow : (0, 1) \rightarrow \mathbb{R} \quad F^\leftarrow(t) = \inf\{r \in \mathbb{R} : F(r) \geq t\} \quad (56)$$

Proposizione 3.4.1. Valgono:

- Se F è strettamente crescente $F^\leftarrow = F^{-1}$
- F^\leftarrow è sempre **non decrescente**
- $F^\leftarrow(F(t)) \leq t \quad \forall t \in \mathbb{R}$
- $F(F^\leftarrow(t)) \geq t \quad \forall t \in \mathbb{R}$
- $F^\leftarrow(t) \leq s \iff F(s) \geq t$

3.5 Variabili discrete notevoli

3.5.1 Binomiali

$$B(n, p) \quad (57)$$

Date n prove ripetute di un esperimento con **due esiti**, chiamiamo uno di questi *successo* con probabilità $0 < p < 1$. Sia X la variabile che conta il numero di successi $(0, 1, \dots, n)$. Vale:

$$\mathbb{P}(X = h) = \binom{n}{h} p^h (1-p)^{n-h} \quad 0 \leq h \leq n \quad (58)$$

Ovvero dati h successi e $n - h$ insuccessi, calcoliamo il numero di modi di disporre i successi.

Osservazione 3.5.1. Date due successioni $x_1, x_2, \dots \in \mathbb{R}$ e $p_1, p_2, \dots \in [0, \infty)$ tale che $\sum_{i=1}^{\infty} p_i = 1$, possiamo definire una variabile discreta tramite

$$\Omega = \mathbb{N} \quad \mathbb{P}(\{k\}) = p_k \quad X(k) = x_k \quad (59)$$

ovvero dove

$$\mathbb{P}_X(k) = \mathbb{P}(X = x_k) = p_k$$

Un caso particolare delle variabili binomiali è quando $n = 1$, ovvero le variabili di **Bernoulli**.

3.5.2 Geometriche

$$G(p) \quad (60)$$

Consideriamo la stessa situazione delle variabili binomiali ma definiamo X come l'istante del primo successo, ovvero il numero h tale che alla prova h -esima si verifichi il primo successo. Vale:

$$P(X = h) = (1-p)^{h-1} p \quad h \in \mathbb{N}_0 \quad (61)$$

Questo corrisponde a dire, dato l'evento A_i successo della prova i -esima,

$$\mathbb{P}(X = h) = \mathbb{P}(A_1^c \cap A_2^c \cap \dots \cap A_{h-1}^c \cap A_h) = \mathbb{P}(A_1^c) \cdot \mathbb{P}(A_2^c) \cdot \dots \cdot \mathbb{P}(A_{h-1}^c) \cdot \mathbb{P}(A_h) = (1-p)^{h-1} p$$

Osservazione 3.5.2 (Assenza di memoria). Le variabili geometriche hanno assenza di memoria, ovvero

$$\mathbb{P}\{X = n + h | X > n\} = \mathbb{P}\{X = h\} \quad (62)$$

3.5.3 Ipergeometriche

$$I(n, h, r) \quad (63)$$

Prendiamo ad esempio un'urna con n biglie di cui $0 \leq h \leq n$ sono bianche e $n - h$ nere. Estraiamo $r \leq n$ biglie senza reinserirle. La variabile che conta quante biglie estratte k sono bianche ha funzione di massa

$$\mathbb{P}(X = k) = \frac{\binom{h}{k} \binom{n-h}{r-k}}{\binom{n}{r}} \quad k = 0, \dots, h \quad (64)$$

Proposizione 3.5.1 (Identità di Vandermonde). Date k biglie bianche e $r - k$ nere, il numero di scelte possibili è

$$\binom{h}{k} \binom{n-h}{r-k}$$

mentre il numero totale di scelte è

$$\binom{n}{r}$$

Otteniamo quindi

$$\sum_{k=0}^h \binom{h}{k} \binom{n-h}{r-k} = \binom{n}{r} \quad (65)$$

che mostra anche $\sum_{k=0}^h \mathbb{P}(X = k) = 1$

3.5.4 Poisson

$$P(\lambda) \quad (66)$$

Una variabile è di Poisson quando

$$\mathbb{P}(X = h) = e^{-\lambda} \frac{\lambda^h}{h!} \quad h \in \mathbb{N}, \lambda > 0 \quad (67)$$

Dato che è una buona approssimazione di una distribuzione binomiale quando n è grande, p è piccolo np è circa λ , possiamo dire che conta il numero di successi quando il numero di prove è alto e la probabilità è bassa. Viene anche detta degli **eventi rari** (eruzioni vulcaniche, particelle α emesse da una sorgente radioattiva).

3.6 Variabili con densità notevoli

Consideriamo i casi in cui esiste una funzione di densità non negativa di integrale unitario su tutto \mathbb{R} f_X tale che

$$\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in (a, b)) = \int_a^b f_X(t) dt \quad (68)$$

3.6.1 Uniformi su intervalli

Dati due numeri reali $a < b$, la densità uniforme sull'intervallo $[a, b]$ è

$$f(t) = \begin{cases} \frac{1}{b-a} & a < t < b \\ 0 & \text{altrove} \end{cases} \quad (69)$$

La c.d.f. è

$$F(t) = \begin{cases} 0 & t \leq a \\ \frac{t-a}{b-a} & a < t < b \\ 1 & t \geq b \end{cases} \quad (70)$$

Ad esempio un numero preso a caso tra 0 e 1.

3.6.2 Esponenziali

Dato il parametro $\lambda > 0$ la densità è

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (71)$$

La c.d.f. è

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases} \quad (72)$$

Descrive ad esempio il tempo di attesa tra due eventi aleatori, come tra le chiamate di un call center.

Osservazione 3.6.1. Questa variabile prende solo valori positivi

$$\mathbb{P}\{X \leq 0\} = 0$$

3.6.3 Gamma

$$\Gamma(r, \lambda) \quad (73)$$

Definizione 3.6.1 (Gamma di Eulero). *Rappresenta l'estensione del fattoriale ai numeri non interi.*

$$\Gamma(r) = \int_0^{+\infty} x^{r-1} e^{-x} dx \quad r > 0 \quad (74)$$

Valgono

$$\Gamma(1) = 1 \quad \Gamma(r) = (r-1)\Gamma(r-1)$$

Definizione 3.6.2. La densità Gamma di parametri $r, \lambda > 0$ è

$$f(x) = \begin{cases} \frac{1}{\Gamma(r)} \lambda^r x^{r-1} e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (75)$$

Proposizione 3.6.1. Date due variabili **indipendenti** X e Y con densità $\Gamma(r, \lambda)$ e $\Gamma(s, \lambda)$ allora $X + Y$ ha densità $\Gamma(r + s, \lambda)$.

Osservazione 3.6.2 (Parametri). Il parametro r rappresenta la forma e λ il tasso di decadimento. In alcuni casi invece di λ viene usato $s = \frac{1}{\lambda}$, ovvero il fattore di scala. In quel caso la densità diventa

$$f(x) = \begin{cases} \frac{1}{\Gamma(r)} \frac{1}{s} \left(\frac{x}{s}\right)^{r-1} e^{-\frac{x}{s}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Osservazione 3.6.3. La densità **esponenziale** di parametro λ corrisponde a $\Gamma(1, \lambda)$.

Proposizione 3.6.2. Una variabile X con densità Gamma ha tutti i momenti e vale

$$\mathbb{E}[X^\beta] = \frac{\Gamma(r + \beta)}{\Gamma(r) \lambda^\beta} \quad \forall \beta > 0 \quad (76)$$

3.6.4 Pareto

Dati $x_m, \alpha > 0$ la densità è

$$f(t) = \begin{cases} \alpha x_m^\alpha t^{-1-\alpha} & t > x_m \\ 0 & t \leq x_m \end{cases} \quad (77)$$

La densità è non nulla dopo la soglia x_m e al diminuire di α ha una coda sempre più pesante. La c.d.f. è

$$F(t) = \begin{cases} 1 & t < x_m \\ 1 - \left(\frac{x_m}{t}\right)^\alpha & t \geq x_m \end{cases} \quad (78)$$

Chiamata anche **power law**, serve a descrivere fenomeni in cui eventi estremi hanno una buona probabilità di avvenire, come la distribuzione della ricchezza nella società.

3.6.5 Gaussiane standard

Viene indicata con $N(0, 1)$ e ha densità

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (79)$$

e c.d.f.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad (80)$$

Osservazione 3.6.4. Questa densità è una funzione pari ($\varphi(x) = \varphi(-x)$). Di conseguenza, dati $x \in \mathbb{R}$ e $0 < \alpha < 1$, si ha

$$\Phi(-x) = 1 - \Phi(x) \quad q_{1-\alpha} = -q_\alpha \quad (81)$$

Di conseguenza, se X è una variabile aleatoria $N(0, 1)$, valgono

$$\mathbb{P}\{-t \leq X \leq t\} = \Phi(t) - \Phi(-t) = 1 - 2\Phi(-t) \quad (82)$$

$$\Phi(0) = \mathbb{P}\{X \geq 0\} = \mathbb{P}\{X \leq 0\} = \frac{1}{2} \quad (83)$$

3.6.6 Chi-Quadro

Definizione 3.6.3 (Densità Chi-Quadro). Se X_1, \dots, X_n sono variabili Gaussiane Standard indipendenti, la variabile $(X_1^2 + \dots + X_n^2)$ ha densità $\Gamma(\frac{n}{2}, \frac{1}{2})$, ovvero Chi-Quadro $\mathcal{X}^2(n)$ con n gradi di libertà.

Proposizione 3.6.3 (Approssimazioni). Definiamo una variabile $C_n = X_1^2 + \dots + X_n^2$ con X_i Gaussiani standard indipendenti. Per $n \rightarrow \infty$ (in generale $n \geq 80$) valgono:

- Per la LGN $C_n \rightarrow 1$, quindi $\frac{C_n}{n} \approx 1$
- Per il TCL $\frac{C_n - n}{\sqrt{2n}}$ converge alla variabile Gaussiana Standard

3.6.7 Student

Definizione 3.6.4. Date due variabili X e C_n con densità $N(0, 1)$ e $\mathcal{X}^2(n)$, definiamo la densità di Student come

$$T_n = \frac{X}{\sqrt{\frac{C_n}{n}}} = \sqrt{n} \frac{X}{\sqrt{C_n}} \quad (84)$$

con densità

$$f_{T_n}(t) = \frac{\Gamma(\frac{n}{2} + \frac{1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n}{2} - \frac{1}{2}} \quad (85)$$

Proposizione 3.6.4. La densità di Student è una funzione pari, di conseguenza valgono

$$F_n(-x) = 1 - F_n(x) \quad \tau_{(\alpha, n)} = -\tau_{(1-\alpha, n)} \quad (86)$$

Teorema 3.6.1. Consideriamo la successione $(T_n)_{n \geq 1}$ di variabili di Student. Questa converge in distribuzione alla variabile $N(0, 1)$. Può quindi essere usata per sostituire una Gaussiana, prestando attenzione al fatto che però tende a 0 più lentamente per $|x| \rightarrow \infty$ (**code pesanti**).

3.6.8 Gaussiani non standard

$$N(m, \sigma^2) \quad (87)$$

Data X una variabile Gaussiana Standard, dati $\sigma > 0$ e $m \in \mathbb{R}$, consideriamo la variabile aleatoria $Y = \sigma X + m$. La sua densità è

$$f_Y(t) = \frac{1}{\sigma} f_X\left(\frac{t-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-m)^2}{2\sigma^2}} \quad (88)$$

mentre la sua c.d.f. è

$$F_Y(t) = \mathbb{P}\{Y \leq t\} = \mathbb{P}\{\sigma X + m \leq t\} = \mathbb{P}\left(X \leq \frac{t-m}{\sigma}\right) = \Phi\left(\frac{t-m}{\sigma}\right) \quad (89)$$

Teorema 3.6.2 (Teorema di Cochran). Siano X_1, \dots, X_n variabili Gaussiani i.i.d. $N(m, \sigma^2)$. Valgono:

- \bar{X}_n e S_n^2 sono **indipendenti**
- \bar{X} ha densità $N(m, \frac{\sigma^2}{n})$
- $\frac{n-1}{\sigma^2} S_n^2$ ha densità $\mathcal{X}^2(n-1)$
- $T = \sqrt{n} \frac{(\bar{X}_n - m)}{S}$ ha densità $T(n-1)$

3.7 Trasformazioni di variabili con densità

Data la variabile aleatoria $X : \Omega \rightarrow \mathbb{R}$ con densità f e una funzione $h : \mathbb{R} \rightarrow \mathbb{R}$, vogliamo la densità della variabile aleatoria composta

$$Y : \Omega \rightarrow \mathbb{R} \quad Y = h \circ X$$

Se è possibile calcolare la c.d.f. di Y

$$F_Y(y) = \mathbb{P}\{Y \leq y\} = \mathbb{P}\{h(X) \leq y\}$$

ed è **continua** e differenziabile, allora è sufficiente derivarla per ottenere la densità di Y .

Proposizione 3.7.1 (Cambio di variabile). *Data X una variabile aleatoria con densità f_X , supportata su un intervallo aperto A (f_X nulla su A^c). Data una funzione $h : A \rightarrow B$, con B un intervallo aperto, biunivoca, differenziabile e con **inversa differenziabile**. Allora $Y = h \circ X$ ha densità*

$$f_Y(y) = \begin{cases} f_X(h^{-1}(y)) \cdot \left| \frac{dh^{-1}(y)}{dy} \right| & y \in B \\ 0 & y \notin B \end{cases} \quad (90)$$

3.8 Valore atteso

Applichiamo il concetto di *media campionaria* e di *varianza campionaria* anche alle variabili aleatorie.

Definizione 3.8.1 (Valore atteso). *Data una variabile discreta X con funzione di massa p_X , si dice che questa ha valore atteso se*

$$\sum_i |x_i| p_X(x_i) < +\infty$$

e vale

$$\mathbb{E}[X] = \sum_i x_i p_X(x_i) \quad (91)$$

Se X è con densità e

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

allora il valore atteso è

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} t f_X(t) dt \quad (92)$$

Note 3.8.1. Il valore atteso è anche chiamato **momento primo** o speranza matematica.

Osservazione 3.8.1. Dato che il valore atteso dipende solo dalla funzione di massa o dalla densità, ovvero solo dalla legge \mathbb{P}_X di X , allora se due variabili sono **equi distribuite** hanno anche lo stesso valore atteso.

Osservazione 3.8.2. Se X prende solo valori positivi, possiamo ammettere che $E[X]$ possa assumere il valore $+\infty$.

Nel caso discreto vuol dire che x_1, x_2, \dots sono sempre positivi e quindi ha senso

$$\mathbb{E}[X] = \sum_{i=1}^{+\infty} x_i p(x_i)$$

Nel caso con densità significa che $f(x) = 0 \quad x < o$ e quindi ha senso

$$\mathbb{E}[X] = \int_0^{+\infty} x f(x) dx \in [0, +\infty]$$

In generale

$$\mathbb{E}[|X|] < +\infty \quad (93)$$

Proposizione 3.8.1. *Valgono:*

- $\forall a, b \in \mathbb{R}$ valgono $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ e $\mathbb{E}[b] = b$
- $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$
- $\mathbb{P}(X \geq 0) = 1 \implies \mathbb{E}[X] \geq 0$

3.8.1 Valore atteso di trasformazioni

Supponiamo di voler calcolare il valore atteso di trasformazioni di una variabile aleatoria X , ovvero $Y = g(x)$ $g: \mathbb{R} \rightarrow \mathbb{R}$.

Proposizione 3.8.2 (Valore atteso di trasformazioni discrete). *Se X è discreta e*

$$\sum_i |g(x_i)|p(x_i) < +\infty$$

allora

$$\mathbb{E}[g(X)] = \sum_i g(x_i)p(x_i) \quad (94)$$

Proposizione 3.8.3 (Valore atteso di trasformazioni con densità). *Se X è con densità e*

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

allora

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx \quad (95)$$

3.8.2 Momenti

Definizione 3.8.2 (Momento). *La variabile aleatoria X ammette momento di ordine $n = 1, 2, \dots$ se*

$$\mathbb{E}[|X|] < +\infty$$

e in quel caso si chiama $\mathbb{E}[X^n]$ il momento di ordine n .

Osservazione 3.8.3. Se una variabile discreta assume solo valori finiti, tutti i momenti sono finiti. Se una variabile con densità è diversa da 0 solo su un intervallo limitato, tutti i momenti sono finiti.

Proposizione 3.8.4. *Siano $1 \leq m < n$*

$$\mathbb{E}[|X|^n] < +\infty \implies \mathbb{E}[|X|^m] < +\infty \quad (96)$$

*Overo se una variabile aleatoria ammette momenti fino a n , ammetterà anche tutti i suoi precedenti. In particolare vale la **disuguaglianza di Jensen**:*

$$\mathbb{E}[|X|^m]^{\frac{1}{m}} \leq \mathbb{E}[|X|^n]^{\frac{1}{n}} \quad (97)$$

Proposizione 3.8.5 (Disuguaglianza di Markov). *Se X è una variabile aleatoria a valori positivi e $a > 0$ vale*

$$a\mathbb{P}\{X \geq a\} \leq \mathbb{E}[X] \quad (98)$$

3.8.3 Varianza di una variabile aleatoria

Definizione 3.8.3 (Varianza). *Se X ammette momento secondo, la sua varianza è*

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \quad (99)$$

e lo scarto quadratico medio o deviazione standard è

$$\sigma(X) = \sqrt{\text{Var}(X)} \quad (100)$$

Proposizione 3.8.6 (Disuguaglianza di Chebyshev). *Data X una variabile aleatoria e $d > 0$, vale*

$$\mathbb{P}\{|X - \mathbb{E}| > d\} \leq \frac{\text{Var}(X)}{d^2} \quad (101)$$

Osservazione 3.8.4. La varianza di una variabile X vale 0 solo se questa è costante tranne che per un insieme trascurabile

$$\mathbb{P}\{|X - \mathbb{E}[X]| \neq 0\} = 0$$

3.8.4 Momenti notevoli

Vediamo i momenti delle variabili notevoli.

3.8.4.1 Variabili binomiali

Per una variabile di Bernoulli, che può valere 0 o 1, vale

$$\mathbb{E}[X^k] = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = 0 \cdot (1 - p) + 1 \cdot p = p \quad (102)$$

$$Var(X) = p - p^2 = p(1 - p) \quad k \geq 1 \quad (103)$$

Dato che una variabile Binomiale può essere vista come somma di variabili di Bernoulli, vale

$$\mathbb{E}[X] = np \quad Var(X) = np(1 - p) \quad (104)$$

3.8.4.2 Variabili di Poisson

Dato che assumono solo valori positivi:

$$\mathbb{E}[X] = \sum_{h=0}^{+\infty} h e^{-\lambda} \frac{\lambda^h}{h!} = \lambda e^{-\lambda} \sum_{h=1}^{+\infty} \frac{\lambda h - 1}{(h-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = \lambda \quad (105)$$

$$\mathbb{E}[X^2] = \lambda + \lambda^2 \quad Var(X) = \lambda \quad (106)$$

3.8.4.3 Variabili uniformi su intervalli finiti

Data una variabile X con densità uniforme su $[a, b]$, vale:

$$\mathbb{E}[X] = \int_a^b \frac{x}{b-a} dx = \frac{x^2}{2(b-a)} \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} \quad (107)$$

$$\mathbb{E}[X^2] = \int_a^b \frac{x^2}{b-a} dx = \frac{x^3}{3(b-a)} \Big|_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3} \quad Var(X) = \frac{(b-a)^2}{12} \quad (108)$$

3.8.4.4 Variabili esponenziali

Vale:

$$\mathbb{E}[X] = \int_0^{+\infty} \lambda x e^{-\lambda x} dx = \frac{1}{\lambda} \quad (109)$$

e più in generale

$$\mathbb{E}[X^n] = \frac{n!}{\lambda^n} \quad (110)$$

Quindi anche

$$Var(X) = \frac{1}{\lambda^2} \quad (111)$$

3.8.4.5 Densità Gamma

Vale:

$$\mathbb{P}\{a < Y < b\} = \mathbb{P}\left\{\frac{a-m}{\sigma} < X < \frac{b-m}{\sigma}\right\} \quad (112)$$

E in particolare:

$$\mathbb{E}[X] = \frac{r}{\lambda} \quad \mathbb{E}[X^2] = \frac{(r+1)r}{\lambda^2} \quad Var(X) = \frac{r}{\lambda^2} \quad (113)$$

3.8.4.6 Variabili Gaussian standard

Se X è Gaussiana Standard, notiamo che possiede tutti i momenti

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |x|^n e^{-\frac{x^2}{2}} dx < +\infty \quad (114)$$

I momenti dispari valgono:

$$\mathbb{E}[X^{2h+1}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2h+1} e^{-\frac{x^2}{2}} dx = \lim_{M \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-M}^{+M} x^{2h+1} e^{-\frac{x^2}{2}} dx = 0 \quad (115)$$

mentre quelli pari, guardando ad esempio il momento secondo

$$\mathbb{E}[X^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx = \left. \frac{-x e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \right|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1 \quad \text{Var}(X) = 1 \quad (116)$$

e più in generale

$$\mathbb{E}[X^{2h+2}] = (2h+1)\mathbb{E}[X^{2h}] \quad (117)$$

3.8.4.7 Variabili di Student

Una variabile di Student ha momenti finiti fino all'ordine $(n-1)$ e quelli di ordine dispari quando esistono sono nulli.

3.8.4.8 Variabili Gaussiane

Data $Y = \sigma X + m$, per linearità del valore atteso

$$\mathbb{E}[Y] = \mathbb{E}[\sigma X + m] = m \quad \text{Var}(Y) = \text{Var}(\sigma X + m) = \sigma^2 \text{Var}(X) = \sigma^2 \quad (118)$$

3.9 Variabili doppie

Dato Ω uno spazio di probabilità e $X, Y : \Omega \rightarrow \mathbb{R}$, il vettore (X, Y) può essere visto come una funzione

$$\Omega \ni \omega \mapsto (X(\omega), Y(\omega)) \in \mathbb{R}^2 \quad (119)$$

La sua legge è una probabilità sui sottoinsiemi $A \subseteq \mathbb{R}^2$

$$\mathbb{P}_{(X,Y)}(A) = \mathbb{P}((X, Y) \in A) = \mathbb{P}\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A\} \quad (120)$$

Osservazione 3.9.1 (Insieme rettangolare). Se $A = A_1 \times A_2$ è un sottoinsieme rettangolare vale

$$\{(X, Y) \in A\} = \{X \in A_1, Y \in A_2\} \quad (121)$$

Note 3.9.1. Con la virgola indichiamo l'intersezione di due condizioni

$$\{X \in A_1, Y \in A_2\} = \{X \in A_1\} \cap \{Y \in A_2\} = X^{-1}(A_1) \cap Y^{-1}(A_2) = (X, Y)^{-1}(A_1 \times A_2)$$

3.9.1 Distribuzioni marginali

Data una variabile doppia (X, Y) possiamo considerare separatamente le leggi delle due componenti \mathbb{P}_X e \mathbb{P}_Y . Queste sono dette **leggi marginali**.

Se $I \subseteq \mathbb{R}$, valgono

$$\mathbb{P}_X(I) = \mathbb{P}_{(X,Y)}(I \times \mathbb{R}) \quad \mathbb{P}_Y(I) = \mathbb{P}_{(X,Y)}(\mathbb{R} \times I) \quad (122)$$

Le distribuzioni marginali non contengono tutta l'informazione della legge $\mathbb{P}_{(X,Y)}$ e di conseguenza non si può ricostruire univocamente dalle prime. L'idea è che la legge totale codifica anche le relazioni tra le due leggi, cosa che le marginali non fanno

3.9.2 Variabili doppie discrete

Una variabile doppia (X, Y) è discreta se la sua immagine è concentrata in un insieme finito o numerabile di punti (x_i, y_j) . La sua **distribuzione di probabilità** è

$$p(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j) \quad (123)$$

e se $A \subseteq \mathbb{R}^2$

$$\mathbb{P}_{(X,Y)}(A) = \mathbb{P}\{(X, Y) \in A\} = \sum_{(x_i, y_j) \in A} p(x_i, y_j) \quad (124)$$

Proposizione 3.9.1. *Se una variabile doppia è discreta con funzione di massa, lo sono anche le sue componenti*

$$p_X(x_i) = \sum_{j=1}^{+\infty} p(x_i, y_j) \quad p_Y(y_j) = \sum_{i=1}^{+\infty} p(x_i, y_j) \quad (125)$$

3.9.3 Variabili doppie con densità

Una variabile doppia (X, Y) è con densità se esiste una funzione $f : \mathbb{R}^2 \rightarrow [0, \infty)$ integrabile e con $\int \int_{\mathbb{R}^2} f(x, y) dx dy = 1$ tale che valga

$$\mathbb{P}_{(X,Y)}(A) = \mathbb{P}\{(X, Y) \in A\} = \int \int_A f(x, y) dx dy \quad A \subseteq \mathbb{R}^2 \quad (126)$$

Teorema 3.9.1 (Teorema di Fubini-Tonelli). Dato un insieme rettangolare $A = A_1 \times A_2$ vale

$$\int \int_A f(x, y) dx dy = \int_{A_1} \left(\int_{A_2} f(x, y) dy \right) dx = \int_{A_2} \left(\int_{A_1} f(x, y) dx \right) dy \quad (127)$$

Proposizione 3.9.2. *Se una variabile doppia ha densità, anche le sue componenti la hanno*

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx \quad (128)$$

Osservazione 3.9.2. A differenza del caso discreto se X e Y sono con densità non è detto che anche X, Y la abbia. Ad esempio (X, X) .

3.10 Indipendenza di variabili aleatorie

Definizione 3.10.1 (Variabili aleatorie indipendenti). *Le variabili aleatorie $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ si dicono indipendenti se, presi comunque $A_1, \dots, A_n \subseteq \mathbb{R}$ vale*

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdot \dots \cdot \mathbb{P}(X_n \in A_n) \quad (129)$$

3.10.1 Indipendenza di variabili doppie

Proposizione 3.10.1 (Indipendenza di variabili doppie discrete). *Date due variabili discrete X e Y con immagine nei punti x_i e y_j , sono indipendenti se e solo se vale*

$$p(x_i, y_j) = p_X(x_i) \cdot p_Y(y_j) \quad \forall (x_i, y_j) \quad (130)$$

Proposizione 3.10.2 (Indipendenza di variabili doppie con densità). *Date due variabili X e Y tale che (X, Y) abbia densità, sono indipendenti se e solo se vale*

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad \forall (x, y) \quad (131)$$

Osservazione 3.10.1. Due variabili aleatorie doppie possono avere le stesse distribuzioni marginali ma essere diverse, ad esempio perché in un caso le componenti sono indipendenti e nell'altro no.

3.10.2 Indipendenza di funzioni di variabili indipendenti

Funzioni di più variabili indipendenti sono indipendenti se la stessa variabile non compare in due funzioni diverse.

Proposizione 3.10.3. *Se $X, Y : \Omega \rightarrow \mathbb{N}$ sono variabili discrete a valori naturali e indipendenti e sia $Z = X + Y$ si ha*

$$p_Z(n) = \sum_{h=0}^n p_X(h) \cdot p_Y(n-h) \quad (132)$$

In particolare se X e Y sono binomiali $B(n, p)$ e $B(m, p)$, allora $Z = X + Y$ è binomiale $B(n+m, p)$.

Proposizione 3.10.4. *Se X e Y sono variabili con densità e indipendenti e sia $Z = X + Y$ si ha*

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{+\infty} f_Y(y) f_X(z-y) dy \quad (133)$$

In particolare se X e Y sono Gaussiane $N(m_1, \sigma_1^2)$ e $N(m_2, \sigma_2^2)$, allora $Z = X + Y$ è Gaussiana $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Proposizione 3.10.5. *Se X e Y sono variabili aleatorie indipendenti, allora per tutte le funzioni $h, k : \mathbb{R} \rightarrow \mathbb{R}$, anche $h(X)$ e $k(Y)$ lo sono.*

3.11 Correlazione

Proposizione 3.11.1. *Date due variabili aleatorie X e Y con valore atteso, allora $X + Y$ ha valore atteso e valgono:*

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

Proposizione 3.11.2. *Date due variabili aleatorie X e Y con valore atteso e **indipendenti**, allora XY ha valore atteso e vale:*

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \quad (134)$$

Proposizione 3.11.3. *Se X e Y sono variabili aleatorie con valore atteso e **indipendenti**, allora per tutte le funzioni $h, k : \mathbb{R} \rightarrow \mathbb{R}$, vale*

$$\mathbb{E}[h(X)k(Y)] = \mathbb{E}[h(X)] \cdot \mathbb{E}[k(Y)] \quad (135)$$

Proposizione 3.11.4 (Disuguaglianza di Schwartz). *Se X e Y hanno valore atteso, non è detto che il loro prodotto XY lo abbia ma se hanno momento secondo allora il loro prodotto ha valore atteso. Questo deriva da*

$$\mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}[X^2]} \cdot \sqrt{\mathbb{E}[Y^2]} \quad (136)$$

3.12 Covarianza

Definizione 3.12.1 (Covarianza). *La covarianza tra due variabili aleatorie X e Y con momento secondo finito è una misura della presenza di una relazione lineare tra le due e vale*

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \quad (137)$$

*Quando vale 0 le variabili sono **scorrelate**.*

3.13 Teoremi limite

Definizione 3.13.1. Data una famiglia di variabili aleatorie X_1, \dots, X_n, \dots , sono i.i.d. se sono indipendenti ed equidistribuite.

Equivalentemente lo sono se hanno tutte la stessa funzione di ripartizione

$$\mathbb{P}_{X_n}(t) = \mathbb{P}(X_n \leq t) = F(t) \quad t \in \mathbb{R}, n = 1, 2, \dots \quad (138)$$

e se vale

$$\mathbb{P}(X_{k_1} \leq t_1, \dots, X_{k_n} \leq t_n) = F_{X_1}(t_1) \cdot \dots \cdot F_{X_n}(t_n) \quad \forall k_1, \dots, k_n \in \mathbb{N}, t_1, \dots, t_n \in \mathbb{R} \quad (139)$$

Definizione 3.13.2 (Media di variabili aleatorie). La media aritmetica di n variabili aleatorie è

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} \quad (140)$$

ed è essa stessa una variabile aleatoria.

3.13.1 Legge debole dei grandi numeri

Definizione 3.13.3 (Convergenza in probabilità). Date variabili aleatorie X e X_1, X_2, \dots definite sullo stesso spazio di probabilità, X_n converge in probabilità a X se

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \epsilon) = 0 \quad \forall \epsilon > 0 \quad (141)$$

In pratica la successione tende in probabilità ad X se per n grande X_n è vicina a X con un'alta probabilità.

Proposizione 3.13.1 (Convergenza ad una costante). Se valgono

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = c \in \mathbb{R} \quad \lim_{n \rightarrow \infty} \text{Var}(X_n) = 0 \quad (142)$$

allora la successione $(X_n)_{n \geq 1}$ converge a c .

Teorema 3.13.1 (Legge debole dei Grandi Numeri). Sia X_1, X_2, \dots una successione di variabili aleatorie i.i.d. dotate di momento secondo finito e sia $\mu = \mathbb{E}[X_i]$ il loro valore atteso. Allora \bar{X}_n converge in probabilità a μ per $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| > \epsilon\right) = 0 \quad \forall \epsilon > 0 \quad (143)$$

Proposizione 3.13.2. Data una successione di variabili i.i.d. X_1, X_2, \dots dotate di momento quarto e sia $\sigma^2 = \text{Var}(X_i)$ la loro varianza,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|S_n^2 - \sigma^2| > \epsilon) = 0 \quad (144)$$

3.13.2 Teorema centrale del limite

Proposizione 3.13.3. Siano $(X_n)_{n \geq 1}$ una successione di variabili aleatorie e X una variabile aleatoria. Date le loro funzioni di ripartizione F_n e F con quest'ultima continua. La successione converge ad X se vale

$$\lim_{n \rightarrow \infty} F_n(t) = F(t) \quad \forall t \in \mathbb{R} \quad (145)$$

Teorema 3.13.2 (Teorema Centrale del Limite). Sia X_1, X_2, \dots una successione di variabili i.i.d. con momento secondo finito e con valore atteso $\mathbb{E}[X_i] = \mu$ e varianza $\sigma^2(X_i) = \sigma^2 > 0$. Vale

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(a \leq \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx = \Phi(b) - \Phi(a) \quad -\infty \leq a < b \leq +\infty \quad (146)$$

In pratica vuol dire che per n grande ($n \geq 50$), la variabile

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \quad (147)$$

è approssimativamente una Gaussiana standard.

Osservazione 3.13.1. Se la successione è composta da variabili Binomiali, possiamo riscriverla come somma di variabili di Bernoulli. Applicando poi il precedente teorema vediamo che è sufficiente che $np(1-p) \geq 15$ per ottenere la convergenza.

Proposizione 3.13.4. *Data una successione X_1, X_2, \dots di variabili i.i.d. con valore atteso $\mathbb{E}[X_i] = \mu$ e varianza $\text{Var}(X_i) = \sigma^2 > 0$*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(a \leq \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx = \Phi(b) - \Phi(a) \quad -\infty \leq a < b \leq +\infty \quad (148)$$

Quindi possiamo usare la varianza campionaria S_n^2 invece di quella teorica σ^2 .

3.13.2.1 Approssimazione binomiale

Definizione 3.13.4. *Data una variabile aleatoria X vale*

$$\mathbb{P}(X \leq t) = \mathbb{P}\left(\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{t - np}{\sqrt{np(1-p)}}\right) \cong \Phi\left(\frac{t - np}{\sqrt{np(1-p)}}\right) \quad (149)$$

4 Stima parametrica

La statistica inferenziale assume che si possa descrivere un carattere da indagare tramite variabili aleatorie di cui vogliamo determinare la legge. Sono quindi fondamentali le seguenti assunzioni (variabili i.i.d.):

- **Indipendenza:** ogni nuova estrazione non deve essere condizionata dalla precedente
- **Equidistribuzione:** l'estrazione di ogni nuovo individuo per il campione deve essere effettuata analogamente alle precedenti

4.1 Campioni

Definizione 4.1.1 (Campione statistico). *Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, una c.d.f. $F = F_X$ di una variabile aleatoria X . Una famiglia finita X_1, \dots, X_n di variabili aleatorie i.i.d.. di legge F si dice campione statistico o aleatorio delle variabili X di numerosità n .*

Osservazione 4.1.1. Assumiamo che la distribuzione di probabilità \mathbb{P}_X sia parzialmente specificata, ovvero sia identificabile in una famiglia di probabilità di parametro $\theta \in \Theta \subseteq \mathbb{R}$ (o più parametri).

4.2 Stimatori

Definizione 4.2.1 (Statistica campionaria). *Una funzione $g(X_1, \dots, X_n)$ di un campione statistico è chiamata statistica. Ad esempio*

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} \quad (\text{Media campionaria})$$

$$S_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad (\text{Varianza campionaria})$$

Definizione 4.2.2 (Stimatore). *Uno stimatore di parametro θ della distribuzione è una statistica che approssima il suo valore. Dato che è in funzione del campione è esso stesso una variabile aleatoria.*

Definizione 4.2.3 (Stimatore corretto). *Uno stimatore si dice **corretto** se ammette momento primo e*

$$\mathbb{E}_\theta[g(X_1, \dots, X_n)] = \theta \quad (150)$$

cioè se la media dello stimatore è il parametro θ .

Osservazione 4.2.1. Quando il parametro coincide con il valore atteso (e.g. Bernoulli), si usa la media campionaria come stimatore. La varianza campionaria si usa invece quando i parametri coincidono con la varianza del campione (e.g. Gaussiana).

Proposizione 4.2.1. *Dato un campione statistico X_1, \dots, X_n con momento secondo. Siano $\mu = \mathbb{E}[X_i]$ e $\sigma^2 = \text{Var}(X_i)$. Valgono:*

$$\mathbb{E}[\bar{X}] = \mu \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \quad \mathbb{E}[S_n^2] = \sigma^2 \quad (151)$$

Osservazione 4.2.2 (Correzione di Bessel). Si noti che $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ non è uno stimatore corretto della varianza. È infatti necessario usare $n-1$ come denominatore.

Definizione 4.2.4 (Stimatore consistente). *Uno stimatore si dice **consistente** se*

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta\{|g_n(X_1, \dots, X_n) - \theta| > \epsilon\} = 0 \quad \forall \epsilon > 0 \quad (152)$$

cioè se, all'aumentare della taglia del campione, lo stimatore si avvicina con alta probabilità a θ .

Osservazione 4.2.3. Media e varianza campionaria sono stimatori consistenti per la legge dei Grandi Numeri.

Definizione 4.2.5 (Stimatore efficiente). *Dato un campione e due stimatori corretti $g(X_1, \dots, X_n)$ e $h(X_1, \dots, X_m)$ che ammettono momento secondo, diciamo che il primo è più efficiente del secondo se*

$$\text{Var}_\theta(g(X_1, \dots, X_n)) \leq \text{Var}_\theta(h(X_1, \dots, X_m)) \quad (153)$$

Ovvero la dispersione del primo stimatore attorno al valore medio θ è minore o uguale alla quella del secondo.

Osservazione 4.2.4. La media campionaria è sempre più efficiente al crescere di n dato che $\text{Var}(\bar{X}_n) = \frac{\text{Var}(X_1)}{n}$ è decrescente per n . Lo stesso vale per la varianza campionaria.

4.2.1 Scelta di uno stimatore

4.2.1.1 Metodo di verosimiglianza

Il primo metodo è tramite la verosimiglianza, ovvero cercare il parametro che meglio approssima al caso effettivamente ottenuto.

Definizione 4.2.6 (Funzione di verosimiglianza). *Si chiama funzione di verosimiglianza la funzione $L: \Theta \times \mathbb{R}^n \rightarrow [0, 1]$ definita nel caso discreto da*

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n p_\theta(x_i) \quad (154)$$

e nel caso con densità da

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i) \quad (155)$$

Ovvero la funzione di massa o la densità congiunta delle variabili aleatorie in questione.

Definizione 4.2.7 (Stima di massima verosimiglianza). *Si chiama stima di massima verosimiglianza una statistica campionaria $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ tale che*

$$L(\hat{\theta}; x_1, \dots, x_n) = \max_{\theta \in \Theta} L(\theta; x_1, \dots, x_n) \quad \forall x_1, \dots, x_n \quad (156)$$

In pratica si fa la scelta del parametro che massimizzi la probabilità dell'esito effettivamente ottenuto.

4.2.1.2 Metodo dei momenti

Un altro metodo per la scelta di uno stimatore è tramite il confronto dei momenti **teorici** con quelli **empirici**.

$$m_k(\theta) = \mathbb{E}_\theta[X^k] \quad (\text{Momenti teorici})$$

$$\sum_{i=1}^n \frac{x_i^k}{n} \quad (\text{Momenti empirici})$$

Definizione 4.2.8 (Stima con il metodo dei momenti). *Si chiama stima con il metodo dei momenti una statistica campionaria $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ che permette di eguagliare alcuni k momenti teorici con quelli empirici*

$$\mathbb{E}_{\hat{\theta}}[X^k] = \frac{1}{n} \sum_{i=1}^n x_i^k \quad \forall x_1, \dots, x_n \quad (157)$$

5 Intervalli di fiducia

Un intervallo di fiducia è un intervallo i cui estremi sono calcolati a partire dai valori assunti dalle variabili X_1, \dots, X_n e nel quale ci aspettiamo che il parametro θ sia contenuto con una certa probabilità (non è detto che ci sia).

Definizione 5.0.1 (Intervallo di fiducia). *Dato un campione statistico X_1, \dots, X_n di legge \mathbb{P}_θ con $\theta \in \Theta \subseteq \mathbb{R}$ e un numero $\alpha \in (0, 1)$, un intervallo **aleatorio***

$$I = [\alpha(X_1, \dots, X_n), b(X_1, \dots, X_n)] \quad (158)$$

è un intervallo di fiducia per θ a **livello** $1 - \alpha$ se vale

$$1 - \alpha \leq \mathbb{P}_\theta(\theta \in I) = \mathbb{P}_\theta(a(X_1, \dots, X_n) \leq \theta \leq b(X_1, \dots, X_n)) \quad \forall \theta \in \Theta \quad (159)$$

Note 5.0.1. Di solito α è un numero piccolo (e.g. 0.05) in modo che il livello sia vicino a 1.

Osservazione 5.0.1. La scelta di un intervallo è un compromesso tra:

- Deve essere il più piccolo possibile per identificare θ con precisione
- Vogliamo un livello di fiducia alto, ovvero vogliamo che sia molto probabile che θ sia nell'intervallo

Queste due necessità si scontrano in quanto all'aumentare del livello di fiducia, diventa più grande l'intervallo.

5.1 Campione Gaussiano

Dato un campione statistico X_1, \dots, X_n di legge Gaussiana $N(m, \sigma^2)$, vogliamo trovare un intervallo per il parametro $\theta = m \in \mathbb{R}$. Dato che la media campionaria è uno stimatore corretto e consistente di m , vogliamo trovare un intervallo della forma

$$I = [\bar{X}_n - d, \bar{X}_n + d] = [\bar{X}_n \pm d] \quad d > 0 \quad (160)$$

in cui d è un numero da determinare per avere un buon compromesso.

Definizione 5.1.1 (Precisione della stima). *Il valore d è detta precisione della stima e $\frac{d}{\bar{X}_n}$ è la precisione **relativa**.*

5.1.1 Bilateri

5.1.1.1 Varianza nota

Definizione 5.1.2. *Dato $\alpha \in (0, 1)$ e $\sigma > 0$ noto, l'intervallo*

$$\left[\bar{X}_n \pm \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \right] \quad (161)$$

è un intervallo di fiducia per m con livello $1 - \alpha$.

Osservazione 5.1.1. La precisione della stima:

- Cresce con il livello
- Cresce con σ^2
- Decresce con n

5.1.1.2 Varianza non nota

Quando non è nota la varianza, possiamo utilizzare il suo stimatore, la varianza campionaria.

Definizione 5.1.3. *Dato $\alpha \in (0, 1)$, l'intervallo*

$$\left[\bar{X}_n \pm \frac{S_n}{\sqrt{n}} \tau_{1-\frac{\alpha}{2}, n-1} \right] \quad (162)$$

è un intervallo di fiducia per m con livello $1 - \alpha$.

Note 5.1.1. Quando $n \geq 60$ si può usare il quantile Gaussiano q invece di quello di Student τ .

5.1.2 Unilateri

A differenza di quelli bilateri, che hanno per entrambi gli estremi variabili aleatorie, quelli unilateri no. Questo è utile ad esempio per capire se la media è troppo alta o bassa.

5.1.2.1 Varianza nota

Definizione 5.1.4. Dato $\alpha \in (0, 1)$, gli intervalli

$$\left(-\infty, \bar{X}_n + \frac{\sigma}{\sqrt{n}} q_{1-\alpha} \right] \quad \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} q_{1-\alpha}, +\infty \right) \quad (163)$$

sono intervalli di fiducia per m con livello $1 - \alpha$.

5.1.2.2 Varianza non nota

Definizione 5.1.5. Dato $\alpha \in (0, 1)$, gli intervalli

$$\left(-\infty, \bar{X}_n + \frac{S_n}{\sqrt{n}} \tau_{1-\frac{\alpha}{2}, n-1} \right] \quad \left[\bar{X}_n - \frac{S_n}{\sqrt{n}} \tau_{1-\frac{\alpha}{2}, n-1}, +\infty \right) \quad (164)$$

sono intervalli di fiducia per m con livello $1 - \alpha$.

5.1.3 Intervalli per la varianza

Se vogliamo trovare un intervallo per la varianza di un campione Gaussiano, non ci interessa che m sia noto o meno.

Definizione 5.1.6 (Intervalli unilateri per la varianza). Dato $\alpha \in (0, 1)$, gli intervalli

$$\left(0, \frac{(n-1)S_n^2}{\chi_{\alpha, n-1}^2} \right] \quad \left[\frac{(n-1)S_n^2}{\chi_{1-\alpha, n-1}^2}, +\infty \right) \quad (165)$$

sono intervalli di fiducia per σ^2 di livello $1 - \alpha$.

5.2 Campione di Bernoulli

Considerando un campione X_1, \dots, X_n di Bernoulli di parametro $p \in (0, 1)$ vogliamo un intervallo della forma $[\bar{X}_n \pm d]$. Quando n è grande, per il Teorema Centrale del Limite, sappiamo che l'equazione 147 è approssimativamente Gaussiana Standard. Inoltre, dato che la varianza $\sigma^2 = p(1-p)$ è in funzione del parametro incognito, possiamo approssimarla come $\bar{X}_n(1 - \bar{X}_n)$.

Definizione 5.2.1. Dato $\alpha \in (0, 1)$, l'intervallo

$$\left[\bar{X}_n \pm \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}} q_{1-\frac{\alpha}{2}} \right] \quad (166)$$

è un intervallo di fiducia per p con livello **approssimativamente** $1 - \alpha$. Precisamente:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(p \in \left[\bar{X}_n \pm \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}} q_{1-\frac{\alpha}{2}} \right] \right) = 1 - \alpha \quad (167)$$

Note 5.2.1. Di conseguenza anche la precisione della stima d è approssimata.

6 Test statistici

Un test statistico è una procedura per verificare ipotesi su uno o più parametri incogniti della distribuzione di probabilità con cui vogliamo descrivere un esperimento ripetuto di cui conosciamo gli esiti.

6.1 Formulazione

6.1.1 Ipotesi

Definizione 6.1.1 (Ipotesi statistica). *Un'affermazione sul parametro $\theta \in \Theta \subseteq \mathbb{R}$ che governa la legge di un campione statistico X_1, \dots, X_n . Data una partizione $\Theta = \Theta_0 \cup \Theta_1$ in due sottoinsiemi disgiunti, l'ipotesi **nulla** H_0 è l'affermazione logica $\theta \in \Theta_0$, mentre quella **alternativa** è l'affermazione logica $\theta \in \Theta_1$. I due insiemi sono rispettivamente dei parametri dell'ipotesi nulla e alternativa.*

Definizione 6.1.2 (Test statistico). *È una procedura per decidere se accettare o rifiutare l'ipotesi nulla H_0 a partire dai valori assunti dal campione:*

- Si **accetta** se i valori assunti sono con essa compatibili (esito **negativo**)
- Si **rifiuta**, in favore di H_1 , se con un alto grado di fiducia, i valori non sono compatibili (esito **positivo**)

6.1.2 Regione critica

Fissata l'ipotesi è necessario determinare un insieme di risultati che portano a rifiutare l'ipotesi nulla.

Definizione 6.1.3 (Regione critica). *La regione critica per l'ipotesi nulla H_0 è un evento $C \subset \Omega$. Il suo complementare $A = \Omega \setminus C$ è quella di accettazione.*

Data una realizzazione (x_1, \dots, x_n) del campione (X_1, \dots, X_n) , l'ipotesi nulla viene:

- **Respinta** se si verifica l'evento C , ovvero se $X_1(\omega), \dots, X_n(\omega)$ assume valore (x_1, \dots, x_n) per un qualche $\omega \in C$
- **Accettata** se non si verifica l'evento C , ovvero se all'interno della regione critica $\omega \in C$, $X_1(\omega), \dots, X_n(\omega)$ non coincide mai con (x_1, \dots, x_n)

6.1.3 Valutazione del test

Il risultato del test è soggetto a due tipi di errore:

- **Prima specie:** consiste nel rifiutare l'ipotesi nulla quando questa è soddisfatta (falso positivo). Se $\theta \in \Theta_0$ la probabilità di commettere questo tipo di errore è $\mathbb{P}_\theta(C)$
- **Seconda specie:** consiste nell'accettare l'ipotesi nulla quando questa non è soddisfatta (falso negativo). Se $\theta \in \Theta_1$ la probabilità di commettere questo tipo di errore è $\mathbb{P}_\theta(A)$

Definizione 6.1.4 (Livello del test). *Dato $0 < \alpha < 1$ si dice che il test è a livello se*

$$\mathbb{P}_\theta(X) \leq \alpha \quad \forall \theta \in \Theta_0 \quad (168)$$

Fissare un livello significa fissare un limite superiore per la probabilità dell'errore di prima specie, scegliendo opportunamente la regione critica.

Definizione 6.1.5 (Potenza del test). *Si chiama potenza del test la funzione*

$$\Theta_1 \ni \theta \mapsto \mathbb{P}_\theta(C) \in [0, 1] \quad (169)$$

Rappresenta la probabilità di rifiutare correttamente l'ipotesi nulla, quindi di accorgersi che non è soddisfatta.

Osservazione 6.1.1. Il test ideale ha livello basso (bassa probabilità dell'errore di prima specie) e potenza alta (bassa probabilità dell'errore di seconda specie). I due fattori sono però in contrapposizione.

Per ovviare alla dipendenza del risultato del test dal livello, si usa spesso il p-value, che dipende dalla realizzazione (x_1, \dots, x_n) del campione (X_1, \dots, X_n) . È la probabilità di ottenere dati più estremi rispetto all'ipotesi nulla di quelli già osservati.

Definizione 6.1.6 (p-value). *Data una famiglia di regioni critiche $\{C(\alpha)\}_{\alpha \in (0,1)}$ tali che il test con regione critica $C(\alpha)$ abbia livello α . Data una realizzazione (x_1, \dots, x_n) del campione (X_1, \dots, X_n) , il p-value è il numero $\bar{\alpha} = \bar{\alpha}(x_1, \dots, x_n)$ tale che:*

- se $\alpha < \bar{\alpha}$ l'ipotesi viene accettata dal test
- se $\alpha > \bar{\alpha}$ l'ipotesi viene rifiutata dal test

6.2 Tipologie di test

6.2.1 Z-Test

È il test sulla **media** di un campione **Gaussiano** con varianza nota.

6.2.1.1 Ipotesi

Dato un campione X_1, \dots, X_n di legge $N(m, \sigma^2)$ con $\sigma > 0$ nota, vogliamo effettuare un test per decidere se la media m coincide o meno con un valore m_0 . Scegliamo quindi $m = \theta \in \Theta = \mathbb{R}$ e $\Theta_0 = \{m_0\}$ $\Theta_1 = \mathbb{R} \setminus \{m_0\}$. Le ipotesi sono:

- H_0) $m = m_0$
- H_1) $m \neq m_0$

6.2.1.2 Regione critica

Dato che il campione è Gaussiano con varianza nota, possiamo sfruttare la variabile Z definita come

$$Z = \sqrt{n} \frac{\bar{X}_n - m}{\sigma}$$

Dato che la media campionaria \bar{X}_n è uno **stimatore corretto e consistente** della media m , rifiutiamo l'ipotesi se la media campionaria si allontana troppo, ovvero una regione critica del tipo

$$C = \{|\bar{X}_n - m_0| > d\}$$

Il valore d deve essere determinato in funzione del livello α

$$\mathbb{P}_{m_0}\{|\bar{X}_n - m_0| > d\} \leq \alpha$$

Per massimizzare la regione critica imponiamo quindi

$$\alpha = \mathbb{P}_{m_0}\{|\bar{X}_n - m_0| > d\} = \mathbb{P}_{m_0}\left\{\frac{\sqrt{n}}{\sigma}|\bar{X}_n - m_0| > d\frac{\sqrt{n}}{\sigma}\right\} = \mathbb{P}_{m_0}\left\{|Z| > \frac{d\sqrt{n}}{\sigma}\right\}$$

e definiamo quindi la regione critica di livello α scegliendo $\frac{d\sqrt{n}}{\sigma} = q_{1-\frac{\alpha}{2}}$

$$C = \left\{ \sqrt{n} \frac{|\bar{X}_n - m_0|}{\sigma} > q_{1-\frac{\alpha}{2}} \right\} = \left\{ |\bar{X}_n - m_0| > \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \right\} \quad (170)$$

Osservazione 6.2.1. L'ipotesi H_0 è accettata a livello α se e solo se m_0 appartiene all'intervallo di fiducia $1 - \alpha$.

6.2.1.3 p-value

Dato il significato di p-value, consideriamo i dati più estremi quelli (y_1, \dots, y_n) che verificano

$$|\bar{y}_n - m_0| > |\bar{x}_n - m_0|$$

Quindi il p-value sarà

$$\bar{\alpha} = \bar{\alpha}(x_1, \dots, x_n) = \mathbb{P}_{m_0} \left(\sqrt{n} \frac{|\bar{X}_n - m_0|}{\sigma} > \frac{\sqrt{n}}{\sigma} |\bar{X}_n - m_0| \right) = 2 \left[1 - \phi \left(\frac{\sqrt{n}}{\sigma} |\bar{X}_n - m_0| \right) \right] \quad (171)$$

in cui l'ultimo passaggio segue perché Z è Gaussiana standard.