

BOZZA

Appunti di Probabilità e Statistica per l'Informatica

Federico Zotti

2° A.A. 2024-25, 2° Semestre

16 May 2025

Università degli Studi di Milano - Bicocca
CdL Informatica

Prof. FRANCESCO CARAVENNA & FEDERICA MASIERO

Indice

0.1. Note e Todos	1
1. Statistica descrittiva	2
1.1. Introduzione	2
1.2. Descrivere i dati	2
1.2.1. Dati a coppie (bivariati)	2
1.3. Riassumere i dati	2
1.3.1. Indici di posizione	2
1.3.1.1. Media campionaria	2
1.3.1.2. Mediana campionaria	3
1.3.1.3. k-esimo percentile campionario	4
1.3.1.4. Moda	5
1.3.2. Indici di dispersione	5
1.3.3. Correlazione	7
2. Spazi di probabilità	8
2.1. Assiomi delle probabilità	8
2.2. Proprietà di base	10
2.3. Calcolo combinatorio	10
2.4. Probabilità condizionata	12
2.5. Indipendenza di eventi	14
3. Variabili aleatorie	16
3.1. Variabili aleatorie discrete	17
3.2. Distribuzioni notevoli discrete	24
3.2.1. Bernoulli	25
3.2.2. Binomiale	25
3.2.3. Poisson	27
3.2.4. Geometrica	28
3.3. Variabili aleatorie assolutamente continue	29
3.3.1. Valore medio e varianza di v.a. assolutamente continue	33
3.4. Funzione di ripartizione	34
3.5. Variabili aleatorie normali	36
3.6. Vettori aleatori (cenni)	38
3.6.1. Vettori aleatori discreti	39
3.6.2. Vettori aleatori assolutamente continui	39
3.6.3. Indipendenza	40
3.6.4. Covarianza e correlazione	40
3.7. Verso la statistica	41
3.7.1. La legge dei grandi numeri	42
4. Teoremi limite	46
5. Statistica inferenziale	49
5.1. Introduzione	49
5.2. Statistica parametrica	49
5.3. Distribuzioni delle statistiche campionarie	51

5.4. Stima per intervalli - Intervalli di confidenza	53
5.4.1. Estremi inferiori e superiori di confidenza	56
5.4.2. Intervalli di confidenza per la media di una popolazione normale con varianza incognita	56
5.4.3. Stima intervallare della media per grandi campioni	57
5.4.4. Stima per intervalli di una frequenza per grandi campioni . .	57
5.5. Intervalli di confidenza per la varianza di una popolazione normale	59
6. Verifica di ipotesi	62
6.1. Nozioni introduttive	62
6.2. Test Z per la media di una popolazione normale con varianza nota	63
6.3. Test t per la media di una popolazione normale con varianza incognita	68
6.4. Test su una proporzione (grandi campioni)	69
6.5. Test su due campioni normali	71
6.5.1. Confronto tra medie per dati accoppiati	71

0.1. Note e Todos

- ☐ Aggiungere degli esempi 13
- ☐ Controllare formula densità 55
- ☐ Controllare che la t di student vada messa con `op()` o senza 71

1. Statistica descrittiva

1.1. Introduzione

Statistica arte di “imparare dai dati”

Si divide in due parti:

1. La **statistica descrittiva** descrive e riassume i dati
2. La **statistica inferenziale** trae conclusioni dai dati

1.2. Descrivere i dati

Misuriamo una certa variabile (qualitativa o quantitativa) in un campione, ottenendo un insieme di dati:

$$x_1, x_2, x_3, \dots, x_n$$

con n il numero dei dati.

Se i dati sono distinti si possono rappresentare in una tabella.

Frequenza assoluta f_i è il numero di volte in cui compare un valore nell'insieme.

Frequenza relativa $p_i = \frac{f_i}{N}$.

I dati possono essere **quantitativi** se sono categorie o nomi, oppure **quantitativi** se sono numeri.

Per rappresentare le frequenze si può utilizzare un **istogramma** (grafico a barre). Esso è una rappresentazione equivalente a una tabella.

Se i valori distinti dei dati sono in numero elevato, conviene suddividere i valori in intervalli detti **classi**.

1.2.1. Dati a coppie (bivariati)

Generalmente gli insiemi di dati si riferiscono a una singola variabile. Se si misurano due dati al posto di uno, ogni dato è una coppia di numeri. Questi vengono detti **dati bivariati**:

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$$

In questo caso al posto di usare un istogramma è meglio utilizzare un **diagramma a dispersione**, rappresentando le coppie in un piano cartesiano.

1.3. Riassumere i dati

1.3.1. Indici di posizione

1.3.1.1. Media campionaria

Per descrivere il **centro** dell'insieme dei dati, definiamo la

Media campionaria

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{i=1}^N x_i$$

Linearità della media

Se si applica una trasformazione lineare ai dati

$$(y_i = ax_i + b)_{i=1, \dots, N} \quad a, b \in \mathbb{R}$$

la media rimane lineare

$$\bar{y} = a\bar{x} + b$$

Con i valori z_i e le loro relative frequenze f_i la formula diventa

$$\bar{x} = \frac{z_1 f_1 + z_2 f_2 + \dots + z_M f_M}{M}$$

1.3.1.2. Mediana campionaria

Un'altra misura del *centro* dell'insieme dei dati alternativa alla media è la **mediana campionaria**.

Mediana campionaria

Avendo i dati in ordine crescente, la mediana è il valore in posizione centrale.

Se i dati sono dispari:

$$m = x_{\frac{N+1}{2}}$$

ovvero l'intero in posizione $\frac{N}{2}$.

Se i dati sono pari:

$$m = \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2}$$

ovvero la media tra i due dati in posizione centrale.

1.3.1.3. **k-esimo percentile campionario****k-esimo percentile campionario**

Fissando un numero $k \in [0, 100]$, il k -esimo percentile campionario è il valore t per il quale:

- almeno il $k\%$ dei dati è $\leq t$
- almeno il $(100 - k)\%$ dei dati è $\geq t$

Casi più importanti: $k = 25 \quad 50 \quad 75$

Scriviamo $k = 100p$, con $p = \frac{k}{100} \in [0, 1]$.

Dunque possiamo definire:

- $p = \frac{1}{4}$: $k = 25$ -esimo percentile (*primo quartile*)
- $p = \frac{1}{2}$: $k = 50$ -esimo percentile (*secondo quartile* o mediana)
- $p = \frac{3}{4}$: $k = 75$ -esimo percentile (*terzo quartile*)

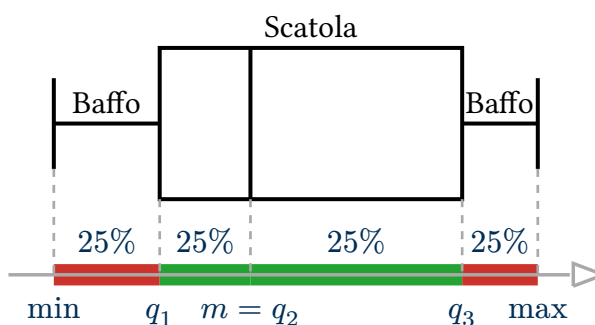
Per calcolare il k -esimo percentile t si ordina l'insieme dei dati

$$x_1 \leq x_2 \leq \dots \leq x_N$$

- Se Np non è intera, $t = x_i$ è il dato la cui posizione i è l'intero successivo a Np .
- Se Np è intera, $t = \frac{x_{Np} + x_{Np+1}}{2}$ è la media aritmetica del dato in posizione Np e il dato successivo.

Attenzione: esistono definizioni alternative di percentile nel caso in cui Np è intero (per esempio nel linguaggio R).

Per rappresentare graficamente la mediana e i quartili si può usare un *box plot*.



Attenzione: esistono convenzioni diverse per i "baffi": si possono escludere valori "estremi" (*outliers*) prima di calcolare \min e \max .

1.3.1.4. Moda

Definizione:

Se tra i dati c'è un valore che compare con frequenza maggiore di tutti, esso si dice **moda**.

La moda non è sempre definita e non la considereremo.

1.3.2. Indici di dispersione

Fissiamo un insieme di dati x_1, x_2, \dots, x_N .

Media campionaria:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

Consideriamo gli "scarti" $x_i - \bar{x}$ rispetto alla media.

Osservazione:

La somma di tutti gli scarti è nulla:

$$\sum_{i=1}^N (x_i - \bar{x}) = 0$$

Consideriamo dunque il quadrato degli scarti $(x_i - \bar{x})^2$.

Facendone la media si ottiene la varianza campionaria

Varianza campionaria

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

Osservazione:

Manipolando la formula si ottiene

$$S^2 = \frac{1}{N-1} \left(\sum_{i=1}^N x_i^2 - N\bar{x}^2 \right)$$

Se i dati x_i sono espressi in una unità di misura (es. metri), S^2 è espressa nel quadrato dell'unità di misura (es. metri quadrati).

Per ottenere una statistica omogenea ai dati, si definisce la deviazione standard.

Deviazione standard (campionaria)

$$S = \sqrt{S^2} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Essa misura la **dispersione** dei dati rispetto alla media \bar{x} .

Osservazione:

$S^2 \geq 0$, inoltre $S^2 = 0 \Leftrightarrow x_1 = x_2 = \dots = x_N$.

Vedremo una disuguaglianza (*Chebyshev*) da cui segue:

Teorema:

L'intervallo attorno a \bar{x} di ampiezza proporzionale a S

$$(\bar{x} - c \cdot s, \bar{x} + c \cdot s)$$

con $c > 1$, contiene una frazione $\alpha \geq 1 - \frac{1}{c^2}$ di dati.

- $c = 2$: $(\bar{x} - 2s, \bar{x} + 2s)$ contiene almeno il 75% dei dati
- $c = 3$: $(\bar{x} - 3s, \bar{x} + 3s)$ contiene almeno l'89% dei dati
- ...

Osservazione:

Applicando trasformazioni lineari-affini ai dati

$$a, b \in \mathbb{R} \quad (y_i = ax_i + b)_{i=1, \dots, N}$$

La varianza e la deviazione standard **non sono lineari**.

$$S_y^2 = a^2 \cdot S_x^2$$

$$S_y = |a| \cdot S_x$$

Un altro indicatore di variabilità, che misura la dispersione dei dati rispetto alla mediana m , è la differenza tra terzo e primo quartile.

Scarto interquartile

$$\text{IQR} = \Delta = q_3 - q_1$$

Per costruzione, l'intervallo $[q_1, q_3]$ contiene almeno il 50% dei dati.

1.3.3. Correlazione

Consideriamo un insieme di N dati a coppie (*bivariati*)

$$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$$

Essi rappresentano misurazioni di due variabili (x e y) in un campione di N individui.

Vogliamo quantificare la correlazione tra le due variabili x e y , ossia la tendenza per cui a valori di x più grandi corrispondono valori di y più grandi (**correlazione positiva**) o piccoli (**correlazione negativa**).

Coefficiente di correlazione lineare (campionario)

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{(N-1) \cdot S_x \cdot S_y}$$

Si può mostrare che $-1 \leq r \leq +1$.

- $r = +1 \Leftrightarrow y_i = ax_i + b$ con $a > 0$ (i dati sono disposti su una retta con coeff. positivo)
- $r = -1 \Leftrightarrow y_i = ax_i + b$ con $a < 0$ (i dati sono disposti su una retta con coeff. negativo)

In generale $r > 0$ indica una correlazione positiva e $r < 0$ indica una correlazione negativa.

- $|r| \gtrsim 0.7$ indica una correlazione **significativa**
- $|r| \lesssim 0.3$ indica una correlazione **debole**

Osservazione:

Manipolando la formula si ottiene

$$r = \frac{\sum_{i=1}^N (x_i y_i) - N \cdot \bar{x} \cdot \bar{y}}{(N-1) \cdot S_x \cdot S_y}$$

2. Spazi di probabilità

Il calcolo delle probabilità è una teoria matematica che permette di descrivere e studiare gli **esperimenti aleatori**, ovvero fenomeni il cui esito non è prevedibile con certezza a priori.

Non è affatto evidente che i fenomeni casuali possano essere descritti mediante una **teoria matematica**.

La teoria moderna della probabilità fu assiomatizzata nel 1933 da Andrey Nikolaevic Kolmogorov.

2.1. Assiomi delle probabilità

La descrizione matematica di un esperimento aleatorio si articola in tre passi.

1. **Spazio campionario**: insieme Ω che contiene tutti i possibili esiti dell'esperimento
2. **Eventi**: sottoinsiemi $A \subseteq \Omega$ dello spazio campionario (affermazioni sull'esito dell'esperimento aleatorio)

Le *operazioni insiemistiche sugli eventi* equivalgono a *operazioni logiche sulle affermazioni*:

- Unione $A \cup B \leftrightarrow$ Si verifica A o B
- Intersezione $A \cap B \leftrightarrow$ Si verifica A e B
- Complementare $A^c \leftrightarrow$ Non si verifica A

Valgono dunque le seguenti regole:

- $(A^c)^c = A$
- Leggi di de Morgan: $\begin{cases} (A \cup B)^c = A^c \cap B^c \\ (A \cap B)^c = A^c \cup B^c \end{cases}$

3. **Probabilità**: funzione $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ che soddisfa opportune proprietà

(Regola che assegna, in modo coerente, a ogni evento $A \subseteq \Omega$ un "grado di fiducia" $P(A)$, tra 0 e 1, che attribuiamo al verificarsi di A)

Sono possibili diverse interpretazioni della probabilità:

- **Soggettivista**: $P(A)$ rappresenta il prezzo "equo" di una scommessa che paga 1 se si verifica A (altrimenti 0)
- **Frequentista**: $P(A)$ è la *funzione asintotica* di volte in cui si verifica A , ripetendo l'esperimento

Definizione:

Sia Ω un insieme (spazio campionario).

Si dice **probabilità** qualsiasi funzione $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ che soddisfa

- $P(\Omega) = 1$
- $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

La coppia (Ω, P) è detta **spazio di probabilità**.

Osservazione:

Se A_1, A_2, \dots, A_k sono disgiunti ($A_m \cap A_n = \emptyset, m \neq n$)

$$P(A_1 \cup A_2 \cup \dots \cup A_k) = P(A_1) + P(A_2) + \dots + P(A_k)$$

Esempio:

Lancio di un dado a 6 facce.

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$A = \text{esce un numero pari} = \{2, 4, 6\}$$

$$B = \text{esce un numero multiplo di 3} = \{3, 6\}$$

$$C = \text{esce un numero multiplo di 3 o pari} = A \cup B = \{2, 3, 4, 6\}$$

$$D = \text{esce un numero multiplo di 3 e pari} = A \cap B = \{6\}$$

$$E = \text{esce 7} = \emptyset$$

Indichiamo la cardinalità (numero di elementi) di un insieme A con $|A|$.

Probabilità uniforme

La probabilità uniforme su un insieme finito Ω è

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

In particolare $P(\{\omega\}) = \frac{1}{n}$ per ogni $\omega \in \Omega$ e $n = |\Omega|$.

Questa formula è l'equivalente di $\frac{\text{casi favorevoli}}{\text{casi possibili}}$.

La probabilità uniforme non è l'unica probabilità.

Supponiamo ad esempio $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$. Possiamo fissare la probabilità dei singoli esiti come

$$p_1 = P(\{\omega_1\}) \geq 0, \quad p_2 = P(\{\omega_2\}) \geq 0, \quad \dots, \quad p_n = P(\{\omega_n\}) \geq 0$$

in modo che $p_1 + p_2 + \dots + p_n = 1$.

Questo determina la probabilità di ogni evento $A \subseteq \Omega$ come

$$P(A) = \sum_{\omega_i \in A} P(\{\omega_i\}) = \sum_{\omega_i \in A} p_i$$

Si può mostrare che P è una probabilità (soddisfa le due proprietà).

Scegliendo $p_1 = p_2 = \dots = p_n$ si ritrova la probabilità uniforme.

2.2. Proprietà di base

Fissiamo uno spazio di probabilità (Ω, P) .

1. $P(\Omega) = 1$
2. $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

Da questi assiomi si deducono molte altre proprietà.

Proposizione:

- $P(\emptyset) = 0$
- Regola del complementare: $P(A^c) = 1 - P(A) \quad \forall A$
- Addizione di probabilità: $P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \forall A, B$
- Monotonia: $A \subseteq B \Rightarrow P(A) \leq P(B)$

Aggiungere degli esempi

2.3. Calcolo combinatorio

Consideriamo uno spazio di probabilità uniforme (Ω, P) . È il modello appropriato per descrivere esperimenti aleatori i cui esiti $\omega \in \Omega$ siano tutti equiprobabili.

Quando si dice “scegliamo casualmente una persona/oggetto in un insieme finito”, senza ulteriori specifiche, si sott'intende che la scelta è effettuata *in modo uniforme*.

Affinché la probabilità uniforme sia ben definita, **lo spazio campionario Ω deve essere finito**.

Osservazione:

Se Ω è finito numerabile (es. $\Omega = \mathbb{N}$) la probabilità uniforme su Ω non esiste.

In uno spazio di probabilità uniforme, calcolare una probabilità significa contare gli elementi di un insieme:

$$P(A) = \frac{|A|}{|\Omega|}$$

Le tecniche di conteggio formano il **calcolo combinatorio**.

Principio fondamentale:

Consideriamo un esperimento costituito da due parti:

- Prima parte: n esiti possibili ($\{a_1, \dots, a_n\}$)
- Seconda parte: m esiti possibili ($\{b_1, \dots, b_m\}$)

Allora l'esperimento totale può avere $n \cdot m$ esiti possibili.

Disposizioni con ripetizione:

Le disposizioni con ripetizione sono sequenze ordinate di k elementi (anche ripetuti) scelti tra n possibili. Sono in numero

$$\underbrace{n \cdot n \cdot \dots \cdot n}_{k \text{ volte}} = n^k$$

Disposizioni semplici:

Le disposizioni semplici sono sequenze ordinate di k elementi **distinti** scelti tra n possibili (con $k \leq n$). Sono in numero

$$n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

Nel caso in cui $k = n$ si hanno **permutazioni** di n oggetti. Sono in numero $n!$.

In molti casi non interessa l'ordine dei risultati.

Combinazioni:

Le combinazioni sono insiemi (o sottoinsiemi) come **collezioni non ordinate** di k elementi distinti scelti tra n possibili (con $k \leq n$). Sono in numero

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

2.4. Probabilità condizionata

Consideriamo un esperimento aleatorio che descriviamo con uno spazio di probabilità (Ω, P) . Consideriamo un evento $A \subseteq \Omega$, che ha una probabilità $P(A)$.

Supponiamo di ricevere l'informazione che un altro evento B si è verificato (cioè l'esito ω dell'esperimento è in B).

Come è ragionevole aggiornare la probabilità di A per tenere conto di questa informazione aggiuntiva?

Probabilità condizionata

La **probabilità condizionata di A dato B** è

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

assumendo $P(B) \neq 0$.

Proprietà della probabilità condizionata:

- **Regola del prodotto:** $P(A \cap B) = P(B) \cdot P(A|B)$
- **Formula di disintegrazione:** $P(A) = P(A \cap B) + P(A \cap B^c)$
- **Formula delle probabilità totali:** $P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)$
- **$P(\cdot | B)$ è una probabilità**, in particolare $P(A^c|B) = 1 - P(A|B)$

Esempio:

Per rilevare la presenza di un virus, viene effettuato un test con le seguenti caratteristiche:

- **Sensibilità:** se il virus è presente, il test dà esito positivo (correttamente) nel 99% dei casi
- **Specificità:** se il virus è assente, il test dà esito negativo (correttamente) nel 99,7% dei casi

È noto che 4 persone su 1000 hanno il virus (**prevalenza**).

Se un individuo estratto a caso dalla popolazione viene sottoposto al test, qual è la probabilità che il test dia esito positivo?

Introduciamo gli eventi:

- A = il test dà esito positivo
- B = l'individuo ha il virus

$$P(A|B) = 0.99 \quad P(A^c|B^c) = 0.997 \quad P(B) = 0.004$$

dunque

$$P(A|B^c) = 1 - P(A^c|B^c) = 0.003 \quad P(B^c) = 1 - P(B) = 0.996$$

Probabilità totali

$$\begin{aligned} P(A) &= P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c) \\ &= 0.99 \cdot 0.004 + 0.003 \cdot 0.996 \\ &\approx 0.004 + 0.003 \\ &= 0.007 = 0.7\% \end{aligned}$$

Un'ultima proprietà, fondamentale:

Formula di Bayes:

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

Esempio:

Tornando all'esempio precedente, se il test dà esito positivo, qual è la probabilità che l'individuo abbia effettivamente il virus?

$$P(B|A) = \frac{0.99 \cdot 0.004}{0.007} \approx \frac{4}{7} \approx 57\%$$

Le proprietà appena viste si generalizzano nel caso di più di due eventi.

Regola del prodotto:

Per ogni famiglia di eventi $B_1, B_2, \dots, B_n \subseteq \Omega$ (con $P(B_i) \neq 0$).

Nel caso di tre eventi:

$$P(B_1 \cap B_2 \cap B_3) = P(B_1) \cdot P(B_2|B_1) \cdot P(B_3|B_1 \cap B_2)$$

Nel caso di n eventi:

$$P(B_1 \cap \dots \cap B_n) = P(B_1) \cdot P(B_2 | B_1) \cdot \dots \cdot P(B_n | B_1 \cap \dots \cap B_{n-1})$$

Formule di disintegrazione e probabilità totali:

B_1, B_2, \dots, B_n partizione di Ω :

$$\begin{cases} B_1 \cup B_2 \cup \dots \cup B_n = \Omega \\ B_i \cap B_j = \emptyset \quad \text{per } i \neq j \end{cases}$$

Formula di disintegrazione

Allora per ogni evento $A \subseteq \Omega$, nel caso di 3 eventi B_i :

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$$

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)$$

Nel caso di n eventi B_i :

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_n)$$

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n)$$

2.5. Indipendenza di eventi

Può capitare che, per un evento A , l'informazione che un altro evento B si è verificato non ne cambi la probabilità:

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

Che equivale a $P(A \cap B) = P(A) \cdot P(B)$.

In questo caso, gli eventi A e B si dicono **indipendenti**.

Gli eventi indipendenti sono differenti dagli eventi disgiunti.

Siano $A, B \subseteq \Omega$ eventi in uno spazio di probabilità (Ω, P) .

- A e B disgiunti: $A \cap B = \emptyset \Rightarrow P(A \cap B) = 0$
- A e B indipendenti: $P(A \cap B) = P(A) \cdot P(B)$

Quindi *due eventi indipendenti non possono essere disgiunti*, tranne nel caso banale in cui uno dei due abbia probabilità nulla.

Generalizzazione:

- Tre eventi A, B, C si dicono indipendenti se valgono:

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

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

$$P(B \cap C) = P(B) \cdot P(C)$$

$$P(A \cap C) = P(A) \cdot P(C)$$

- n eventi A_1, A_2, \dots, A_n si dicono indipendenti se non solo

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n)$$

ma la stessa formula vale per ogni sottogruppo di eventi presi tra $\{A_1, A_2, \dots, A_n\}$.

Anche in questo caso possiamo scambiare qualcuno degli A_i con A_i^c e abbiamo ancora eventi indipendenti.

3. Variabili aleatorie

Consideriamo un esperimento aleatorio, descritto da uno **spazio di probabilità** (Ω, P) .

Spesso si è interessati solo a una *quantità (tipicamente numerica) determinata dall'esito dell'esperimento*. Una tale quantità è detta **variabile aleatoria**.

Esempio:

Lancio due dadi a sei facce. Esito dell'esperimento: coppia (x, y) dei risultati dei dadi.

$$\Omega = \{(x, y) : 1 \leq x, y \leq 6\}$$

Sono variabili aleatorie:

- X : risultato del primo dado
- Y : risultato del secondo dado
- S : somma dei risultati dei due dadi $(X + Y)$

Osserviamo che X, Y, S sono funzioni da Ω in \mathbb{R} :

$$\omega = (x, y) \implies X(\omega) = x, \quad Y(\omega) = y, \quad S(\omega) = x + y$$

Una **variabile aleatoria** è una quantità che dipende dal caso, ovvero una funzione definita sullo spazio campionario:

$$X : \Omega \rightarrow \mathbb{R}$$

Sia X una variabile aleatoria e x un suo possibile valore:

$$\{X = x\} \text{ è un evento } \rightarrow \begin{cases} \rightarrow X \text{ assume il valore } x \\ \rightarrow \{\omega \in \Omega : X(\omega) = x\} \end{cases}$$

Ogni variabile aleatoria X determina molti eventi. Possiamo calcolarne la probabilità:

$$P(X = x), \quad P(X \geq 5), \quad P(X \neq 0), \quad \dots$$

Esempio:

Estraggo casualmente una famiglia con due figli. Indichiamo con X il numero di figli maschi. Che valori può assumere X ? Con quali probabilità?

$$\Omega = \{MM, MF, FM, FF\} \quad P = \text{probabilità uniforme}$$

Allora $X : \Omega \rightarrow \mathbb{R}$ è definita da

$$X(MM) = 2, \quad X(MF) = 1, \quad X(FM) = 1, \quad X(FF) = 0$$

Quindi X assume i valori $X(\Omega) = \{0, 1, 2\}$ con

$$P(X = 0) = P(\{FF\}) = \frac{1}{4}$$

$$P(X = 1) = P(\{\text{MF}, \text{FM}\}) = \frac{2}{4} = \frac{1}{2}$$

$$P(X = 2) = P(\{\text{MM}\}) = \frac{1}{4}$$

Attenzione a non confondere variabili aleatorie ed eventi!

3.1. Variabili aleatorie discrete

Una variabile aleatoria X (reale) si dice **discreta** se i valori che può assumere sono **un insieme finito**

$$X(\Omega) = \{x_1, x_2, \dots, x_n\} \subseteq \mathbb{R}$$

oppure **un insieme infinito numerabile**

$$X(\Omega) = \{x_1, x_2, \dots\} = \{x_i\}_{i \in \mathbb{N}} \subseteq \mathbb{R}$$

Densità discreta:

A ogni variabile aleatoria discreta X possiamo associare una **densità discreta**

$$p_X(x_i) = P(X = x_i)$$

Se x non è un valore assunto da X , si pone $p_X(x) = 0$

Proprietà della densità discreta:

- p_X è una funzione da \mathbb{R} in $[0, 1]$
- $p_X(x) = P(X = x) = 0$ se x non è uno dei valori x_i assunti da X
- $p_X(x_i) \geq 0 \quad \forall i$
- $\sum_{i \geq 1} p_X(x_i) = 1$ perché gli eventi $\{X = x_i\}$ per $i \geq 1$ sono disgiunti (partizione di Ω)

Sapendo la densità discreta è possibile calcolare la probabilità di qualsiasi evento collegata alla variabile aleatoria.

$$P(X \in B) = \sum_{x_i \in B} \underbrace{p_X(x_i)}_{P(X=x_i)} \quad \forall B \subseteq \mathbb{R}$$

Esempio:

Sia X il numero dei figli maschi in una famiglia estratta a caso con due figli/e. Densità discreta?

Ricordiamo che $X(\Omega) = \{0, 1, 2\}$ e $p_X(x) = P(X = x)$. Dunque

$$p_X(0) = \frac{1}{4}, \quad p_X(1) = \frac{1}{2}, \quad p_X(2) = \frac{1}{4}$$

Esempio:

Lancio due monete “truccate”, che danno testa con probabilità $p \in [0, 1]$ fissata. Sia X il numero di teste.

$\Omega = \{TT, TC, CT, CC\}$ con

$$P(\{TT\}) = p^2, \quad P(\{TC\}) = P(\{CT\}) = p(1-p), \quad P(\{CC\}) = (1-p)^2$$

Possiamo rappresentare $X : \Omega \rightarrow \mathbb{R}$ con

$$X(TT) = 2, \quad X(TC) = X(CT) = 1, \quad X(CC) = 0$$

pertanto $X(\Omega) = \{0, 1, 2\}$ e la densità discreta è

$$p_X(2) = P(X = 2) = P(\{TT\}) = p^2$$

$$p_X(1) = P(X = 1) = P(\{TC, CT\}) = P(\{TC\}) + P(\{CT\}) = 2p(1-p)$$

$$p_X(0) = P(X = 0) = P(\{CC\}) = (1-p)^2$$

È concettualmente importante sapere che una variabile aleatoria X è rappresentata matematicamente da una funzione definita sullo spazio campionario Ω di un esperimento aleatorio. Allo stesso tempo, possiamo pensare alla v.a. X come a un numero che dipende dal caso. Se siamo interessati a una v.a. discreta X , spesso non è necessario scrivere lo spazio campionario Ω ed esprimere X come funzione: basta conoscere la densità discreta.

Esempio:

Una pasticceria prepara ogni giorno 3 torte Sacher. In base all'analisi delle vendite passate, è noto che

- Il 20% dei giorni *nessun* cliente richiede una Sacher
- Il 30% dei giorni *un* cliente richiede una Sacher
- Il 35% dei giorni *due* clienti richiedono una Sacher
- I restanti giorni, *tre o più* clienti richiedono una Sacher

Sia X il numero di torte Sacher invendute a fine giornata. Qual è la densità discreta di X ?

Per costruzione $X(\Omega) = \{0, 1, 2, 3\}$.

$$p_X(3) = P(X = 3) = P(0 \text{ clienti vogliono una Sacher}) = 0.2$$

$$p_X(2) = P(X = 2) = P(1 \text{ cliente vuole una Sacher}) = 0.3$$

$$p_X(1) = P(X = 1) = P(2 \text{ clienti vogliono una Sacher}) = 0.35$$

Bisogna determinare $p_X(0)$. Sappiamo che

$$p_X(0) + p_X(1) + p_X(2) + p_X(3) = 1$$

dunque

$$\begin{aligned} p_X(0) &= 1 - p_X(1) - p_X(2) - p_X(3) \\ &= 1 - 0.35 - 0.3 - 0.2 = 0.15 \end{aligned}$$

Concludiamo con una domanda: qual è la probabilità q che il numero di torte invendute sia pari?

$$\begin{aligned} q &= P(X \text{ è pari}) = P(X = 0 \vee X = 2) \\ &= P(\{X = 0\} \cup \{X = 2\}) \\ &= P(X = 0) + P(X = 2) \\ &= p_X(0) + p_X(2) \\ &= 0.15 + 0.3 = 0.45 \end{aligned}$$

Valore medio

Sia X una variabile aleatoria discreta (reale) che assume una quantità finita di valori x_1, x_2, \dots, x_n . Si definisce **valore medio di X**

$$\begin{aligned} E[X] &= \sum_{i=1}^n x_i \cdot p_X(x_i) \\ &= \sum_{i=1}^n x_i \cdot P(X = x_i) \end{aligned}$$

Ovvero è la somma dei valori assunti da X “pesati” con le rispettive probabilità.

Esempio:

Sia X il numero dei figli maschi di una coppia con due figli/e.

$$X(\Omega) = \{0, 1, 2\} \quad \text{con} \quad p_X(0) = p_X(2) = \frac{1}{4}, \quad p_X(1) = \frac{1}{2}$$

Allora

$$E[X] = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$$

Esempio:

Sia X il risultato del lancio di un dado regolare (6 facce).

$$X(\Omega) = \{1, 2, 3, 4, 5, 6\} \quad p_X(x) = P(X = x) = \frac{1}{6} \quad \forall x \in X(\Omega)$$

Allora

$$E[X] = \sum_{i=1}^6 x_i \cdot \frac{1}{6} = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3.5$$

Esempio:

Scegliamo casualmente un bimbo nato in Italia nel 1900 e indichiamo con X gli anni di vita.

Allora $X(\Omega) = \{0, 1, 2, \dots, 125(?)\}$.

La quantità $E[X]$ è la speranza di vita (alla nascita) relativa al 1900. Si stima $E[X] \approx 62.5$.

Proprietà del valore medio:

Per ogni variabile aleatoria (reale) X

- $E[c + X] = c + E[X]$
- $E[c \cdot X] = c \cdot E[X]$

$\forall c \in \mathbb{R}$

Se X e Y sono due variabili aleatorie, che dipendono entrambe dallo stesso esperimento aleatorio, allora

- $E[X + Y] = E[X] + E[Y]$

Si dice che **il valore medio è un operatore lineare**.

Esempio:

Sia X il numero di figli maschi. Abbiamo visto che $E[X] = 1$.

Siano $Z = X - 1$ e $W = 2X$. Calcoliamo $E[Z]$ e $E[W]$:

$$E[Z] = E[X] - 1 = 1 - 1 = 0$$

$$E[W] = 2 E[X] = 2$$

Osservazione:

Ogni costante reale $c \in \mathbb{R}$ può essere pensata come una variabile aleatoria “degenera” $X = c$, che assume sempre il valore c . ($X(\Omega) = \{c\}, p_X(c) = 1$).

Se $X = c$ costante allora $E[X] = E[c] = c$.

Un'altra proprietà importante del valore medio è

$$X \geq 0 \Rightarrow E[X] \geq 0$$

Formula di trasferimento:

$$\begin{aligned} E[f(x)] &= \sum_{i=1}^n f(x_i) \cdot p_X(x_i) \\ &= \sum_{i=1}^n f(x_i) \cdot P(X = x_i) \end{aligned}$$

con $f : \mathbb{R} \rightarrow \mathbb{R}$ e $f(x)$ variabile aleatoria che assume i valori $f(x_1), f(x_2), \dots, f(x_n)$.

Questo è valido per ogni funzione $f : \mathbb{R} \rightarrow \mathbb{R}$.

In particolare

$$E[X^2] = \sum_{i=1}^n x_i^2 \cdot p_X(x_i) = \sum_{i=1}^n x_i^2 \cdot P(X = x_i)$$

Varianza:

La **varianza** è definita come

$$\begin{aligned} \text{Var}[X] &= E[(X - \mu)^2] \geq 0 \\ \text{con } \mu &= E[X] \in \mathbb{R} \end{aligned}$$

Deviazione standard:

$$\text{SD}[X] = \sqrt{\text{Var}[X]}$$

Formula alternativa:

$$\text{Var}[X] = E[X^2] - E[X]^2$$

Esempio:

Sia X il numero di figli maschi.

Abbiamo già calcolato $E[X] = 1$.

Calcoliamo

$$E[X^2] = \cancel{0^2 \cdot p_X(0)} + 1^2 \cdot p_X(1) + 2^2 \cdot p_X(2) = 1 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} = \frac{3}{2}$$

dunque

$$\text{Var}[X] = \text{E}[X^2] - \text{E}[X]^2 = \frac{3}{2} - 1^2 = \frac{1}{2}$$

$$\text{SD}[X] = \frac{1}{\sqrt{2}}$$

La deviazione standard $\text{SD}[X]$ ha la stessa “unità di misura” di X e fornisce una misura della larghezza (o dispersione) dei valori x_i assunti da X rispetto al valore medio $\text{E}[X]$.

Disuguaglianza di Chebyshev:

Per ogni $t > 0$

$$P\left(|X - \text{E}[X]| \geq t \cdot \text{SD}[X]\right) \leq \frac{1}{t^2}$$

Per semplificare poniamo

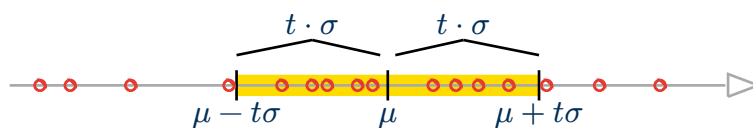
$$\left(|X - \text{E}[X]| \geq t \cdot \text{SD}[X]\right) = \left(|X - \mu| \geq t \cdot \sigma\right)$$

con

- $\mu = \text{E}[X]$
- $\sigma^2 = \text{Var}[X]$
- $\sigma = \text{SD}[X]$

Per interpretarla meglio utilizziamo il complementare dell'evento:

$$P\left(X \in (\mu - t\sigma, \mu + t\sigma)\right) \geq 1 - \frac{1}{t^2}$$



Quindi possiamo scegliere un t , ad esempio

$$t = 2 \quad P\left(X \in (\mu - 2\sigma, \mu + 2\sigma)\right) \geq 75\%$$

$$t = 3 \quad P\left(X \in (\mu - 3\sigma, \mu + 3\sigma)\right) \geq 89\%$$

Il valore medio $\text{E}[X]$ e la varianza $\text{Var}[X]$ sono due numeri reali che *riassumono le caratteristiche salienti di una v.a. X* (meglio della sua densità discreta p_X). Sono importanti anche perché talvolta possono essere calcolati senza conoscere in dettaglio la densità discreta p_X , ma sfruttando le proprietà di valore medio e varianza.

Proprietà della varianza:

Per ogni variabile aleatoria (reale) X

- $\text{Var}[c + X] = \text{Var}[X]$
- $\text{Var}[cX] = c^2 \cdot \text{Var}[X]$

$\forall c \in \mathbb{R}$

Notare che le proprietà sono diverse da quelle del valore medio.

L'intuizione è $\text{Var} = (\text{SD})^2 \approx (\text{larghezza della distribuzione})^2$.

Inoltre

$$X = c \text{ costante} \iff \text{Var}[X] = 0$$

Siano ora X e Y due variabili aleatorie, che dipendono entrambe dallo stesso esperimento aleatorio. $\text{Var}[X + Y]$ dipende da “come sono legate” X e Y .

Esempio:

Consideriamo $Y = X$. Allora $\text{Var}[Y] = \text{Var}[X]$ e

$$\text{Var}[X + Y] = \text{Var}[2X] = 4 \text{Var}[X]$$

Consideriamo invece $Y = -X$. Ancora $\text{Var}[Y] = \text{Var}[X]$, ma

$$\text{Var}[X + Y] = \text{Var}[X - X] = \text{Var}[0] = 0$$

Gli esempi $Y = X$ e $Y = -X$ sono casi estremi di “dipendenza”. Un caso totalmente diverso è quello in cui invece X e Y sono *indipendenti*, nel senso seguente.

Definizione:

Due v.a. discrete X e Y si dicono **indipendenti** se gli eventi $\{X = x\}$ e $\{Y = y\}$ sono indipendenti, ossia

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

per ogni scelta di x e y (tra i valori assunti da X e Y).

Se X e Y sono costruite come funzioni esplicite su uno spazio campionario Ω , è possibile mostrare se sono indipendenti. Molto spesso, l'indipendenza è assunta in partenza.

Teorema:

Se X e Y sono v.a. **indipendenti**, allora

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$$

e di conseguenza

$$\text{SD}[X + Y] = \sqrt{\text{SD}[X]^2 + \text{SD}[Y]^2}$$

Questa relazione di “addittività” della varianza non è intuitiva!

Se X e Y sono indipendenti e $\text{Var}[X] = \text{Var}[Y] = \sigma^2$, allora

$$\text{Var}[X + Y] = 2\sigma^2 \quad \implies \quad \text{SD}[X + Y] = \sqrt{2}\sigma$$

Se la distribuzione dei valori di X ha “larghezza” (SD) σ e lo stesso vale per Y , la “larghezza” dei valori di $X + Y$ è “solo” $\sqrt{2}\sigma \approx 1.41\sigma$ (non 2σ !). Questa “concentrazione” dei valori di $X + Y$ è dovuta alla **indipendenza di X e Y** . Il fenomeno diventa ancora più interessante quando si sommano v.a. indipendenti X_1, X_2, \dots, X_N con N molto grande (*legge dei grandi numeri*).

3.2. Distribuzioni notevoli discrete

Consideriamo una variabile aleatoria X , definita nello spazio di probabilità (Ω, P) di un certo esperimento aleatorio:

$$X : \Omega \rightarrow \mathbb{R}$$

Possiamo calcolare la probabilità $P(X \in A)$ per ogni $A \subseteq \mathbb{R}$ (ad es. per ogni intervallo $A = (a, b)$). L'insieme di tali probabilità definisce la **distribuzione (di probabilità) della v.a. X** .

$$\mu_X(A) = P(X \in A) \quad \mu_X : \overbrace{\mathcal{P}(\mathbb{R})}^{A \in} \rightarrow [0, 1]$$

È dunque una probabilità su \mathbb{R} .

Per v.a. discrete, la distribuzione di X è determinata dalla densità discreta p_X :

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

Per tale ragione, con abuso di notazione, per una v.a. discreta si può chiamare distribuzione la densità discreta.

Classifichiamo ora le **distribuzioni discrete più importanti**:

3.2.1. Bernoulli

Bernoulli:

Si chiama **Bernoulli** una v.a. X che può assumere soltanto i valori 0 e 1, cioè $X(\Omega) = \{0, 1\}$. Sia $p = P(X = 1)$. Dato che $\sum_{i=1}^n p_X(x_i) = p_X(0) + p_X(1) = 1$, si ottiene

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

Quindi X è Bernoulli \Leftrightarrow la sua densità discreta è di questa forma, per un $p \in [0, 1]$. Scriveremo $X \sim \text{Be}(p)$.

Il **valore medio** di una Bernoulli è

$$E[X] = 0 \cdot p_X(0) + 1 \cdot p_X(1) = p$$

Dato che $X(\Omega) = \{0, 1\} \Rightarrow X^2 = X \Rightarrow E[X^2] = E[X] = p$, la **varianza** è

$$\text{Var}[X] = E[X^2] - E^2 = p - p^2 = p(1 - p)$$

A dispetto della loro semplicità, le v.a. di Bernoulli sono importanti.

3.2.2. Binomiale

Consideriamo un esperimento aleatorio costituito da “prove ripetute e indipendenti”, dove ciascuna prova può avere due soli esiti (successo = 1, insuccesso = 0) con una probabilità di successo $p \in [0, 1]$ fissata. (La stessa per ogni prova).

Binomiale:

Siano $n \in \mathbb{N}$ il numero totale di prove e $p \in [0, 1]$ la probabilità di successo di ciascuna prova.

Consideriamo quindi la v.a. X come il numero di “successi” che si verificano nelle n prove.

La distribuzione di X è detta **binomiale** di parametri n e p e indicata con $X \sim \text{Bin}(n, p)$.

Osservazione:

Per $n = 1$ ritroviamo Bernoulli: $\text{Bin}(1, p) = \text{Be}(p)$.

Calcoliamo la **distribuzione** di X . Per costruzione

$$X(\Omega) = \{0, 1, 2, \dots, n\}$$

Inoltre la densità discreta è data da

$$p_X(k) = \overbrace{P(X=k)}^{k \text{ successi in } n \text{ prove}} = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

per $k = 0, 1, \dots, n$.

Nella formula

- $\binom{n}{k}$ indica le scelte di quali prove hanno successo (combinazioni)
- p^k indica la probabilità di k successi fissati
- $(1-p)^{n-k}$ indica la probabilità di $(n-k)$ insuccessi fissati

In definitiva, una v.a. X è binomiale di parametri n e p , $X \sim \text{Bin}(n, p)$, se ha questa densità discreta.

Introduciamo ora le variabili aleatorie X_1, X_2, \dots, X_n con

$$X_i = \begin{cases} 1 & \text{se la } i\text{-esima prova ha successo} \\ 0 & \text{altrimenti} \end{cases}$$

Possiamo allora scrivere

$$X = X_1 + X_2 + \dots + X_n$$

Per la natura dell'esperimento aleatorio, assumiamo che

- X_1, X_2, \dots, X_n siano v.a. indipendenti
- $X_i \sim \text{Be}(p)$

Abbiamo mostrato che $E[X_i] = p$ e $\text{Var}[X_i] = p(1-p)$.

Ricordando le proprietà di valore medio e varianza:

$$\begin{aligned} E[X] &= E[X_1] + \dots + E[X_n] \\ \text{Var}[X] &= \text{Var}[X_1] + \dots + \text{Var}[X_n] \end{aligned}$$

otteniamo che se $X \sim \text{Bin}(n, p)$:

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

Esempio:

Sia X il numero di figli maschi (coppia con due figli/e).

- $X(\Omega) = \{0, 1, 2\}$
- $p_X(0) = \frac{1}{4}$, $p_X(1) = \frac{1}{2}$, $p_X(2) = \frac{1}{4}$

Allora $X \sim \text{Bin}[2, \frac{1}{2}]$:

$$p_X(k) = \binom{2}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{2-k} = \frac{2}{k!(2-k)!} \cdot \frac{1}{2^2} = \frac{1}{2} \cdot \frac{1}{k!(2-k)!}$$

Avevamo calcolato

$$E[X] = 1 \quad \text{Var}[X] = \frac{1}{2}$$

valori che combaciano con quelli di una v.a. $\text{Bin}(2, \frac{1}{2})$.

Non è sorprendente che $X \sim \text{Bin}(2, \frac{1}{2})$ perché possiamo vedere X come il numero di “successi” in $n = 2$ prove ripetute e indipendenti con probabilità di successo $p = \frac{1}{2}$.

Infatti

$$X = X_1 + X_2 \quad \text{dove} \quad X_i = \begin{cases} 1 & \text{se } i\text{-esimo figlio è M} \\ 0 & \text{altrimenti} \end{cases}$$

e le v.a. X_1 e X_2 sono $\text{Be}(\frac{1}{2})$ indipendenti.

3.2.3. Poisson

Poisson:

Una v.a. X si dice **Poisson** di parametro $\lambda \in (0, \infty)$, scritta $X \sim \text{Pois}(\lambda)$, se $X(\Omega) = \mathbb{N}_0 = \{0, 1, 2, \dots\}$ e

$$p_X(k) = P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

per $k = 0, 1, 2, \dots$

Si può ottenere una v.a. di Poisson $X \sim \text{Pois}(\lambda)$ come opportuno **limite di una v.a. binomiale** $Y \sim \text{Bin}(n, p)$ **quando**

$$n \rightarrow \infty, \quad p \rightarrow 0 \quad \text{con } np = \lambda$$

Si può infatti mostrare che in questo limite

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} \simeq e^{-np} \frac{(np)^k}{k!} = P(X = k)$$

Si può inoltre mostrare che se $X \sim \text{Pois}(\lambda)$

$$E[X] = \lambda \quad \text{Var}[X] = \lambda$$

Le v.a. di Poisson sono approssimazioni per v.a. che contano il “numero di successi” quando si considera una grande quantità di prove la cui probabilità di successo è “piccola”.

Esempi:

- Numero di accessi a una pagina web in un'ora

- Numero di nascite in un ospedale in una giornata
- Numero di clienti in un ufficio postale in una mattinata.
- ...

Esempio:

In un ospedale nascono mediamente 2.2 bambini/e ogni giorno. Qual è la probabilità che un dato giorno non nasca nessuno, assumendo una distribuzione di Poisson? Qual è la probabilità che nascano più di 3 bambini/e?

Sia X il numero di nascite nel giorno in esame.

Assumiamo $X \sim \text{Pois}(\lambda)$. Sappiamo che $E[X] = \lambda$.

Per ipotesi $E[X] = 2.2 \Rightarrow \lambda = 2.2$.

Calcoliamo

$$P(X = 0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda} \approx 0.11 = 11\%$$

$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) = 1 - (P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)) \\ &= 1 - e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6} \right) \approx 0.18 = 18\% \end{aligned}$$

3.2.4. Geometrica

Geometrica:

Una v.a. T si dice **geometrica** di parametro $p \in (0, 1]$, e si scrive $T \sim \text{Geo}(p)$, se $T(\Omega) = \mathbb{N} = \{1, 2, 3, \dots\}$ e

$$p_T(k) = P(T = k) = p(1 - p)^{k-1}$$

per $k = 1, 2, 3, \dots$

Si può ottenere una v.a. $\text{Geo}(p)$ a partire da una successione (infinita) di prove ripetute e indipendenti, con probabilità di successo p , e considerando la v.a.

T = istante (numero della prova) del primo successo

È possibile calcolare valore medio e varianza: se $X \sim \text{Geo}(p)$

$$E[X] = \frac{1}{p} \quad \text{Var}[X] = \frac{1-p}{p^2}$$

Osservazione:

$$P(T > n) = (1 - p)^n \quad \forall n \in \mathbb{N}$$

$$\Leftrightarrow$$

$$P(T \leq n) = \sum_{k=1}^n p_T(k) = 1 - (1 - p)^n$$

In particolare

$$P(T < +\infty) = \sum_{k=1}^{\infty} p_T(k) = \lim_{n \rightarrow \infty} \sum_{k=1}^n p_T(k) = 1$$

3.3. Variabili aleatorie assolutamente continue

Esperimento aleatorio \rightsquigarrow spazio di probabilità (Ω, P)

Variabile aleatoria \rightsquigarrow funzione $X : \Omega \rightarrow \mathbb{R}$

Abbiamo studiato le v.a. *discrete*, che assumono un insieme finito oppure infinito numerabile di valori $X(\Omega) = \{x_1, x_2, \dots\}$. La distribuzione di X è determinata dalla densità discreta:

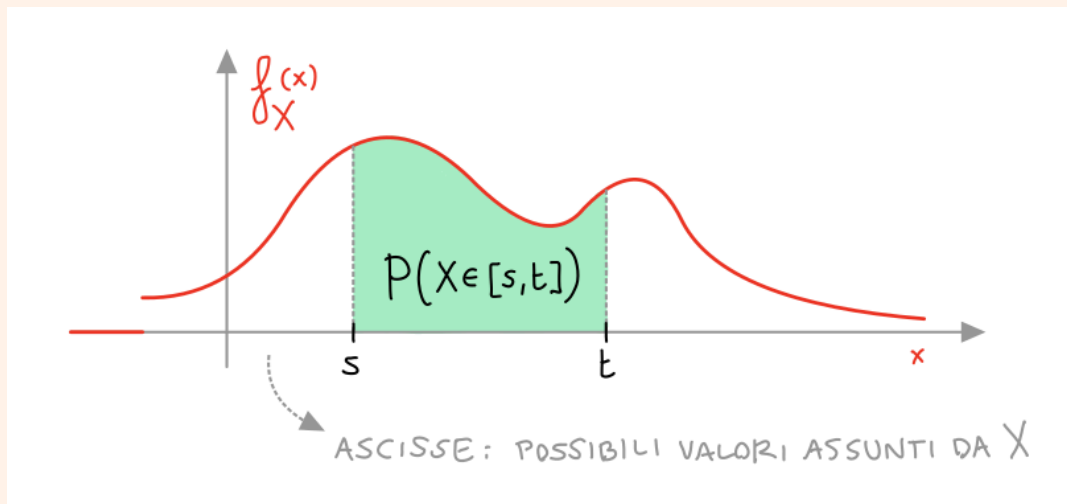
$$\begin{aligned} p_X(x_i) &= P(X = x_i) & p_X(x) &= 0 \text{ se } x \notin \{x_i\} \\ P(X \in A) &= \sum_{x_i \in A} p_X(x_i) & \text{con } A &\subseteq \mathbb{R} \end{aligned}$$

Consideriamo ora una classe “complementare” di v.a., dette **assolutamente continue**, che assumono un insieme **infinito più che numerabile di valori**, come ad es. un intervallo di \mathbb{R} :

$$[0, 1], \quad [0, +\infty), \quad (-\infty, +\infty), \quad (s, t)$$

Definizione:

Una v.a. X è assolutamente continua se la sua distribuzione $\{P(X \in A) : A \subseteq \mathbb{R}\}$ è determinata da una funzione $f_X(x)$, a valori positivi, detta **densità della v.a. X** , nel modo seguente:



$$P(X \in [s, t]) = \int_s^t f_X(x) dx$$

con $-\infty \leq s \leq t \leq +\infty$.

Ovvero l'area sotto il grafico di f_X tra i punti s e t .

In generale

$$P(X \in A) = \int_A f_X(x) dx \quad \text{con } A \subseteq \mathbb{R}$$

Proprietà: la densità di una v.a. assolutamente continua X è una funzione $f_X : \mathbb{R} \rightarrow \mathbb{R}$ (integrabile) tale che

- $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$
- $\int_{-\infty}^{+\infty} f_X(x) dx = 1$ (ovvero l'area totale sotto il grafico)

Ci sono analogie formali tra v.a. discrete e assolutamente continue, come

$$P(X \in [s, t]) = \int_s^t f_X(x) dx \iff P(X \in [s, t]) = \sum_{x_i \in [s, t]} p_X(x_i)$$

ma anche importanti differenze, infatti se X è assolutamente continua

$$P(X = x) = 0 \quad \forall x \in \mathbb{R} \quad \rightsquigarrow \quad f_X(x) \neq P(X = x)$$

Quindi

$$P(X \in [s, t]) = P(X \in (s, t)) = P(X \in [s, t)) = P(X \in (s, t])$$

Esempio: v.a. uniforme continua

Estraggo un numero aleatorio X “uniformemente” nell’intervallo

$$[0, 1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$$

Descriviamo X con una v.a. assolutamente continua:

$$f_X(x) = \begin{cases} c & \text{se } x \in [0, 1] \\ 0 & \text{se } x \notin [0, 1] \end{cases}$$

con $c = 1$.

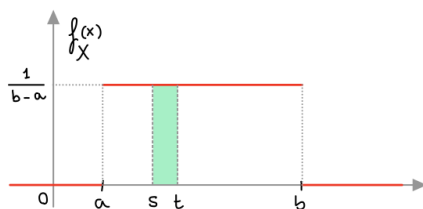
Una v.a. X con tale densità è detta **uniforme continua in $[0, 1]$** e si scrive $X \sim U(0, 1)$.

Fissiamo ora un intervallo limitato $[a, b]$.

Una v.a. X si dice uniforme continua in $[a, b]$ sse $X \sim U(a, b)$ se X è ass. cont. con densità

$$f_X(x) = \begin{cases} c & \text{se } x \in [a, b] \\ 0 & \text{se } x \notin [a, b] \end{cases} \quad \text{con } c = \frac{1}{b-a}$$

Dato un intervallo $[s, t] \subseteq [a, b]$:



$$\begin{aligned} P(X \in [s, t]) &= \int_s^t f_X(x) dx \\ &= \frac{1}{b-a} \int_s^t 1 dx = \frac{t-s}{b-a} \end{aligned}$$

Osservazione:

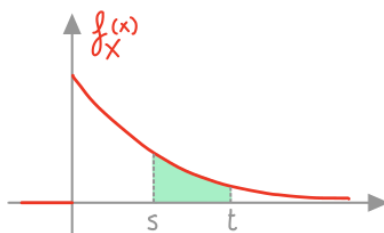
Si può mostrare che, a partire da una v.a. $U(0, 1)$, è possibile generare una v.a. con distribuzione arbitraria.

Esempio:

Misuro il tempo di emissione X di una particella radioattiva da un atomo, con “tempo medio di emissione” τ . Sia $\lambda = \frac{1}{\tau}$.

Descriviamo X con una v.a. assolutamente continua:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$



Una v.a. X con tale densità è detta **esponenziale di parametro $\lambda \in (0, \infty)$** e si scrive $X \sim \text{Exp}(\lambda)$.

Per ogni intervallo $[s, t] \subseteq [0, \infty)$:

$$\begin{aligned} P(X \in [s, t]) &= \int_s^t f_X(x) dx = \int_s^t \underbrace{\lambda e^{-\lambda x}}_{\frac{d}{dx}(-e^{-\lambda x})} dx \\ &= [-e^{-\lambda x}]_s^t = e^{-\lambda s} - e^{-\lambda t} \end{aligned}$$

Osservazione:

Si può mostrare che se $Y \sim U(0, 1)$ allora

$$X = \frac{1}{\lambda} \cdot (-\log Y) \sim \text{Exp}(\lambda)$$

I **valori assunti** da una v.a. assolutamente continua sono

$$X(\Omega) = \{x \in \mathbb{R} : f_X(x) > 0\}$$

- $X \sim U(0, 1) \Rightarrow X(\Omega) = [a, b]$
- $X \sim \text{Exp}(\lambda) \Rightarrow X(\Omega) = [0, \infty)$

Si può avere $f_X(x) > 1$ in un intervallo di punti, ad esempio $X \sim U(0, \frac{1}{2})$ ha $f_X(x) = 2$ per ogni $x \in [0, \frac{1}{2}]$.

La densità $f_X(x)$ non è univocamente definita: si può modificare in un numero finito di punti

Esempio:

X unif. cont. in $[a, b] \iff X$ unif. cont. in (a, b)

Spesso c'è una versione “canonica” di $f_X(x)$ **continua a tratti**.

Le v.a. assolutamente continue sono necessariamente definite su uno spazio campionario Ω **infinito più che numerabile**.

3.3.1. Valore medio e varianza di v.a. assolutamente continue

Le definizioni di $E[X]$ e $\text{Var}[X]$ per X assolutamente continua ricalcano quelle date per v.a. discrete:

$$E[X] = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx \qquad \text{Var}[X] = E[X^2] - E[X]^2$$

Si definisce

$$\text{SD}[X] = \sqrt{\text{Var}[X]}$$

Continuano a valere tutte le proprietà:

- $E[X + c] = E[X] + c$
- $E[cX] = c \cdot E[X]$
- $E[X + Y] = E[X] + E[Y]$
- $\text{Var}[X + c] = \text{Var}[X]$
- $\text{Var}[cX] = c^2 \cdot \text{Var}[X]$
- Se X e Y sono indipendenti: $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$

Esempio:

Sia $X \sim U(a, b)$:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in [a, b] \\ 0 & \text{se } x \notin [a, b] \end{cases}$$

Quindi

$$E[X] = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} = \frac{a+b}{2}$$

$$\text{Var}[X] = E[X^2] - E[X]^2 = \dots = \frac{(b-a)^2}{12}$$

$$\text{SD}[X] = \sqrt{\text{Var}[X]} = \frac{b-a}{2\sqrt{3}}$$

Esempio:

Sia $X \sim \text{Exp}(\lambda)$:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$

Quindi

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx = \int_0^{+\infty} x \cdot \lambda e^{-\lambda x} dx = \dots = \frac{1}{\lambda}$$

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \dots = \frac{1}{\lambda^2}$$

$$\text{SD}[X] = \sqrt{\text{Var}[X]} = \frac{1}{\lambda}$$

3.4. Funzione di ripartizione

Finora abbiamo studiato le v.a. discrete e le v.a. assolutamente continue.

Introduciamo un nuovo oggetto, per una v.a. $X : \Omega \rightarrow \mathbb{R}$ generativa:

Funzione di ripartizione:

$$F_X(x) = P(X \leq x)$$

con $x \in \mathbb{R}$.

- F_X è ben definita per ogni v.a. $X : \Omega \rightarrow \mathbb{R}$
- F_X determina la distribuzione della v.a. X . Ad esempio

$$\underbrace{P(X \in (s, t])}_{P(X \leq t) - P(X \leq s)} = F_X(t) - F_X(s) \quad \forall -\infty \leq s < t \leq +\infty$$

- F_X è legata alla densità discreta / densità di X

$$F_X(x) = \begin{cases} \sum_{x_i \in (-\infty, x]} p_X(x_i) & \text{se } X \text{ è discreta} \\ \int_{-\infty}^x f_X(t) dt & \text{se } X \text{ è assol. continua} \end{cases}$$

La funzione di ripartizione F_X è utile soprattutto per v.a. assolutamente continue, su cui ci concentreremo nel seguito. Mostriamo comunque un esempio per una v.a. discreta.

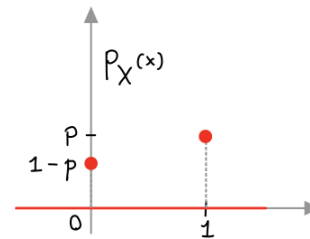
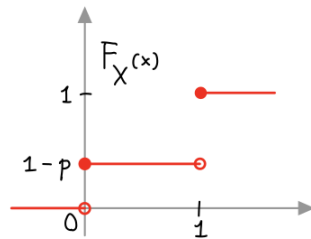
Esempio:

Sia $X \sim \text{Be}(p)$ con $p \in (0, 1)$:

$$X(\Omega) = \{0, 1\} \quad p_X(0) = 1 - p \quad p_X(1) = p$$

Allora $F_X(x) = P(X \leq x)$ vale

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - p & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases}$$



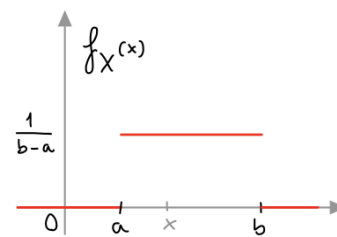
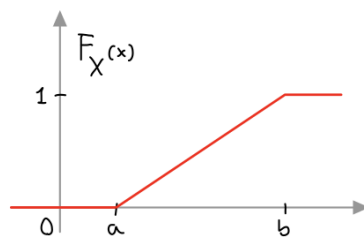
Esempio:

Sia $X \sim U(a, b)$:

$$X(\Omega) = [a, b] \quad f_X(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{se } x < a \vee x > b \end{cases}$$

Allora $F_X(x) = P(X \leq x)$ vale

$$F_X(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{se } a \leq x \leq b \\ 1 & \text{se } x > b \end{cases}$$



Funzione di ripartizione di v.a. discrete:

X è v.a. discreta $\iff F_X$ è costante a tratti

Valori assunti $\{x_i\}$ \iff Punti di discontinuità di F_X

Densità discreta \iff Ampiezze dei salti

$$p_X(x_i) = F_X(x_i) - F_X(x_i^-)$$

Funzioni di ripartizione di v.a. assolutamente continue

X è una v.a. assolutamente continua sse F_X è una **funzione continua derivabile a tratti**.

La densità è

$$f_X(x) = (F_X)'(x) = \frac{dF_X(x)}{dx}$$

dove è derivabile.

3.5. Variabili aleatorie normali

Abbiamo visto due classi notevoli assolutamente continue:

- Uniforme continua $U(a, b)$
- Esponenziale $\text{Exp}(\lambda)$

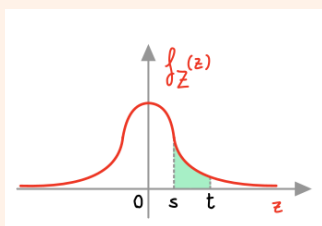
Consideriamo ora l'ultima classe, la più importante: le **variabili aleatorie normali** (o *gaussiane*).

Definizione:

Una v.a. Z si dice **normale standard**, e si scrive $Z \sim N(0, 1)$, se è assolutamente continua con densità

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \forall z \in \mathbb{R}$$

$$\rightsquigarrow Z(\Omega) = (-\infty, +\infty)$$



Ha una forma “a campana” simmetrica rispetto all'origine.

$$\text{Standard} \begin{cases} E[Z] = 0 \\ \text{Var}[Z] = 1 \end{cases}$$

Dato un intervallo $[s, t] \subseteq \mathbb{R}$

$$P(Z \in [s, t]) = \int_s^t f_Z(z) dz$$

come per ogni v.a. assolutamente continua.

Purtroppo questo integrale non si può calcolare esattamente (la densità $f_Z(z)$ non ammette primitiva esplicita). Introduciamo la **funzione di ripartizione di Z** , indicata con Φ .

$$\Phi(z) = F_Z(z) = P(Z \leq z) = \int_{-\infty}^z f_Z(t) dt \quad \forall z \in \mathbb{R}$$

Anche questa funzione non si può calcolare esplicitamente, ma i valori di $\Phi(z)$ per $z \geq 0$ sono riportati in una tavola.

I valori $\Phi(z)$ per $z < 0$ si ricavano con la formula

$$\Phi(z) = 1 - \Phi(-z) \quad \forall z \in \mathbb{R}$$

Grazie alla tavola, è come se “conoscessimo” $\Phi(z) = F_Z(z)$.

Possiamo dunque calcolare le probabilità degli intervalli

$$P(Z \in [s, t]) = F_Z(t) - F_Z(s) = \Phi(t) - \Phi(s)$$

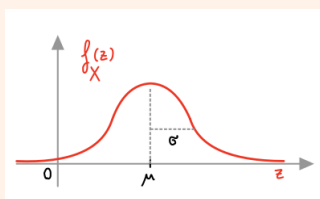
Definizione:

Siano ora $\mu \in \mathbb{R}, \sigma \in (0, \infty)$.

Una v.a. X si dice **normale con media μ e varianza σ^2** , e si scrive $X \sim N(\mu, \sigma^2)$, se X è assol. cont. con

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

$$X(\Omega) = (-\infty, +\infty)$$



La densità f_X di X si ottiene dalla densità f_Z di $Z \sim N(0, 1)$ mediante una traslazione e un riscalamento. Ci si può sempre ricondurre a una v.a. normale standard Z

$$X \sim N(\mu, \sigma^2) \implies Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

e viceversa

$$Z \sim N(0, 1) \implies X = \sigma Z + \mu \sim N(\mu, \sigma^2)$$

Si deduce in particolare che μ e σ^2 sono media e varianza:

$$X \sim N(\mu, \sigma^2) \implies E[X] = \mu \quad \text{Var}[X] = \sigma^2$$

Teorema:

Se X è normale, allora $Y = aX + b$ è normale, $\forall a \neq 0, b \in \mathbb{R}$.

$$X \sim N(\mu, \sigma^2) \implies Y \sim N(a\mu + b, a^2\sigma^2)$$

Teorema:

Se X e Y sono normali indipendenti allora $X + Y$ è normale.

$$X \sim N(\mu_x, \sigma_x^2), Y \sim N(\mu_y, \sigma_y^2) \text{ indep.} \implies X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

Esempio:

$X \sim N(0, 1)$ e $Y \sim N(0, 1)$ indipendenti. Allora

$$X + Y \sim N(0, 2) \quad X - Y \sim N(0, 2)$$

3.6. Vettori aleatori (cenni)

Finora abbiamo studiato le variabili aleatorie “individualmente”. In molte circostanze siamo interessati allo studio congiunto di due (o più) variabili aleatorie relative allo stesso esperimento aleatorio.

$$(\Omega, P) \text{ spazio di prob.} \quad \begin{cases} X : \Omega \rightarrow \mathbb{R} \text{ variabile aleatoria} \\ Y : \Omega \rightarrow \mathbb{R} \text{ variabile aleatoria} \end{cases}$$

La coppia (X, Y) è detta **vettore aleatorio**:

$$(X, Y) : \Omega \rightarrow \mathbb{R}^2$$

Esempio:

Lancio due monete (1 = testa, 0 = croce) e considero

X = faccia della prima moneta

Y = faccia della seconda moneta

Z = numero totale di teste

Esempio:

Intervisto un elettore estratto casualmente, considero

X = candidato preferito

$Y = \text{età dell'elettore}$

3.6.1. Vettori aleatori discreti

(X, Y) si dice **discreto** se i valori che può assumere sono contenuti in un insieme finito o numerabile $\{(x_i, y_i)\}$.

Ovvero le v.a. X e Y sono entrambe discrete:

$$X(\Omega) = \{x_i\} \quad Y(\Omega) = \{y_i\}$$

Si definisce la **densità discreta congiunta**

$$p_{(X,Y)}(x_i, y_i) = P(X = x_i, Y = y_i)$$

Con essa possiamo definire le **densità discrete marginali** $p_X(x_i)$ e $p_Y(y_i)$

$$p_X(x_i) = \sum_{y_i} p_{(X,Y)}(x_i, y_i)$$

$$p_Y(y_i) = \sum_{x_i} p_{(X,Y)}(x_i, y_i)$$

Esempio:

$$E[X \cdot Y] = \sum_{x_i} \sum_{y_i} x_i \cdot y_i \cdot p_{(X,Y)}(x_i, y_i)$$

3.6.2. Vettori aleatori assolutamente continui

(X, Y) si dice assolutamente continuo se esiste una funzione $f_{(X,Y)}(x, y) \geq 0$ detta **densità congiunta** di X e Y tale che

$$P(X \in [s, t], Y \in [u, v]) = \int_s^t \left(\int_u^v f_{(X,Y)}(x, y) dy \right) dx$$

Se (X, Y) è assolutamente continuo allora X e Y sono v.a. assolutamente continue (ma non viceversa!).

Le **densità marginali** $f_X(x)$ e $f_Y(y)$ si ricavano dalla densità congiunta:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y) dy \quad \forall x$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y) dx \quad \forall y$$

Esempio:

$$E[X \cdot Y] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x \cdot y \cdot f_{(X,Y)}(x, y) \, dx \, dy$$

3.6.3. Indipendenza

Consideriamo due v.a. X e Y (relative allo stesso esperimento aleatorio) che insieme formano un vettore aleatorio (X, Y) .

Si dice che X e Y sono **indipendenti** se

$$P(X \in [s, t], Y \in [u, v]) = P(X \in [s, t]) \cdot P(Y \in [u, v])$$

$$\forall [s, t], [u, v] \subseteq \mathbb{R}$$

Ovvero

$$P(X \in [s, t] \mid Y \in [u, v]) = P(X \in [s, t])$$

o “conoscere il valore che assume Y non cambia la distribuzione di X ”.

Per vettori aleatori discreti, oppure assolutamente continui, l'indipendenza ha una *riformulazione equivalente*.

(X, Y) discreto

(X, Y) ass. continuo

X e Y sono indipendenti \iff

$$p_{(X,Y)}(x_i, y_i) = p_X(x_i) p_Y(y_i)$$

$$f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$$

$$\forall x_i, y_i$$

$$\forall x, y$$

“La densità (discreta) congiunta è il prodotto delle marginali”.

Infine la proprietà fondamentale dell'indipendenza:

$$X \text{ e } Y \text{ indipendenti} \implies E[X \cdot Y] = E[X] \cdot E[Y]$$

3.6.4. Covarianza e correlazione

Consideriamo due v.a. X e Y (relative allo stesso esperimento aleatorio) che insieme formano un vettore aleatorio (X, Y) .

Indichiamo i valori medi con $\mu_X = E[X]$ e $\mu_Y = E[Y]$.

Si definisce **covarianza di X e Y**

$$\text{Cov}[X, Y] = E[(X - \mu_X) \cdot (Y - \mu_Y)]$$

La covarianza misura il grado di “associazione” tra X e Y .

$$\text{Cov}[X, Y] \geq 0 \text{ quando } X > \mu_X \text{ "tipicamente" si ha anche } Y \geq \mu_Y$$

La formula alternativa per la covarianza è

$$\text{Cov}[X, Y] = E[X \cdot Y] - E[X] \cdot E[Y]$$

In particolare se X e Y sono indipendenti, si ha $\text{Cov}[X, Y] = 0$ (ma non viceversa).

Proprietà della covarianza:

- Per ogni v.a. X : $\text{Var}[X] = \text{Cov}[X, X]$
- **Simmetria**: $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
- **Costanti**: $\text{Cov}[X, c] = 0 \quad \forall c \in \mathbb{R}$
- **Bilinearità**:
 - $\text{Cov}[aX, Y] = a \cdot \text{Cov}[X, Y] \quad \forall a \in \mathbb{R}$
 - $\text{Cov}[X + Z, Y] = \text{Cov}[X, Y] + \text{Cov}[Z, Y]$
- **Formula della somma**:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \text{Cov}[X, Y]$$

- X e Y indipendenti $\implies \text{Cov}[X, Y] = 0 \implies \text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$

Supponiamo che $\text{Var}[X] > 0$ e $\text{Var}[Y] > 0$. Si definisce **coefficiente di correlazione lineare**

$$\rho[X, Y] = \frac{\text{Cov}[X, Y]}{\text{SD}[X] \cdot \text{SD}[Y]} = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]} \cdot \sqrt{\text{Var}[Y]}}$$

Si tratta di una “*versione normalizzata della covarianza*”. Si può infatti mostrare che

- $-1 \leq \rho[X, Y] \leq +1$
- $\rho[X, Y] = \begin{cases} +1 \\ -1 \end{cases} \iff Y = aX + b \quad \text{con } a \begin{cases} >0 \\ <0 \end{cases} \quad b \in \mathbb{R}$

Se $\text{Cov}[X, Y] = 0$, X e Y si dicono **scorrelate**.

Notiamo che $\text{Cov}[X, Y] = 0 \iff \rho[X, Y] = 0$.

Dunque X e Y indipendenti \implies ~~X e Y~~ si dicono scorrelate.

3.7. Verso la statistica

Modello probabilistico fondamentale per la statistica:

Successione X_1, X_2, X_3, \dots di variabili aleatorie

I.I.D. = Indipendenti e Identicamente Distribuite

ovvero con la stessa distribuzione.

Concretamente, le v.a. X_1, X_2, X_3, \dots possono rappresentare “misure” o “rilevazioni” indipendenti, ripetute nel tempo, di una stessa quantità che si vuole studiare.

Tipicamente le v.a. (X_i) sono:

- **Discrete**, con la stessa densità discreta fissata $p(x)$:

$$p_{X_i} = p(x)$$

- **Assolutamente continue**, con la stessa densità $f(x)$:

$$f_{X_i} = f(x)$$

Esempio:

Lancio ripetutamente un dado a sei facce:

$$X_i = \text{risultato del lancio } i\text{-esimo}$$

In un esperimento concreto si osserva una successione di dati numerici x_1, x_2, x_3, \dots , che vengono interpretati come realizzazioni delle v.a. X_1, X_2, X_3, \dots :

$$x_i = X_i(\omega) \quad \text{per un certo } \omega \in \Omega$$

Il problema fondamentale è, a partire dai dati x_1, x_2, x_3, \dots , **dedurre la distribuzione comune** delle v.a. X_i .

3.7.1. La legge dei grandi numeri

Siano X_1, X_2, X_3, \dots v.a. reali i.i.d..

Sia $\mu = E[X_i], \sigma^2 = \text{Var}[X_i]$. Introduciamo la **media campionaria**

$$\overline{X}_N = \frac{1}{N} \sum_{i=1}^N X_i = \frac{X_1 + X_2 + \dots + X_N}{N}$$

Questa è una nuova v.a., diversa dalle X_i .

Teorema: Legge dei grandi numeri

Per ogni $\varepsilon > 0$

$$\lim_{N \rightarrow \infty} P\left(|\overline{X}_N - \mu| \geq \varepsilon\right) = 0 \iff \lim_{N \rightarrow \infty} P\left(\overline{X}_N \in (\mu - \varepsilon, \mu + \varepsilon)\right) = 1$$

A partire dalla sequenza di dati osservati $x_1, x_2, x_3, \dots, x_N$ si può essere fiduciosi che la media dei dati $\bar{x} = \frac{x_1 + \dots + x_N}{N}$ (ovvero la realizzazione di \overline{X}_N) sia vicina a $\mu = E[X]$.

Dimostrazione:

Siano X_1, X_2, X_3, \dots v.a. reali i.i.d. con

$$\mu = E[X_i] \quad \sigma^2 = \text{Var}[X_i]$$

Sia

$$\overline{X}_N = \frac{X_1 + X_2 + \dots + X_N}{N}$$

Si ha

$$\begin{aligned} E[\overline{X}_N] &= \frac{\overbrace{E[X_1]}^{\mu}, \overbrace{E[X_2]}^{\mu}, \dots, \overbrace{E[X_N]}^{\mu}}{N} = \frac{N\mu}{N} = \mu \\ \text{Var}[\overline{X}_N] &= \frac{\overbrace{\text{Var}[X_1]}^{\sigma^2}, \overbrace{\text{Var}[X_2]}^{\sigma^2}, \dots, \overbrace{\text{Var}[X_N]}^{\sigma^2}}{N^2} = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N} \end{aligned}$$

Per il teo di Chebyshev:

$$P(|Y - E[Y]| \geq \varepsilon) \leq \frac{\text{Var}[Y]}{\varepsilon^2} \quad \forall \varepsilon > 0$$

Dunque applicando a $Y = \overline{X}_N$ si ottiene la legge dei grandi numeri:

$$0 \leq P(|\overline{X}_N - \mu| \geq \varepsilon) \leq \frac{\frac{\sigma^2}{N}}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 N} \xrightarrow{N \rightarrow \infty} 0$$

Osservazione:

La legge dei grandi numeri è un legame tra media campionaria \overline{X}_n e media come valore atteso μ .

Osservazione:

Supponiamo di avere v.a. i.i.d. X_1, \dots, X_n con distribuzione discreta avente densità discreta p .

$F_k^{(n)}$ = frequenza relativa del valore k nelle realizzazioni X_1, \dots, X_n delle v.a.

ossia

$$F_k^{(n)} = \frac{|\{i = 1, 2, \dots, n \text{ t.c. } X_i = k\}|}{n}$$

Allora dalla L.G.N. $\forall c > 0$

$$P(|F_k^{(n)} - p(k)| \geq c) \rightarrow 0 \quad \text{se } n \rightarrow \infty$$

con $p(k) = P(X_i = k)$.

Uno dei pochi casi facili in cui si riesce a calcolare la distribuzione di \overline{X}_n è il caso di v.a. normali:

$$\begin{aligned} X_1, \dots, X_n \text{ v.a. i.i.d. } &\sim N(\mu, \sigma^2) \\ X_1, \dots, X_n \text{ indep. } &\Rightarrow X_1 + \dots + X_n \text{ normale} \end{aligned}$$

Dunque

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Ricordiamo anche che

$$\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right) \leftrightarrow \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

Osservazione:

X_1, \dots, X_n i.i.d. $\sim N(\mu, \sigma^2)$. Sia $S_n = X_1 + \dots + X_n$.

Allora $S_n \sim N(n\mu, n\sigma^2)$, quindi

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \sim N(0, 1)$$

Con le tavole della $Z \sim N(0, 1)$ si riesce a calcolare la distribuzione di \overline{X}_n e di $S_n \forall a < b$.

$$\begin{aligned} P(a < \overline{X}_n \leq b) &= P\left(\frac{a - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{b - \mu}{\frac{\sigma}{\sqrt{n}}}\right) \\ &= P\left(Z \leq \frac{b - \mu}{\frac{\sigma}{\sqrt{n}}}\right) - P\left(Z \leq \frac{a - \mu}{\frac{\sigma}{\sqrt{n}}}\right) \end{aligned}$$

Esempio:

Siano X_1, \dots, X_{100} v.a. i.i.d. $\sim N(4, 25)$. Calcolare

$$P(3 < \overline{X}_{100} \leq 5) \quad \text{e} \quad P(|S_{100}| > 100)$$

Dunque

$$\overline{X}_{100} \sim N\left(4, \frac{1}{4}\right)$$

$$P(3 < \overline{X}_{100} \leq 5) = \frac{P\left(\frac{3-4}{\frac{1}{2}} < \overline{X}_{100} - 4\right)}{\frac{1}{2}} \leq \frac{5-4}{\frac{1}{2}}$$

$$\begin{aligned}
 &= P(-2 < Z \leq 2) = P(Z \leq 2) - P(Z \leq -2) \\
 &= \Phi(2) - \Phi(-2) \approx 0.95544
 \end{aligned}$$

$$S_{100} \sim N(400, 25 \cdot 100)$$

$$\begin{aligned}
 P(|S_{100}| > 100) &= P(\{S_{100} > 100\} \cup \{S_{100} < -100\}) \\
 &= P(S_{100} > 100) + P(S_{100} < -100) \\
 &= P\left(\frac{S_{100} - 400}{50} > \frac{100 - 400}{50}\right) + P\left(\frac{S_{100} - 400}{50} < \frac{-100 - 400}{50}\right) \\
 &= P(Z > -6) + P(Z < -5) \approx 1
 \end{aligned}$$

Osservazione:

In generale $\forall x \in \mathbb{R}$ per simmetria si ha $\Phi(-x) = 1 - \Phi(x)$.

Osservazione:

Data una v.a. qualunque Y di media μ e varianza σ^2 :

$\frac{Y-\mu}{\sigma}$ è una v.a. di media 0 e varianza 1, dunque si dice **standardizzata di Y** .

Se X_1, \dots, X_n i.i.d. con $E[X_i] = \mu$ e $\text{Var}[X_i] = \sigma^2$, allora

- $\frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}}$ è la standardizzata di $\overline{X_n}$
- $\frac{S_n - n\mu}{\sqrt{n}\sigma}$ è la standardizzata di S_n

Se inoltre X_1, \dots, X_n sono normali:

- $\frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1) \leftrightarrow \overline{X_n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
- $\frac{S_n - n\mu}{\sqrt{n}\sigma} \sim N(0, 1) \leftrightarrow S_n \sim N(n\mu, \sqrt{n}\sigma)$

4. Teoremi limite

Teorema del limite centrale:

Siano X_1, \dots, X_n, \dots v.a. i.i.d. con $E[X_k] = \mu$ e $\text{Var}[X_k] = \sigma^2$. Allora

$$P\left(\frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq t\right) \rightarrow \Phi(t) \quad \text{se } n \rightarrow \infty$$

Equivalentemente, per n grande, $\frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}}$ ha approssimativamente distribuzione $N(0, 1)$.

Solitamente si richiede $n \geq 30$.

Osservazione:

La precedente affermazione si riscrive anche come

$$\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} \text{ approssimativamente } \sim N(0, 1)$$

Esempio:

Si lanci 100 volte una moneta equilibrata. Qual è la probabilità che il numero di teste sia compreso tra 40 e 70 (inclusi gli estremi)?

$$X_k = \begin{cases} 1 & \text{se il } k\text{-esimo lancio dà testa} \\ 0 & \text{altrimenti} \end{cases}$$

$$k = 1, \dots, 100 \quad X_k \sim \text{Be}\left(\frac{1}{2}\right), \text{ i.i.d.}$$

Potremmo calcolare precisamente

$$P(40 \leq X_1 + \dots + X_{100} \leq 70) = \dots \approx 0.98238$$

Possiamo anche utilizzare l'approssimazione normale che deriva dal T.L.C.

$$E[X_1] = p = \frac{1}{2} = \mu \quad \text{Var}[X_1] = p(1-p) = \frac{1}{4} = \sigma^2 \rightarrow \sigma = \frac{1}{2}$$

standardizziamo la somma $X_1 + \dots + X_{100}$

$$\frac{X_1 + \dots + X_{100} - 100 \cdot \frac{1}{2}}{\frac{1}{2} \cdot \sqrt{100}} = \frac{X_1 + \dots + X_{100} - 50}{5}$$

approssimativamente ha distribuzione $N(0, 1)$

$$P(40 \leq X_1 + \dots + X_{100} \leq 70) = \dots = \Phi(4) - \Phi(-2) \approx 0.9772$$

Gran parte dell'errore è dovuto dal fatto che $X_1 + \dots + X_n$ ha distribuzione discreta e la approssimiamo con una distribuzione continua.

Osserviamo che

$$P(40 \leq X_1 + \dots + X_{100} \leq 70) = P(39 < X_1 + \dots + X_{100} < 71)$$

Correzioni di continuità: scegliere una via di mezzo tra le precedenti alternative e calcolare

$$\begin{aligned} &P(39.5 \leq X_1 + \dots + X_{100} \leq 70.5) \\ &= P\left(\frac{39.5 - 50}{5} \leq \frac{X_1 + \dots + X_{100} - 50}{5} \leq \frac{70.5 - 50}{5}\right) \\ &\approx 0.982 \end{aligned}$$

È come trasformare una variabile discreta Y a valori interi sapendo che $P(Y \leq k) = P(Y < k + 1) = P(Y \leq k + \varepsilon)$, con $0 \leq \varepsilon < 1$

Osservazione:

La correzione di continuità va applicata tutte le volte che abbiamo una v.a. discreta.

Oss: approssimazione della binomiale con la normale

Se $X \sim \text{Bin}(n, p)$ e $n \geq 30, np \geq 5, n(1-p) \geq 5$, allora X si può approssimare (scritto \sim) come $N(np, np(1-p))$.

$$\begin{aligned} P(X = k) &= P(k - 0.5 \leq X \leq k + 0.5) \\ &= P\left(\frac{k - 0.5 - np}{\sqrt{np(1-p)}} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{k + 0.5 - np}{\sqrt{np(1-p)}}\right) \\ &\approx \Phi\left(\frac{k + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - 0.5 - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

Se $np < 5$ allora $X \sim \text{Bin}(n, p) \sim \text{Pois}(np)$.

Se $n(1-p) < 5$ allora $n - X \sim \text{Bin}(n, 1-p) \sim \text{Pois}(n(1-p))$.

Esempio:

Siano X_1, \dots, X_n, \dots i.i.d. uniformi discrete su $\{-1, 0, 1\}$.

1. Calcolare $P(|\overline{X}_{100}| > 0.01)$
2. Quanto deve essere grande n in modo che $P(|\overline{X}_4| > 0.01) < 0.1$?

$$E[X_1] = 0$$

$$\text{Var}[X_1] = E[X_1^2] = 0^2 \cdot \frac{1}{3} + (-1)^2 \cdot \frac{1}{3} + (1)^2 \cdot \frac{1}{3} = \frac{2}{3}$$

Dunque

$$\begin{aligned} P(|\overline{X}_{100}| > 0.01) &= P\left(\left|\frac{\overline{X}_{100}}{\sqrt{\frac{2}{3} \cdot \frac{1}{10}}}\right| > \frac{0.01}{\sqrt{\frac{2}{3} \cdot \frac{1}{10}}}\right) \\ &\underset{n \geq 30}{\approx} P(|Z| > 0.12) = 2P(Z > 0.12) \\ &= 2(1 - \Phi(0.12)) \approx 0.9044 \end{aligned}$$

E per il secondo punto n t.c. $P(|\overline{X}_n| > 0.01) < 0.1$:

$$\text{Var}[\overline{X}_n] = E[|\overline{X}_n - \mu|^2] = \frac{\sigma^2}{n} \rightarrow 0 \quad \text{se } n \rightarrow +\infty$$

Dunque

$$\begin{aligned} P(|\overline{X}_n| > 0.01) &= P\left(\frac{|\overline{X}_n|}{\sqrt{\frac{2}{3} \cdot \frac{1}{\sqrt{n}}}} > \frac{0.01}{\sqrt{\frac{2}{3} \cdot \frac{1}{\sqrt{n}}}}\right) \\ &\approx P\left(|Z| > \frac{0.01}{\sqrt{\frac{2}{3} \cdot \frac{1}{\sqrt{n}}}}\right) = \dots \\ &= 2\left(1 - \Phi\left(\frac{0.01}{\sqrt{\frac{2}{3} \cdot \frac{1}{\sqrt{n}}}}\right)\right) \end{aligned}$$

e

$$\begin{aligned} 2\left(1 - \Phi\left(\frac{0.01}{\sqrt{\frac{2}{3} \cdot \frac{1}{\sqrt{n}}}}\right)\right) &\leq 0.1 \\ \Phi\left(\frac{0.01}{\sqrt{\frac{2}{3} \cdot \frac{1}{\sqrt{n}}}}\right) &\geq 0.95 \\ \frac{0.01}{\sqrt{\frac{2}{3} \cdot \frac{1}{\sqrt{n}}}} &\geq 1.645 \\ n &\geq 18049 \end{aligned}$$

5. Statistica inferenziale

5.1. Introduzione

Mentre la statistica descrittiva si occupa di organizzare e presentare i dati, la statistica inferenziale cerca di trarre conclusioni da un'analisi dei dati.

Definizione:

- **Popolazione:** X_1, X_2, \dots v.a. i.i.d.
- **Campione casuale di ampiezza n :** X_1, \dots, X_n v.a. i.i.d.

Nella statistica inferenziale, la legge delle v.a. X_1, X_2, \dots non è del tutto nota e noi dovremo fare *inferenze* (deduzioni a partire dalle osservazioni) sulla legge.

$$\begin{array}{ll} X_1, \dots, X_n & \text{v.a.} \\ x_1, \dots, x_n & \text{osservazioni} \end{array}$$

5.2. Statistica parametrica

X_1, \dots, X_n campione casuale.

La legge delle v.a. è nota a meno di uno o più parametri.

Definizione:

Dati (X_1, \dots, X_n) campione casuale, si definisce **statistica** una funzione del campione, ossia una v.a. T della forma

$$T = g(X_1, \dots, X_n)$$

Esempio:

Esempi di statistiche:

- $\overline{X}_n = \frac{X_1 + \dots + X_n}{n}$ (*media campionaria*)
- $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ (*varianza campionaria*)

Entrambe queste statistiche sono **stimatori**.

Stimatori:

Gli stimatori sono particolari statistiche che ci servono a stimare i parametri incogniti.

X_1, \dots, X_n campione con parametro incognito θ da stimare.

Definizione:

Uno stimatore T si dice **non distorto** o **corretto** se

$$E[T] = E[g(X_1, \dots, X_n)] = \theta$$

con θ il parametro incognito.

Qualunque campione casuale X_1, \dots, X_n con media μ , $\overline{X_n}$ è lo stimatore non distorto di μ .

Osservazione:

Per la linearità del valore atteso se T è uno stimatore non distorto di θ , allora dati $a, b \in \mathbb{R}$, $aT + b$ è uno stimatore non distorto di $a\theta + b$. La proprietà di essere non distorto è “stabile” per trasformazioni lineari, non è “stabile” per trasformazioni non lineari.

Definizione:

Uno stimatore non distorto T di θ si dice **consistente** se

$$\text{Var}[T] \rightarrow 0 \quad \text{per } n \rightarrow +\infty$$

Quando abbiamo un campione casuale estratto da una popolazione con media μ e varianza σ^2 finite si ha sempre che $\overline{X_n}$ è uno stimatore consistente di μ .

$$\text{Var}[\overline{X_n}] = \frac{\sigma^2}{n} \rightarrow 0 \quad \text{per } n \rightarrow +\infty$$

Se lo stimatore è una generica funzione del campione, la **stima** di θ è il “risultato” della funzione quando vengono applicati i dati del campione.

La stima di un campione casuale X_1, \dots, X_n di cui osserviamo x_1, \dots, x_n è

$$\hat{\mu} = \overline{x_n} = \frac{x_1 + \dots + x_n}{n}$$

Osservazione:

Supponiamo di considerare un campione casuale con media μ incognita e varianza σ nota.

$$\text{Var}[\overline{X}_n] = \frac{\sigma^2}{n}$$

$$\text{SD}[\overline{X}_n] = \frac{\sigma}{\sqrt{n}}$$

Se si pensa \overline{X}_n come stimatore di μ , $\text{SD}[\overline{X}_n]$ prende il nome di **errore standard** e rappresenta l'errore commesso stimando μ con \overline{X}_n .

Ora consideriamo un campione con varianza incognita.

Dalla statistica descrittiva

$$s_n^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \overline{x}_n)^2$$

È sensato introdurre, in un modello statistico con varianza σ^2 incognita lo stimatore

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \overline{X}_n)^2$$

Si può verificare che $E[S_n^2] = \sigma^2$ ed è dunque corretto.

5.3. Distribuzioni delle statistiche campionarie

Campione normale

X_1, \dots, X_n campione casuale $N(\mu, \sigma^2)$.

- $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$
- $\overline{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$
- $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$
- $\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Per caratterizzare la legge di S_n^2 e di \overline{S}_n^2 dobbiamo introdurre una nuova distribuzione continua: $X^2(n)$. Si dice legge di **X (chi) quadrato con n gradi di libertà** la legge di una v.a.

$$Y = \sum_{i=1}^n Z_i^2 = Z_1^2 + \dots + Z_n^2$$

con Z_1, \dots, Z_n i.i.d. $N(0, 1)$.

Y v.a. continua, ≥ 0 .

Per $n = 2$ è la legge $\text{Exp}(\frac{1}{2})$. Per n grande vale l'approssimazione della legge $X^2(n)$ con una $N(n, 2n)$.

Proposizione

Sia (X_1, \dots, X_n) campione casuale estratto da una popolazione $N(\mu, \sigma^2)$.

- $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim X^2(n)$
- $\sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2 \sim X^2(n-1)$
- Se $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$, $(n-1) \frac{S_n^2}{\sigma^2} \sim X^2(n-1)$
- S_n^2 e \bar{X}_n sono indipendenti

Definizione:

Date $Y \sim X^2(n)$, $Z \sim N(0, 1)$ indipendenti, si dice legge **t di student con n gradi di libertà** la legge di una v.a.

$$T = \frac{Z}{\sqrt{\frac{Y}{n}}}$$

$$E[T] = 0 \quad \text{Var}[T] = 1$$

T è una v.a. continua con densità $f_T(t) = cn \left(1 + \frac{t^2}{n} \right)^{-\frac{n+1}{2}}$ per $t \in \mathbb{R}$.

Controllare formula densità

Per $n \rightarrow +\infty$, $T \sim N(0, 1)$.

Proposizione

Sia X_1, \dots, X_n un campione casuale estratto da una popolazione $N(\mu, \sigma^2)$. Allora

$$\frac{\bar{X}_n - \mu}{\sqrt{\frac{S_n^2}{n}}} \sim t(n-1)$$

Di queste variabili aleatorie ci interessa calcolare i **percentili**. X v.a., $P(X \leq q_\alpha) = \alpha$. Fissato α (**probabilità**), trovare q_α (**α -esimo quartile o 100α -percentile di X**).

- Con $Z \sim N(0, 1)$, z_α t.c. $P(Z > z_\alpha) = \alpha$ (rispetto al percentile, che considera la coda a sinistra della distribuzione, z_α considera la coda a destra)
- Con $T \sim t(n)$, $t_{\alpha,n}$ t.c. $P(T > t_{\alpha,n}) = \alpha$
- Con $Y \sim \chi^2(n)$, $\chi_{\alpha,n}^2$ t.c. $P(Y > \chi_{\alpha,n}^2) = \alpha$

Dunque $z_\alpha, t_{\alpha,n}, \chi_{\alpha,n}^2$ sono i $100(1 - \alpha)$ percentili (di Z, T, Y).

Per la normale 0, 1 ovvero $Z \sim N(0, 1)$, $z_\alpha = \Phi^{-1}(1 - \alpha)$.

Osservazione:

È importante notare la relazione

$$z_{1-\alpha} = -z_\alpha$$

data dalla simmetria della normale.

Questo vale anche per la t di student

$$t_{n,1-\alpha} = -t_{n,\alpha}$$

ma non per la χ^2 , non essendo simmetrica.

- $P(|Z| > x) = \alpha \quad x = z_{\frac{\alpha}{2}}$
- $P(|T| > x) = \alpha \quad x = t_{\frac{\alpha}{2},n}$
- $P(|Z| < x) = \alpha \quad x = z_{\frac{1-\alpha}{2}}$
- $P(|T| < x) = \alpha \quad x = t_{\frac{1-\alpha}{2},n}$

5.4. Stima per intervalli - Intervalli di confidenza

Stima per intervalli dalla media – campione normale $N(\mu, \sigma^2)$ con σ^2 nota.

\overline{X}_n stimatore di μ : per L.G.N. $\overline{X}_n \rightarrow \mu$ se $n \rightarrow +\infty$ (nel senso che $\forall c > 0, P(|\overline{X}_n - \mu| > c) \xrightarrow{n \rightarrow +\infty} 0$). Lo si deduce dal fatto che

$$\text{Var}[\overline{X}_n] = \frac{\sigma^2}{n} \quad \text{SD}[\overline{X}_n] = \frac{\sigma}{\sqrt{n}}$$

ovvero al crescere di n la stima di μ con \overline{x}_n è più affidabile. Vogliamo quantificare questa affidabilità. Si vuole costruire un intervallo centrato in \overline{X}_n a cui μ appartenga con prob. $1 - \alpha$. Ovvero si vuole cercare l'ampiezza $2E$ di questo intervallo:

$$P(|\overline{X}_n - \mu| < E) = P\left(\mu \in (\overline{X}_n - E, \overline{X}_n + E)\right) = 1 - \alpha \quad (\text{con } \alpha \text{ piccolo})$$

si vuole usare $\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$.

$$\begin{aligned} 1 - \alpha &= P\left(\frac{|\overline{X}_n - \mu|}{\frac{\sigma}{\sqrt{n}}} < \frac{E}{\frac{\sigma}{\sqrt{n}}}\right) \\ &= \text{campione normale con } \mu \text{ incognita e } \sigma \text{ nota} \\ &= P\left(|Z| < \frac{E}{\frac{\sigma}{\sqrt{n}}}\right) \end{aligned}$$

$$= P(|Z| < z_{\frac{\alpha}{2}})$$

dunque

$$E = \frac{\sigma}{\sqrt{n}} \cdot z_{\frac{\alpha}{2}}$$

Definizione:

$$\left(\overline{X}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

è l'**intervallo di confidenza per μ di livello $100(1 - \alpha)\%$** .

Prima di eseguire le osservazioni, questo è un intervallo aleatorio, quindi i suoi estremi sono v.a. e allora ha senso parlare di probabilità che il valore del parametro μ appartenga a questo intervallo.

Eseguite le osservazioni si ottiene un intervallo numerico

$$\left(\overline{x}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{x}_n + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

a cui il valore del parametro μ appartiene con una **confidenza** del $100(1 - \alpha)\%$.

Questo è un intervallo di confidenza **simmetrico** rispetto a \overline{x}_n .

Esempio:

Supponiamo che la statura dei piloti sia distribuita secondo una legge normale $N(\mu, \sigma^2)$. Si vuole stimare μ a partire dalle osservazioni della statura di 100 piloti. In prima approssimazione si suppone $\sigma = 6.1\text{cm}$ (dev. st. della popolazione adulta complessiva). Fornire una stima puntuale di μ e calcolare un **I.C.** nel caso in cui la statura media rilevata dalle 100 osservazioni sia pari a 178.5cm , con livello di confidenza del 95%.

Soluzione:

\overline{X}_n stimatore di μ stima puntuale della media.

$$\hat{\mu} = \overline{x}_n = 178.5\text{cm}$$

$$\text{Intervallo di confidenza: } \left(\overline{x}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{x}_n + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

$$95\% = 100(1 - \alpha)\% \Rightarrow \alpha = 0.05$$

$$z_{\frac{\alpha}{2}} = z_{0.025} = \Phi^{-1}(0.975) = 1.96$$

$$\text{I.C.: } (178.5 - 1.1956, \quad 178.5 + 1.1956) \approx (177.3, 179.7)$$

Livello di confidenza $100(1 - \alpha)\%$	Valore di α	Valore di $z_{\frac{\alpha}{2}}$
90	0.10	$z_{0.05} = 1.645$
95	0.05	$z_{0.025} = 1.960$
99	0.01	$z_{0.005} = 2.576$

Ampiezza dell'intervallo di confidenza: $2E = 2z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$.

E è l'errore massimo commesso stimando μ con \bar{x}_n .

La "bontà" della stima dipende da:

- Livello di confidenza: maggiore è, più affidabile è la stima
- Ampiezza dell'intervallo ($2E$): più è piccola, più precisa è la stima

Vogliamo ora calcolare un intervallo di confidenza del 99%.

$$n = 100, \quad 1 - \alpha = 0.99, \quad \alpha = 0.01, \quad z_{\frac{\alpha}{2}} = z_{0.005} = 2.578$$

$$\bar{x}_n = 178.5, \quad \sigma = 6.1$$

$$\text{I.C.: } \bar{x}_n \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 178 \pm 2.578 \cdot 0.61 \approx (176.9, 180.1)$$

Supponiamo ora di aver estratto un campione di 500 individui, anziché di 100, e di aver trovato ancora $\bar{x}_n = 178.5$. Calcolare l'intervallo di confidenza al 95%.

$$\bar{x}_n \pm \frac{6.1}{\sqrt{500}} z_{\frac{\alpha}{2}} = 178.5 \pm 1.96 \frac{6.1}{\sqrt{500}} \approx (177.9, 179.1)$$

$$E = 0.61 \cdot 1.96 \cdot \frac{1}{\sqrt{5}}$$

Il campione si è quintuplicato, E è diminuito di $\frac{1}{\sqrt{5}}$.

Calcolo della numerosità n del campione affinché l'errore (semiampiezza dell'I.C.) sia uguale (o minore uguale) a un certo E_0 assegnato (α dato).

$$E_0 = \sigma \cdot \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}} \Rightarrow \sqrt{n} = \frac{\sigma z_{\frac{\alpha}{2}}}{E_0} \Rightarrow n = \left(\frac{\sigma z_{\frac{\alpha}{2}}}{E_0} \right)^2$$

$$\frac{\sigma z_{\frac{\alpha}{2}}}{\sqrt{n}} \leq E_0 \Rightarrow \frac{\sigma z_{\frac{\alpha}{2}}}{E_0} \leq \sqrt{n} \Rightarrow n \geq \left[\left(\frac{\sigma z_{\frac{\alpha}{2}}}{E_0} \right)^2 \right]$$

Osservazione:

L'ampiezza $A = 2E$.

5.4.1. Estremi inferiori e superiori di confidenza

Estremi inferiori e superiori di confidenza per la media di una popolazione normale con varianza nota, ovvero determinare se la media di una popolazione è maggiore/minore di un certo valore.

Useremo ancora $\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n} \frac{\overline{X}_n - \mu}{\sigma} \sim N(0, 1)$.

Confidenza: $100(1 - \alpha)\%$.

Per l'estremo inferiore:

$$P\left(\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} < z_\alpha\right) = 1 - \alpha = P\left(\mu > \overline{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

Un **estremo inferiore di confidenza** (*coda destra del grafico*) al $100(1 - \alpha)\%$ per la media di una popolazione normale con varianza nota è dato da $\overline{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}}$. La sua realizzazione (dai dati campionari) è $\overline{x}_n - z_\alpha \frac{\sigma}{\sqrt{n}}$.

L'**intervallo destro di confidenza** per μ al $100(1 - \alpha)\%$ è dunque $\left(\overline{x}_n - z_\alpha \frac{\sigma}{\sqrt{n}}, +\infty\right)$.

Per l'estremo superiore:

$$P(\mu < \overline{X}_n + E) = 1 - \alpha = P\left(Z > -\frac{E}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$-\frac{E}{\frac{\sigma}{\sqrt{n}}} = z_{1-\alpha} = -z_\alpha$$

$$E = z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

Un **estremo superiore di confidenza** (*coda sinistra del grafico*) al $100(1 - \alpha)\%$ per la media di una popolazione normale con varianza nota è dato da $\overline{X}_n + z_\alpha \frac{\sigma}{\sqrt{n}}$. La sua realizzazione (dai dati campionari) è $\overline{x}_n + z_\alpha \frac{\sigma}{\sqrt{n}}$.

L'**intervallo sinistro di confidenza** per μ al $100(1 - \alpha)\%$ è dunque $\left(-\infty, \overline{x}_n + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$.

5.4.2. Intervalli di confidenza per la media di una popolazione normale con varianza incognita

X_1, \dots, X_n campione casuale proveniente da popolazione normale con media e varianza incognite $N(\mu, \sigma^2)$ e μ, σ incognite.

Con σ incognita

$$T_{n-1} = \frac{\overline{X}_n - \mu}{\sqrt{S_n^2}} \cdot \sqrt{n} \sim t(n-1)$$

σ^2 stimata con $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Ricordiamo che T_{n-1} ha le stesse simmetrie di Z .

Costruzione dell'intervallo: cerco E (semiampiezza)

$$P(|\bar{X}_n - \mu| < E) = 1 - \alpha = P\left(\frac{|\bar{X}_n - \mu|}{\sqrt{S_n^2}} \cdot \sqrt{n} < t_{n-1, \frac{\alpha}{2}}\right)$$

$$\bar{X}_n \pm t_{n-1, \frac{\alpha}{2}} \sqrt{\frac{S_n^2}{n}}$$

$$\left(\bar{X}_n - t_{n-1, \frac{\alpha}{2}} \sqrt{\frac{S_n^2}{n}}, \quad \bar{X}_n + t_{n-1, \frac{\alpha}{2}} \sqrt{\frac{S_n^2}{n}} \right)$$

→ I.C. di livello $100(1 - \alpha)\%$ per media di pop. normale con varianza incognita

Realizzazione dell'I.C. dai dati campionari

$$\left(\bar{x}_n - t_{n-1, \frac{\alpha}{2}} \sqrt{\frac{s_n^2}{n}}, \quad \bar{x}_n + t_{n-1, \frac{\alpha}{2}} \sqrt{\frac{s_n^2}{n}} \right)$$

$$E = t_{n-1, \frac{\alpha}{2}} \cdot \sqrt{\frac{s_n^2}{n}}$$

Possiamo dire:

- Con confidenza maggiore, E aumenta (a parità del campione) però la stima è più affidabile
- Estremo inferiore di confidenza al $100(1 - \alpha)\%$

$$\bar{X}_n - t_{n-1, \alpha} \frac{S_n}{\sqrt{n}} : P\left(\mu > \bar{X}_n - t_{n-1, \alpha} \frac{S_n}{\sqrt{n}}\right) = 1 - \alpha$$

- Estremo superiore di confidenza al $100(1 - \alpha)\%$

$$\bar{X}_n + t_{n-1, \alpha} \frac{S_n}{\sqrt{n}} : P\left(\mu < \bar{X}_n + t_{n-1, \alpha} \frac{S_n}{\sqrt{n}}\right) = 1 - \alpha$$

Con notazione $S_n = \sqrt{S_n^2}$

5.4.3. Stima intervallare della media per grandi campioni

Grazie al T.L.C., $\frac{\bar{X}_n - \mu}{\sigma} \sqrt{n}$ si distribuisce *circa* come $Z \sim N(0, 1)$ e $\frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \sqrt{n}$ si distribuisce *circa* come $T_{n-1} \sim t(n-1)$. I risultati precedenti rimangono validi.

5.4.4. Stima per intervalli di una frequenza per grandi campioni

Situazioni:

- Pezzi difettosi

- Sondaggi d'opinione (stimare la proporzione della popolazione che è d'accordo con un'opinione osservando il valore di questa proporzione su un campione di n individui)
- ...

X_1, \dots, X_n campione casuale $\text{Be}(p)$ t.c. valga l'approssimazione normale: $n \geq 30, np > 5, n(1-p) > 5$ (essendo che p è da stimare, utilizzo \hat{p} , ovvero la sua stima, che nel caso della bernoulli è $\hat{p} = \bar{x}_n$).

\bar{X}_n approssimativamente ha distribuzione $N\left(p, \frac{p(1-p)}{n}\right)$. Quindi fissato α trovare E : $P(|\bar{X}_n - p| < E) = 1 - \alpha$

$$p \in \bar{X}_n \pm \sqrt{\frac{p(1-p)}{n}} z_{\frac{\alpha}{2}} \text{ con una confidenza del } 100(1-\alpha)\%$$

La sua realizzazione è

$$\left(\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

dove $\hat{p} = \bar{x}_n$ è la proporzione degli individui del campione che hanno la caratteristica di interesse.

$$E = z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} \leq \frac{1}{2} \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}$$

perché $\bar{X}_n(1-\bar{X}_n) \leq \frac{1}{4}$ $\bar{X}_n \in (0, 1)$.

Esempio:

Si vuole stimare la proporzione di elettori che approva l'operato del loro governo. Su un campione di 130 persone intervistate a maggio 75 erano favorevoli. Su un secondo campione di 1056 persone intervistate a ottobre, 642 erano favorevoli.

1. Per ogni campione, si costruisca un I.C. al livello del 95% per la proporzione di elettori favorevoli
2. Si confronti la precisione delle stime effettuate
3. Insoddisfatti delle stime precedenti, si decide di procedere a un nuovo campionamento. Come si deve scegliere n se si vuole ottenere un I.C. di ampiezza non superiore all'1%?

Soluzione:

$$\underbrace{\bar{x}_n}_{\leadsto \hat{p}} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}$$

$$100(1 - \alpha) = 95 \quad \alpha = 0.05 \quad z_{\frac{\alpha}{2}} = z_{0.025} = \Phi^{-1}(0.975) = 1.96$$

1.	$n = 130$ $\bar{x}_n = \frac{75}{130} = 0.577$ $\text{I.C.: } 0.577 \pm 1.96 \sqrt{\frac{0.577(1 - 0.577)}{130}}$ $\approx 0.577 \pm 1.96 \cdot 0.043329$ $\approx 0.577 \pm 0.08$	$n = 1056$ $\bar{x}_n = \frac{642}{1056} = 0.608$ $\text{I.C.: } 0.608 \pm 1.96 \sqrt{\frac{0.608(1 - 0.608)}{1056}}$ $\approx 0.608 \pm 1.96 \cdot 0.015024$ $\approx 0.608 \pm 0.03$
----	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

2. Col secondo campione, l'ampiezza dell'I.C. è più piccola, dunque la stima è più precisa. Ce lo aspettavamo perché il campione è più numeroso.

3.

$$n \text{ t.c. } 2E = A \leq 0.01$$

$$2 \cdot z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\bar{x}_n(1 - \bar{x}_n)}{n}} \leq 2 \cdot z_{\frac{\alpha}{2}} \cdot \frac{1/\sqrt{2}}{\sqrt{n}} \leq 0.01$$

$$\Rightarrow \sqrt{n} \geq \frac{1.96}{0.01} \Rightarrow n \geq (196)^2 = 38416$$

5.5. Intervalli di confidenza per la varianza di una popolazione normale

X_1, \dots, X_n campione casuale $N(\mu, \sigma^2)$.

$$(n - 1) \frac{S_n^2}{\sigma^2} \sim \chi^2(n - 1)$$

con μ, σ^2 incognite e $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \rightsquigarrow \sigma^2$.

Questa è una statistica che ha legge nota che non dipende da μ e σ^2 incogniti.

I.C. di livello $100(1 - \alpha)\%$ costruito ponendo $1 - \alpha = P\left(a < (n - 1) \frac{S_n^2}{\sigma^2} < b\right)$

Devo calcolare a e b t.c. $P(a < Y < b) = 1 - \alpha$ con $Y = (n - 1) \frac{S_n^2}{\sigma^2} \sim \chi^2(n - 1)$.

$$P(Y > b) = P(0 < Y < a) = \frac{\alpha}{2}$$

$$b = \chi_{n-1, \frac{\alpha}{2}}^2 \quad a = ?$$

$$P(Y > a) = 1 - \frac{\alpha}{2} \Rightarrow a = \chi_{n-1, 1-\frac{\alpha}{2}}^2$$

$$P\left(\underbrace{\chi_{n-1, 1-\frac{\alpha}{2}}^2}_a < (n - 1) \frac{S_n^2}{\sigma^2} < \underbrace{\chi_{n-1, \frac{\alpha}{2}}^2}_b\right) = 1 - \alpha$$

$$\frac{(n-1)S_n^2}{\chi_{n-1, \frac{\alpha}{2}}^2} < \sigma^2 < \frac{(n-1)S_n^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}$$

Quindi la realizzazione dell'I.C. al $100(1-\alpha)\%$ è

$$\left(\frac{(n-1)s_n^2}{\chi_{n-1, \frac{\alpha}{2}}^2}, \frac{(n-1)s_n^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2} \right)$$

detto **I.C. bilatero per la varianza**.

Se volessimo calcolare intervalli sinistri e destri (ossia estremi superiori o inferiori di confidenza) procederemo in modo analogo.

Confidenza: $1 - \alpha$.

$$P\left((n-1)\frac{S_n^2}{\sigma^2} > \chi_{n-1, 1-\alpha}^2\right) = P\left(\sigma^2 < \frac{(n-1)S_n^2}{\chi_{n-1, 1-\alpha}^2}\right) = 1 - \alpha$$

Dunque $\frac{s_n^2(n-1)}{\chi_{n-1, 1-\alpha}^2}$ è l'estremo superiore di confidenza di livello $1 - \alpha$, ossia $\left[0, \frac{(n-1)s_n^2}{\chi_{n-1, 1-\alpha}^2}\right]$ è l'intervallo sinistro di confidenza.

Analogamente partendo da

$$P\left(\frac{(n-1)S_n^2}{\sigma^2} < \chi_{n-1, \alpha}^2\right) = 1 - \alpha$$

Si ha $\frac{(n-1)s_n^2}{\chi_{n-1, \alpha}^2}$ come estremo inferiore di confidenza $1 - \alpha$, ossia $\left[\frac{(n-1)s_n^2}{\chi_{n-1, \alpha}^2}, +\infty\right]$ è l'intervallo destro di confidenza per σ^2 .

Se la media fosse stata nota (e pari a μ) si sarebbe potuto usare come statistica

$$\frac{n\overline{S_n^2}}{\sigma^2} \sim \chi^2(n) \quad \text{con} \quad \overline{S_n^2} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

tutto come sopra con i gradi di libertà che passano da $n-1$ a n .

Esempio:

Avendo osservato il campione

$$\{1, 2, 4, 5, 3, 5, 6, 0, 4, 6, 7\}$$

proveniente da una legge normale;

1. Fornire una stima puntuale della varianza usando uno stimatore non distorto.
2. Calcolare l'I.C. della varianza al livello del 90%.
3. Rispondere alle stesse domande nel caso in cui μ sia nota e sia $\mu = 4$.

Soluzione:

1. $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ stimatore non distorto. $n = 11$.

$$\text{stima} \leftarrow s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x}_n)^2$$

$$\overline{x}_n = \frac{x_1 + \dots + x_{11}}{11} = 3.909$$

$$s_n^2 = \frac{1}{10} \sum_{i=1}^{11} (x_i - \overline{x}_n)^2 = 4.899 \rightarrow \hat{\sigma}^2$$

2. I.C. al 90% $\rightsquigarrow 1 - \alpha = 0.9 \rightarrow \alpha = 0.1$.

$$\left(\frac{10 \cdot s_n^2}{\chi_{10,0.05}^2}, \frac{10 \cdot s_n^2}{\chi_{10,0.95}^2} \right) = \left(\frac{48.91}{18.31}, \frac{48.91}{3.94} \right) = (2.672, 12.413) \rightsquigarrow \text{non centrato in } \sigma^2$$

3. Se fosse $\mu = 4$ si ha stimatore non distorto $\overline{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ e stima $\hat{\sigma}^2 = \overline{s}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - 4)^2 = \frac{49}{11} = 4.45$.

Si ha un'I.C.

$$\left(\frac{n\overline{s}_n^2}{\chi_{11,0.05}^2}, \frac{n\overline{s}_n^2}{\chi_{11,0.95}^2} \right) = \left(\frac{49}{19.69}, \frac{49}{4.57} \right) = (2.49, 10.711)$$

Entrambi gli estremi sono più piccoli, con varianza stimata più piccola nel secondo caso anche tramite stima per intervalli.

6. Verifica di ipotesi

6.1. Nozioni introduttive

Ipotesi statistica affermazione sulla distribuzione della popolazione in esame.

Un'ipotesi statistica può essere espressa in termini di un parametro (*test parametrici*) (ad es. $\mu = 2, \sigma^2 = 4, \mu < 0, \dots$), oppure può riguardare la natura della distribuzione della popolazione o altre caratteristiche (*test non parametrici*) (ad es. verificare se una popolazione ha distr. normale, verificare l'indipendenza, ...). Verificare un'ipotesi statistica equivale a verificare se è compatibile con i dati del campione.

Denotiamo l'**ipotesi nulla** con H_0 , mentre denotiamo la sua negazione, l'**ipotesi alternativa** con H_1 . L'ipotesi nulla viene rifiutata se “risulta” incompatibile con i dati del campione, altrimenti non viene rifiutata.

Lo scopo della **verifica di ipotesi** è trovare una *regola* che sulla base dei dati campionari permetta di rifiutare o meno H_0 . Si userà un'opportuna statistica (*statistica del test*) e a seconda del suo valore assunto sui dati campionari si rifiuterà o no.

Un test per la verifica dell'ipotesi nulla H_0 contro l'ipotesi alternativa H_1 consiste nel trovare una regione C , detta **regione critica** (o *regione di rifiuto*) tale che se $(x_1, \dots, x_n) \in C$ si rifiuta H_0 , quindi si accetta H_1 . Tale regione sarà calcolata utilizzando la statistica del test.

Esempio:

In un impianto industriale si produce una certa sostanza chimica, con impurità che fino a 150 ppm si ritengono accettabili. Si preleva un campione e si misura l'impurità X (in ppm). $X \sim N(\mu, 20^2)$. Stabilire se la produzione è sotto controllo:

$$H_0 : \mu \leq 150 \qquad H_1 : \mu > 150$$

(*ipotesi unilaterale*).

\bar{X}_n è lo stimatore di μ . Dunque la regione critica costruita a partire da \bar{x}_n plausibilmente sarà del tipo: rifiuto H_0 se $\bar{x}_n > k$ con k da trovare.

“Non basta $k = 150$ ”: se prendessi $C = \{\bar{x}_n > 150\}$, in C rifiuto H_0 .

$$P(\text{rifiutare } H_0, H_0 \text{ vera}) = P_{H_0}(\bar{X}_n - 150 > 0) = \frac{1}{2}$$

Infatti esistono due tipi di errore:

- **Errore di prima specie:** si rifiuta H_0 e H_0 è vera
- **Errore di seconda specie:** si accetta H_0 e H_0 è falsa

La regione critica “ideale” dovrebbe rendere piccola la probabilità di commettere entrambi gli errori, ma questo in genere è impossibile: restringendo la regione critica si

diminuisce la prob. di commettere un errore di prima specie ma può aumentare la prob. di commettere un errore di seconda specie e viceversa.

Si tende a diminuire la prob. di errore di prima specie perchè considerato più grave. Bisogna fissare un α piccolo ($\alpha = 0.10, 0.05, 0.01, \dots$) e chiedere che la prob. di rifiutare H_0 quando è vera sia $\leq \alpha$. In questo caso α è il **livello di significatività del test**.

Dunque un test per la verifica di H_0 con regione critica C ha livello di significatività α se

$$P_{H_0}((x_1, \dots, x_n) \in C) \leq \alpha$$

(P_{H_0} indica la prob. se vale H_0).

Il controllare l'errore di prima specie crea una "asimmetria" tra H_0 e H_1 .

- Se $(x_1, \dots, x_n) \in C$, ossia si rifiuta H_0 , i dati sperimentali sono in **contraddizione significativa con H_0** .
- Se $(x_1, \dots, x_n) \notin C$ i dati sperimentali **non** sono in **contraddizione significativa con H_0** . Non è detto che siano in contraddizione con H_1 , ma solo che non escludono in modo significativo che H_0 sia vera.

La conclusione "forte" è il rifiuto di H_0 .

L'implicazione è che se si vuole "verificare" con dati sperimentali una certa ipotesi sulla distribuzione della popolazione, ossia di una variabile, si adotterà questa ipotesi come **ipotesi alternativa**.

Esistono vari tipi di ipotesi, ad esempio

- $\mu = \mu_0$ è un'**ipotesi semplice**
- $\mu > \mu_0$ è un'**ipotesi composta**
- $H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$ è un'**ipotesi bilatera**
- $H_0 : \mu \leq \mu_0$ $H_1 : \mu > \mu_0$ è un'**ipotesi unilatera**

Definizione:

La **potenza** di un test è definita come $1 -$ la probabilità di errore di secondo tipo.

Tra 2 di livello α per la verifica di H_0 vs H_1 va scelto il più potente. I test che vengono presentati sono i più potenti.

6.2. Test Z per la media di una popolazione normale con varianza nota

X_1, \dots, X_n campione casuale proveniente da una popolazione $N(\mu, \sigma^2)$, σ^2 nota.

Verifica di $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$ (**test bilatero**).

\overline{X}_n stimatore di μ (\overline{x}_n stima di μ), dunque ragionevole che la regione critica sia del tipo

$$C = \{(x_1, \dots, x_n) \in \mathbb{R}^n : |\bar{x}_n - \mu_0| > c\}$$

con c da trovare.

α = livello di significatività del test.

$$P_{\mu_0}(|\bar{x}_n - \mu_0| > c) = \alpha$$

come per gli I.C. ho una statistica la cui legge è nota:

$$\sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma} = \frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1) \quad \text{se vale } H_0$$

Dunque P_{μ_0} si può riscrivere come

$$P_{\mu_0} \left(\sqrt{n} \frac{|\bar{X}_n - \mu_0|}{\sigma} > \frac{c\sqrt{n}}{\sigma} \right) = \alpha$$

ovvero rifiuto H_0 a livello α se

$$\left| \sqrt{n} \frac{\bar{x}_n - \mu_0}{\sigma} \right| > z_{\frac{\alpha}{2}}$$

altrimenti non rifiuto H_0 .

Osservazione:

$(x_1, \dots, x_n) \in C$, ossia rifiuto H_0 a livello α , solo se

$$\mu \notin \left(\bar{x}_n - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}, \bar{x}_n + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}} \right)$$

cioè μ_0 non appartiene all'I.C. per μ di livello $100(1 - \alpha)\%$.

Osservazione:

L'appartenenza di (x_1, \dots, x_n) alla regione critica dipende dalla scelta del livello di significatività α . Esiste un α t.c.

- Per $\alpha > \bar{\alpha}$ si rifiuta H_0
- Per $\alpha \leq \bar{\alpha}$ si accetta H_0

$\bar{\alpha}$ viene detto **p-value** (o **p dei dati** o **valore p**). In questo test per calcolare $\bar{\alpha}$ si pone

$$\frac{|\bar{x}_n - \mu_0|}{\frac{\sigma}{\sqrt{n}}} = z_{\frac{\bar{\alpha}}{2}} \quad \text{ossia} \quad 1 - \frac{\bar{\alpha}}{2} = \Phi\left(\frac{|\bar{x}_n - \mu_0|}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$\bar{\alpha} = 2\left(1 - \Phi\left(\left|\frac{\bar{x}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right|\right)\right)$$

Più piccolo è il p-value, più i dati sono in contraddizione con H_0 .

Esempio:

Per una variabile aleatoria normale con media incognita e deviazione standard $\sigma = 2$ si raccoglie un campione di 10 dati, che forniscono $\bar{x}_{10} = 18.58$.

1. Si verifichi l'ipotesi $H_0 : \mu = 20$ vs $H_1 : \mu \neq 20$ a l.s. $\alpha = 5\%$ e si calcoli il p-value del test

Statistica del test: $Z = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma} \underset{H_0}{\sim} N(0, 1)$. Rifiuto H_0 se $|Z| > z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$.

$$|Z| = \left| \frac{\bar{x}_n - 20}{\sigma} \sqrt{10} \right| = \left| \frac{18.58 - 20}{2} \sqrt{10} \right| = 2.24 > 1.96$$

rifiuto H_0 a livello $\alpha = 5\%$.

Quindi il p-value $\bar{\alpha}$ sarà più piccolo di 0.05. Lo trovo ponendo

$$z_{\frac{\bar{\alpha}}{2}} = \left| \frac{\bar{x}_n - 20}{\frac{2}{\sqrt{10}}} \right| = 2.24 \quad \text{ossia}$$

$$1 - \frac{\bar{\alpha}}{2} = \Phi(2.24) \quad \bar{\alpha} = 2(1 - \Phi(2.24)) \approx 0.0248 = 2.48\%$$

$\alpha = 1\% \Rightarrow$ accetto H_0 perchè $\alpha < \bar{\alpha}$, ovvero il p-value.

2. Qual è la probabilità che, assumendo un livello di significatività $\alpha = 5\%$, l'ipotesi nulla non venga rifiutata quando $\mu = 21$ (" $\mu \in H_1$ " $\rightarrow H_0$ è falsa)?

Si tratta di calcolare la probabilità di errore di secondi tipo per un determinato valore di μ .

Statistica $\sqrt{10} \frac{\bar{X}_n - 20}{2}$ e calcolo

$$\begin{aligned}
 & P_{\mu=21} \left(\underbrace{\left| \sqrt{10} \frac{\bar{X}_n - 20}{2} \right|}_{\text{non rifiuto } H_0} < Z_{0.025} \right) \\
 &= P_{\mu=21} \left(-1.96 < \sqrt{10} \frac{\bar{X}_n - 20}{2} < 1.96 \right) \\
 &= P_{\mu=21} \left(-1.96 - \frac{\sqrt{10}}{2} < \underbrace{\frac{\bar{X}_n - 20 - 1}{2} \sqrt{10}}_{\substack{\frac{\bar{X}_n - 21}{2} \sim N(0,1) \text{ se } \mu=21}} < 1.96 + \frac{\sqrt{10}}{2} \right) \\
 &= P(-3.54 < Z < 0.37) \quad \text{con } Z \sim N(0, 1) \\
 &= \Phi(0.37) - \Phi(-3.54) \approx 0.6443
 \end{aligned}$$

Quindi si ha una probabilità elevata di errore di secondo tipo.

Verifica di ipotesi unilateri:

1. $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$
2. $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$

Risultano uguali alla verifica di:

1. $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$
2. $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$

Ripetiamo il ragionamento sulla statistica fatto per il test bilatero.

Se H_0 è vera, $\frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$. Quindi la regione critica sensata sarà del tipo: rifiuto H_0 se $\bar{x}_n > \mu_0 + k$ ($k > 0$).

La costruzione della regione critica è legata ad α . Si dovrà avere

$$\begin{aligned}
 P_{\mu_0} \left(\frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} > \dots \right) &= \alpha \quad \dots = z_\alpha \\
 C &= \left\{ (x_1, \dots, x_n) : \sqrt{n} \frac{\bar{x}_n - \mu_0}{\sigma} > z_\alpha \right\}
 \end{aligned}$$

ossia rifiuto H_0 a livello α se $\sqrt{n} \frac{\bar{x}_n - \mu_0}{\sigma} > z_\alpha$.

Il p -value del test è $\bar{\alpha} = 1 - \Phi\left(\frac{\bar{x}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right)$, ottenuto ponendo $\sqrt{n} \frac{\bar{x}_n - \mu_0}{\sigma} = z_{\bar{\alpha}}$; $\Phi(z_{\bar{\alpha}}) = 1 - \bar{\alpha}$.

Per verificare $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$ a livello di significatività α si usa la regione critica

$$C = \left\{ (x_1, \dots, x_n) : \frac{x_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -z_\alpha \right\}$$

Il p -value è $\bar{\alpha} = \Phi\left(\frac{\bar{x}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right)$.

Osservazione:

Il test Z per la media di una popolazione normale con varianza nota si può usare su una popolazione con legge qualunque, con varianza nota, purché il campione sia sufficientemente numeroso: con $n \geq 30$ si usa l'approssimazione normale che deriva dal TLC.

$$\frac{\bar{X}_n - \mu_0}{\sigma} \cdot \sqrt{n} \approx N(0, 1)$$

Osservazione:

La regione critica dei test unilateri è il complementare degli intervalli destri/sinistri di confidenza $100(1 - \alpha)\%$.

Esempio:

Il preside di una scuola elementare sospetta che i suoi alunni abbiano un QI superiore alla media italiana. Dopo aver selezionato casualmente 64 bambini tra i suoi alunni e misurati i loro QI, il preside riscontra un QI medio pari a 106, contro una media italiana pari a 100. Supponiamo che il QI di un alunno si distribuisca come una $N(\mu, 256)$.

1. A livello di significatività $\alpha = 0.025$ può il preside concludere che i suoi studenti siano più intelligenti della media?
2. Ripetere il test a livello di significatività $\alpha = 0.025$ nel caso in cui la varianza sia 900 e calcolare il p -value del test.
3. L'ipotesi di normalità è essenziale?

Soluzione:

3. $n = 64 > 30$, quindi l'ipotesi di normalità non è essenziale.

$H_0 : \mu = 100$ vs $H_1 : \mu > 100$. Statistica $\frac{\bar{X}_n - \mu_0}{\sigma} \sqrt{n} = Z$. $\mu_0 = 100$, $n = 64$.

1. $\sigma^2 = 256 \rightarrow \sigma = 16 \rightarrow z = \frac{106 - 100}{\frac{16}{8}} = 3$. Rifiuto H_0 a livello $\alpha = 2.5\%$ se

$$\frac{\bar{x}_n - \mu_0}{\sigma} \sqrt{n} > z_\alpha$$

$$z_\alpha = z_{0.025} = 1.96$$

2. $\sigma = 900$

$$z = \frac{106 - 100}{\frac{30}{8}} = \frac{24}{15} = 1.6$$

$16 < 1.96 \Rightarrow$ i dati non consentono di rifiutare l'ipotesi nulla \Rightarrow a livello $\alpha = 0.025$ non si può concludere che nella scuola il QI sia maggiore della media.

p -value: $\bar{\alpha} = 1 - \Phi(1.6) = 0.0548$.

6.3. Test t per la media di una popolazione normale con varianza incognita

Verificare

$$\begin{array}{ll} \text{test bilatero } \{H_0 : \mu = \mu_0 & \text{vs } H_1 : \mu \neq \mu_0 \\ \text{test unilatero } \begin{cases} H_0 : \mu \leq \mu_0 & \text{vs } H_1 : \mu > \mu_0 \\ H_0 : \mu \geq \mu_0 & \text{vs } H_1 : \mu < \mu_0 \end{cases} \end{array}$$

Si ripetono i ragionamenti precedenti, ma alla statistica $Z = \frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ si sostituisce in totale analogia a quanto fatto con gli intervalli di confidenza la statistica

$$T = \frac{\bar{X}_n - \mu_0}{\frac{S_n}{\sqrt{n}}} \quad T \sim t_{n-1} \quad \text{se } \mu = \mu_0$$

(con $S_n = \sqrt{S_n^2}$) e ai quantili di una $N(0, 1)$ i quantili della $t(n-1)$.

Controllare che la t di student vada messa con `op()` o senza

Si ottengono quindi le seguenti regioni critiche a livello di significatività α :

- $H_0 : \mu = \mu_0$

$$C = \left\{ (x_1, \dots, x_n) : \left| \frac{\bar{x}_n - \mu_0}{s_n} \sqrt{n} \right| > t_{n-1, \frac{\alpha}{2}} \right\}$$

- $H_0 : \mu \leq \mu_0$

$$C = \left\{ (x_1, \dots, x_n) : \frac{\bar{x}_n - \mu_0}{s_n} \sqrt{n} > t_{n-1, \alpha} \right\}$$

Osservazione:

Con le tavole a disposizione abbiamo solo i percentili della t -student, non la f.d.r., quindi non si può calcolare direttamente il p -value. Bisogna stimarlo eseguendo il test a l.s. $\alpha_1 \rightarrow$ accetto H_0 e a l.s. $\alpha_2 \rightarrow$ rifiuto H_0 . A quel punto $\alpha_1 < p\text{-value} < \alpha_2$.

Esempio:

Il contenuto dichiarato di certe lattine è di 330 ml. Scegliendo un campione casuale di 20 bottiglie si trova un contenuto medio di 328 ml e una deviazione standard pari a 3.2 ml. In base a questi dati si può ritenere che si tratti di una frode deliberata?

Visto che il campione è di misurazioni si può supporre sia normale, dunque si deve fare il test t .

Conclusione “forte”: “si tratta di frode” $\leadsto H_1$. Test unilatero $H_0 : \mu \geq 330$ vs $H_1 : \mu < 330$.

$$T = \frac{\overline{X}_n - 330}{\frac{S_n}{\sqrt{n}}} \sim t(n-1)$$

Si rifiuta H_0 a livello α se

$$\frac{\overline{x}_n - 330}{S_n} \sqrt{n} < -t_{n-1, \alpha}$$

ossia nella situazione specifica se

$$t = \frac{\overline{x}_n - 330}{s_n} \sqrt{n} = \frac{328 - 330}{3.2} \cdot \sqrt{20} = -2.7951$$

Scelgo $\alpha = 5\%$

$$t_{19, 0.05} = 1.729 - 2.79 = t < -1.729 = -t_{10, 0.05}$$

A livello $\alpha = 5\%$ si rifiuta H_0 . Anche a livello 1% e meno. Il p -value $\bar{\alpha} = 0.0058$.

6.4. Test su una proporzione (grandi campioni)

X_1, \dots, X_n campione estratto da una popolazione $B(p)$ con $p \in (0, 1)$. Si vuole sottoporre a test l'ipotesi nulla

$$H_0 : p = p_0 \quad \text{vs} \quad H_1 : p \neq p_0$$

Statistica test: $Z = \frac{\overline{X}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$. Se H_0 è vera approssimativamente Z è distribuita come una $N(0, 1)$, se $n \geq 30, np_0 > 5, n(1-p_0) > 5$.

La regione critica per la verifica di H_0 a livello α è

$$C = \left\{ (x_1, \dots, x_n) : \left| \frac{\overline{x}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right| > z_{\frac{\alpha}{2}} \right\}$$

$$p\text{-value } \bar{\alpha} : 1 - \frac{\bar{\alpha}}{2} = \Phi \left(\frac{|\overline{x}_n - p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right).$$

Analogamente, e con le osservazioni fatte anche per i test Z sulla media di una popolazione normale, si determinano le regioni critiche per i test unilaterali per la verifica di

$$\bullet H_0 : p \leq p_0 \quad \text{vs} \quad H_1 : p > p_0$$

$$C = \left\{ (x_1, \dots, x_n) : \frac{\bar{x}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z_\alpha \right\}$$

$$1 - \bar{\alpha} = \Phi \left(\frac{\bar{x}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right)$$

$$\bullet H_0 : p \geq p_0 \quad \text{vs} \quad H_1 : p < p_0$$

$$C = \left\{ (x_1, \dots, x_n) : \frac{\bar{x}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < -z_\alpha \right\}$$

$$\bar{\alpha} = \Phi \left(\frac{\bar{x}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right)$$

Esempio:

Un lotto di 500 pezzi viene ritenuto inaccettabile se contiene almeno l'8% di pezzi difettosi. Per decidere se accettare o no il lotto si esamina un campione di 100 pezzi e se ne trovano 9 difettosi. Il lotto va rigettato o no?

Popolazione $\text{Be}(p)$.

$$H_0 : p \leq 0.08 \quad \text{vs} \quad H_1 : p > 0.08$$

Test unilatero. $n = 100 > 30$, $np = 8 > 5$, $n(1-p) = 92 > 5$. Utilizzo dunque il test Z sulla proporzione.

Statistica: $\frac{\bar{x}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ che con buona approssimazione ha distr. $N(0, 1)$.

Regione critica: rifiuto H_0 a livello α se

$$Z = \frac{\bar{x}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{100}}} > z_\alpha$$

Scelgo $\alpha = 5\%$

$$z_\alpha = 1.645$$

$$Z = \frac{0.09 - 0.08}{\sqrt{0.08 \cdot 0.92}} \cdot 10 = 0.3686$$

Il lotto non va rifiutato a livello $\alpha = 5\%$.

6.5. Test su due campioni normali

Molti problemi in statistica consistono nel confronto tra 2 o più variabili. Noi ci occuperemo del confronto tra 2 popolazioni.

6.5.1. Confronto tra medie per dati accoppiati

X_1, \dots, X_n e Y_1, \dots, Y_n campioni della stessa numerosità. Supponiamo che provengano da popolazioni normali. Il campione X ha media μ_X e il campione Y ha media μ_Y . Si suppone che per $i = 1, \dots, n$ ci sia un legame. Ad es. si sta andando a misurare la stessa variabile nell'individuo i -esimo prima e dopo un certo trattamento. Per misurare l'efficacia del trattamento, sottoponiamo a verifica una delle seguenti ipotesi

1. $H_0 : \mu_X = \mu_Y$ vs $H_1 : \mu_X \neq \mu_Y$
2. $H_0 : \mu_X \leq \mu_Y$ vs $H_1 : \mu_X > \mu_Y$
3. $H_0 : \mu_X \geq \mu_Y$ vs $H_1 : \mu_X < \mu_Y$

Consideriamo il campione D_1, \dots, D_n il campione delle differenze

$$D_i = X_i - Y_i \quad i = 1, \dots, n$$

Assumiamo che anche questo sia normale di media $\mu_D = \mu_X - \mu_Y$ per la linearità del valore atteso. La varianza è incognita. Le ipotesi si riscrivono come

1. $\mu_D = 0$
2. $\mu_D \leq 0$
3. $\mu_D \geq 0$

Il test sarà

$$T = \frac{\overline{D_n}}{\frac{S_n}{\sqrt{n}}} \quad T \sim t(n-1) \quad \text{se } H_0 \text{ è vera}$$

1. Rifiuto H_0 a livello α se $|t| = \left| \frac{\overline{d_n}}{S_n} \sqrt{n} \right| > t_{n-1, \frac{\alpha}{2}}$
2. Rifiuto H_0 a livello α se $t > t_{n-1, \frac{\alpha}{2}}$
3. Rifiuto H_0 a livello α se $t < t_{n-1, \frac{\alpha}{2}}$

Osservazione:

Non c'è un legame “preciso” tra regione critica del test per la popolazione e I.C..

Test bilatero:

$$C = \left\{ \left| \frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}} \right| > z_\alpha \right\}$$

I.C. simmetrico:

$$\bar{x}_n \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}$$