PES - Probabilità e Statistica

Elia Ronchetti
@ulerich

Fabio Ferrario
@fefabo

2021/2022

Indice

Introduzione

Il corso di probabilità e statistica per l'informatica è diviso in 2 parti

1. Stastica Descrittiva - Descrivere e riassumere i dati

   1.1 Probabilità - Descrivere matematicamente i fenomeni casuali

2. Statistica inferenziale - Trarre conclusioni dai dati

Programma Esteso

1. **Statistica Descrittiva**
   - Introduzione all'analisi dei Dati
   - Statistiche Campionarie (media, mediana, quantili, varianza, correlazione)
   - Rappresentazioni grafiche
2. **Spazi di Probabilità**
   - Fenomeni Aleatori, Spazi di probabilità ed eventi
   - Proprietà di base delle probabilità
   - Probabilità condizionata
   - Elementi di calcolo Combinatorio
   - Indipendenza di Eventi
3. **Variabili Aleatorie**
   - Variabili aleatorie discrete
   - Valore medio, momenti, varianza e covarianza
   - Variabili aleatorie assolutamente continue
   - Distribuzioni notevoli discrete e assolutamente continue

- Variabili aleatorie normali

4. **Teoremi di Convergenza**
   - Convergenza di variaibli aleatorie e distribuzioni (cenni)
   - Legge dei grandi numeri
   - Teorema limite centrale
5. **Stima di Parametri**
   - Campioni e Statistiche
   - Stimatori (media e varianza campionarie)
   - Intervalli di confidenza
6. **Verifica di ipotesi**
   - Test per la verifica di un'ipotesi, errori di I e II specie
   - Test parametrici per media e varianza
   - Test non parametrici di buon adattamento ed indipendenza
7. **Regressione Lineare**
   - Introduzione alla Regressione
   - Inferenza statistica sui parametri
   - analisi dei residui

Esame

L'esame è costituito da una prova scritta e da una eventuale prova orale e riceve un voto in trentesimi.
La prova scritta è costituita da due parti:

- **Parte 1 - Teoria** 8 Domande a risposta multipla - Punteggio 10/30
- **Parte 2 - Pratica** 4 Esercizi a risposta aperta - Punteggio 20/30

La prova orale è facoltativa (su richiesta dello studente e/o docente) e può contribuire sia in maniera positiva che in maniera negativa al voto finale.

Progetto è possibile fare un progetto con il software "R", da consegnare prima dell'esame, può fornire un massimo di 2 punti.

Capitolo 1

Statistica Descrittiva

Che cos'è la statistica? La statistica è l'arte di imparare dai dati. La statistica si divide in tre rami: **Statistica Descrittiva**, Porbabilità e Statistica inferenziale. La statistica descrittiva é il ramo della statistica che ci permette di **Descrivere i dati**.

1.1 Descrivere i dati

Se misuriamo una variabile in un campione otteniamo un insieme di dati:

$$x_1, x_2, ..., x_N \text{ con N Numero di dati}$$

Se questo insieme contiene un *numero ridotto di valori distinti*, per descrivere questi dati in maniera chiara e immediata è utile riassumerli in una **tabella delle frequenze**:

Valori	Frequenza Assoluta F_i	Frequenza Relativa P_i

Notiamo che abbiamo due tipi di frequenza:

- Frequenza Assoluta F_i: Numero di volte in cui compare i nell'insieme di dati.

- Frequenza Relativa P_i :Frazione di volte in cui compare i nell'insieme di dati ($P_i = F_i/N$)

Il dato che compare con frequenza più alta è detto **Moda**.

Tipi di dati I dati possono essere di due tipi:

- Qualitativi, ovvero "categorie".
- Quantitativi, ovvero numerici.

In questo corso useremo i dati **quantitativi**.

Rappresentare i dati

Per rappresentare i dati attraverso le frequenze risulta efficace e immediato l'utilizzo di un **istogramma**, ovvero un grafico a barre che rappresenta la tabella, da cui chiaramamente è possibile risalire alla tabella stessa.

Raggruppamento dei Dati Capita spesso di avere degli insiemi di dati che assumono un numero elevato di **valori distinti**, in questi casi puó essere conveniente suddividerli in classi e determinare la frequenza di ciascuna classe.

In questo modo c'è una perdita d'informazioni (sui valori specifici),ma spesso non è un problema e così facendo possiamo calcolare le frequenze delle classi e avere un'idea migliore della distribuzione dei dati.

Dati Bivariati

Quando per ciascun individuo vengono misurate due variabili ci troviamo un insieme di N dati a coppie detti **dati bivariati**.

$$(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)$$

Queste coppie di dati sono *inseparabili.*
Anche in questo caso è possibile calcolare le frequenze, sia assolute che relative, dette **frequenze congiunte**.

Correlazione Se facciamo un *Diagramma di dispersione* (o scatterplot), possiamo evidenziare se c'è una correlazione tra i dati osservandone la tendenza.

Osservazione:

$$\text{CORRELATION} \neq \text{CAUSATION}$$

Correlazione non significa causalità! Non è detto che l'aumento di una variabile causi la diminuzione dell'altra o viceversa, potrebbe esserci una causa comune.

1.2 Riassumere i dati

Dopo aver rappresentato i dati vogliamo ora riassumerli mediante quantità numeriche, dette **Statistiche Campionarie**, al fine di sintetizzare le proprietà salienti dei dati.

1.2.1 Media e Mediana

Media Campionaria

La Media Campionaria $\overline{x}$ é una metrica utile per caratterizzare l'insieme di dati:

DEFINIZIONE

La Media Campionaria dei dati x_i contenuti in un insieme di N valori è:

$$\overline{x} := \frac{\sum_{i=1}^{N} x_i}{N}$$

Media da una tabella Se ho una tabella delle frequenze è intuitivamente facile fare una media campionaria partendo da essa:

Valori	Freq
z_1	f_1
z_2	f_2
$\vdots$	$\vdots$
z_N	f_N

$$\Longrightarrow \overline{x} := \frac{\sum_{i=1}^{N} z_i \cdot f_i}{\sum_{i=1}^{N} f_i}$$

Trasformazioni Lineari Affini dei dati Quando faccio una trasformazione lineare, per esempio nei cambi di unitá di misura, anche la media è lineare:

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

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

Mediana Campionaria

La Mediana é il valore centrale dell'insieme ordinato.

DEFINIZIONE

Dato un insieme **Ordinato** di N dati, la Mediana m é cosí definita:

- se N dispari: $m := x_{(\frac{N+1}{2})}$
- se N pari $m := \frac{x_{(\frac{N}{2})}+x_{(\frac{N}{2}+1)}}{2}$

Ovvero se abbiamo un numero dispari di elementi, la mediana é il valore $(\frac{N+1}{2})$ esimo della lista (ovvero il valore centrale), se invece abbiamo un numero pari di elementi, la mediana è la media dei due valori centrali ($\frac{N}{2}$ e $\frac{N}{2}+1$).

Osservazione: La mediana è insesibile alle code, se per esempio quindi aumento anche di molto il valore dell'ultima cifra lasciando invariate le altre la mediana non cambierà (a differenza della media).

Indici di Centralitá Media e Mediana rappresentano "Indici di Centralitá", ovvero descrivono il 'centro' dell'insieme di dati.

Moda

La moda é un indice meno importante per la rappresentazione di un insieme e indica il dato con la frequenza assoluta Maggiore.

1.2.2 Percentili e quantili

Per analizzare la distribuzione dei dati è utile usare una generalizzazione della mediana chiamato **Percentile Campionario**.

DEFINIZIONE

Il K-Esimo Percentile Campionario di un insieme di dati é il valore t per cui:

- almeno il $k\%$ dei dati è $<= t$

- almeno il $(100-k)\%$ dei dati è $>= t$

In pratica, il k-esimo percentile indica che il $k\%$ dei dati sará a sinistra dell'indice, il restante a destra.

Quantili

Eistono dei casi importanti (in quanto piú utilizzati) di percentili detti **Quantili**, questi casi più importanti sono per $k = 25$, 50 e 75.

Definiamo $k = 100p$, quindi $p = \frac{k}{100} \in [0, 1]$. I tre Quantili importanti sono:

- Primo Quartile $p = \frac{1}{4} : k = 100p = 25$-esimo percentile (q_1)
- Secondo Quartile $p = \frac{1}{2} : k = 100p = 50$-esimo percentile (q_2)
  Equivalente alla mediana m.
- Terzo Quartile $p = \frac{3}{4} : k = 100p = 75$-esimo percentile (q_3)

Questi Quartili partizionano i dati in tre parti, dandomi una idea della distribuzione dei dati.

Il Boxplot Il boxplot, o grafico 'scatola e baffi' é la rappresentazione grafica dei quantili. In questo grafico é anche possibile eliminare gli estremi *Outliers*.

N.B. Esistono diverse definizioni di quantile, R per esempio ne utilizza una diversa di default.

Calcolo del k-esimo percentile

Avendo N dati, per calcolare il k-esimo percentile t con $p = \frac{k}{100}$:

1. Ordino l'insieme di dati $x_1 <= x_2 <= \cdots <= x_n$
2. Calcolo t:
   - Se $N \cdot p$ non è intero $\implies t = x_i$, con $i = \lceil N \cdot p \rceil$
     (t è il dato la cui posizione i è l'intero successivo a $N \cdot p$)
   - Se $N \cdot p$ è intero $\implies t = \frac{x_{(N \cdot p)} x_{(N \cdot p+1)}}{2}$
     (t è la media aritmetica fra il dato in posizione $N \cdot p$ e il successivo)

1.2.3 Indici di Dispersione: Varianza e Deviazione

Ora vogliamo descrivere la dispersione dei dati di un insieme (ovvero l'ampiezza dei dati).

Fissiamo un insieme di dati e la sua media $\overline{x}$. Consideriamo gli "scarti" $(x_i - \overline{x})$ rispetto alla media. Se sommiamo tutti gli scarti otterremo 0, perché gli scarti positivi e negativi si compensano.
Quindi, se vogliamo considerare la dispersione non uso solo gli scarti ma il loro **Quadrato** $(x_i - \overline{x})^2$, ottenendo cosí solo valori positivi. Se faccio una sorta di media di questi valori ottengo la **Varianza Campionaria**

DEFINIZIONE

Si definisce **Varianza Campionaria** la 'media' del quadrato degli scarti rispetto alla media:

Varianza Campionaria: $S^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \overline{x})^2$

Come mai faccio la 'media' con $N - 1$? lo scopriremo piú avanti.
Una *Forma Alternativa* della varianza che é piú facile da calcolare é:

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

Deviazione Standard La varianza campionaria quindi mi rappresenta una media degli scarti elevati al quadrato, di conseguenza se i dati hanno una unitá di misura la varianza S^2 avrá la stessa unitá ma al quadrato.

Per avere una statistica omogenea ai dati utilizzo la **Deviazione Standard**:

DEFINIZIONE

Si definisce **Deviazione Standard** la radice quadrata della varianza campionaria:

Deviazione Standard: $S = \sqrt{S^2} = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \overline{x})^2}$

Essa rappresenta la dispersione dei dati rispetto a $\overline{x}$, ovvero la 'larghezza' della distribuzione.

Osservazione: Sia La varianza che la deviazione standard sono sempre maggiori o uguali a zero $S^2 \geq 0, S \geq 0$. Inoltre, $S^2 = S = 0$ sse tutti i dati sono uguali, quindi ogni $x_i = \overline{x}$

Utilitá Esiste una disuguaglianza, detta di Chebyschev, da cui segue il Teorema:
Dato un $c > 0$, l'intervallo intorno ad $\overline{x}$ di ampiezza proporzionale ad S:

$$(\overline{x} - c \cdot S, \overline{x} + c \cdot S)$$

Contiene una frazione $\alpha \geq 1 - \frac{1}{c^2}$ di dati.
esempio: Almeno il 75% dei dati sono contenuti nell'intervallo $(\overline{x} - 2 \cdot S, \overline{x} + 2 \cdot S)$ con $c = 2$.

Trasformazioni lineri affini e Varianza/Deviazione Se abbiamo una trasformazione lineare affine ($y_i = a \cdot x_i + b$) sappiamo che la Media é lineare, peró:

- la Varianza NON é lineare : $S_y^2 = a^2 \cdot S_x^2$
- la Deviazione Standard NON é lineare : $S_y = |a| \cdot S_x$

Notiamo che b non appare in varianza e deviazione perché "trasla" la dispersione ma non la modifica.

Scarto Interquartile La deviazione standard va a misurare la dispersione dei dati in base alla *media*, Definiamo quindi ora una dispersione per la **Mediana** come lo scarto Interquartile:

Scarto Interquartile: $\Delta = q_3 - q_1$

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

1.3 Coefficiente di Correlazione Lineare

Se abbiamo dei dati bivariati possiamo, attraverso la deviazione standard, quantificarne la **Correlazione**. Posso misurare il grado e la 'direzione' di correlazione di una coppia di dati attraverso il coefficiente di correlazione lineare.

Coefficiente di Correlazione Lineare: $r = \frac{\sum_{k=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{(N-1) \cdot S_x \cdot S_y}$

Si può mostrare che: $-1 <= r <= 1$
Dal coefficiente r si possono ottenere il tipo di correlazione:

- $r > 0$ indica una **Correlazione Positiva**
- $r < 0$ indica una **Correlazione Negativa**

E anche il grado di correlazione:

- $|r| > 0.7$ indica una **Correlazione Significativa**
- $|r| < 0.3$ indica una **Correlazione Debole**

Si nota che questi valori sono arbitrari, vanno solo a indicare che piú r si avvicina a ± 1 e piú la correlazione é forte.

Capitolo 2

Probabilità

Il calcolo delle probabilità è una teoria della matematica che permette di descrivere e studiare Fenomeni Aleatori.
Un **Fenomeno Aleaotrio** (o casuale) é un fenomeno il cui esito *non è prevedibile con certezza* a priori.

2.1 Assiomi della Probabilità

La probabilitá si basa sullo studio di **Esperimenti Aleatori**, ovvero esperimenti il cui risultto non è prevedibile con certezza. Per poter studiare gli eseprimenti aleatori bisogna poterli descrivere matematicamente. La loro descrizione matematica si articola in tre passi:

1. **Spazio campionario** (o spazio degli esiti)
   → Insieme Ω che contiene tutti i possibili esiti dell'esperimento
   es. Tiro un dado a sei facce $\Omega = 1, 2, 3, 4, 5, 6$

2. **Eventi**
   → I sottinsiemi dello spazio campionario $A \subseteq \Omega$ che riassumono affermazioni sull'esito dell'esperimento aleatorio.
   es. Tiro un dado a sei facce: esce un numero pari $A = 2, 4, 6$

3. **Probabilità**
   → 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.

Matematicamente la **Probabilità** é una funzione $P : \rho(\Omega) \rightarrow [0, 1]$ che soddisfa opportune proprietá. In questo caso $\rho(\Omega)$ sono tutti i sottoinsiemi di Ω.

Operazioni insiemistiche e Logiche

Nella definizione degli eventi, che sono concetti logici tradotti in concetti insiemistici, ci sono dei parallelismi tra le operazioni logiche e quelle insiemistiche:

- Unione $A \cup B \longleftrightarrow$ Si verifica A o B o entrambi.
- Intersezione $A \cap B \longleftrightarrow$ Si verificano A e B.
- Complementare $A^c \longleftrightarrow$ Non si verifica A.

Ci sono poi anche due interessanti proprietá dei complementari:

- $(A^c)^c = A$ (doppia negazione)
- Leggi di Demorgan: $(A \cup B)^c = A^c \cap B^c$ e $(A \cap B)^c = A^c \cup B^c$

Interpretazioni di $P(A)$

Che cosa significa grado di fiducia? cosé $P(A)$?
Esistono due diverse interpretazioni sul grado di fiducia, entrambe equamente valide:

Interpretazione Soggettivista In questa interpretazione $P(A)$ é il prezzo equo di una scommessa che paga 1 se si verifica A e 0 altrimenti.

Intrepretazione Frequentista In questa interpretazione invece $P(A)$ è la frazione asintotica di volte in cui si verifica A ripetendo l'esperimento.

Quale scelgo? Entrambe queste interpretazioni sono valide, quindi si puó scegliere quella che si vuole. Tutte e due queste interpretazioni peró devono soddisfare le due Proprietà di Base.

2.2 Proprietà di base

Ogni interpretazione probabilistica deve rispettare queste due proprietà fondamentali:

1. $P(\Omega) = 1$, ovvero la probabilitá dello spazio campionario deve essere 1 (massima).

2. Se A e B sono eventi disgiunti, cioè $A \cap B \neq \emptyset$, allora deve valere che $P(A \cup B) = P(A) + P(B)$

Da queste proprietá si deriva la Definizione dell'approccio moderno alla probabilitá, definito da Kolmogorov nel 1933:

DEFINIZIONE

Sia Ω un insieme (spazio campionario). Si dice **Probabilità** qualsiasi funzione $P : \rho(\Omega) \to [0, 1]$ che soddisfa:

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

La coppia (Ω, P) é detta **Spazio di Probabilità**

Osservazione: Se Ω è più che numerabile (ad esempio $\mathbb{R}$ o $[0, \infty]$) Una probabilità può essere definita solo per una classe $A \subseteq P(\Omega)$ di elementi.

Proprietà di Base Fissiamo uno spazio di probabilità (Ω, P) e consideriamo le proprietà 1 e 2 della definizione. Da queste proprietà si deducono le seguenti altre proprietà.
Fissiamo la prpoposizione $P(\emptyset) = 0$:

- **Regola del complementare**

$$P(A^c) = 1 - P(A) \quad \forall A$$

- **Regola della addizione di probabilità**

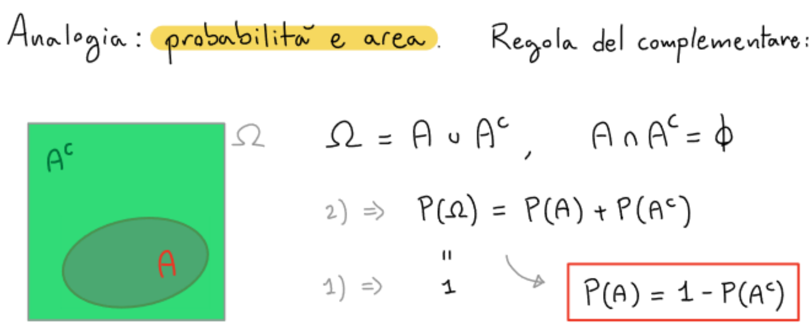
$$\to P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \forall A, B$$

Vale anche $A \cap B \neq \emptyset$

- **Monotonia**

$$\text{Se } A \subseteq B \text{ allora } P(A) <= P(B)$$

Analogia c'è un analogia tra probabilità e area



Insiemi disgiunti Multipli Dalla seconda proprietà (Se $A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)$) segue che:
Se $A_1, A_2, ..., A_k$ sono insimi disgiunti, per cui $A_m \cap A_n = \emptyset, m \neq n$, allora varrà che:

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

Per sviluppare la **Teoria Matematica** però è richiesto qualcosa in più.

2.3 Spazio di Probabilità Uniforme

Introduciamo ora il concetto di Spazio di Probabilità uniforme:

DEFINIZIONE

Indichiamo con $|A|$ la Cardinalità, ovvero il numero di elementi, dell'insieme A.
La **Probabilità Uniforme** su un insieme finito Ω è:

$$P(A) := \frac{|A|}{|\Omega|} = \frac{\text{Casi Favorevoli}}{\text{Casi Possibili}} \text{ per ogni } A \subseteq \Omega$$

In particolare, per un singolo elemento varrà:

$$P(w) = \frac{1}{\Omega} = \frac{1}{n} \text{ per ogni } \in \Omega$$

La probabilità uniforme è l'unica che mi da **la stessa probabilità per ogni elemento**.

N.B. La P.Uniforme non è l'unico tipo di probabilità esistente

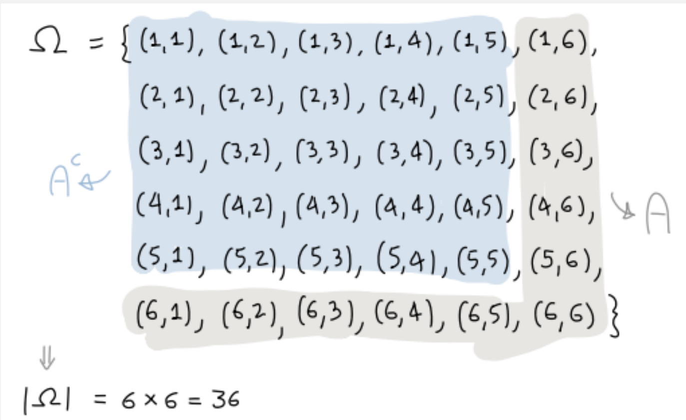
Questo è il modello appropriato per descrivere esperimenti aleatori i cui esiti siano tutti **equiprobabili**. Quando scegliamo casualmente una persona/oggetto in un insieme finito senza ulteriori specifiche, si sottintende che la scelta è effettuata in modo uniforme. Affinchè la probabilità uniforme sia ben definita, lo spazio campionario Ω deve essere finito (se così non fosse la probabilità uniforme su Ω non esiste)

Esempio: **Almeno Uno** - Qual'è la probabilità di ottenere almeno un 6 lancaindo 2 dadi regolari a sei facce?

Impostazione del Problema

- $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} = \{(x, y) : 1 \leq x, y \leq 6, x, y \in \mathbb{N}\}$
- P è una probabilità uniforme su $\Omega \rightarrow P(A) = \frac{|A|}{|\Omega|}$.
- A="Esce almeno un 6" $= \{w = (x, y) \in \Omega : x = 6 \vee y = 6\}$.

Rappresentazione Grafica



Soluzioni Contando gli elementi abbiamo due possibili soluzioni:

1. $|A| = 11 \implies P(A) = \frac{|A|}{|\Omega|} = \frac{11}{36}$

2. $A^c =$"Non esce neanche un 6$= \{1,2,3,4,5\} \times \{1,2,3,4,5\}$. $|A^c| = 5 \times 5 = 25$.
   $P(A^c) = \frac{|A^c|}{|\Omega|} = \frac{25}{36} \implies P(A) = 1 - P(A^c) = \frac{11}{36}$

2.4 Calcolo combinatorio

Consideriamo uno spazio di Probabilità Uniforme, in cui $P(A) = \frac{|A|}{|\Omega|}$. In questo tipo di spazio calcolare una probabilità significa **contare gli elementi di un insieme**.

Dato che contare non è banale per insiemi grandi sono nate tecniche di conteggio, esse formano il **Calcolo Combinatorio**

DEFINIZIONE

Principio Fondamentale: Consideriamo un esperimento costituito da due parti:

1. Prima parte con n esiti possibili $\{a_1, ..., a_n\}$.
2. Seconda parte con m esiti possibili $\{b_1, ..., b_m\}$.

L'esperimento totale può avere $n \cdot m$ **esiti possibili**.

$$\Omega = \{(a_i, b_j) : 1 \leq i \leq n, 1 \leq j \leq n\} \implies |\Omega| = n \cdot m$$

Abbiamo visto l'esempio del lancio di due dadi a sei facce, in cui:

$$\Omega = \{(x, y) : x, y \in \{1, 2, 3, 4, 5, 6\}\} \implies |\Omega| = 6 \cdot 6 = 36$$

Se lanciassi tre dadi $|\Omega| = 6 \cdot 6 \cdot 6 = 216$ esiti possibili e così via.

2.4.1 Disposizioni con ripetizione

Le disposizioni con ripetizione possono essere definite come segue:

DEFINIZIONE

Disposizioni con Ripetizione: Sequenze ordinate di k elementi (anche ripetuti) scelti tra n possibili. Il numero totale è:

$$n * n \ldots n = n^k \tag{2.1}$$

Esempio: Estraggo casualmente 3 persone: Qual'è la probabilità che siano tutte nate in primavera? In questo caso lo **Spazio Campionario** sono i compleanni delle tre persone, quindi:

$$\Omega = \{(x_1, x_2, x_3) : x_1, x_2, x_3 \in \text{ Calendario}\}$$

Questa è una disposizione con ripetizione di 3 elementi estratti dal calendario.

$$|\Omega| = 365 * 365 * 365 = 365^3$$

Abbiamo una **Probabilità Uniforme**: $P(A) = \frac{|A|}{|\Omega|}$
Troviamo $|A|$: Consideriamo tutte le persone nate tra il 20 marzo e il 21 giugno:

A = "tutti nati in primavera" = [20 marzo, 21 giugno) → 92 giorni

Si tratta anche qua di una disposizione con ripetizione di 3 elementi.

$$|A| = 92 * 92 * 92 = 92^3 (= 778.688)$$

Per calcolare la probabilità è sufficiente dividere A per Ω

$$P(A) = \frac{|A|}{|\Omega|} = \frac{92^3}{365^3} = 0,016 = 1,6\%$$

Se avessi estratto k persone sarebbe stato sufficiente sostituire l'esponente con k.

Osservazione: Esperimenti molto diversi, come lanciare 3 dadi o estrarre 3 persone e osservarne i compleanni, ammettono la **Stessa descrizione matematica**:

$$\Omega = \{(x_1, x_2, x_3) : x_1, x_2, x_3 \in E\}$$

$$\Omega = \{\text{Disposizione con Ripetizione di tre elementi estratti da E}\}$$

Con $E = \{1, 2, 3, 4, 5, 6\}$ oppure $E =$ Calendario, in ogni caso avremo sempre che $|\Omega| = |E|^3$.

2.4.2 Disposizioni semplici

Ora chiediamoci: Quanti sono gli esiti del lancio di due dadi i cui numeri ottenuti sono **distinti**?
Quante coppie (x, y) con $x \neq y$, se $x, y \in \{1, 2, 3, 4, 5, 6\}$?

DEFINIZIONE

Disposizioni Semplici[a]: Sequenze ordinate di k elementi distinti scelti tra n possibili (con $k <= n$). Sono in numero:

$$n * (n-1) * (n-2) \dots (n-k+1) = \frac{n!}{(n-k)!} \tag{2.2}$$

[a] Semplici si intende "Senza Ripetizione"

N.B. E preferibile utilizzare la prima formula su R dato che il fattoriale scala molto male, su carta spesso si semplifica, ma su Computer si calcolerebbe tutto il fattoriale e spesso richiede molto tempo.

Un caso speciale è se $k = n$: in questo caso si parla di **Permutazioni** di n oggetti. Sono in numero:

$$n! = n * (n-1) * (n-2) \dots 2 * 1 \tag{2.3}$$

Esempio: Quanti sono i possibili ordini di arrivo di 3 squadre, a, b e c?
Si tratta di una **permutazione** di 3 elementi:

$$abc, acb, bac, bca, cab, cba$$

E sono esattamente $3! = 3 \cdot 2 \cdot 1$

2.4.3 Combinazioni

In molti casi non siamo interessati all'ordine. Per esempio, se dobbiamo scegliere un comitato di 2 persone non ci interessa l'ordine dei candidati. Si parla in questo caso di **combinazioni**, esse si possono ottenere dalle disposizioni semplici "dimenticando" l'ordine degli elementi.

DEFINIZIONE

Combinazioni: Insiemi = Collezioni

$$\binom{n}{x} = \frac{n!}{k!(n-k)!} \tag{2.4}$$

Insiemi = collezioni (non ordinate) di k elementi distinti scelti tra n possibili (con $k <= n$).

Esempio . Mano di carte a Poker, un giocatore riceve 5 carte estratte da un mazzo che ne contiene 52. Il numero di possibili mani è

$$\binom{52}{5} = \frac{52!}{5!47!} \tag{2.5}$$

2.5 Probabilità Condizionata

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

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

Come è ragionevole **aggiornare la probabilità di A** per tenere conto di questa **informazione aggiuntiva**? La soluzione è data dal concetto di probabilità Condizionata.

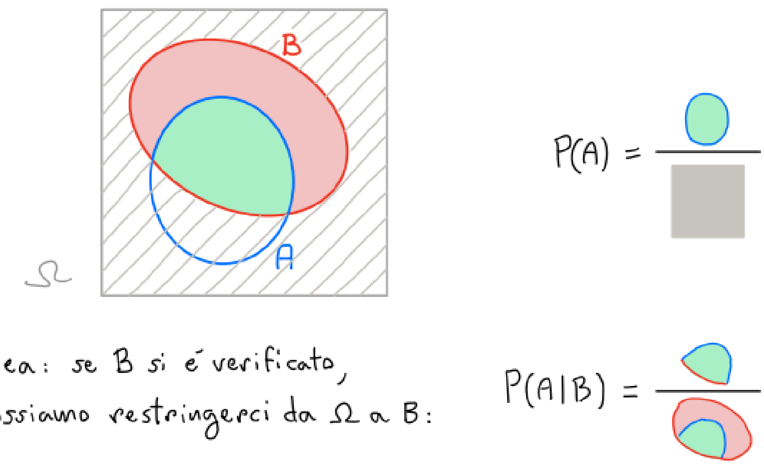
DEFINIZIONE

Probabilità Condizionata di A dato B:

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

Ovvero, la probabilità che l'evento A si verifichi **sapendo che l'evento** B **è verificato** si calcola tramite la Probabilità Condizionata.

Rappresentazione Grafica Rappresentando graficamente questo concetto abbiamo:



Ovvero quando si verifica un evento B posso ridurre lo spazio di probabilità su di esso.

2.5.1 Proprietà della Probabilità Condizionata

Elenchiamo ora alcune proprietà importanti della Probabilità Condizionata:

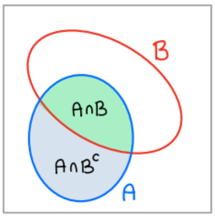
- La probabiltà dell'intersezione di A e B è data da:

  Regola del prodotto: $P(A \cap B) = P(B) \cdot P(A|B)$

  Infatti: $P(B) \cdot \frac{P(A \cap B)}{P(B)} = P(A \cap B)$

- La probabilità di A è data da:

  Formula di Disintegrazione: $P(A) = P(A \cap B) + P(A \cap B^c)$



Segue dall'additività della probabilità:

(assioma 2))

$A = (A \cap B) \cup (A \cap B^c)$

(EVENTI DISGIUNTI)

- La probabilità totale, se $P(B) \neq 0, P(B) \neq 1$, è data da:

  Formula delle probabilità totali:
  $P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)$

- La probabilità $P(\cdot|B)$ è una probabilità, perchè soddisfa gli assiomi 1 e 2.

  $P(\cdot|B)$ **è una probabilità**:

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

- Una proprietà fondamentale è la **Formula di Bayes**. Dati $P(A|B), P(A), P(B)$ la probabilità di $P(B|A)$ è:

  Formula di Bayes: $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$

Esempio: La classe 1^b è composta da 12 femmine e 4 maschi. la classe 1^c è composta da 10 femmine e 10 maschi.

Scelgo una classe *a caso*[a], quindi estraggo una persona *a caso* dalla classe scelta.

1. Qual'è la probabilità di scegliere la 1^B ed estrarre una femmina? $(P(A \cap B))$
2. Qual'è la probabilità di estrarre una femmina? $(P(A))$
3. Se estraggo una femmina, qual'è la probabilità che sia della 1^B? $(P(B|A))$

Soluzione Chiamo: A="Estraggo una femmina", e B="Scelgo la classe 1^B". Dati del problema:

1. $P(B)$ è probabilità uniforme, quindi $P(B) = \frac{1}{2} \implies P(B^c) = 1 - P(B) = \frac{1}{2}$
2. $P(A|B) = P(\text{estraggo una femmina} \mid \text{Ho scelto la } 1^B) = \frac{12}{16}$[b]$= \frac{3}{4}$
3. $P(A|B^c) = P(\text{estraggo una femmina} \mid \text{Ho scelto la } 1^C) = \frac{10}{20} = \frac{1}{2}$

Risposte alle domande:

1. $P(A \cap B) = P(B) \cdot P(A|B) = \frac{3}{8} = 37.5\%$
2. $P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c) = 62.5\%$
3. $P(B|A) = \frac{P(A|B)P(B)}{P(A)} = 60\%$, ovvero "il 60% delle femmine estratte provengono dalla 1^B".

[a] uniformemente

[b] 12 femmine su 16 studenti nella 1^B

2.6 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**.

DEFINIZIONE

Se abbiamo due eventi A e B, e la probabilità di A dato B è uguale a quella di A:

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

Allora i due eventi si dicono **Indipendenti**

Dire che gli eventi A e B sono indipendenti equivale a dire che la probabilità della loro **intersezione** è uguale al prodotto delle loro probabilità:

Intersezione di Eventi Indipendenti: $P(A \cap B) = P(A) \cdot P(B)$

Perchè se :

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

allora implica che:

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

Complementari Se A e B sono indipendenti, allora lo sono anche A e B^c, e dunque anche (A^c e B) e (A^c e B^c).
Infatti, per la formula di disintegrazione $P(A) = P(A \cap B) + P(A \cap B^c)$ implica che:

$$\begin{aligned} \implies P(A \cap B^c) &= P(A) - P(A \cap B) = P(A) - P(A) \cdot P(B) \\ &= P(A) - (1 - P(B)) = P(A) \cdot P(B^c) \end{aligned}$$

Osservazione:

Eventi indipendenti $\neq$ Eventi disgiunti!

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

- A e B disgiunti: $A \cap B = \phi \Rightarrow P(A \cap B) = 0$
- 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.

Esempio: Una famiglia ha due figli/e descritti da
$\Omega = \{MM, FF, FM, FF\}$ e P = Probabilità uniforme = $\frac{1}{4}$
Consideriamo gli eventi:

- A := "il primo genito è maschio" = {MM, FF}
- B := "il secondo genito è maschio" = {MM, FM}
- C := "la primogenita è femmina" ={FM, FF}

In questo caso A e B sono **indipendenti**, ma **NON disgiunti**;
A e C sono **disgiunti**, ma **NON indipendenti**.

Estensioni Tre eventi A, B, C si dicono indipendenti se valgono

$$\begin{gathered} P(A \cap B \cap C) = P(A)P(B)P(C) \\ P(A \cap B) = P(A)P(B) \\ P(B \cap C) = P(B)P(C) \\ P(A \cap C) = P(A)P(C) \end{gathered}$$

Capitolo 3

Variabili aleatorie

Consideriamo un esperimeno aleatorio, descritto da uno spazio di probabilità (Ω, P)[1]. Spesso non siamo interessati a tutti i dettagli dell'esito dell'esperimento, ma oslo a una *quantità, tipicamente numerica, determinata dall'esito dell'esperimento*. Una tale qunatità è detta **Variabile Aleatoria**.

Una Variabile Aleatoria[2], è una variabile che può assumere valori diversi in dipendenza da qualche fenomeno aleatorio. Il termine "aleatorio" deriva dal latino *alea* (gioco di dadi), ed esprime il concetto di rischio calcolato.

"alea iacta est" - "il dado è tratto"

DEFINIZIONE

Una **Variabile Aleatoria** ha ha due possibili definizioni, una intuitiva e una più matematica:

- Intuitivamente, è una **quantità che dipende dal "caso"**, ovvero in funzione dell'esito di un esperimento aleatorio.
- Matematicamente, è una **funzione definita sullo spazio campionario**: $X : \Omega \to R$.

Ricodiamo che un evento è una affermazione sull'esito dell'esperimento aleatorio, o matematicamente un sottoinsieme dello spazio campionario $A \subset \Omega$.

[1] Ω = l'insieme di tutti i possibili esiti dell'esperimento, P = probabilità scelta.
[2] detta anche casuale o stocastica.

Gli Eventi e la Probabilità Sia X una Variabile Aleatoria e x un suo possibile valore:
$\{X = x\}$ è un **Evento** in cui X assume il valore x, ovvero:

$$\{X = x\} = \{w \in \Omega : X(w) = x\} \subseteq \Omega$$

Sono possibili eventi anche $\{X \geq x\}$, $\{X \neq x\}$, ...

Ogni variabile aleatoria X determina molti eventi, di cui possiamo calcolare la **Probabilità**:

$$P(X = x), P(X \geq x), P(X \neq x), ...$$

Esempio: Estraggo casualmente una famiglia con due figli/e. indichiamo con X il **numero di figli maschi**.

Che valori può assumere X? Con quali probabilità?

$$\Omega = \{\text{MM,MF,FM,FF}\} \quad P = \text{Probabilità Uniforme}$$

Allora $X : \omega \to \mathbb{R}$ è definita da:

$$X(\text{MM})\text{=}2,\ X(\text{MF})\text{=}1,\ X(\text{FM})\text{=}1,\ X(\text{FF})\text{=}0,$$

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

$$P(X = 0) = P(\{\text{FF}\}) = \tfrac{1}{4},\ P(X = 2) = P(\{\text{MM}\}) = \tfrac{1}{4},$$
$$P(X = 2) = P(\{\text{MF, FM}\}) = \tfrac{1}{2}$$

N.B. Non bisogna confondere variabili aleatorie ed eventi.

3.1 Variabili Aleatorie Discrete

DEFINIZIONE

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[a]:

$$X(\Omega) = x_1, X_2, \dots = x_{i i \in \mathbb{N}} \subseteq \mathbb{R}$$

[a] Di punti "separati".

3.1.1 Densità Discreta

E' concettualmente importante sapere che una Variabile Aleatoria X è rappresentata matematicamente da una una funzione definita sullo spazio campionario Ω di un esperimento aleatorio.

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

Allo stesso tempo possiamo pensare 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, ci basta conoscere la **Densità Discreta**:

DEFINIZIONE

Ad ogni variabile aleatoria discreta X possiamo associare una **Densità discreta** così definita:

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

Anche detta distribuzione di probabilità, è una funzione che assegna ad ogni valore possibile di X la probabilità dell'evento $(X = x)$

Proprietà della densità discreta

Una densità discreta p_X è una funzione da $\mathbb{R} \in [0, 1]$ e gode delle seguenti proprietà:

- $p_X(x) = P(X = x) = 0$ Se x non è uno dei valori x_i assunti da X.
- $p_X(x_i) \geq 0 \forall i$ $(p_X(x_i) \in [0, 1])$
- $\sum_{i \geq 1} p_X(x_i) = 1$, perchè gli eventi $\{X = x_i\}$ per $i \geq 1$ sono **Disgiunti**, ovvero una partizione di Ω.

Esempio: Riprendiamo l'esempio precedente in cui X era il numero di **figli maschi** in una famiglia estratta a caso con due figli/e. Ricordiamo che $X(\Omega) = \{0, 1, 2\}$.

Sapendo le probabilità precedenti, quali sono le densità discrete $p_X(x) = P(X = x)$?

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

3.1.2 Valore Medio

DEFINIZIONE

Sia X una variabile aleatoria discreta (reale) che assume una quantità finita di valori $x_1, x_2, \ldots, x_n$. Si definisce **Valore Medio** di X $E[X]$:

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

Ovvero la somma dei valori assunti da X "pesati" con le rispettive probabilità.

Il valore medio è anche detto valore atteso, media, espettativa o speranza matematica. Si indica come $E[X]$ dall'inglese *Expected Value.*

Esempio: Riprendendo sempre l'esempio del numero di **Figli Maschi**:

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

Allora $E[X] = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2) = 1$

In poche parole se ripetiamo tante volte l'esperimento e ne calcoliamo la media otteniamo il valore medio, per esempio il valore medio del numero di figli maschi su una coppia con 2 figli è 1, perchè mediamente una coppia con 2 figli ha almeno 1 figlio maschio.

Osservazione: Si può notare che il valore medio E[X] non è necessariamente uno dei valori x_i assunti da X!

A maggior ragione, E[X] non è un valore tipico di X, nè un valore che necessariamente ci aspettiamo di osservare.
Ma allora qual è l'interpretazione del valore medio $E[X]$? A cosa serve? Riportiamo l'interpretazione frequentista e quella geometrica di EX:

Interpretazione frequentista del valore medio Supponendo di ripetere l'esperimento aleatorio un numero elevato di volte $N >> 1$ e indicando con $X_1, X_2, \ldots, X_n$ le variabili aleatorie che rappresentano X nelle ripetizioni dell'esperimento si ha con grande probabilità che:

$$\frac{X_1 + X_2 + \cdots + X_n}{N} \backsimeq E[X]$$

Anche detta Legge dei Grandi Numeri.

Interpretazione Geometrica Il valore medio $E[X]$ rappresenta il "baricentro" dei valori x_i assunti dalla variabile aleatoria X, pesati con le rispettive probabilità.

Proprietà del valore medio

Per ogni variabile aleatoria (reale) X valgono le seguenti proprietà:
Per ogni costante $c \in R$

- $E[X + c] = E[X] + c$
- $E[cX] = cE[X]$

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**.

Osservazione: Se $Z = X + c$ (oppure $Z = cX$, o $Z = X + Y$) è una variabile aleatoria, per calcolare il valore medio $E[Z]$ non c'è bisogno di determinare la densità discreta.

Altre importanti proprietà

- Se $X = c$ (costante) allora $E[X] = E[c] = c$
- Se $X >= 0$ allora $E[X] >= 0$

Formula di trasferimento

$$E[f(x)] = \sum_{i=1}^{n} f(x_i)p_X(x_i) = \sum_{i=1}^{n} f(x_i)P(X = x_i) \tag{3.1}$$

Valida per ogni funzione $f : R \to R$. In particolare

$$E[X^2] = \sum_{i=1}^{n} x_i^2 p_X^{x_i} = \sum_{i=1}^{n} x_i^2 P(X = x_i) \tag{3.2}$$

Nota per gli esercizi Queste proprietà servono nel caso in cui ci venga richiesto di calcolare una nuova media Y in funzione di X, in questo modo applicando le proprietà non sarà necessario ricalcolare il tutto, ma partendo da Y e applicando le proprietà di possiamo ricondurre a X e avere già il risultato.

3.1.3 Varianza e Deviazione Standard per le Variabili Aleatorie

Definiamo ora Varianza e Deviazione Standard per le variabili aleatorie:

DEFINIZIONE

Definiamo la Varianza $VAR[X]$ come:

$$VAR[X] := E[(X - \mu)^2] >= 0$$

con $\mu := E[X]$.

E la Deviazione Standard $SD[X]$ come:

$$SD[X] := \sqrt{VAR[X]}$$

Alternativamente, la deviazione standard può essere definita come:

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

Questa è una formula più facile da calcolare.

La deviazione standard 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 E[X].
Valore medio e varianza sono due numeri reali che riassumono le caratteristiche salienti di una v.a. X (meglio della sua densità discreta). Sono importanti anche perchè talvolta possono essere calcolati senza conoscere in dettaglio la densità descrita p_X, ma sfruttando le proprietà di valore medio e varianza.

Proprietà della varianza

Per ogni variabile aleatoria (reale) X

- $VAR[X + c] = VAR[X]$
- $VAR[cX] = c^2 VAR[X]$

Per ogni costante reale $c \in R$

Osservazione Diverse dalle proprietà del valore media!
Varianza $= (SD)^2 \simeq$ (larghezza della distribuzione)^2

$$\text{Inoltre } X = c \Leftrightarrow VAR[X] = 0$$

Osservazione: Negli esericzi, le proprietà sono super utili nel caso ci viene richiesto di calcolare una varianza Y in funzione di X e già calcolata in precedenza.

3.1.4 Indipendenza di Variabili Aleatorie

Siano X e Y due variabili aleatorie, che dipendono entrambe dallo stesso esperimento aleatorio.

$$\text{Quanto vale } VAR[X + Y]?$$

La risposta dipende da **come solo legate X e Y!**

DEFINIZIONE

Due Variabili Aleatorie discrete X e Y si dicono **indipendenti** se gli eventi $\{X = x\}$ e $\{Y = y\}$ sono indipendenti, ossia:

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

Per ogni scelta di x e y tra i valori assunti da X e da Y.

Intuitivamente conoscere il valore assunto da X non modifica la proabilità dei valori di Y e viceversa.

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, ad esempio quando X e Y "non si influenzano".

Additività della Varianza Introduciamo ora un teorema molto importante: quello di additività della varianza.

Se X e Y sono variabili aleatorie **indipendenti**, allora:

$$\mathrm{VAR}[X + Y] = \mathrm{VAR}[X] + \mathrm{VAR}[Y]$$

Ne segue che la deviazione standard:

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

Osservazione: La relazione di additività della varianza non è affatto scontata e nemmeno intuitiva. Infatti la varianza non è lineare, ma la sua somma si comporta linearmente.

3.1.5 Distribuzioni Notevoli Discrete

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

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

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

$\mu_X(A) := P(X \in A)$ $\quad\quad$ $\mu_X : \underset{\overset{\cap}{A}}{}\mathcal{P}(\mathbb{R}) \to [0,1]$

(È UNA PROBABILITÀ SU $\mathbb{R}$)

Per Variabili Aleatorie 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:

3.1.6 Bernoulli

Si chiama Bernoulli una Variabile Aleatoria X che può **assumere soltanto i valori 0 e 1** cioè

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

Sia $p := P(X = 1)$, dato che la somma di tutti i valori che può assumere è 1 si ottiene:

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

Ovvero:

$$\begin{cases} P(X = 1) = p \\ P(X = 0) = 1 - p \end{cases}$$

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

Valore Medio e Varianza Se $X \sim \mathrm{Be}(p)$ è una variabile di Bernoulli, allora:

- Valore Medio: $E[X] = p$
- Varianza: $\mathrm{VAR}[X] = E[X^2] - E[X]^2 = p - p^2 = p(1 - p)$

A dispetto della loro semplicità, le variabili aleatorie di bernoulli sono importanti, perchè spesso vengono utilizzate per costruire variabili aleatorie più complesse.

3.1.7 Binomiale

Consideriamo un esperimento aleatorio costituito da "prove ripetute e indipendenti", dove ciascua 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.

Esempio:

- Lancio ripetutamente una moneta o un dado
- Guardo se i figli/e di una coppia sono M o F
- estraggo persona da una popolazione molto ampia (successo = elettore del candidato A)

Siano

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

Consideriamo quindi la Variabile Aleatoria:

$$X := \text{numero di "successi" che si verificano nelle } n \text{ prove.}$$

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

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

Calcoliamo la distribuzione di X. Per costruzione:

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

Inoltre la densità discreta è data da:

$$p_X(k) = P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad \text{per } k = 0, 1, \ldots, n$$

Dove:

- $\binom{n}{k}$ sono le scelte di quali prove hanno successo, ovvero la combinazione di n prove in cui k hanno successo.

- p^k è la probabilità di k successi fissati.
- $(1-p)^{n-k}$ è la probabilità di $(n-k)$ insuccessi fissati.

In definitiva, una variabile aleatoria X è **binomiale** di parametri n e p, $X \sim \mathrm{Bin}(n,p)$, se ha questa densità discreta.

Valore Medio $E[X_i] = np$

Varianza $Var[X_i] = np(1-p)$

Esempio: Riprendendo l'esempio dei figli maschi, sia $X :=$ numero di figli maschi in una coppia con due figli:

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

Allora $X \sim \mathrm{BIN}(2, \frac{1}{2})$:

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

k=0 k=1 k=2

Non è soprendente che X ~Bin$(2, \frac{1}{2})$ perchè possiamo vedere X come il numero di "successi" in $n = 2$ prover ripetute e indipendenti con probabilità di successo $p = \frac{1}{2}$. Infatti

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

E le variabili aleatorie X_1 e X_2 sono Be$(\frac{1}{2})$ indpendenti.

3.1.8 Poisson

Una variabile aleatoria X si dice Poisson di parametro $\lambda \in (0, \infty)$, e si scrive $X \sim \mathrm{Pois}(\lambda)$, se :

$$X(\Omega) = \mathbb{N}_0 = 0, 1, 2, \ldots$$

E la sua densità discreta è:

$$p_X(k) = P(X = k) = e^{-\lambda}\cdot\frac{\lambda^k}{k!} \quad \text{per } k = 0, 1, 2, \ldots$$

Osservazione: Si può ottenere una variabile aleatoria di Posson $X \sim \text{Pois}(\lambda)$ come opportuno limite di una variabile aleatoria binomiale $Y \sim \text{Bin}(n, p)$ quando

$$n \to \infty, p \to 0 \text{ con } np = \lambda \text{ cioè } p = \frac{\lambda}{n}$$

Valore Medio e Varianza Se $X \sim \text{Pois}(\lambda)$ è una variabile di Poisson, allora:

- Valore Medio: $E[X] = \lambda$
- Varianza: $\text{VAR}[X] = \lambda$

Esempi Le variabili aleatori di Poisson sono approssimazoni per variabili aleatorie che contano il "numero di successi" quando si considera una grande quantità di prova la cui probabilità di successo è "piccola".
Qui di seguito alcuni esempi:

- Numero di accessi a una pagina web in un'ora
- Numero di nascite in un ospedale in una giornata
- Numero di cliente in un ufficio postale in una mattinata

Geometrica

Una variabile aleatoria X si dice geometrica di parametro $p \in (0, 1]$, e si scrive $X \sim \text{Geo } p$, se:

$$X(\Omega) = \mathbb{N} = 1, 2, 3, \ldots$$

E la sua densità discreta è:

$$p_X(k) = P(X = k) = p(1-p)^{k-1} \text{per} k = 1, 2, 3, \ldots$$

Si può ottenere una variabile aleatoria Geo(p) a partire da una successione infinita di prove ripetute e indipendenti, con probabilità di successo p, e considerando la variabile aleatoria:

$$\text{T} := \text{istante del primo successo}$$

Dove l'istante è il numero della prova.
Ad esempio, indicando con $X_i \sim \text{Be}(p)$ per $i = 1, 2, 3, \ldots$ la v.a. che vale 1 se la i-esima prova ha successo, si ha

$$P(T = 1) = P(X_1 = 1) = p$$

$$P(T = 2) = P(X_1 = 0, X_2 = 1) = P(X_1 = 0)P(X_2 = 1) = (1 - p)p$$

e in generale

$$P(T = k) = P(X_1 = 0, \ldots, X_{k-1} = 0, X_k = 1) = (1 - p)^{k-1}p$$

Valore Medio e Varianza se $X \sim \text{Geo}(p)$

- Valore Medio: $E[X] = \frac{1}{p}$
- Varianza: $\text{VAR}[X] = \frac{1-p}{p^2}$

Note per gli esercizi Quando ho un esercizio che riguarda l'estrazione di un elemento n volte, tramite una serie di prove ripetute e indipendenti, con probabilità di successo p, come lancio di dadi o estrazioni di una pallina colorata **con reimmissione** allora si tratta di una v.a. con distribuzione geometrica.

3.2 Variabili Aleatorie Assolutamente Continue

Fin'ora abbiamo studiato le Variabili Aleatorie Discrete, che assumono un **insieme finito** oppure **infinito numerabile di valori** $X(\Omega) = x_1, x_2, \ldots$, dove la distribuzione X è determinata dalla **Densità Discreta**:

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

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

Consideriamo ora una classe "complementare" di Variabili Aleatorie, dette **Assolutamente Continue** che **assumono un insieme infinito più che numerabile di valori**, come ad esempio un intervallo di $\mathbb{R}$: $[0, 1]$ $[0, +\infty)$ $(-\infty, +\infty)$

DEFINIZIONE

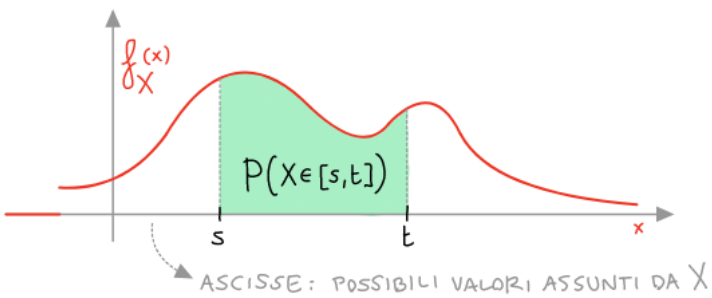
Una Variabile Aleatoria è **Assolutamente Continua** se la sua distribuzione è determinata da una funzione $f_x(x)$, a valori positivi, detta **densità della variabile aleatoria** X nel modo seguente:

$$P(X \in A) = \int_A f_x(x)\, dx$$

Con $A \subseteq \mathbb{R}$.

In particolare, quando consideriamo X compresa in un intervallo $[s, t]$, con $-\infty \leq s \leq t \leq +\infty$, avremo:

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



La Densità di X corrisponde all'area sotto il grafico di f_x tra i punti s e t.

In altre parole Una variabile aleatoria assolutamente continua può assumere qualunque valore in un certo intervallo, ovvero infiniti valori all'interno di ogni intervallo. Alcuni esempi di variabile aleatoria continua possono essere il tempo impiegato a portare a termine un esperimento scientifico o il peso di un individuo.

Ogni v.a. assolutamente continua X ha una curva associata, chiamata funzione di densità di probabilità, che può essere usata per ottenere le probabilità associate a una v.a.

V.A continue vs discrete Tra le Variabili Aleatorie Assolutamente Continue e quelle Discrete ci sono alcune analogie, infatti:

X Assolutamente Continua	X Discreta
$P(X \in [s, t]) = \int_s^t f_x(x)dx$ Integrale	$P(X \in [s, t]) = \sum_{x_i \in [s,t]} p_X(x_i)$ Somma

Ma c'è anche un'importante proprietà che le differenzia: Se X è assolutamente continua:

$$\forall x \in R : P(X = x) = 0$$

Quindi $f_x(x)$ **NON è** $P(X = x)$, tranne dove $f_x(x) = 0$.

Osservazione: Ne segue che nella scelta degli intervalli non cambia niente se includiamo gli estremi o no:

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

Proprietà della Densità: La densità di una Variabile Aleatoria Assolutamente Continua X è una funzione $f_x : R \to R$ (integrabile) tale che:

- $f_x(x) \geq 0 \ \forall x \in R$
- $\int_{-\infty}^{+\infty} f_x(x)\, dx = 1$ Area totale sotto il grafico di f_x

Osservazione: $1 = P(X \in (-\infty, +\infty)) \neq \sum_{x \in (+\infty, -\infty)} P(X = x) = 0$ Dato che $X \in (+\infty, -\infty)$ più che numberabile!

Valori Assunti I valori assunti da una variabile aleatoria assolutamente continua generica sono:

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

Con $P(X \in X(\omega)) = 1$.

3.2.1 Variabile Uniforme Continua

Fissiamo un intervallo limitato $[a,b]$: $(-\infty < a < b < -\infty)$.

DEFINIZIONE

Una variabile aleatoria X si dice uniforme continua in $[a,b]$ e si scrive $X \sim U(a,b)$ se essa è assolutamente continua con densità:

$$f_x(x) = \begin{cases} c \text{ se } x \in [a,b] \\ 0 \text{ se } x \notin [a,b] \end{cases}$$

con $c = \frac{1}{b-a}$, in modo da avere $\int_{-\infty}^{+\infty} f_x(x)dx = 1$.

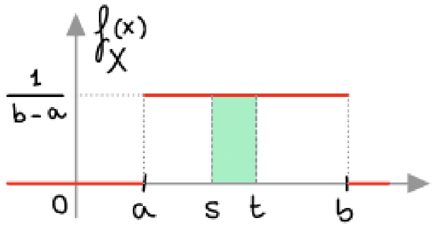
Probabilità di U Dato un intervallo $[s,t] \subseteq [a,b]$, avremo che la probabilità $P(X \in [s,t])$ sarà:

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

Quindi:

$$\rightarrow \frac{1}{b-a} \int_s^t 1\, dx = \frac{t-s}{b-a}$$

Ovvero la lunghezza di $[s,t]$ diviso la lunghezza di $[a,b]$.



Osservazione: Si può mostrare che a partire da una variabile aleatoria $U(0,1)$ è possibile generare una variabile aleatoria con distribuzione arbitraria.
Ad esempio, se $X \sim U(0,1)$ allora $Y := a + (b-a)X \sim U(a,b)$.

Valori Assunti Da una Variabile Aleatoria Uniforme Continua sono:

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

Valore Medio e Varianza

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

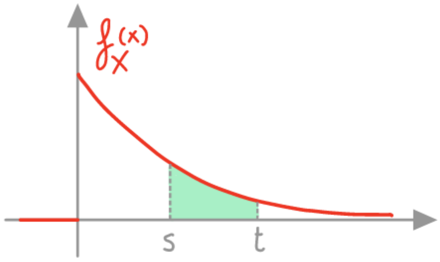
- $E[X] = ... = \frac{a+b}{2}$
- $\text{VAR}[X] = ... = \frac{(b-a)^2}{12}$

3.2.2 Variabile Esponenziale

DEFINIZIONE

Una variabile aleatoria è detta **Esponenziale** di parametro $\lambda \in (0, \infty)$ e si scrive $X \sim Exp(\lambda)$ se la sua densità è del tipo:

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



Per ogni intervallo $[s, t] \subseteq [0, \infty]$, la probabilità che la variabile esponenziale X sia nell'intervallo è:

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

$$\rightarrow \int_s^t \lambda e^{-\lambda x}\, dx = [-e^{-\lambda x}]_s^t = e^{-\lambda s} - e^{-\lambda t}$$

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 variabile aleatoria assolutamente continua esponenziale di parametro λ.

Valori Assunti Da una Variabile Aleatoria Esponenziale sono:

$$X \sim Exp(\lambda) : X(\Omega) = [0, \infty)$$

Osservazione: Si può mostrare che se $Y \sim U(0,1)$ allora $X := \frac{1}{\lambda} \cdot (-\log Y) \sim Exp(\lambda)$.

Valore Medio e Varianza

Sia $X \sim Exp(\lambda)$, allora vale:

- $E[X] = ... = \frac{1}{\lambda}$
- $\mathrm{VAR}[X] = ... = \frac{1}{\lambda^2}$

Alcune Proprietà

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

Inoltre, la densità $f_X(x)$ non è univocamente definita, si può modificare in un numero finito di punti e $\int_s^t f_X(x)dx$ non cambia.

NB $f_x(x)$ NON è $P(X = x)$, la densità di una V.A. Continua non è la densità discreta e soprattutto NON è la probabilità di assumere il valore x.

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

3.2.3 Valore medio e Varianza di Variabili Assolutamente Continue

Le definizioni di $E[X]$ e $\mathrm{VAR}[X]$ per X Assolutamente Continue ricalcano quelle date per Per le Variabili Aleatorie Discrete:

- **Valore Medio**: $E[X] = \int_{-\infty}^{+\infty} x \cdot f_x(x)\, dx$
- **Varianza** : $\mathrm{VAR}[X] = E[X^2] - E[X]^2$
- Distribuzione Standard: $SD[X] := \sqrt{\mathrm{VAR}}[X]$

Si nota che il momento secondo $E[X^2]$ si calcola:

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 \cdot f_x(x)\, dx$$

Proprietà Tutte le proprietà di Valore Medio e Varianza valide per le Variabili Aleatorie Discrete continuano a valere. In particolare:

- $E[X+c] = E[X]+c \qquad E[cX] = cE[X] \qquad E[X+Y] = E[X]+E[Y]$

- $\mathrm{VAR}[X+c] = \mathrm{VAR}[X] \qquad \mathrm{VAR}[cX] = c^2\mathrm{VAR}[X]$

- Se X e Y sono indipendenti: $\mathrm{VAR}[X+Y] = \mathrm{VAR}[X] + \mathrm{VAR}[Y]$

3.3 Funzione di Ripartizione

Le funzioni di ripartizione ci permettono di capire di che tipo di variabile aleatoria stiamo parlando.

$$\text{Funzione di Ripartizione: } F_X(x) = P(X \leq x)$$

A livello probabilistico, $F_X(x)$ è **la probabilità che X sia minore o uguale a x.**
Perchè studiare la funzione di ripartizione?

- **F_x è ben definita per ogni variabile aleatoria** $X : \Omega \to \mathbb{R}$, sia che questa sia discreta, assolutamente continua, o nè l'una nè l'altra.
- **F_x determina la distribuzione della variabile aleatoria**. Ad esempio: $P(X \in (s,t]) = F_X(t) - F_X(s)$.
- F_x è **legata alla densità discreta/densità** di X:
$$F_X(x) = \begin{cases} \sum_{x_i} p_X(x_i) & \text{se X è discreta} \\ \int_{-\infty}^{x} f_x(t)\,dt & \text{se X è assolut. continua} \end{cases}$$
  con $x_i \in (-\infty, x]$

La funzione di ripartizione è utile soprattutto per variabili assolutamente continue, su cui ci concentreremo nel seguito.

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

$$X(\Omega) = 0,1 \qquad p_X 0 = 1 - p, \qquad p_X(1) = p$$

Allora $F_x(x) = P(X \leq x)$ vale:

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

Teorema

di ripartzione di V.A. Discrete

- X è v.a. discreta $\Leftrightarrow$ F_x è costante a tratti.
- Valori assuntix_i $\Leftrightarrow$ Punti di discontinuità di F_x

- Densità discreta $\Leftrightarrow$ Ampiezze dei salti

$$p_X(x_i) = F_X(x_i) - F_X(X_i^-)$$
$$F_X(x_i^-) = \lim_{t \to x_i^- F_X(t)}$$

Teorema

di ripartizione di V.A Assol. Continue X v.a. assolutamente continua $\Leftrightarrow$ F_X è una funzione continua ed è derivabile a tratti.

$$\text{Densità} \qquad f_X(x) = (F_X)'(x) \tag{3.3}$$

3.4 Variabili aleatorie Normali

Abbiamo visto due classi notevoli assolutamente continue, ovvero Uniforme Continua $U(a, b)$ e Esponenziale $Exp(\lambda)$.
L'ultima classe notevole per le Variabili Aleatorie Assolutamente Continue che vedremo è quella delle variabili aleatorie Normali o Gaussiane. Questa è anche la più importante, di fatti merita una sezione a sè.

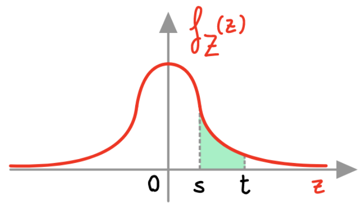
DEFINIZIONE

Una Variabile Aleatoria Z si dice **Normale Standard** e si scrive:

$$Z \sim N(0, 1)$$

Se è assolutamente continua con densità:

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



In questo caso i valori assunti sono: $Z(\Omega) = (-\infty, +\infty)$

Come si può notare la densità ha forma a *campana* simmetrica rispetto all'orgine.

$$\text{STANDARD} = \begin{cases} E[Z] = 0 \\ \text{VAR}[Z] = 1 \end{cases}$$

3.4.1 Probabilità di Una Variabile Normale

Dato un intervallo $[s, t] \subseteq \mathbb{R}$, la probabilità P si calcola, come per ogni variabile assolutamente continua:

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

Purtroppo questo integrale **non si può calcolare esattamente** perchè la densità $f_Z(z)$ non ammette primitiva esplicita.

Funzione di Ripartizione Introduciamo quindi la **funzione di ripartizione** di Z, indicata con Φ.

$$\Phi(z) = F_Z(z) = P(Z \leq z) = \int_{-\infty}^{z} f_z(t)\,dt$$

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

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

$$\Phi(z) = 1 - \Phi(-z)$$

Grazie alla tavola, è come se conoscessimo $\Phi(z) = F_Z(z)$. Possiamo allora calcolare la **probabilità degli intervalli**:

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

3.4.2 Variabili Normali Centrate

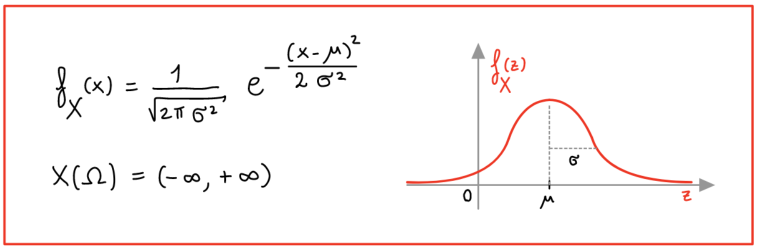
Siano ora $\mu \in \mathbb{R}$, $\sigma \in (0, \infty)$.

DEFINIZIONE

Una Variabile Aleatoria X si dice **Normale con media μ e varianza σ^2** e si scrive

$$X \sim N(\mu, \sigma^2)$$

, se X è assolutamente continua e:



La densità di f_X di X ha un grafico a *campana* centrata in μ e con ampiezza σ. Essa si ottiene dalla densità di f_Z di $Z \sim N(0,1)$ mediante una traslazione e un riscalamento.
Ci si può sempre ricondurre a una Variabile Aleatoria Normale Standard Z:

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

E viceversa: $Z \sim N(0,1) \to X = \sigma Z + \mu \sim N(\mu, \sigma^2)$.

Media e Varianza Si deduce, in particolare che μ e σ^2 sono rispettivamente media e varianza:

$$X \sim N(\mu, \sigma^2) \to E[X] = \mu \quad \mathrm{VAR}[X] = \sigma^2$$

Teoremi Il fatto che ci si può ricondurre a una Variabile Aleatoria normale standard è un caso particolare della seguente proprietà:

Teorema

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

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

E avremo:

- $E[Y] = aE[X] + b$
- $Var[Y] = a^2 Var[X]$

Infine concludiamo con un ultimo teorema imporante:

Teorema

X e Y normali **indipendenti** $\to X + Y$ è normale.

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

Esempio: $X \sim N(0,1)$ e $Y \sim N(0,1)$ Indipendenti. Allora:

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

Capitolo 4

Teoremi di convergenza

4.1 Verso la statistica

Introduciamo ora un modello probabilistico fondamentale per la **statistica**:

Successione $X_1, X_2, X_3, ...$ di Variabili Aleatorie
I.I.D = Indipendenti e Identicamente Distribuite[1].

Concretamente, le variabili X_i possono rappresentare misure o rilevazioni indipendenti, ripetute nel tempo di una sterssa quantità che si vuole studiare. Tipicamente, queste varibili sono:

- Discrete, con la stessa densità discreta fissata: $\rightarrow p_{X_i}(x) = p(x)$
- Assolutamente continue, con la stessa densità fissata: $\rightarrow f_{x_i}(x) = f(x)$

Esempio: Lancio ripetutamente un dado a sei faccie:

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

Esempio: Rilevo la pressione arteriosa in un campione di pazienti:

$$X_i := \text{valore della pressione misurata nel paziente } i\text{-esimo}$$

Problema $p(x), f(x)$ non sono completamente note, come faccio quindi a trovare la distribuzione comune delle variabili? Uso al Statistica Inferenziale.

[1] Con la stessa distribuzione.

4.2 Statistica Inferenziale

Dalle osservazioni posso fare delle deduzioni, ovvero delle Inferenze, sulla distribuzione comune di X_1, X_2 ossia su $p(x)$ e $f(x)$.

Non sempre posso osservare tutte le V.A e per questo ne scelgo casualmente n formando un campione:

$X_1, \ldots, X_n$ è **Campione Aleatorio** di ampiezza $n \to n$ V.A. I.I.D.

Noi lavoreremo con variabili aleatorie che siano funzioni del campione aleatorio, cioè del tipo $g(x_1, \ldots, X_n)$:

DEFINIZIONE

Dato un campione aleatorio $X_1, \ldots, X_n$ una **Statistica Campionaria** è una qualunque funzione del campione:

$$g(X_1, \ldots, X_n) \ (g : \mathbb{R}^n \to \mathbb{R})$$

Se $X_1, ..., X_n$ sono variabili I.I.D. di media μ e varianza σ^2, allora:

$$E[X_i] = \mu \quad var(X_i) = \sigma^2$$

$$\to E(\bar{X}_n) = \mu \quad var(\bar{X}_n) = \frac{\sigma^2}{n}$$

4.3 Legge dei Grandi Numeri

Siano $X_1, X_2, ...$ Variabili Aleatorie reali indipendenti e identicamente distribuite.
Sia $\mu = E[X_i]$, $\sigma^2 = Var[X_i]$

Osservazione: Siccome le variabili sono identicamente distribuite, μ e σ^2 sono uguali per ogni variabile.

Introduciamo la **Media Campionaria**:

$$\bar{X}_N := \frac{1}{N} \sum_{i=1}^{N} X_i = \frac{X_1 + X_2 + ... + X_N}{N}$$

Questa è una *nuova variabile aleatoria* distinta dalle X_i.

Possiamo ora introdurre la legge dei grandi numeri:

Teorema Legge dei Grandi Numeri

Per ogni $\epsilon > 0$ avremo che:

$$\lim_{N \to \infty} P(|\bar{X}_N - \mu| \geq \epsilon) = 0$$

$$\iff \lim_{N \to \infty} P(\bar{X}_N \in (\mu - \epsilon, \mu + \epsilon)) = 1$$

A partire dalla sequenza di dati osservati $X_1, X_2, ..., X_N$, posso essere fiducioso che la media dei dati $\bar{X}$ sia vicina a $\mu = E[X]$, e che la distribuzione di $\bar{X}_n$ è concentrata attorno alla media μ.

Osservazione: La Legge dei Grandi Numeri implica che il legame tra media campionaria e media = valore atteso di μ.

La legge dei grandi numeri si dimostra con la **disuguaglianza di Chebishev**.

4.4 Distribuzione di X

$$\mathbb{P}(a < \bar{X_n x} \leq b) = \mathbb{P}(\frac{a - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{X_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{b - \mu}{\frac{\sigma}{\sqrt{n}}})$$

$$\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

$$\frac{\sigma}{\sqrt{n}} = \sqrt{\text{Var}}$$

4.5 Teorema del limite centrale

Siano $X_1, ..., X_n$ v.a i.i.d. (Variabili aleatorie indipendenti identicamente distribuite) $E(i) = \mu$ e $\text{Var}X_i = \sigma^2$ (con media e varianza finite).

$$\mathbb{P}(\frac{X_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq t) \to \Phi(t) \quad n \to +\infty$$

Dove Φ è la funzione di ripartizione di una V.A $Z \sim N(0, 1)$

$$\Phi(t) = \mathbb{P}(Z \leq t)$$

Condizione per usare l'approssimazione normale Abbiamo definito che solo con $n \geq 30$ si può usare l'approx per semplicità, in realtà non è semplice dare un criterio univoco dato che dipende dall'asimmetria del campione di partenza.

Nota per i campioni discreti Quando ho un campione discreto è sempre necessario applicare la correzione di continuità

Note per esercizi

- Riconosci la distribuzione notevole
- Ricava media e Var
- Scrivi la formula che ti serve (es P che una va sia $\leq$ di n)
- Effettua la correzione di continuità se si tratta di una VA discreta
- Verifica se $n \geq 30$, questo per verificare se si può applicare il TLC (teorema del limite centrale)
- Sottrai media e dividi per $\sqrt{\text{Var}}$
- Se hai $>$ come segno, anzichè $\mathbb{P}$ dovrai trovare $1 - \mathbb{P}$ e invertire il segno in $<$
- Se trovi un valore z negativo all'interno di $\Phi(z)$ dovrai calcolare $1 - \Phi(-z)$, quindi rendere positivo z, calcolare il suo Φ e sottrarlo a 1

4.6 Le Domande di Teoria

Di seguito riporto alcune domande di teoria sui teoremi di convergenza:

Domanda 1

Sia $X_1, X_2, \ldots$ una successione di variabili aleatoriei.i.d. con distribuzione di Poisson $P(\lambda)$. Quale delle seguenti variabili casuali ha una distribuzione che, per n grande, approssima quella di una Normale Standard?

Scegli un'alternativa:

- a. $X_1 + \ldots . X_n$
- b. $\frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \sqrt{n}$
- c. $\frac{\bar{X}_n}{\sqrt{\lambda}} \sqrt{n}$
- d. $\frac{\bar{X}_n - \lambda}{\sqrt{n\lambda}}$

Capitolo 5

Stima di parametri

In questo capitolo vedremo come usare i dati campionari per stimare una media, una varianza o una proporzione della popolazione. Verranno discusse le **stime puntuali**, ovvero le stime a valore singolo del parametro, di cui considereremo poi l'errore standard. Inoltre considereremo gli intervalli di confidenza, che contengono il parametro con un certo livello di confidenza.

5.1 Statistica inferenziale

La **statistica inferenziale** ha lo scopo di definire in modo non ambiguo e quantitativo la plausibilità di un'inferenza.

Plausibilità La plausibilità di un'inferenza dipende dal modo con cui è stato selezionato un campione di n individui della popolazione. La corretta metodologia di campionamento è la **scelta casuale**.
Per la casualità del campionamento utilizzo il calcolo delle probabilità.
Quando parleremo di popolazione tratteremo sempre N molto grandi rispetto all'ampiezza del campione. Potremmo anche avere a che fare con popolazioni infinite, as esempio quando si parla di processi produttivi.

La **popolazione** è una V.A. i.i.d. con la stessa legge F non sempre completamente nota.

5.1.1 Modello Statistico Parametrico

Il modello statistico parametrico è una famiglia di leggi note a meno di uno o più parametri: $\underline{\theta} \in I \subseteq R^d, d \geq 1$.

- Caso discreto: $p(x;\Theta)$ Densità discreta $p(x;\Theta) : \mathbb{P}_\Theta(X = x)$

- Caso continuo: densità $f(x;\Theta)$

- CAMPIONE CASUALE DI AMPIEZZA n

  $X_1, \dots, X_n$ v.a. i.i.d con densità data

  da $\begin{cases} p(x;\underline{\theta}) & \text{CASO DISCRETO} \\ f(x;\underline{\theta}) & \text{CASO CONTINUO} \end{cases}$

> ***N.B.*** Attenzione a non confondere le V.A. con le osservazioni:
> $X_1, ..., X_n$ sono le Variabili Aleatorie, mentre $x_1, ..., x_n$ sono le Osservazioni

Esempi di modelli statistici Alcuni esempi di modelli statistici sono:

- Modello di Bernoulli
- Modello esponenziale
- Modello normale

Le Statistiche Diamo la definizione di statistica:

DEFINIZIONE

Dato $(X_1, ..., X_n)$ campione casuale, si definisce **Statistica** una funzione del campione, ossia una v.a. T della forma

$$T = f(X_1, ..., X_n)$$

Alcuni esempi di statistica sono:

- Media campionaria (v.a.)$\bar{X}_n = \frac{X_1 + ... + X_n}{n}$
- Varianza campionaria (v.a.): $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

5.2 Statistica Parametrica e Stimatori

Il primo obiettivo della statistica inferenziale è fornire una *stima dei parametri incogniti.*

Stimatori Uno **Stimatore** è una statistica che serve a stimare i parametri incogniti:

DEFINIZIONE

Uno stimatore T si dice **Non Distorto** o corretto se:

$$E_\theta(T) = E_\theta(g(X_1, ..., X_n)) = \theta$$

In cui E_θ è il valore medio rispetto alla probabilità P_θ.

In parole povere uno stimatore il cui valore atteso è uguale al parametro che si vuole stimare si dice **corretto** per quel parametro.

Ad esempio, qualunque sia il campione casuale $X_1, ..., X_n$, con $\mu = E(X_1) \geq +\infty$:

$$\bar{X}_n \text{ è uno stimatore non distorto di } \mu\text{, quindi } E(\bar{X}_n = \mu)$$

Quindi $\bar{X}_n$ è lo stimatore non distorto di:

- p in un modello $Be(p)$
- λ in un modello $P(\lambda)$ (Poisson)
- $\frac{1}{\lambda}$ in un modello $exp(\lambda)$, quindi $X \sim exp(\lambda) \implies E(X) = \frac{1}{\lambda}$
- μ in un modello $N(\mu, \sigma^2)$

Osservazione: La proprietà di essere non distorto NON è stabile per trasformazioni (non lineari), ad esempio nel modlelo esponenziale X_n è stimatore non distorto di $\frac{1}{\lambda}$, ma si ha che $\frac{1}{(X)_n}$ NON è uno stimatore non distorto di λ.

DEFINIZIONE

Uno stimatore non distorto di Θ si dice **Consistente** se, quando $n \to +\infty$

$$\text{VAR}_\Theta(T) \to 0$$

Quando abbiamo un campione casuale estratto da una popolazione con media μ e varianza σ^2 finite si ha sempre che $\bar{X}_n$ **è uno stimatore consistente di** μ:

$$\mathrm{VAR}_\sigma(\bar{X}_n) = \frac{\sigma^2}{n} \to 0 \quad n \to +\infty$$

Stimatore di Θ: $T = g(X_1, ..., X_n) \to$ V.A. Stima di Θ : $f(x_1, ..., x_n) \to$ Numero.
Stima di μ per un campione casuale $X_1, ..., X_n$ di cui osserviamo $x_1, ..., x_n$.

$$\hat{\mu} = \bar{x}_n = \frac{x_1 + ... + x_n}{n}$$

Osservazione: Supponiamo di considerare un campione casuale con media μ incognita e varianza σ^2 nota:

$$\mathrm{VAR}\bar{X}_n = \frac{\sigma^2}{n}$$

$$\mathrm{SD}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$$

Se si pensa $\bar{X}_n$ come **stimatore** di μ, $\mathrm{SD}(\bar{X}_n)$ prende il nome Di **errore standard**, esso rappresenta l'errore commesso stimando μ con $\bar{X}_n$.

Ora consideriamo un modello statistico con **varianza incognita**.
Dalla statistica descrittiva: siano n osservazioni $(x_1, ..., x_n)$, la varianza campionaria è:

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

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

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

Si può verificare che $E(S_n^2) = \sigma^2$, ovvero che S_n^2 è uno stimatore corretto.

Se nel modello statistico con varianza σ^2 incognita **la media è nota**, si ha che lo stimatore:

$$\bar{S}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$$

è uno stimatore corretto di σ^2.

Esempio: Nel liceo XXX su un campione di 100 studenti, 40 sono ragazze. Fornire una stima della proporzione di ragazze frequentanti il liceo XXX. Si è utilizzato uno stimatore non distorto?

Siano $X_1, ..., X_n$ V.A. i.i.d $\sim Be(p)$, con $p =$ proporzione di ragazze. Allora;

$$\hat{p} = \bar{x}_n = \frac{40}{100} = \frac{2}{5}$$

5.3 Distribuzione delle Statistiche Campionarie

Usiamo un campione *Normale*: $X_1, ..., X_n$ campione casuale $\mathcal{N}(\mu, \sigma^2)$, allora:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to \bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

Che, una volta standardizzato diventa:

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$$

Possiamo quindi ottenere la legge di X_n, e invece per S_n^2 e $\bar{S}_n^2$?

Per caratterizzare la legge di S_n^2 e di $\bar{S}_n^2$ dobbiamo introdurre una nuova distribuzione continua $\chi^2(n)$ (*chi*):

DEFINIZIONE

Si dice legge **chi quadrato con** n **gradi di libertà** la legge di una v.a.

$$Y = \sum_{i=1}^{n} Z_i^2$$

con $Z_1, ..., Z_n$ V.A. i.i.d $\mathcal{N}(0, 1)$

Dove Y è una v.a. ≥ 0 con densità $f_Y(t) = c_n t^{\frac{n}{2}-1} e^{-\frac{t}{2}}$ per $t > 0$
Per essa si ha $\mathbb{E}(Y) = n$ e $\mathrm{Var}(Y) = 2n$
Per n grande vale l'approssimazione della legge $\chi^2(n)$ con una $\mathcal{N}(n, 2n)$

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

1. $\sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2 \sim \chi^2(n)$
2. $\sum_{i=1}^{n}(\frac{X_i-\bar{X}_n}{\sigma}) \sim \chi^2(n-1)$
3. se $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$
   $(n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2(n-1)$
4. S_n^2 e $\bar{X}_n$ sono indipendenti

Riassumendo, χ^2 ci serve per la distribuzione della legge delle varianze campionarie

5.4 Distribuzione t di student

Wikipedia definisce la Distribuzione t di Student come: *"Una distribuzione di probabilità continua che governa il rapporto tra due variabili aleatorie, la prima con distribuzione normale e la seconda, al quadrato, segue una distribuzione chi quadrato. Questa distribuzione interviene nella stima della media di una popolazione che segue la distribuzione normale, e viene utilizzata negli omonimi test t di Student per la significatività e per ogni intervallo di confidenza della differenza tra due medie."*

Matematicamente, si definisce:

DEFINIZIONE

Si dice legge di **t di student** con n gradi di libertà la legge di una v.a.:

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

In cui: $z \sim \mathcal{N}(0,1)$ e $Y \sim \chi^2(n)$ Sono due V.A. indipendenti.

La Media di T è: $\mathbb{E}(T) = 0$ e la sua varianza è: $var[T] = 1$ La sua densità è:

$$f_T(t) = c_n(1 + \frac{t^2}{n})^{\frac{-(n+1)}{2}}$$

in cui c_n è una costante che fa risultare l'integrale della $f = 1$.

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

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

5.5 Percentili

Cosa dovremo saper calcolare di queste variabili aleatorie? I loro percentili.
sia X una v.a., e $P(X \leq q_\alpha) = \alpha$.
Fissato α, trovare q_α.
q_α è α - esimo quantile o $100 - \alpha$ percentile di X.

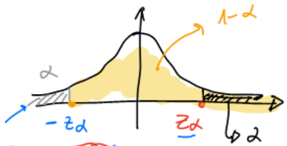
- $Z \sim \mathcal{N}(0,1)$ $\quad z_\alpha$ t.c. $\mathbb{P}(Z > z_\alpha) = \alpha$
- $T \sim t(n)$ $\quad t_{\alpha,n}$ t.c. $\mathbb{P}(T > t_{\alpha,n}) = \alpha$
- $Y \sim \chi^2 n$ $\quad \chi^2_{\alpha,n}$ t.c. $\mathbb{P}(Y > \chi^2_{\alpha,n}) = \alpha$
- z_α, $t_{n,\alpha}$, $\chi^2_{m,\alpha}$ $\quad 100(1-\alpha)$percentili (di Z, T, Y)

$$Z \sim \mathcal{N}(0,1) \qquad \alpha \quad \text{definito da}$$

$$P(Z > z_\alpha) = \alpha$$
$$P(Z \leq Z_\alpha) = 1 - \alpha$$
$$z_\alpha = 100(1-\alpha) \qquad \text{percentili di Z}$$
$$z_\alpha = \Phi^{-1} 1 - \alpha$$
$$P(Y > X^2_{\alpha,n}) = \alpha$$

La simmetria di T student è la stessa della normale.
Mentre per la χ^2 non ci sono simmetrie.

· $Z \sim \mathcal{N}(0,1)$

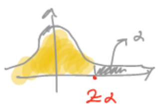


$P(Z > -z_\alpha) = 1-\alpha$

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

· $Y \sim \chi^2(n)$

α definito da



$P(Z > z_\alpha) = \alpha$

$P(Z \leq z_\alpha) = 1-\alpha$

z_α : $100(1-\alpha)$ percentile di Z

$z_\alpha : \Phi^{-1}(1-\alpha)$

$z_{0.025} = \Phi^{-1}(0.975) = 1{,}96$

$P(Y > \chi^2_{\alpha,n}) = \alpha$

5.6 Stima per intervalli - Intervalli di confidenza

Stima per intervalli della media - campione normale $\mathcal{N}(\mu, \sigma^2)$, con σ^2 nota.
La qualità dlela stima dipende Da

- Livelli di confidenza: maggiore è il livello, più è affidabile è la stima
- Ampiezza dell'intervallo = $2E$: più è piccolo, più è precisa la stima

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

E crece se α diminuisce (ossia se la confidenza aumenta) fissati n e σ E diminuisce al crescere di n (come $\frac{1}{\sqrt{n}}$)

Estremi Inferiori e Superiori di Confidenza Per la media di una popolazione normale con varianza nota.

Per la media di una popolazione normale con varianza nota.
Determinare se la media di una popolazione è maggiore o minore di un certo valore.
Useremo ancora $\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1)$
Confidenza: $100(1 - \alpha)\%$

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

Un estremo inferiore di confidenza al $100(1 - \alpha)\%$ per la media di una popolazione normale con varianza nota è dato da

$$\bar{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}}$$

La sua realizzazione (dai dati campionari) è:

$$\bar{x}_n - z_\alpha \frac{\sigma}{\sqrt{n}}$$

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

Completare ———————————— Possiamo dire:

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

Intervalli di confidenza per la varianza di una popolazione normale

Capitolo 6

Verifica di Ipotesi

Introduzione La verifica di ipotesi consiste nel formulare un'ipotesi sui parametri dell'universo e verificarne la sua validità. L'ipotesi che vogliamo sottoporre a verifica è detta ipotesi di base o **ipotesi nulla** e viene indicata con H_0, e se non risulterà valida allora lo sarà l'ipotesi alternativa H_1. Un'ipotesi alternativa può essere bilaterale, ovvero in cui $\mu \neq 0$, o unilaterale ($>$ o $<$).

Ipotesi statistica Un'ipotesi Statistica è un'affermazione sulla distribuzione della popolazione in esame.
Può essere espressa in termini di un parametro (**Test Parametrici**), oppure può riguardare la natura della distribuzione della popolazione o altre caratteristiche (**Test Non Parametrici**): es verificare se una popolazione ha distr. normale, verifica l'indipendenza.
Verificare un'ipotesi statistica significa verificare se è compatibile con i dati del campione.
Un'ipotesi statistica denotata con H_0 è detta **Ipotesi Nulla**.
La negazione di H_0 è denotata con H_1 e viene definita **Ipotesi Alternativa**.
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. Per questo si utilizzerà un'opportuna statistica detta **Statistica del Test**. A seconda del suo valore assunto sui dati campionari si rifiuterà o meno.
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 $(x_1, ..., x_n) \in C$ si rifiuta H_0, quindi si accetta H_1, tale regione sarà calcolata utilizzando la statistica del test.

6.1 Possibili errori durante il calcolo della regione critica

Durante il calcolo della regione critica possiamo incorrere in due tipi di errore

- Errore di prima specie (o tipo): si rifiuta H_0 e H_0 è vera
- Errore di seconda specie (o tipo): 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 diminuisce la prob. di commettere un errore di prima specie, ma non può aumentare la prob. di commettere un errore di seconda specie e viceversa.
Di solito si controlla la prob. di errore di prima specie.

$\alpha =$ **livello di significatività del test** Fissare α piccolo ($\alpha = 0.10, 0.05, 0.01$) e chiedere che la prob. di rifiutare H_0 quando è vera sia $\leq \alpha$. Ossia un test per la verifica di H_0 con regione critica C ha livello significatività α se:

$$\mathbb{P}_H((x_1, ..., x_n) \in C) \leq \alpha \tag{6.1}$$

Controllare errore di I specie $\rightarrow$ assimetria tra H_0 e H_1.
Se $(x_1, ..., x_n) \in C$ ossia si rifiuta H_0 i dati sperimentali sono in contraddizione significativa con H_0.
Se $(x_1, ..., x_n) \notin C$ i dati sperimentali non sono in contraddizione signficativa 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.
paragraph*Implicazione Se si vuole dimostrare con dati sperimentali una certa ipotesi sulla distribuzione della popolazione, ossia di una variabile, si adotterà questa ipotesi come **Ipotesi Alternativa**.
paragraph*Osservazione L'appartenenza di $(x_1, ..., x_n)$ alla regione critica dipende dalla scelta del livello di significatività di α.
Esiste un $\bar{\alpha}$ 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** (σ p dei dati, *sigma* valore p del test.
Più piccolo è il p-value, più i dati sono in contraddizione con H_0.

Definizione p-value Minimo livello di significatività t.c. i dati consentono di rifiutare H_0.
paragraph*Legame tra Test di Ipotesi e I.C. Nel test ipotesi bilatero: rifiuto $H_0\,(\mu = \mu_0)$ a livello α se $\bar{x_n}$ non appartiene all'intervallo di confidenza di livello $100(1-\alpha)\%$ per μ. C'è anche un legame tra estermi superiori e inferiori di confidenza e regione critiche dei test unilateri.

Efficacia di un test 1 - prob di errore di II tipo.

Test Non Parametrici

Non si fanno inferenze nei parametri ma nella distribuzione di una o più popolazioni.

6.1.1 Test Chi quadro di buon adattamento

Test adattamento: verificare se una certa popolazione abbia o meno una certa distribuzione, ipotizzato sulla base di dati sperimentali (incluso verificare che una popolazione abbia distribuzione normale).

6.2 Esame

All'esame potrebbero esserci alcuni esercizi sui Test di Ipotesi. Riporto degli esempi

6.2.1 Test sulla media con varianza Nota.

In alcuni esercizi vengono riportati i seguenti dati:

- Varianza σ
- Campione n
- Media campionaria $\bar{x}$

E si richiede ti testare l'ipotesi, a livello di significatività α (in percentuale), che la media sia in realtà uguale a un μ_0.

Procedimento

- Imposto l'ipotesi nulla H_0, cioè quella che voglio verificare come $H_0 \to \mu = \mu_0$.
- Imposto l'ipotesi alternativa H_1 come le negazione di H_0, quindi $\mu \neq \mu_0$.
- Prendo la formula della regione critica dal formulario (Test z sulla media di una popolazione normale con varianza nota pari a σ^2)
- Ci troveremo quindi con una formula del tipo $T > z_{\alpha/2}$
- Se questa disequazione è vera, allora RIFIUTO H_0 al $\alpha\%$. Altrimenti, la accettiamo.