PES - Probabilità e Statistica

Elia Ronchetti
@ulerich

2021/2022

Indice

Capitolo 1

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

Capitolo 2

Esame

L'esame sarà strutturato nella seguente maniera

Parte 1 - Teoria 8 Domande a risposta multipla - Punteggio 10/30

Parte 2 - Pratica 4 Esercizi a risposta aperta - Punteggio 20/30

Progetto (facoltativo) Progetto R, da consegnare prima dell'esame, può fornire un massimo di 2/30

Capitolo 3

Analisi Descrittiva

3.1 Descrivere i dati

Per descrivere una raccolta dati in maniera chiara e immediata è utile utilizzare una **tabella delle frequenze** all'interno della quale sono contenuti:

- Valori
- Frequenze Assolute - Numero di volte in cui compare "i" nell'insieme di dati
- Frequenze Relative - Frazione di volte in cui compare i nell'insieme di dati
- Percentuali - (Frequenza relativa x 100)

Il dato che compare con frequenza più alta è detto **moda**.
I dati possono essere

- Qualitativi
- Quantitativi

Noi useremo i dati **quantitativi**

3.1.1 Rappresentazione dei dati

Per rappresentare le frequenze (assolute o relative) risulta efficace e immediato l'utilizzo di un grafico a barre detto istogramma, esso rappresenta in graficamente la tabella, chiaramamente da esso è possibile risalire alla tabella stessa. Capita di avere degli insiemi di dati che assumono un valore elevato di valori distini, per questo conviene suddividerli in classi e determinare la frequenza di ciascuna classe. In questo modo c'è una perdita d'informazioni (sui valori specifici), ma così facendo possiamo calcolare le frequenze delle classi e avere un'idea migliore della distribuzione dei dati.

3.1.2 Dati Bivariati

Quando per ciascun individuo vengono misurate due variabili ci troviamo un insieme di N dati a coppie detti **dati bivariati**. Anche in questo caso è possibile calcolare le frequenze, in questo caso detto **frequenze congiunte**.

è possibile, inoltre, misurare la correlazione tra le due variabili attraverso per esempio un diagramma di dispersione (detto anche scatterplot).

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.

3.2 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.

3.2.1 Indici di posizione

Per definire il centro dell'insieme dei dati definiamo la

Media Campionaria $\frac{x_1+x_2+\cdots+x_n}{N}$

Per misurare il valore in posizione centrale (considerando l'insieme di dati ordinato), utilizziamo la

Mediana

- Se N dispari $\rightarrow X_{\frac{N+1}{2}}$
- Se N pari $\rightarrow m = \frac{X_{\frac{N}{2}}+X_{\frac{N}{2}+1}}{2}$

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).

3.3 Coefficiente di correlazione lineare

Posso misurare il grado di correlazione tra una coppia di dati attraverso il coefficiente di correlazione lineare.

$$r = \frac{\sum_{k=1}^{N}(x_i - x)(y_i - y)}{(N-1)S_x S_y} \tag{3.1}$$

Si può mostrare che:

$$-1 <= r <= 1 \tag{3.2}$$

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

3.3.1 Correlazioni significative

$|r| > 0.7$ Correlazione significativa
$|r| < 0.3$ Correlazione debole

3.4 Percentili e quantili

Per analizzare la distribuzione dei dati è utile fissare un numero k che rappresenta la posizione all'interno dato all'interno dell'insieme questo valore percentuale è detto **k-esimo Percentile Campionario**, valore t per cui

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

I casi più importanti sono per k = 25, 50, 75
Risulta pratico scrivere $k = 100p$ dove $p = \frac{k}{100} \in [0, 1]$, dove i casi importanti sono per:

- $p = \frac{1}{4}$: $k = 100p$ = 25-esimo percentile = primo quartile q_1
- $p = \frac{1}{2}$: $k = 100p$ = 50-esimo percentile = secondo quartile q_2 = mediana m
- $p = \frac{3}{4}$: $k = 100p$ = 75-esimo percentile = terzo quartile q_3

Per calcolare il k-esimo percentile t è necessario:

1. Ordinare l'insieme di dati $x_1 <= x_2 <= \cdots <= x_n$

2. Se N_p non è intera $t = x_i$ è il dato la cui posizione i è l'intero successivo a N_p

3. Se N_p è intera $t = \frac{x_(Np) x_(Np+1)}{2}$ è la media aritmetica fra il dato in posizione N e il successivo

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

É possibile utilizzare i **Boxplot** per la rappresentazione dei quantili

Capitolo 4

Probabilità

Il calcolo delle probabilità è la teoria matematica che permette di descrivere e studiare **esperimenti aleatori**

Esperimento aleatorio $\rightarrow$ Fenomeno il cui esito non è prevedibile con certezza a priori

4.1 Introduzione

La descrizione matematica si articola in tre passi

1. Spazio campionario (o spazio degli esiti)$\rightarrow$ 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 $\rightarrow$ sono i sottinsiemi dello spazio campionario $A \subseteq \Omega$
   es. Tiro un dado a sei facce: esce un numero pari $A = 2, 4, 6$

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

4.2 Proprietà di base

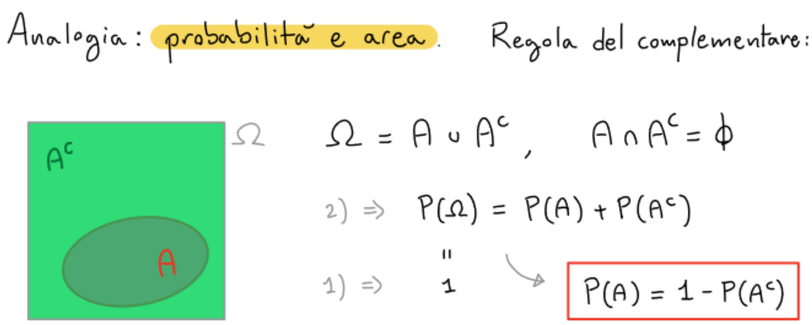
1. $P(\Omega) = 1$

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

La coppia (Ω, P) è detta Spazio di Probabilità .
Fissiamo uno spazio di probabilità (Ω, P). Da queste proprietà si deducono molte altre proprietà

- $P(\emptyset) = 0$
- **Regola del complementare** $\rightarrow P(A^c) = 1 - P(A)$
  Vale per ogni A
- **Regola della addizione di probabilità** $\rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  Vale per ogni A, B (anche $A \cap B \neq \emptyset$)
- Monotonia: se $A \subseteq B$ allora $P(A) <= P(B)$

Analogia c'è un analogia tra probabilità e area



4.3 Calcolo combinatorio

Consideriamo uno spazio di probabilità uniforme (Ω, P)

- $P(A) = \frac{|A|}{|\Omega|} = \frac{Casifavorevoli}{CasiPossibili}$ per ogni $A \subseteq \Omega$
- $P(w) = \frac{1}{\Omega} = \frac{1}{n}$ per ogni $w \in \Omega$

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)

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

$$P(A) = \frac{|A|}{|\Omega|} \tag{4.1}$$

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

Principio Fondamentale Consideriamo un esperimento costituito da due parti:

1. n esiti possibili
2. m esiti possibili

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

Esempio Il lancio dei dadi. Se lancio 2 dadi a sei facce ho Ω esiti possibili

$$|\Omega| = 6 * 6 = 36 \tag{4.2}$$

4.3.1 Disposizioni con ripetizione

Sequenze ordinate di k elementi (anche ripetuti) scelti tra n possibili. Numero totale è:

$$n * n \dots n = n^k \tag{4.3}$$

Esempio Estrazione casuale di 3 persone, calcolare la probabilità che siano tutte nate in primavera. In questo caso lo spazio campionario sono i compleanni delle tre persone quindi

$$\Omega = (x_1, x_2, x_3) : x_1, x_2, x_3 \text{ in Calendario} \tag{4.4}$$

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

$$|\Omega| = 365 * 365 * 365 = 365^3 \tag{4.5}$$

Probabilità uniforme $P(A) = \frac{|A|}{|\Omega|}$
In questo caso tutti i nati in primavera vengono considerati nati A = tutti nati in primavera = [20 marzo, 21 giugno) e in totale sono 92 giorni.
Si tratta anche qua di una disposizione ripetuta di 3 elementi.

$$|A| = 92 * 92 * 92 = 92^3 \tag{4.6}$$

Per calcolare la probabilità è sufficiente dividere A per Ω

$$P(A) = \frac{|A|}{|\Omega|} = \frac{92^3}{365^3} = 0,016 = 1,6\% \tag{4.7}$$

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

4.3.2 Disposizioni semplici

Sequenze ordinate di k elementi distinti scelti tra n possibili (con $k <= n$) **senza ripetizione**

$$n * (n-1) * (n-2) \ldots (n-k+1) = \frac{n!}{(n-k)!} \tag{4.8}$$

É 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.
Se $k = n$ si parla di **Permutazioni** di n oggetti. In numero sono:

$$n! = n * (n-1) * (n-2) \ldots 2 * 1 \tag{4.9}$$

Esempio . Quanti sono i possibili ordini di arrivo di 3 squadre? Si tratta di una permutazione di 3 elementi.

Esempio Paradosso dei compleanni

4.3.3 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.

$$\binom{n}{x} = \frac{n!}{k!(n-k)!} \tag{4.10}$$

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

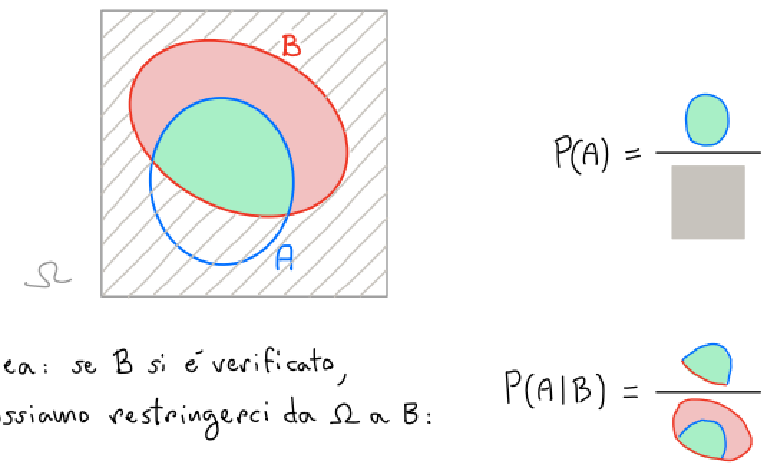
$$\binom{52}{5} = \frac{52!}{5!47!} \tag{4.11}$$

4.4 Probabilità Condizionata

Consideriamo un esperimento aleatorio, che descriviamo con uno spazio di probabilità (Ω, P). Consideriamo un evento $A \subseteq \Omega$, che ha un probabilità $P(A)$. Supponiamo di ricevere l'informazione che un altro evento B si è verificato. Come è ragionevole aggiornare la probabilità di A per tenere conto di questa informazione aggiuntiva?
La soluzione è data dalla probabilità Condizionata.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{4.12}$$

La precedente formula denota la probabilità condizionata di A dato B (o sapendo B). Sto quindi calcolando la probabilità di A.



Quando si verifica un evento lo spazio di probabilità si riduce (vedere esempio sui dadi)

4.4.1 Regola del prodotto

$$P(A \cap B) = P(A) * P(B|A) \tag{4.13}$$

4.4.2 Formula di Disintegrazione

$$P(A) = P(A \cap B) + P(A \cap B^c) \tag{4.14}$$

4.4.3 Formula delle probabilità totali

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) \tag{4.15}$$

Inoltre $P(*|B)$ + una probabilità, in particolare:

$$P(A^c|B) = 1 - P(A|B) \tag{4.16}$$

4.4.4 Formula di Bayes

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{4.17}$$

Esempio . Per vedere la parte pratica andare a vedere l'esempio sui tamponi per rilveare la presenza di un virus. Super interessante e utile.
Il file è **Appunti Lezione 3 - In fondo al PDF**

4.5 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.

$$P(A|B) = P(A) \tag{4.18}$$

che equivale

$$P(A \cap B) = P(A)P(B) \tag{4.19}$$

In questo caso gli eventi A e B si dicono **Indipendenti**

Esempi Lancio di due dadi, i risultati sono eventi indipendenti
Urna contenente 5 palline rosse e 3 palline verdi. Pesco in successione due palline, senza reimmissione. La probabilità che la prima pallina sia rossa e che la seconda sia rossa sono **dipendenti**!

4.5.1 Eventi indipendenti != 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

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

Capitolo 5

Variabili aleatorie

Variabile aleatoria (detta anche casuale o stocastica), 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"

Consideriamo un esperimento aleatorio, descritto da uno spazio di probabilità (Ω, P). Spesso non siamo interessati a tutti i dettagli del'esito dell'esperimento, ma solo a una quantità (tipicamente numerica) determinata dall'esito dell'esperimento. Una tale quantità è detta **Variabile aleatoria**. Possiamo considerare la variabile aleatoria come:

- Intero: Quantità che dipende dal "caso"
- Funzione matematica: Funzione definita sullo spazio campionario: $X : \Omega \to R$

Ricodiamo che un evento è:

- Affermazione sull'esito dell'esperimento aleatorio
- Sottoinsieme dello spazio campionario: $A \subset \Omega$

Se X è una variabile aleatoria, e se x è un suo possibile valore, allora:

- o X assume il valore x
- oppure $w \in \Omega : X(w) = x$

Ogni variabile aleatoria X determina molti eventi!

Attenzione : non confondere variabili aleatorie ed eventi
Le variabili aleatorie rappresentano esiti esprimibili numericamente di esperimenti ancora da effettuare! Dove per esperimento si intende qualsiasi fenomeno o situazione con sviluppi imprevedibili a priori. Essendo imprevedibile a priori il valore assunto da una variabile aleatoria, tutto ciò che si può fare è esprimere delle valutazioni di tipo probabilistico sui valori che essa assumerà. Per questa ragione ad ogni variabile aleatoria X è associata una funzione che esprime in modo chiaro tali valutazioni.
Se la variabile è discreta si parlerà di **Densità discreta**.
Se la variabile è continua di parlerà di **Funzione di ripartizione**.

5.1 Notazione

Con X_i indichiamo la Variabile Aleatoria $X_i : \Omega \to \mathbb{R}$
con x_i indichiamo l'osservazione relativa alla V.A X

5.2 Variabili aleatorie discrete

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

$$X(\Omega) = x_1, x_2, \ldots, x_n \subseteq R \tag{5.1}$$

Oppure un insieme infinito numerabile

$$X(\Omega) = x_1, X_2, \ldots = x_i i \in N \subseteq R \tag{5.2}$$

Ad ogni variabile aleatoria discreta X possiamo associare

Densità discreta $P_x(x_i) = P(X = x_i)$

Definita anche come distribuzione di probabilità, è una funzione che assegna ad ogni valore possibile di X la probabilità dell'evento elementare $(X = x)$

5.2.1 Proprietà

Proprietà. • 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

• $p_X(x_i) \geq 0 \quad \forall i$ • $\sum_{i \geq 1} p_X(x_i) = 1$ $\left\{ \begin{array}{l} \text{PERCHÉ GLI EVENTI } \{X = x_i\} \\ \text{PER } i \geq 1 \text{ SONO DISGIUNTI} \\ (\text{PARTIZIONE DI } \Omega) \end{array} \right.$

Perché la densità discreta p_X? $\rightsquigarrow$ $P(X \in B) = \sum_{x_i \in B} \underbrace{p_X(x_i)}_{P(X = x_i)}$

$\forall B \subseteq \mathbb{R}$

Concettualmente una v.a. (variabile aleatoria) X è rappresentata matematicamente da una una funzione definita sullo spazio campionario Ω di un esperimento aleatorio.

$$X : \Omega \to R$$

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.

5.2.2 Valore medio di X

Sia X una variabile aleatoria discreta (reale) che assume una quantità finita di valori $x_1, x_2, \ldots, x_n$. Si definisce

$$E[X] := \sum_{i=1}^{n} x_i (p_X)^{x_i} = \sum_{i=1}^{n} x_i P(X = x_i) \tag{5.3}$$

Valore Medio è la somma dei valori assunti da X "pesati" con le rispettive probabilità.

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

Esempio dado Sia "X = risultato del lancio di un dado regolare a 6 facce". $E[X] = 3.5$.
Non è un valore assunto dalla v.a. dato che i numeri sul dado sono solo interi.
Si può notare dall'esempio precedente 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?
Di seguito è riportata l'interpretazione frequentista di E che spiega a livello pratica cosa rappresenta.

Interpretazione frequentista di E[X]

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:

$$\frac{X_1 + X_2 + \cdots + X_n}{N} \backsimeq E[X] \tag{5.4}$$

5.2.3 Proprietà del valore medio

Per ogni variabile aleatoria (reale) X

$$\begin{gathered} E[X + c] = E[X] + c \\ E[cX] = cE[X] \end{gathered}$$

Questo vale per ogni costante $c \in R$
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.

Altre importanti proprietà Se $X = c$ (costante) allora $E[X] = E[c] = c$
Un'altra proprietà importante: Se $X >= 0$ allora $E[X] >= 0$

Formula di trasferimento

$$E[f(x)] = \sum_{i=1}^{n} f(x_i) p_x^{x_i} = \sum_{i=1}^{n} f(x_i) P(X = x_i) \tag{5.5}$$

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

$$E[X^2] = \sum_{i=1}^{n} x_i^2 p_X^{x_i} = \sum_{i=1}^{n} x_i^2 P(X = x_i) \tag{5.6}$$

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.

5.3 Varianza e Deviazione standard con valore medio

Varianza $VAR[X] := E[(X - u)^2] >= 0$ con $u := E[X]$

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

Formula alternativa $VAR[X] = E[X^2] - E[X]^2$
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[X] e varianza VAR[X] 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.

5.3.1 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 = $(SD)^2 \backsimeq$ (larghezza della distribuzione)^2

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

Note per gli esercizi Stesso discordo della media, le proprietà sono super utili nel caso ci viene richiesto di calcolare una varianza Y in funzione di X e già calcolata in precedenza.

5.3.2 Dipendenza e indipendenza variabili aleatorie

Siano X e Y due variabili aleatorie, che dipendono entrambe dallo stesso esperimento aleatorio. Quanto valre $VAR[X+Y] =?$ **dipende da come solo legate X e Y!**

Definizione indipendenza Due v.a. discrete X e Y si dicono indipendenti se gli eventi X = x e Y = y sono indipendenti, ossia:

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

Per ogni scelta di x e y.
Intuitivamente conoscere il valore assunto da X non modifica la proabilità dei valori di Y e viceversa.
Molto spesso l'indipendenza è assunta in partenza.

Teorema Se X e Y sono v.a. indipendenti, allora:

$$VAR[X+Y] = VAR[X] + VAR[Y]$$

5.4 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 $A = (a, b)$). L'insieme di tali probabilità definisce la distriuzione (di probabilità) della v.a. X. Per v.a. discrete, la distribuzione di X è determinata dalla 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**.

5.5 Classificazione distribuzioni discrete più importanti

Bernoulli

Si chiama Bernoulli una v.a. X che può **assumere soltanto i valori 0 e 1** cioè

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

Dato che la somma di tutti i valori che può assumere è 1 si ottiene

$$p_x(x) = P(X = x) = \begin{cases} p \text{ se } x = 1 \\ 1 - p \text{ se } x = 0 \end{cases} \quad \text{cioè} \quad \begin{cases} P(X = 1) = p \\ P(X = 0) = 1 - p \end{cases}$$

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

Valore Medio $E[X] = p$

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

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.

Esempi

- 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

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

Consideriamo quindi la v.a. X:= numero di successo 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: $\text{Bin}(1, p) = \text{Be}(p)$
La distribuzione di X per costruzione è

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

Inoltre la densità discreta è data da:

$$p_x(k) = P(X = k) = \binom{n}{k} p^l (1-p)^{n-k} \text{ per } k = 0, 1, \dots, n$$

Dove:

- $\binom{n}{k} \to$ scelte di quali prove hanno successo $\to$ combinazioni
- $p^k \to$ probabilità di k successi fissati
- $(1-p)^{n-k} \to$ 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.

Se $X \sim \text{Bin}(n, p)$

Valore Medio $E[X_i] = np$

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

Poisson

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

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

Si può ottenere una v.a. di Posson $X \sim \text{Pois}(\lambda)$ come opportuno limite di una v.a. 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 $X \sim \text{Pois}(\lambda)$

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

Poisson in pratica ed Esempi Le v.a. di Poisson sono approssimazoni per v.a. che contano il "numero di successi" quando si considera una grande quantità di prova la cui probabilità di accesso 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 v.a. **X si dice geometrica di parametro** $p \in (0, 1]$ e si scrive $X \sim$ **Geo** p, se $X(\Omega) = \mathbb{N} = 1, 2, 3, \dots$:

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

Si può ottenere una v.a. Geo(p) a partire da una successione infinita di prove ripetute e indipendenti, con probabilità di successo p, e considrando la v.a.

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

Dove l'istante è il numero della prova.
Ad esempio, indicando con $X_i \sim \text{Be}(p)$ per $i = 1, 2, 3, \dots$ 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

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

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

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

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

5.6 Variabili aleatorie continue

Esperimento aleatorio → spazio di probabilità (Ω, P)
Variabile aleatoria → funzione $X : \Omega \to R$
Fin'ora abbiamo studiato v.a. 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:

$$P_x^{x_i} = P(X = x_i)$$

$$P(X \in A) = \sum_{x_i \in A} P_x^{x_i} \tag{5.7}$$

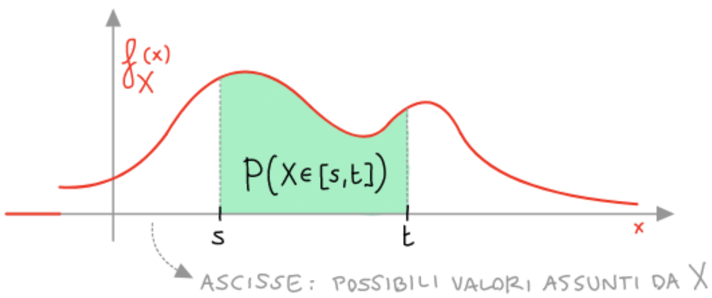
Consideriamo ora una classe "complementare" di v.a., dette **assolutamente continue** che assumono un insieme infinito più che numerabile di valore, come ad es. un intervallo di R: $[0, 1]$ $[0, +\infty)$ $(-\infty, +\infty)$
Una v.a. è **assolutamente continua** se la sua distribuzione è determinata da una funzione $f_x(x)$, a valori positivi, detta densità della v.a. X, nel modo seguente:

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

In particolare:

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

con $-\infty \le s \le t \le +\infty$
Area sotto il grafico di f_x tra i punti s e t



In altre parole una variabile aleatoria continua può assumere qualunque valore in un certo intervallo. 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. X ha una curva associata. Questa curva, nota come funzione di densità di probabilità, può essere usata per ottenere le probabilità associate a una v.a.
Visto che **X assume sempre un valore, otteniamo che l'area totale sottesa dalla curva deve essere uguale a 1.**
Inoltre visto che l'area sotto il grafico di una funzione di densità di probabilità tra i punti a e b non varia se gli estremi sono inclusi o esclusi otteniamo che:

$$P\{a \leq X \leq b\} = P\{a < X < b\}$$

Questo significa che la probabilità che una variabile aleatoria continua rientri in un intervallo non cambia se includiamo o no gli estremi.
La curva di densità di probabilità di una v.a. X non scende mai sotto all'asse x ha la proprietà che l'area delimitata da essa e dall'asse x è sempre uguale a 1.
La curva determina le probabilità di X in questo modo l'area sottesa dalla curva tra i punti a e b è uguale alla probabilità che X assuma un valore compreso tra a e b.

Analogie tra v.a ass. continua e discreta

Ci sono analogie formali tra le 2:

- X ass. continua $P(X \in [s,t]) = \int_s^t f_x(x)\,dx$ **Integrale**
- X discreta $P(X \in [s,t]) = \sum_{x_i \in [s,t]} p_x(x_i)$ **Somma**

Ma anche importanti differenze!
Se X è assolutamente continua:

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

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

In particolare: $f_x(x)$ **NON è** $P(X = x)$, tranne dove $f_x(x) = 0$

Proprietà La densità di una v.a. 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 $X \in (+\infty, -\infty)$ è più che numberabile!

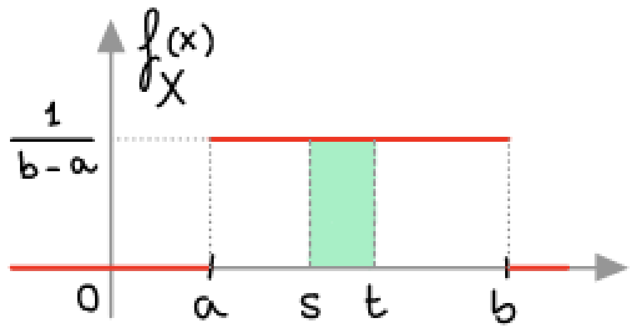
5.6.1 Variabile uniforme continua

Fissiamo un intervallo limitato $[a, b]$ $(-\infty < a < b < -\infty)$.
Una v.a. X si dice uniforme continua in [a, b] e si scrive $X \sim U(a, b)$ se X è ass. cont con densità

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

con $c = \frac{1}{b-a}$
Dato un intervallo $[s, t] \subseteq [a, b]$

$$P(X \in [s, t]) = \int_s^t f_x(x)\, dx = \frac{1}{b-a} \int_s^t 1\, dx = \frac{t-s}{b-a} = \frac{\text{Lungh. di } [s, t]}{\text{Lungh. di } [a, b]}$$



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

Valori assunti da una V.A assolutamente continua

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

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.

5.6.2 Valori assunti da una V.A. assolutamente continua

La densità di una V.A. assolutamente continua X è una funzione $f_x : \mathbb{R} \to \mathbb{R}$ (integrabile) tale che:

$$f_x(x) \geq 0 \ \forall x \in \mathbb{R}$$

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

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

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

5.7 Valore medio e Varianza di V.A. Assolutamente Continue

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

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

Varianza

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

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

Tutte le proprietà valide nelle v.a. discrete continuano a valere In particolare:

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

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

Se X e y **sono indipendenti**

5.8 V.A. Esponenziale

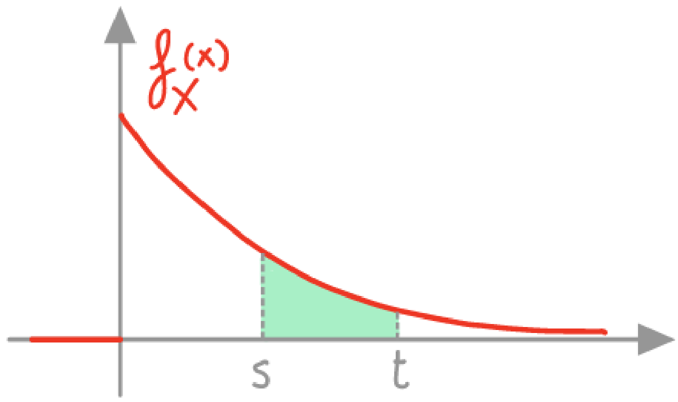
Misuro il tempo di emissione X di una particella radioattiva da un atomo, con "tempo medio di emissione" τ. Sia $\lambda = \frac{1}{\tau}$. Descriviamo X con una v.a. assolutamente continua:

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

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 $[s, t] \subseteq [0, \infty]$:

$$P(X \in [s,t]) = \int_s^t f_x(x)\, dx = \int_s^t \lambda e^{-lambdax}\, dx = [-e^{-\lambda x}]_s^t =$$

$$= e^{-\lambda s} - e^{-\lambda t}$$



5.9 Funzione di Ripartizione

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

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

A livello probabilistico è la probabilità che X sia minore o uguale a x.

- F_x è ben definita per ogni v.a. $X : \Omega \to \mathbb{R}$, sia che la v.a. sia discreta, ass. continua, o nè l'una nè l'altra.

- F_x determina la distribuzione della v.a.
- F_x è legata alla densità discreta/densità di X

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

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

Esempio Bernoulli Sia $X \sim \mathrm{Be}(p)$ 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

$$p_X(x_i) = F_X(x_i) - F_X(X_{\bar{i}})$$

$$F_X(x_{\bar{i}}) = \lim_{t \to x_{\bar{i}} F_X(t)}$$

Funzione di ripartizione di V.A. Assolut. Continue

X v.a. assolutamente continua $\Leftrightarrow$ F_X è una funzione continua ed è derivabile a tratti.

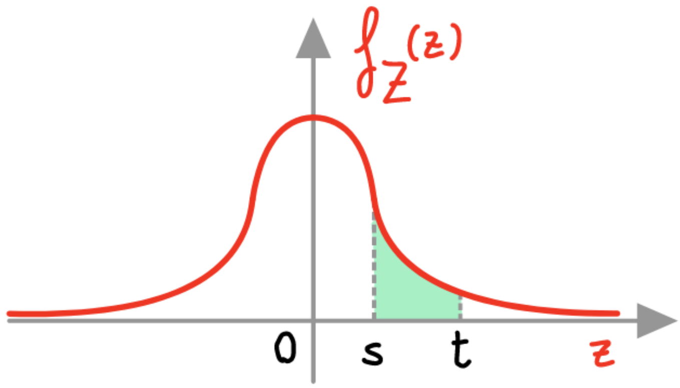
$$\text{Densità} \qquad f_X(x) = (F_X)'(x) \tag{5.8}$$

5.10 Variabili aleatorie Normali

L'ultima classe che vedremo e la più importante, è quella delle variabili aleatorie normali o Gaussiane.
Una v.a. Z si dice **Normale Standard** (si scrive $Z \sim N(0,1)$), se è assolutamente continua con densità:

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

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



Come si può notare la forma "a campana" è simmetrica rispetto all'orgine.

$$\text{"STANDARD"} = \begin{cases} E[Z] = 0 \\ \text{Var}[Z] = 1 \end{cases} \tag{5.9}$$

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

$$P(Z \in [s,t]) = \int_s^t f_z(z)\, dz \qquad \text{(come per ogni v.a. assolutamente continua)}$$

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 Φ.

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

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

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

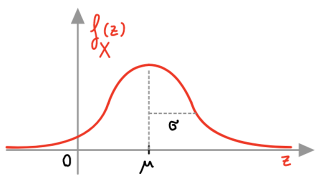
Grazie alla tavola, è come se conoscessimo $\Phi(z) = F_Z(z)$. Possiamo allora calcolare la probabilità degli intervalli:

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

Siano ora $\mu \in \mathbb{R}$, $\sigma \in (0, \infty)$. Una v.a. X si dice **Normale con media μ e varianza σ^2**, si scrive $X \sim N(\mu, \sigma^2)$, se X è assolutamente continua con

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

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



La densità di f_X di X si ottiene dalla densità di f_Z di $Z \sim N(0,1)$ mediante una traslazione e un riscalamento:

$$f_X : \text{grafico "a campana" centrata in } \mu, \text{ "ampiezza" } \sigma$$

Ci si può sempre ricondurre a una v.a. Normale Standard Z:

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

Viceversa

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

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

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

Il fatto che ci si può ricondurre a una v.a. normale standard è un caso particolare della seguente proprietà:

Teorema Se X è normale $\to Y = aX + b$ è normale $\forall a \neq 0, b \in \mathbb{R}$
$X \sim N(\mu, \sigma^2) \to Y \sim N(a\mu + b, a^2\sigma^2)$

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

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

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

Vettori Aleatori - Non presenti agli esami

Completare - vedi commento nel codice

5.11 Legge dei grandi numeri

Completare - vedi commento nel codice

Capitolo 6

Teoremi di convergenza

Siano $X_1, X_2 \dots$ successione di Variabili aleatorie indipendenti identicamente distribuite (V.A I.I.D.).
Discrete $\rightarrow p_x(x) = p(x)$
Assolutamente continue $\rightarrow f_{x_i}(x) = f(x)$

Per osservare X_1, X_2 devo raccogliere dei dati, il problema è che $p(x), f(x)$ non sono completamente note.

Statistica inferenziale Dalle osservazioni posso fare delle deduzioni (Inferenze) sulla distribuzione comune di X_1, X_2 ossia su $p(x)$ e $f(x)$.
Non posso osservare tutte le V.A per questo ne scelgo casualmente n.
$X_1, \dots, X_n$ Campione aleatorio di ampiezza n. Lavoreremo con variabili aleatorie che siano funzioni del campione aleatorio, cioè del tipo $g(x_1, \dots, x : n)$

Statistica campionaria Dato un campione aleatorio $X_1, \dots, X_n$ una statistica campionaria è una qualunque funzione del campione:

$$g(X_1, \dots, X_n) \ (g : \mathbb{R}^n \rightarrow \mathbb{R})$$

Esempio di statistica campionaria Media campionaria

$$\bar{X_n} : \frac{1}{n} \sum_{n=1}^{n} X_i$$

6.1 Distribuzione di X

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

$$\mathrm{Var}(\bar{X_n}) = \frac{\sigma^2}{n}$$
$$\frac{\sigma}{\sqrt{n}} = \sqrt{\mathrm{Var}}$$

6.2 Teorema del limite centrale

Siano $X_1, ..., X_n$ v.a i.i.d. (Variabili aleatorie indipendenti identicamente distribuite) $E(i) = \mu$ e $\mathrm{Var} X_i = \sigma^2$ (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 $\sqrt{\mathrm{Var}}$
- 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

Capitolo 7

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 che sono stime a valore singolo del parametro. Verrà considerato poi l'errore standard di queste stime. Inoltre considereremo gli intervallo di confidenza, che contengono il parametro con un certo livello di confidenza.

7.1 Statistica inferenziale

La **statistica inferenziale** ha lo scopo di definire in modo non ambiguo e quantitativo la plausibilità di un'inferenza.
La plausibilità di un'inferenza dipende dal modo con cui è stato selezionato il campione di n individui della popolazione. La corretta metodologia di campionamento è la **scelta casuale**.
Per la casualità dle campionamento utilizzo il calcolo delle probabilità.
Quando parleremo di popolazione tratteremo sempre N molto grandi rispetto alla numerosità del campione.

Popolazione V.A. i.i.d. con la stessa distribuzione $\leftrightarrow$ legge F.
Consideriamo il caso in cui F è nota a meno di qualche parametro incongito.

Modello statistico parametrico Famiglia di leggi note a meno di uno o più parametri.

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

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

Attenzione Non confondere le V.A. con le osservazioni

- $X_1, ..., X_n$ sono le v.a.
- $x_1, ..., X_n$ sono le osservazioni

Esempi di modelli statistici Esempi di modelli statistici sono:

- Modello di Bernoulli
- Modello esponenziale
- Modello normale

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

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

Esempi di statistiche

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

7.2 Statistica parametrica e Stimatori

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

Stimatori Uno **Stimatore** è una statistica il cui valore dipende dal particolare campione che è stato estratto. Il valore dello stimatore, **la Stima**, viene usato per predire il valore di un parametro della popolazione. particolari statistiche che servono a stimare i parametri incogniti.

Stimatori Puntuali Sono valori singoli che speriamo siano prossimi ai parametri stimati.

Stimatori intervallari Meglio noti come **Intervalli di confidenza**, in questo caso non rappresentano un singolo valore, ma un intervallo in cui ci aspettiamo che il parametro rientri. Ci occupiamo anche di determinare quanta confidenza associare a un dato intervallo, cioè quanto possiamo essere sicuro che il parametro si trovi in questo intervallo.

Definizione Stimatore Corretto Uno **Stimatore** T si dice **Non Distorto o Corretto** se:

$$\mathbb{E}_\Theta T : \mathbb{E}_\Theta(g(X_1, ..., X_n)) = \Theta$$

Dove E_Θ è il valore medio rispetto alla probabilità di 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.

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

- p in un modello di Bernoulli
- λ in un modello di Poisson
- $\frac{1}{\lambda}$ in un modello esponenziale
- μ in un modello normale

Osservazione La proprietà di essere non distorto NON è stabile per trasformazioni, 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.

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

$$\mathrm{Var}_\Theta(T) \to 0$$

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

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

7.3 Errore standard

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

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

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

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**.
In statistica descrittiva: siano n osservazioni $(x_1, ..., x_n)$;

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

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

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

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

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

è uno stimatore corretto di σ^2.

7.4 Chi quadrato

Distribuzione delle statistiche campionarie

Campione normale $X_1, ..., X_n$ campione casuale $\mathcal{N}(\mu, \sigma^2)$

$$\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

Standardizzando otteniamo

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

Per caratterizzare la legge di S_n^2 e di $\bar{S}_n^2$ dobbiamo introdurre una nuova distribuzione continua

$$\chi^2(n)$$

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

$$Y = \sum_{i=1}^{n} Z_i^2 , \quad z_1, ..., z_n \quad \text{i.i.d.} \quad \mathcal{N}(0,1)$$

Dove Y è una v.a. ≥ 0 con densità $f_r(t) = c_n t^{\frac{n}{2}-1} e^{-\frac{t}{2}}$ per $t > 0$
Si ha $\mathbb{E}(Y) = n$ e $\text{Var}(Y) = 2n$
Per $n = 2$ è la legge $\exp(\frac{1}{2})$
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

7.5 Distribuzione t di student

Definizione Wikipedia Nella teoria delle probabilità la distribuzione di Student, o t di Student, è 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.

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

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

$$z \sim \mathcal{N}(0,1)$$

$$Y \sim \chi^2(n)$$

Z e Y sono indipendenti.

$$\mathbb{E}(T) = 0 \quad \mathrm{Var}[T] = 1$$

T è una v.a. continua con densità

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

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

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

7.6 Parte pratica

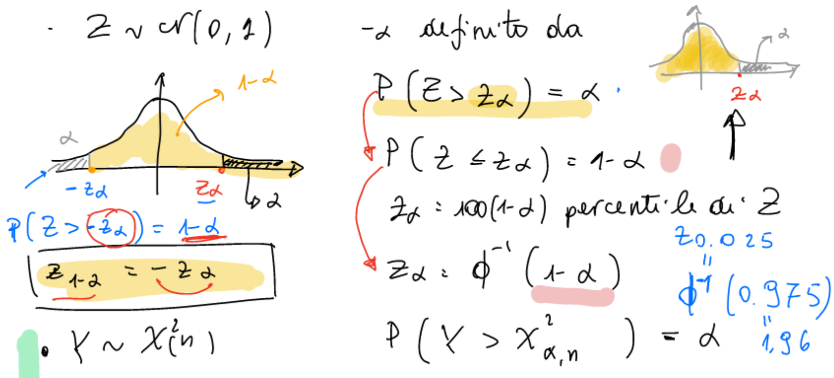
Cosa dovremo saper calcolare di queste variabili aleatorie? I loro percentili.
X v.a.
$P(X \leq q_\alpha) = \alpha$ Fissato α, trovare q_α.
$q_\alpha = \alpha$ - esimo quantile o 100α - percentile di X.

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



7.7 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 $\rightarrow$ 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 8

Verifica di Ipotesi

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.

8.1 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.

α = **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:

$$\mathbb{P}_H((x_1, ..., x_n) \in C) \leq \alpha \tag{8.1}$$

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.

8.1.1 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).