Esercitazione 1 - Basi di Statistica

Non tutto in questo sarà importante, non ci verranno chieste domande di statistica, solo per capire per il corso.

Statistica → analizza i fenomeni incerti

Fenomeni aleatori → fenomeno dove il risultato non è definito e può variare

Misura è processo aleatorio perché ci sono misure che trascuriamo per semplificare il nostro modello, dati questi fattori il nostro risultato varia rispetto al valore vero.

Possibili esiti di fenomeno aleatorio è detto evento elementare

L'insieme dei eventi elementari è spazio campionario

Finito o infinito numerabile

Infinito

↓
Discreto

Non numerabile
↓
Continuo

↳ Sotto insieme è evento.

Probabilità: (Informale)

↳ Probabilità di un evento tra 0 e 1 che misura quanto riteniamo possibile che si verifichi un evento

Come assegnare probabilità:

$$P = \frac{n_A}{N}$$

n_A ← Probabilità individuale evento

N ← Numero eventi

Variabile aleatoria è una funzione che associa ad un fenomeno aleatorio una grandezza o qualità

Variabile aleatoria è risultato possibile di fenomeno

<u>Statistica Descrittiva:</u>

Con fenomeno aleatorio e variabile aleatoria associato, e supponiamo di verificare l'esito del fenomeno per N volte, e N realizzazioni

Statistica Descrittiva ⇒ permette di analizzare la probabilità.

Popolazione ⇒ insieme di tutte le realizzazioni di una variabile aleatoria, cioè l'insieme di tutti i valori che variabile aleatoria assume

→ Tende ad esser ∞

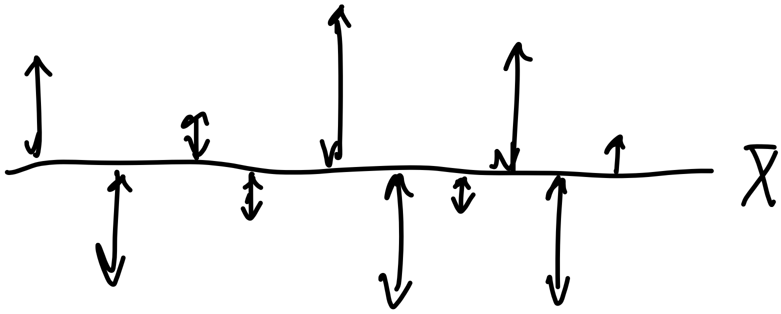
Campione ⇒ porzione finita della popolazione

↳ Quello che guardiamo nella statica descrittiva.

<u>Indici Statistici</u>

Media (Campionaria) ⇒ $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} x_i$

Varianza (Campionaria) ⇒ $\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{X})^2$



Deviazione standard
(Campionario)

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

È possibile calcolare per la popolazione, ma in generale useremo per la stima campionaria

Istogramma
↳ Diagramma descrittivo delle probabilità

$$\text{Frequenza relativa} = \frac{\text{Numero di eventi}}{\text{Eventi Totali}}$$

Gli istogrammi può esser costruito con edge non equi spaziati

Densità di Frequenza Relativa = $\frac{\text{Frequenza relativa}}{\text{Dimensione di edge}}$

Frequenza = Densità di probabilità · Intervallo

<u>Funzioni di densità</u>

Variabile Aleatoria assume valore in base all'evento

Possiamo associare ad ogni evento una probabilità

È possibile definire una ✓ Funzione che associa ad un variabile aleatoria una probabilità

v.a. discrete → funzioni di probabilità
v.a. continue → funzione di densità di probabilità

Caso discreto (interessa meno):

Funzione di probabilità $\qquad\qquad$ v.a. = x

$$f_x(x) = \begin{cases} P[X = x_j] & se \ x = x_j \\ 0 & se \ x \neq x_j \end{cases}$$

Funzione di probabilità può esser posto su diagramma barra

Istogramma usato per stima funzione di densità

$$0 \leq f_x(x) \leq 1$$

$$\sum_j f_x(x) = 1$$

$$P[a < X \leq b] = \sum_{j: a < x_j \leq b} f_x(x_j)$$

Valore Attesto $\qquad E[X] = \sum_i \underbrace{x_i}_{\text{Valore Discreto}} \cdot \underbrace{f_x(x_i)}_{\text{Funzione di prob}}$

↳ È il baricentro può non coincidere con valori

Varianza $\quad Var[X] = \sum_i (x_i - E[X])^2 \cdot f_x(x_i)$

↳ indica dispersione dal baricentro

$$\sigma_X = \sqrt{Var[X]}$$

↳ Deviazione Standard

Funzione di probabilità cumulata

↳ prob. che v.a. sia minore o uguale al valore certo

$$F_X(x) = P[X \leq x] = \sum_{j: x_j \leq x} f_X(x_j)$$

Proprietà

$$\lim_{x \to -\infty} F_X(x) = 0 \qquad \lim_{x \to \infty} F_X(x) = 1$$

$$F_X(b) \geq F_X(a) \quad \forall b > a$$

$$P[a < X \leq b] = F_X(b) - F_X(a)$$ ← ha senso

Variabili Aleatorie Continue

Per intervalli che tendono a 0, la probabilità associata tende a 0, infatti la probabilità associata al singolo valore è 0 (solo per continuo)

Prendendo solo le frequenze tendono a 0, prendendo la densità, la curva tende ad un valore finito non 0, ma non è una probabilità è un probabilità per intervallo.

Probabilità per ogni livello di separazione.

Funzione di densità di probabilità continua o PDF

$$P[X \in J] = \int_J f_X(x)\,dx$$

↳ Probabilità che assuma valore
↳ Va integrata la PDF per trovare la probabilità che cerchio

Per v.a. continue non senso cercare possibilità di valore individuale, serve usare un intervallo, anche piccolo anche infinitesimo

Prop.

$$P[X=a] = \int_a^a f_X(x)\,dx = 0$$

$$P[a \le X \le b] = \int_a^b f_X(x)\,dx$$

$$P[a \le X < b] = P[a < X \le b]$$

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

PDF cumulata o CDF

$$F_X(x) = \int_{-\infty}^{x} f_X(x) \cdot dx$$

Proprietà

$$P[a \le X \le b] = F_X(b) - F_X(a)$$ Useremo molto

$$\lim_{x \to -\infty} F_X(x) = 0 \qquad \lim_{x \to \infty} F_X(x) = 1$$

$$F_X(b) \ge F_X(a) \qquad \forall\, b > a$$

Parametri funzione di densità

Valore Atteso $E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) \cdot dx$

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

Deviazione Standard:

$$\sigma_X = \sqrt{Var[X]}$$

Discreto - $f_X(x)$ per v.a discrete esprimono una probabilità

- $f_X(x)$ per v.a. continue esprimono una densità di probabilità. La probabilità è associata ad intervalli e si determina quindi mediante un'operazione di integrazione

Probabilità dalla distribuzione

$$P(a) = \int_{X_i}^{X_j} f_X(x) dx = \int_{-\infty}^{X_j} f_X(x) dx - \int_{-\infty}^{X_i} f_X(x) dx = F_X(X_j) - F_X(X_i)$$

serve sapere la CDF o avere tabelle

Standardizzazione della distribuzione. convenzionalmente si vuole $\mu = 0$, e $\sigma = 1$

μ valore atteso

σ deviazione standard

Praticamente si opera un cambio di variabile, usando la variabile z al posto della variabile x

$$z = \frac{x-\mu}{\sigma}$$

Riscriviamo CDF:

$$F_x(x) = \int_{-\infty}^{x} f_x(\xi)\, d\xi = \int_{-\infty}^{\frac{x-\mu}{\sigma}} f_z(z)\,\sigma\, dz = F_z\left(\frac{x-\mu}{\sigma}\right)$$

$F_x(x)$: Cumulata non standard

$F_z\left(\frac{x-\mu}{\sigma}\right)$: Cumulata standard → Deviazione standard

Si può usare la tabella di z per trovare $F_x(x)$ perché sono associate

Tornando indietro:

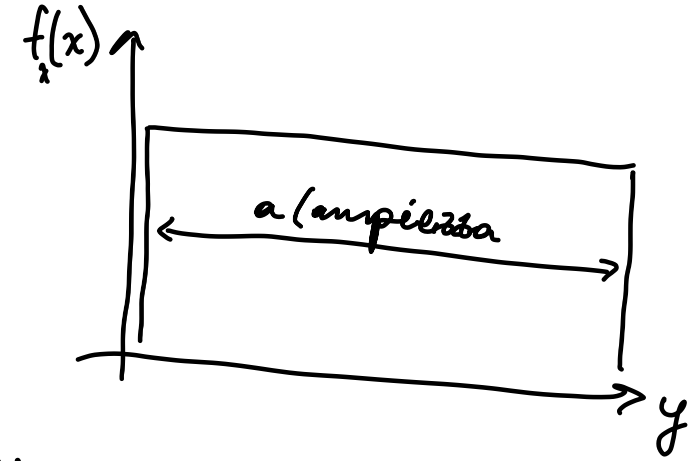
$$P(a) = F_x(X_j) - F_x(X_i) = F_z\left(\frac{X_j-\mu}{\sigma}\right) - F_z\left(\frac{X_i-\mu}{\sigma}\right)$$

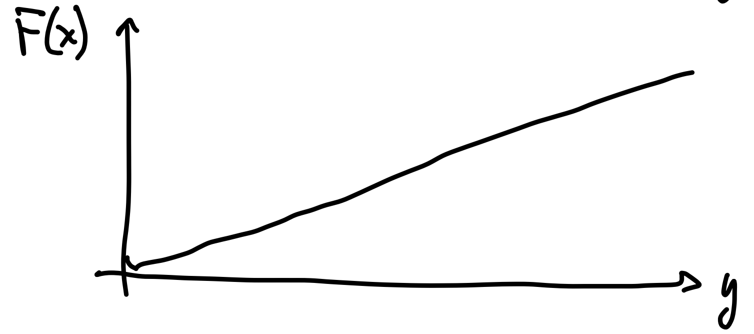
Famiglie di Funzioni di Densità (distribuzione)
(quelle che ci interessano)

- Uniforme
- Normale / Gaussiana
- T-Student

Solo alcune funzioni possono avere densità di probabilità

Distribuzione Uniforme





$$E[X] = \int_0^a x \cdot f_x(x) =$$

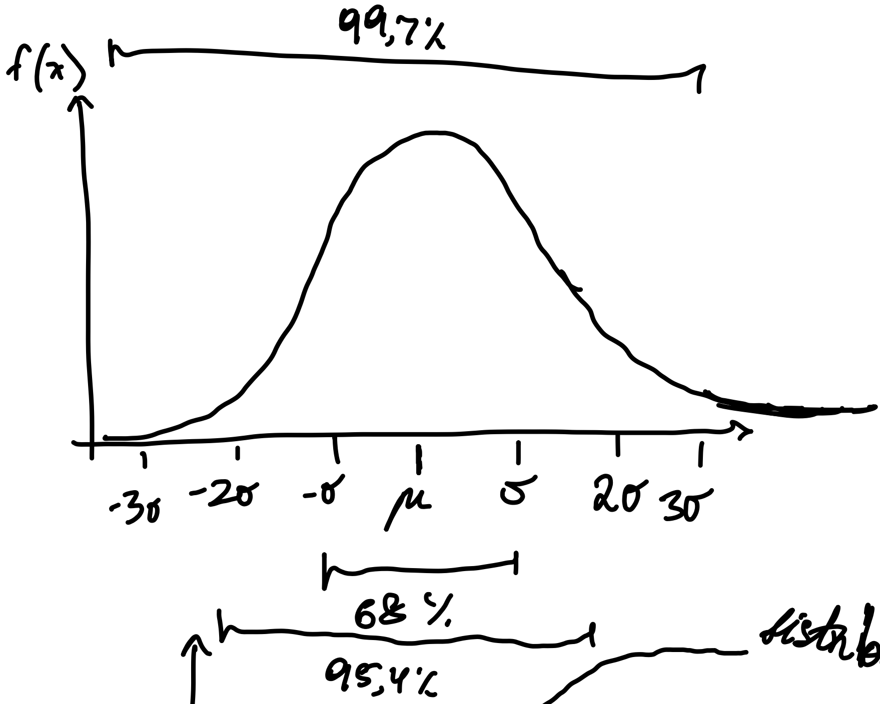
$$= \int_0^a x \cdot \frac{1}{a} = \frac{1}{a} \cdot \frac{a^2}{2} = \frac{a}{2}$$

Varianza e Stand. Dev.

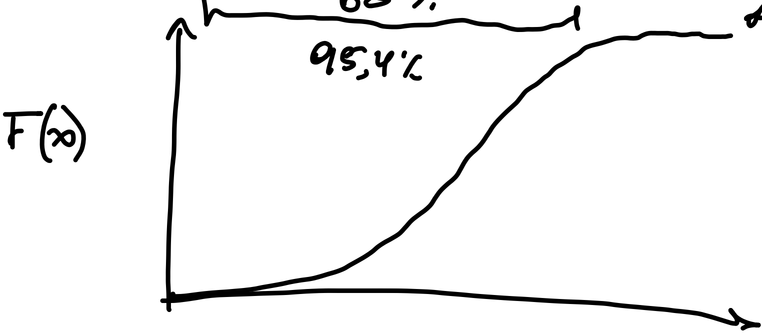
$$Var[X] = \frac{a^2}{12} \qquad \sigma_x = \frac{a}{2 \cdot \sqrt{3}}$$

Normale o Gaussiana

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



Servono tabelle per questo, distribuzione molto comune



La media sposta la curva, è sempre centrata alla media

Con cambio di σ, cambia la altezza.

T-student

Fenomeno aleatorio

n campioni da popolazione normale con media μ e stand.dev σ"

Media campionaria $\bar{X}$ di n-campioni

Allora $\bar{X}$ è una variabile aleatoria perché cambia con n

se $\bar{X}$ è una v.a. allora avrà una sua distribuzione...

Se popolazione ha distribuzione normale

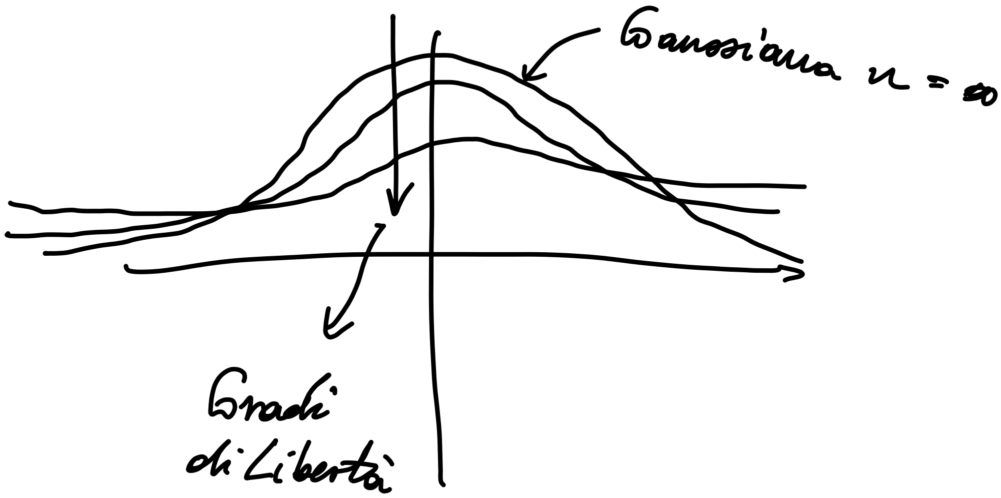
$X \sim N(\mu, \sigma)$, allora $\bar{X}$ si distribuisce come $N(\mu, \frac{\sigma}{\sqrt{n}})$

La distribuzione standardizzata cioè la distribuzione della v.a. $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ è di conseguenza una normale standard.

In molti casi non abbiamo σ, quindi è stimata con la standard campionaria $\underline{S}$.

quindi $\frac{\bar{X} - \mu}{S/\sqrt{n}}$, ma non si distribuisce più come una standard, si distribuisce invece come una t-student con n-1 gradi di libertà

$$X \sim N(\mu, ?) \longrightarrow \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$


Gaussiana $n = \infty$
Gradi di Libertà