UTILIZZO DELLA F. D.

$x \in \mathbb{R}$, $\quad F_X(x) := \mathbb{P}(X \le x)$

la probabilità che X assume un valore nell'intervallo $]-\infty, x]$

E PER GLI ALTRI INTERVALLI?

1) $\quad [x_1, x_2]$, $\quad x_1 < x_2$

$$F_X(x_2) = \mathbb{P}(X \le x_2) = \mathbb{P}(\{X \le x_1\} \cup \{x_1 < X \le x_2\})$$

$$= \mathbb{P}(X \le x_1) + \mathbb{P}(x_1 < X \le x_2) \implies$$

$$\mathbb{P}(x_1 < X \le x_2) = F_X(x_2) - F_X(x_1).$$

---

2) $\{x\}$, $x \in \mathbb{R}$

$$\varepsilon > 0, \quad \mathbb{P}(x - \varepsilon < X \le x) \overset{1)}{=} F_X(x) - F_X(x-\varepsilon)$$

$n \in \mathbb{N}$, $\varepsilon = \frac{1}{n}$ $\quad \mathbb{P}\left(x - \frac{1}{n} < X \le x\right) = F_X(x) - F_X\left(x - \frac{1}{n}\right)$

$$\lim_{n \to \infty} \mathbb{P}\Big(\underbrace{x - \tfrac{1}{n} < X \le x}_{D_n \downarrow \{x\}}\Big) = F_X(x) - \lim_{n} F_X\left(x - \frac{1}{n}\right)$$

$$\mathbb{P}\left(\lim_{n \to \infty} x - \frac{1}{n} < X \le x\right) = F_X(x) - F_X(x^-)$$

$$\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$$

$$\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$$

$$= F_X(x) - \lim_{\varepsilon \to 0^+} F_X(x - \varepsilon).$$

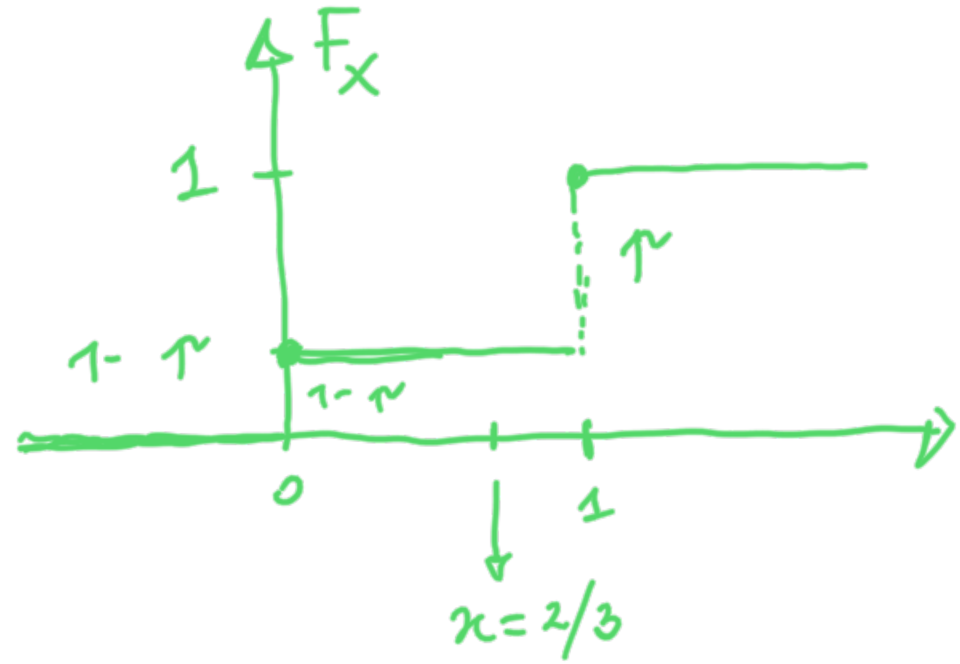
2 bis) nel caso che F_X è continua in x allora

$$\mathbb{P}(X = x) = F_X(x) - F_X(x) = 0.$$

ESEMPIO

$X \sim B(1, p)$

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



$$\mathbb{P}\left(X=\frac{2}{3}\right)=F\left(\frac{2}{3}\right)-F\left(\frac{2^-}{3}\right)=(1-p)-(1-p)=0.$$

---

3) $[x_1, x_2]$, $x_1 < x_2$

$$[x_1, x_2] =]x_1, x_2] \,\dot{\cup}\, \{x_1\}$$

$$\mathbb{P}(x_1 \leq X \leq x_2) = \mathbb{P}(x_1 < X \leq x_2) + \mathbb{P}(X = x_1)$$

$$\overset{1)}{=} F_X(x_2) - F_X(x_1) + \mathbb{P}(X = x_1)$$

$$\overset{2)}{=} F_X(x_2) - F_X(x_1) + \left[F_X(x_1) - F_X(x_1^-)\right]$$

$$= F_X(x_2) - F_X(x_1^-).$$

---

4) $]x_1, x_2[$, $x_1 < x_2$

$$]x_1, x_2] =]x_1, x_2[\,\dot{\cup}\, \{x_2\}$$

$$\mathbb{P}(x_1 < X \leq x_2) = \mathbb{P}(x_1 < X < x_2) + \mathbb{P}(X = x_2)$$

$$\mathbb{P}(x_1 < X < x_2) = \mathbb{P}(x_1 < X \leq x_2) - \mathbb{P}(X = x_2)$$

$$\stackrel{1)}{=} F_X(x_2) - F_X(x_1) - \mathbb{P}(X = x_2)$$

$$\stackrel{2)}{=} F_X(x_2) - F_X(x_1) - \left[F_X(x_2) - F_X(x_2^-)\right]$$

$$= F_X(x_2^-) - F_X(x_1).$$

---

5) $[x_1, x_2[$, $x_1 < x_2$

$$[x_1, x_2] = [x_1, x_2[\,\dot{\cup}\, \{x_2\}$$

$$\mathbb{P}(x_1 \leq X \leq x_2) = \mathbb{P}(x_1 \leq X < x_2) + \mathbb{P}(X = x_2) \Rightarrow$$

$$\mathbb{P}(x_1 \leq X < x_2) = \mathbb{P}(x_1 \leq X \leq x_2) - \mathbb{P}(X = x_2)$$

$$\underset{2)}{\overset{3)}{=}} F_X(x_2) - F_X(x_1^-) - \left[F_X(x_2) - F_X(x_2^-)\right]$$

$$= F_X(x_2) - F_X(x_1^-) - F_X(x_2) + F_X(x_2^-)$$

$$= F_X(x_2^-) - F_X(x_1^-).$$

---

6) $]x, +\infty[$

$$]x, +\infty[= (\,]-\infty, x]\,)^c \implies$$

$$\mathbb{P}(x < X < +\infty) = 1 - \mathbb{P}(-\infty < X \leq x)$$

$$\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leq x)$$

$$= 1 - F_X(x).$$

---

7) $]-\infty, x[$

$$]-\infty, x] =]-\infty, x[\,\dot\cup\, \{x\}$$

$$\mathbb{P}(X \leq x) = \mathbb{P}(X < x) + \mathbb{P}(X = x) \Rightarrow$$

$$\mathbb{P}(X < x) = \mathbb{P}(X \leq x) - \mathbb{P}(X = x)$$

$$= F_X(x) - [F_X(x) - F_X(x^-)]$$

$$= F_X(x) - F_X(x) + F_X(x^-)$$

$$= F_X(x^-).$$

---

8) $[x, +\infty[$

$$[x, +\infty[= (]-\infty, x[)^c \Rightarrow$$

$$\mathbb{P}(X \geq x) = 1 - \mathbb{P}(X < x)$$

$$\overset{7)}{=} 1 - F_X(x^-).$$

$\square$

In statistica descrittiva i dati che vengono rilevati mediante uno strumento di misura si riferiscono a un carattere "quantitativo continuo".

In calcolo delle probabilità l'analogo del carattere quantitativo continuo è rappresentato da una variabile aleatoria con legge (assolutamente) continua.

ge (rispettivamente) continua.

Le variabili discrete sono caratterizzate dallo spettro

$$S_X = \{x_1, x_2, \dots\}$$

e da una funzione che assegna la probabilità a ciascuno dei valori dello spettro

$$x_i \in S_X, \quad \mathbb{P}(X = x_i).$$

Entrambe le informazioni sono utilizzate nella determinazione della funzione di distribuzione (F.D.) di X

$$F_X(x) = \sum_i 1_{\{x_i \le x\}} \mathbb{P}(X = x_i).$$

Per le variabili e legge di probabilità

Per le variabili con legge di probabilità di tipo continuo è necessario introdurre un'altra funzione: la funzione di densità di probabilità (f.d.p.) si designa con il simbolo f_X. Essa integrata da $-\infty$ a un reale x qualsiasi restituisce il valore della funzione di distribuzione in x:

$$x \in \mathbb{R}, \qquad \int_{-\infty}^{x} f_X(t)\,dt = F_X(x)$$

ovvero, la F.D. è la funzione integrale della f.d.p.

e questo implica che

$$f_X(x) = F'_X(x)$$

ovvero, la f.d.p è la derivata della F.D.

La f.d.p. gode delle seguenti proprietà

1) $x \in \mathbb{R}$, $f_X(x) \geq 0$

2) $\lim_{x \to +\infty} F_X(x) = 1 \Rightarrow$

$$1 = \lim_{x \to +\infty} \int_{-\infty}^{x} f_X(t)\,dt = \int_{-\infty}^{+\infty} f_X(t)\,dt = \mathbb{P}(X \in \mathbb{R}).$$

Per le v.a. con legge individuata da una f.d.p. risulta che la F.D. non presenta salti ovvero che la F.D. è continua. Pertanto,

$$\forall x \in \mathbb{R}, \quad \mathbb{P}(X = x) = 0.$$

Inoltre, per una tale v.a. le formule

per il calcolo delle probabilità degli intervalli di $\mathbb{R}$ si semplificano:

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b)$$

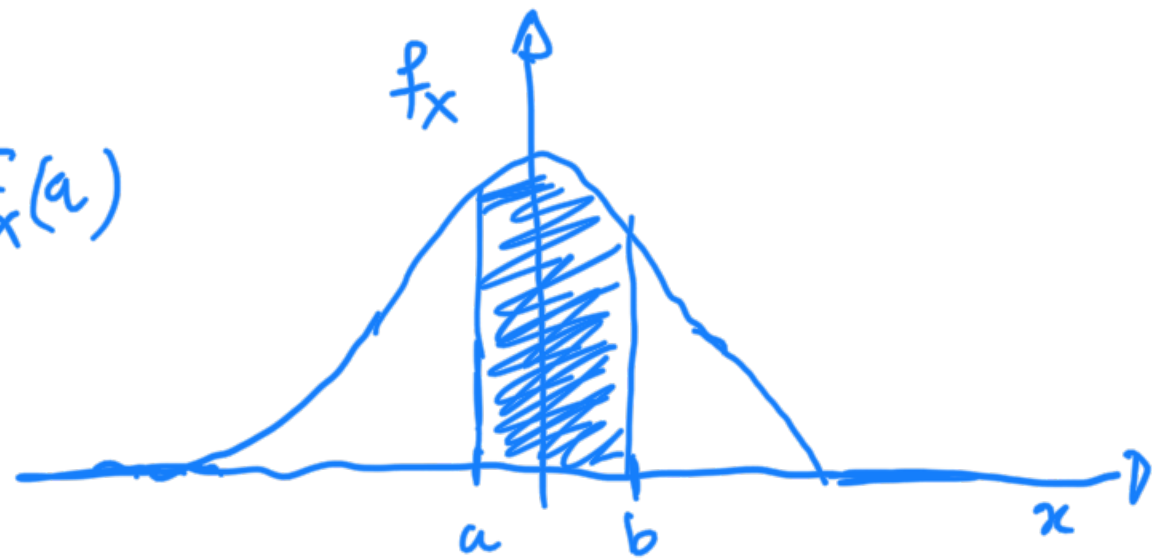
$$= \mathbb{P}(a < X \leq b) = \int_a^b f_X(t)\,dt.$$

Infatti, i singoletti non apportano contributo di probabilità e

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$$

$$= \int_{-\infty}^b f_X(t)\,dt - \int_{-\infty}^a f_X(t)\,dt$$

$$= \int_a^b f_X(t)\cdot dt.$$



Si osservi che l'elemento differenziale $f_X(t)\cdot dt$ rappresenta $\mathbb{P}(a \le X \le b)$ e l'integrale tra a e b fornisce la loro "somma" per t appartente all'intervallo (a,b).

Inoltre,

$$\mathbb{P}(X \le a) = \mathbb{P}(X < a) = \int_{-\infty}^{a} f_X(t)\,dt$$

e

$$\mathbb{P}(X \ge b) = \mathbb{P}(X > b) = \int_{b}^{+\infty} f_X(t)\,dt.$$

GENESI DI UNA fdp

Si consideri una rilevazione dati relativa ad un carattere di tipo quantitativo ne di un carattere di tipo quantitativo continuo. L'istogramma è la rappresentazione grafica adatta a tale rilevazione. Si sa che

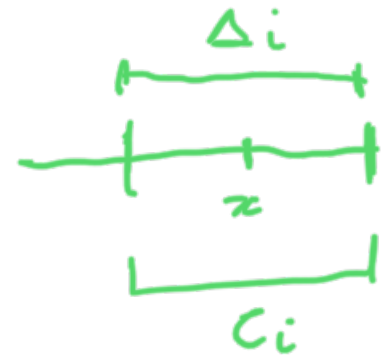
$$i=(1,2,\dots,k)\,, \quad h_i = \frac{f_i}{\Delta_i} = \frac{n_i}{n\,\Delta_i}$$

dove n è la taglia, k è il numero delle classi, f_i è la frequenza relativa della classe i, n_i è la frequenza assoluta della classe i, Δ_i è l'ampiezza della classe i, e h_i è l'altezza dell'i-mo rettangolo. Posto

$$x \equiv \text{punto di mezzo di } C_i$$

allora $h_i = \frac{f_i}{\Delta_i}$ approssima il valore assunto dalla f.d.p. di una variabile aleatoria che "genera" i dati in questione:

$$f_X(x) \approx \frac{h_i}{\Delta_i} = \frac{n_i}{n \cdot \Delta_i}$$



L'approssimazione diventa via via più precisa procedendo in questo modo:

1) si incrementa la taglia n della rilevazione dati

e contestualmente

2) si decrementa la massima ampiezza delle classi.

$\square$

ESEMPIO 5.1.a

Sia X una v.a. continua con f.d.p. data da

$$f_X(x) = \begin{cases} c\cdot(4x-2x^2), & 0<x<2; \\ 0, & \text{altrimenti.} \end{cases}$$

a) Qual è il valore di c?

$$1 = \int_{-\infty}^{+\infty} f_X(t)\,dt = \int_{-\infty}^{0} 0\cdot dt + \int_0^2 c(4t-2t^2)\,dt + \int_2^{+\infty} 0\cdot dt$$

$$= c\int_0^2 (4t-2t^2)\,dt \implies$$

$$c = \frac{1}{4\int_0^2 t\,dt - 2\int_0^2 t^2\,dt} = \frac{1}{4\left.\frac{t^2}{2}\right|_0^2 - 2\left.\frac{t^3}{3}\right|_0^2}$$

$$= \frac{1}{4\cdot\frac{4}{2} - 2\cdot\frac{8}{3}} = \frac{1}{8-\frac{2}{3}\cdot 8} = \frac{1}{8/3} = \frac{3}{8}.$$

(b) Determinare $\mathbb{P}(X>1)$.

$$\mathbb{P}(X>1) = \int_1^{+\infty} \underbrace{f_X(t)\,dt}$$

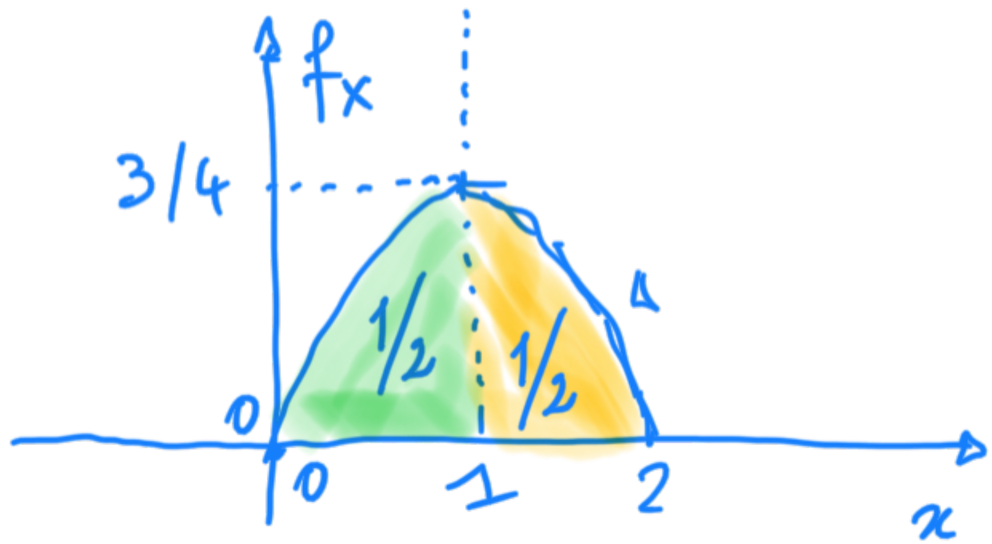
$$= \int_1^2 f_X(t)\,dt + \int_2^{+\infty} f_X(t)\,dt = \int_1^2 \frac{3}{8}\cdot(4t-2t^2)\,dt$$

$$= \frac{3}{8}\left[4\int_1^2 t\,dt - 2\int_1^2 t^2\,dt\right] = \frac{3}{8}\left(2\,t^2\Big|_1^2 - \frac{2}{3}\,t^3\Big|_1^2\right)$$

$$= \frac{3}{8}\left[2\,(4-1) - \frac{2}{3}\,(8-1)\right] = \frac{3}{8}\left(6-\frac{14}{3}\right) = \frac{3}{8}\,\frac{4}{3}$$

$$= \frac{1}{2}.$$

Il risultato ottenuto ha la seguente giustificazione geometrica. Nel supporto $(0,2)$ la f_X ha per grafico un ramo di parabola passante per i punti $(0,0)$ e $(2,0)$ con il vertice in $(1, 3/4)$. Pertanto il grafico è simmetrico rispetto alla retta $x=1$. L'area sotto il grafico di f_X vale $1/2$ prima di 1 e dopo 1.



$\square$