COMBINAZIONI LINEARI DI NORMALI INDIPENDENTI

In generale la somma di due v.a. Normali non ha una distribuzione Normale (si possono costruire controesempi; un po' difficile con quel che vediamo in queste lezioni).

Però, se le v.a. Normali sono indipendenti, la loro somma è Normale. Anzi questa proprietà vale per una qualsiasi combinazione lineare (la somma è una particolare combinazione lineare).

A tal proposito, per contemplare il caso in cui i coefficienti della combinazione lineare sono uguali a zero, è utile considerare le variabili aleatorie costanti (che quindi sono discrete) come particolari v.a. Normali tali che:

- la media è la costante stessa;
- la varianza è zero.

Quindi, se si ha $P(X=c)=1$ per qualche $c \in \mathbb{R}$, possiamo dire che $X \sim N(c,0)$.

Dopo queste premesse possiamo enunciare il seguente risultato.

PROPOSIZIONE (con dimostrazione parziale).

Siano $a_1, \dots, a_n \in \mathbb{R}$ e $X_1, \dots, X_n$ v.a. tali che

$$\left\{\begin{array}{l} X_1 \sim N(\mu_1, \sigma_1^2) \\ \vdots \\ X_n \sim N(\mu_n, \sigma_n^2) \end{array}\right. \quad \text{indipendenti.}$$

Allora la v.a. $a_1 X_1 + \dots + a_n X_n \sim N(a_1\mu_1 + \dots + a_n\mu_n, a_1^2\sigma_1^2 + \dots + a_n^2\sigma_n^2)$.

DIMOSTRAZIONE (PARZIALE)

Non dimostreremo che $a_1 X_1 + \dots + a_n X_n$ ha distribuzione Normale.

Però possiamo calcolare media e varianza di $a_1 X_1 + \dots + a_n X_n$ a partire dalle proprietà di media e varianza viste precedentemente:

$$\mathbb{E}[a_1 X_1 + \dots + a_n X_n] \overset{\text{linearità}}{=} a_1 \underbrace{\mathbb{E}[X_1]}_{=\mu_1} + \dots + a_n \underbrace{\mathbb{E}[X_n]}_{=\mu_n} = a_1 \mu_1 + \dots + a_n \mu_n . \qquad (ok)$$

$$\mathrm{Var}[a_1 X_1 + \dots + a_n X_n] = \underbrace{\mathrm{Var}[a_1 X_1]}_{= a_1^2 \mathrm{Var}[X_1] = a_1^2 \sigma_1^2} + \dots + \underbrace{\mathrm{Var}[a_n X_n]}_{= a_n^2 \mathrm{Var}[X_n] = a_n^2 \sigma_n^2} = a_1^2 \sigma_1^2 + \dots + a_n^2 \sigma_n^2 . \qquad (ok)$$

essendo $X_1, \dots, X_n$ indipendenti, si può dimostrare che lo sono anche $a_1 X_1, \dots, a_n X_n$; quindi...

proprietà della varianza

ESERCIZIO

Siano $X_1 \sim N(\mu_1 = 1, \sigma_1^2 = 1)$ e $X_2 \sim N(\mu_2 = 5, \sigma_2^2 = 2)$ v.a. indipendenti.

Calcolare $P(2X_1 - 3X_2 \geq 0)$.

RISPOSTA

Si ha che $2X_1 - 3X_2 \sim N(\underbrace{2 \cdot 1 + (-3) \cdot 5}_{= 2 - 15 = \underbrace{-13}_{=\mu}}, \underbrace{2^2 \cdot 1 + (-3)^2 \cdot 2}_{4 + 18 = \underbrace{22}_{=\sigma^2}})$.

$$\boxed{(2X_1 - 3X_2)^* \sim N(0,1)}$$

Allora

$$P(2X_1 - 3X_2 \geq 0) = P\left(\frac{2X_1 - 3X_2 - \mu}{\sigma} \geq \underbrace{\frac{0 - \mu}{\sigma}}_{= \frac{0 - (-13)}{\sqrt{22}}}\right) = P\left((2X_1 - 3X_2)^* \geq \frac{13}{\sqrt{22}}\right) = 1 - \Phi\left(\frac{13}{\sqrt{22}}\right).$$

ESERCIZIO

Siano X_1 e X_2 due v.a. Normali standard indipendenti.

Calcolare $P(X_1-X_2 \geq 0)$ e $P(X_1-X_2 > \frac{1}{2})$.

SVOLGIMENTO

Si ha che $X_1 - X_2 \sim N(\underbrace{1\cdot 0+(-1)\cdot 0}_{=0-0=0}, \underbrace{1^2\cdot 1+(-1)^2\cdot 1}_{=1+1=2})$.

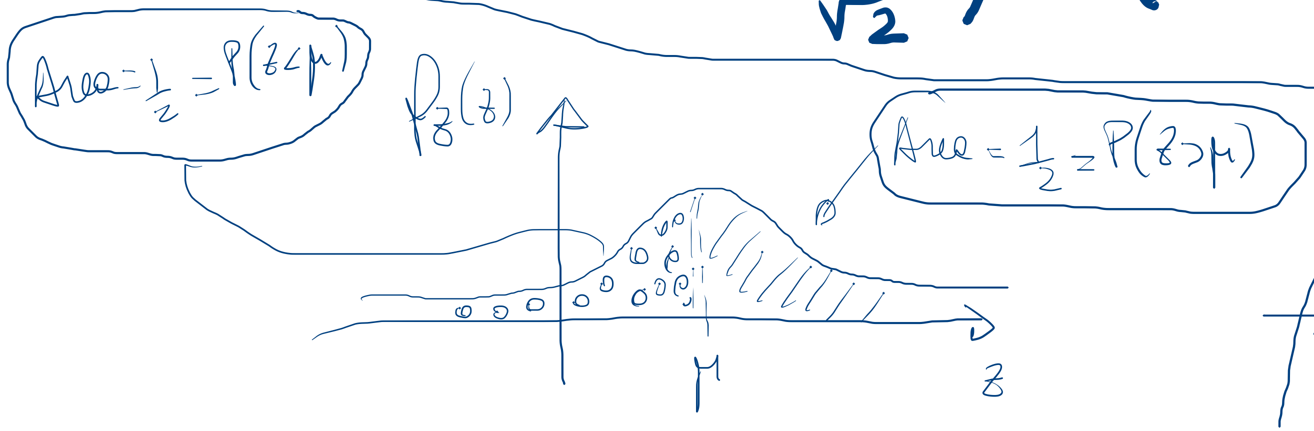
Allora

$$P(X_1-X_2>0) = P\left((X_1-X_2)^* > \frac{0-0}{\sqrt{2}}\right) = P((X_1-X_2)^*>0) = \frac{1}{2}$$

$(X_1-X_2)^* \sim N(0,1)$

$$P\left(X_1-X_2>\frac{1}{2}\right) = P\left((X_1-X_2)^* > \frac{\frac{1}{2}-0}{\sqrt{2}}\right) = P\left((X_1-X_2)^* > \frac{1}{2\sqrt{2}}\right) = 1 - \Phi\left(\frac{1}{2\sqrt{2}}\right).$$

$(X_1-X_2)^* \sim N(0,1)$



OSS. Il risultato ottenuto nella risposta alla prima domanda è esatto (cioè non approssimato) e si poteva dedurre senza usare le Tavole.

Infatti, data $Z \sim N(\mu, \sigma^2)$, in generale (qualunque sia σ) si ha

$P(Z>\mu) = \frac{1}{2}$

(e anche $P(Z<\mu) = \frac{1}{2}$).

Nel nostro caso si ha $Z = X_1 - X_2$ e $\mu = 0$.

ESERCIZIO

Sia X una v.a. con distribuzione di Cauchy, cioè con densità continua

$$f_X(x) = \frac{1}{\pi(1+x^2)}.$$

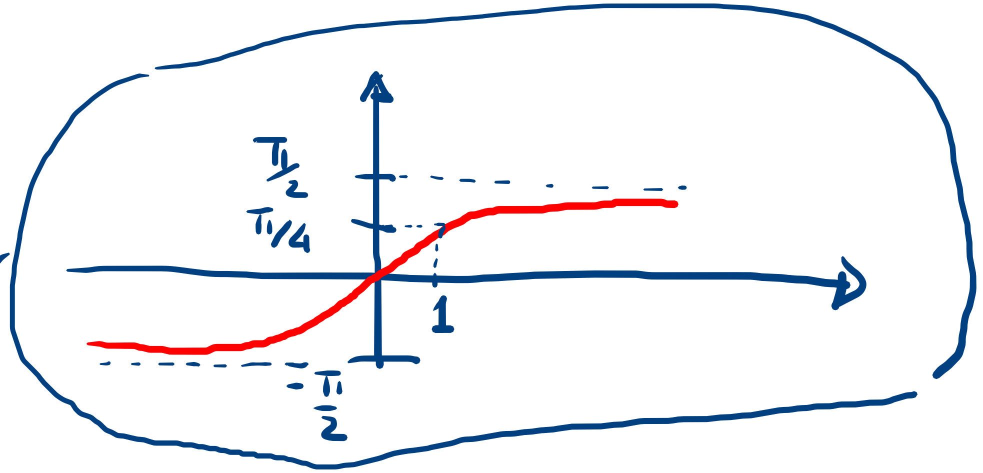
1) Calcolare $P(X>1)$.

2) Calcolare $P(X<0 \mid X>-1)$.

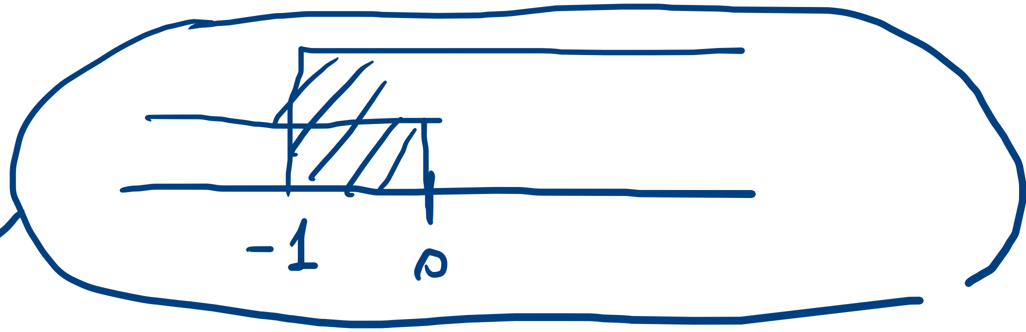
3) Verificare che X <u>non</u> ha speranza matematica finita.

SVOLGIMENTO

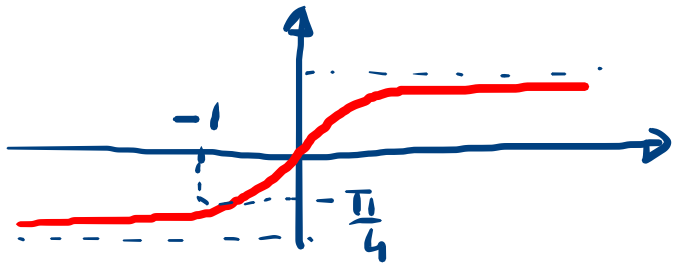
1) $$P(X>1) = \int_1^{\infty} \frac{1}{\pi(1+x^2)}\,dx = \frac{1}{\pi}\Big[\arctan x\Big]_{x=1}^{x=\infty} = \frac{1}{\pi}\left(\frac{\pi}{2} - \frac{\pi}{4}\right) = \frac{1}{2} - \frac{1}{4} = \frac{2-1}{4} = \frac{1}{4}.$$



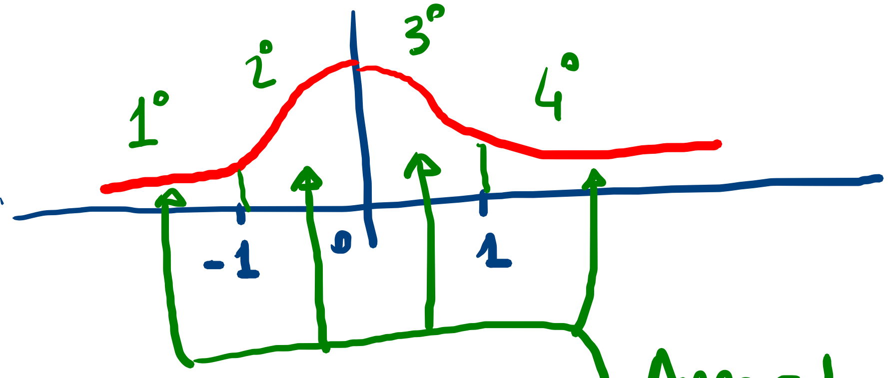
2) $$P(X<0 \mid X>-1) = \frac{P(\{X<0\} \cap \{X>-1\})}{P(X>-1)} = \frac{P(-1<X<0)}{P(X>-1)} =$$



$$= \frac{\cancel{\frac{1}{\pi}} \int_{-1}^{0} \frac{1}{1+x^2}\,dx}{\cancel{\frac{1}{\pi}} \int_{-1}^{\infty} \frac{1}{1+x^2}\,dx} = \frac{\left[\operatorname{arctg} x\right]_{x=-1}^{x=0}}{\left[\operatorname{arctg} x\right]_{x=-1}^{x=\infty}} = \frac{0-\left(-\frac{\pi}{4}\right)}{\frac{\pi}{2}-\left(-\frac{\pi}{4}\right)} = \frac{\frac{\pi}{4}}{\frac{\pi}{2}+\frac{\pi}{4}} = \frac{\frac{1}{4}}{\frac{1}{2}+\frac{1}{4}} = \frac{\frac{1}{\cancel{4}}}{\frac{2+1}{\cancel{4}}} = \frac{1}{3}$$



COMMENTO su 1ª e 2ª domanda



Area $= \frac{1}{4}$ per ciascuno dei 4 settori

- Nella 1ª domanda si chiede l'area del 4° settore
- Nella 2ª domanda si chiede l'area dei 1° e 2° settore sapendo di stare negli ultimi 3; quindi l'area del 2° settore diviso la probabilità di 2°, 3°, 4° settore e quindi non sorprende che si ottenga $\frac{1}{3}$.

3) Si deve verificare che

$$\int_{-\infty}^{\infty} |x| \frac{1}{\pi(1+x^2)}\, dx = +\infty .$$

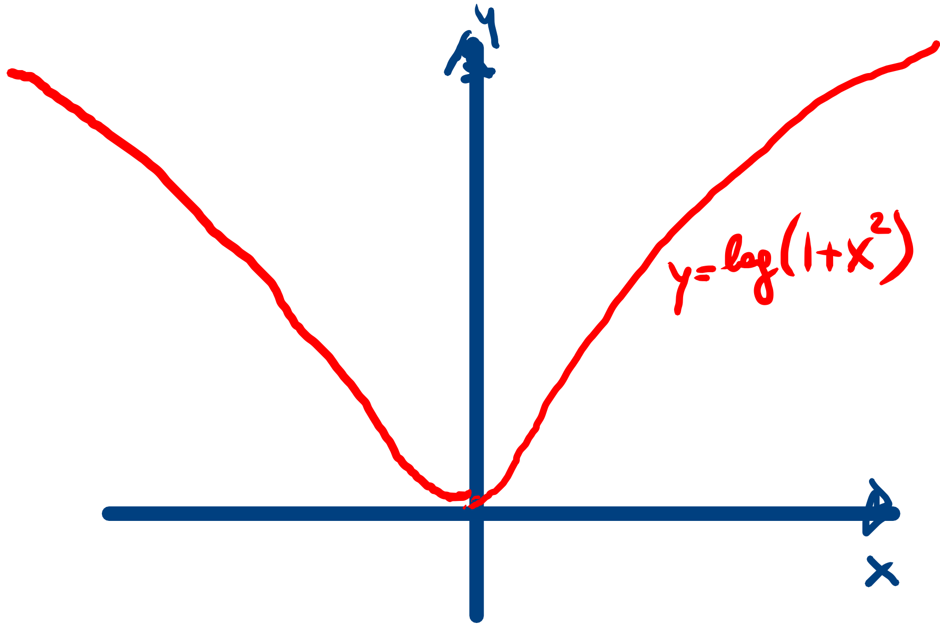
FUNZIONE INTEGRANDA È "PARI"
INTERVALLO DI INTEGRAZIONE $(-\infty,\infty)$ SIMMETRICO

Infatti si ha

$$\int_{-\infty}^{\infty} |x| \frac{1}{\pi(1+x^2)}\, dx = \frac{2}{\pi}\int_0^{\infty} x \frac{1}{1+x^2}\, dx = \frac{1}{\pi}\int_0^{\infty} \frac{2x}{1+x^2}\, dx =$$

$$= \frac{1}{\pi}\Big[\log(1+x^2)\Big]_{x=0}^{x=\infty} = \frac{1}{\pi}\Big(\overbrace{\log(1+\infty^2)}^{=\infty} - \overbrace{\log 1}^{=0}\Big) = +\infty .$$

abuso di notazione; in effetti $\log(1+x^2)$ tende a $+\infty$ quando $x\to+\infty$ (e anche per $x\to-\infty$).



ESERCIZIO

Sia X una v.a. con densità continua $f_X(x) = c \cdot x \, 1_{(0, 9/2)}(x)$, per qualche $c > 0$.

1) Determinare il valore della costante c.

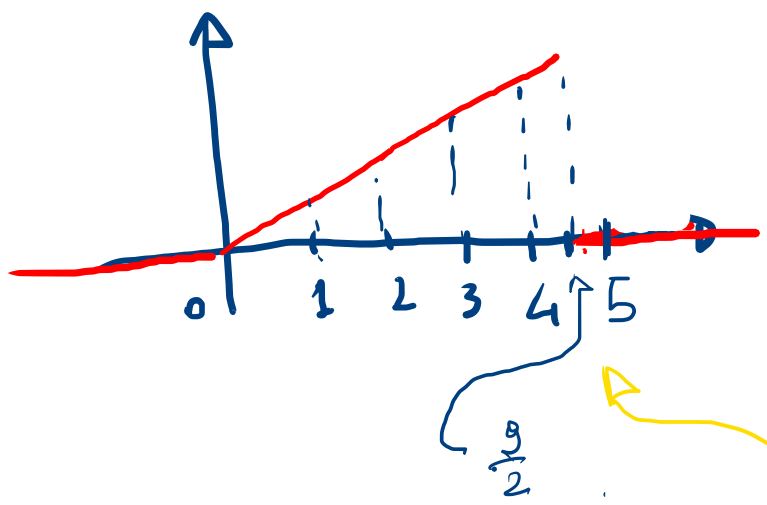
2) Trovare la densità discreta di $Y := [X]$.

SVOLGIMENTO

1) Imponiamo $\int_{-\infty}^{\infty} f_X(x)dx = 1$, e si ha $1 = c\int_0^{9/2} x\,dx = c\left[\frac{x^2}{2}\right]_{x=0}^{x=9/2} = \frac{c}{2}\left(\frac{9}{2}\right)^2 = c \cdot \frac{81}{8}$,

da cui segue $c = \frac{8}{81}$.

2)



In generale si ha

$$\{Y = k\} = \{k \leq X < k+1\} \quad \forall k \in \mathbb{Z}$$

I valori di interesse sono $k \in \{0, 1, 2, 3, 4\}$ perché $\frac{9}{2} = 4.5 \in (4, 5)$.

Allore

$$p_Y(k) = P(k \le X < k+1) = \int_k^{k+1} f_X(x)\,dx =$$

$$= \begin{cases} \text{per } k=0,1,2,3 & \frac{8}{81}\int_k^{k+1} x\,dx = \frac{8}{81}\left[\frac{x^2}{2}\right]_{x=k}^{x=k+1} = \frac{8}{81}\,\frac{(k+1)^2-k^2}{2} = \dots = \frac{8}{81}\,\frac{2k+1}{2} = \frac{4}{81}(2k+1) \\ \text{per } k=4 & \frac{8}{81}\int_4^{9/2} x\,dx = \frac{8}{81}\left[\frac{x^2}{2}\right]_{x=4}^{x=9/2} = \frac{8}{81}\,\frac{\frac{81}{4}-16}{2} = \dots = \frac{81-64}{81} = \frac{17}{81} \\ \text{altrimenti} & 0 \end{cases}$$

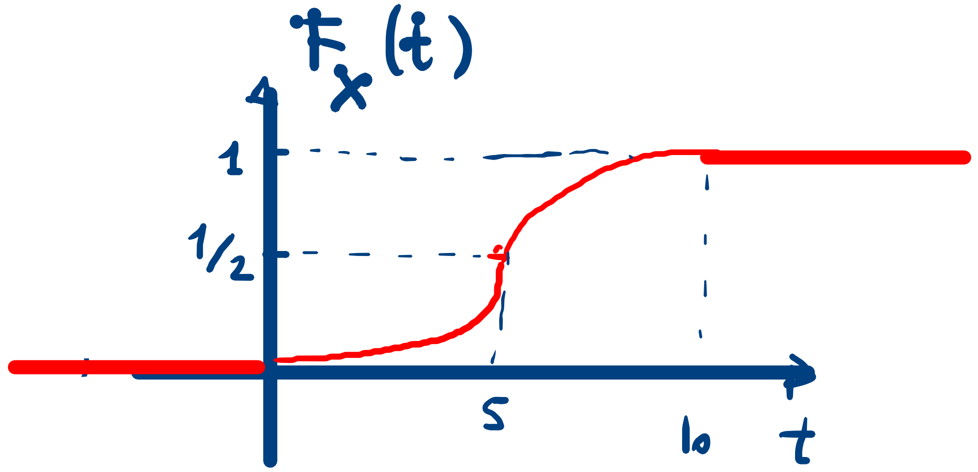
$$= \begin{cases} 4/81 & \text{per } k=0 \\ 12/81 & \text{per } k=1 \\ 20/81 & \text{per } k=2 \\ 28/81 & \text{per } k=3 \\ 17/81 & \text{per } k=4 \\ 0 & \text{altrimenti} \end{cases}$$

SOMMA = 1
OK

ESERCIZIO

Sia X una v.a. continua con la seguente funzione di distribuzione:

$$F_X(t) = \begin{cases} 0 & \text{per } t \leq 0 \\ t^2/50 & \text{per } 0 < t \leq 5 \\ -\frac{t^2}{50} + \frac{2}{5}t - 1 & \text{per } 5 < t \leq 10 \\ 1 & \text{per } t > 10. \end{cases}$$

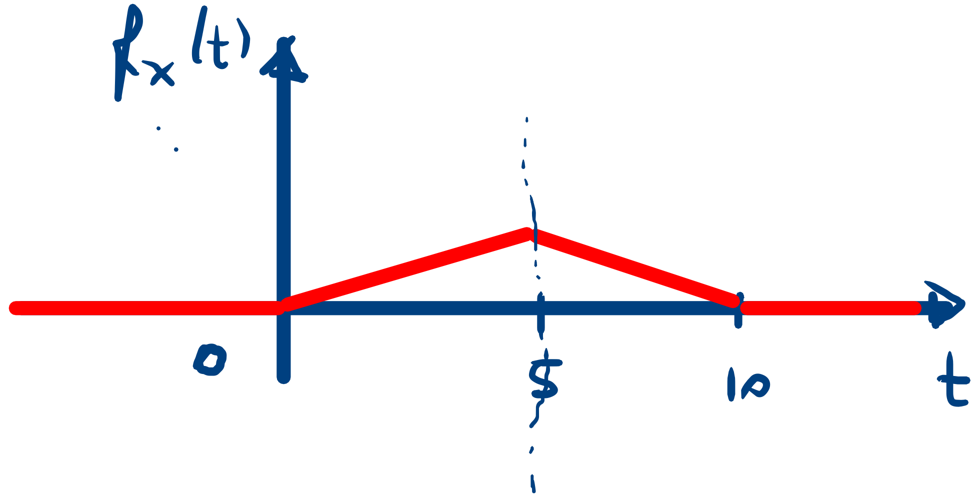


Trovare la densità continua di X e calcolare $\mathbb{E}[X]$.

SVOLGIMENTO

$$f_X(t) = \begin{cases} 0 & \text{per } t < 0 \text{ e } t > 10 \\ t/25 & \text{per } 0 < t < 5 \\ -\frac{t}{25} + \frac{2}{5} & \text{per } 5 < t < 10 \end{cases}$$

OSS. La densità $f_X(t)$ è simmetrica rispetto a $t = 5$. Allora, poiché siamo nella condizione perché vale (*), ci aspettiamo che $\mathbb{E}[X] = 5$.



Calcoliamo $\mathbb{E}[X]$. Si ha

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\,dx = \underbrace{\int_{-\infty}^{0} x\cdot 0\,dx}_{=0} + \int_{0}^{5} x\cdot\frac{x}{25}\,dx + \int_{5}^{10} x\left(-\frac{x}{25}+\frac{2}{5}\right)dx + \underbrace{\int_{10}^{\infty} x\cdot 0\,dx}_{=0} =$$

$$= \frac{1}{25}\int_{0}^{5} x^2\,dx + \int_{5}^{10} x\left(\frac{-x+10}{25}\right)dx = \frac{1}{25}\left[\frac{x^3}{3}\right]_{x=0}^{x=5} + \frac{1}{25}\underbrace{\int_{5}^{10} -x^2+10x\,dx}_{=\left[-\frac{x^3}{3}+10\frac{x^2}{2}\right]_{x=5}^{x=10}} =$$

$$= \frac{1}{25}\,\frac{5^3-0^3}{3} + \frac{1}{25}\left(-\frac{10^3}{3}+\frac{10^3}{2}-\left(-\frac{5^3}{3}+\frac{10\cdot 5^2}{2}\right)\right) =$$

$$= \frac{1}{\cancel{25}}\,\frac{\cancel{125}^{\,5}}{3} + \frac{1}{25}\left(-\frac{1000}{3}+\frac{1000}{2}-\left(-\frac{125}{3}+\frac{250}{2}\right)\right) =$$

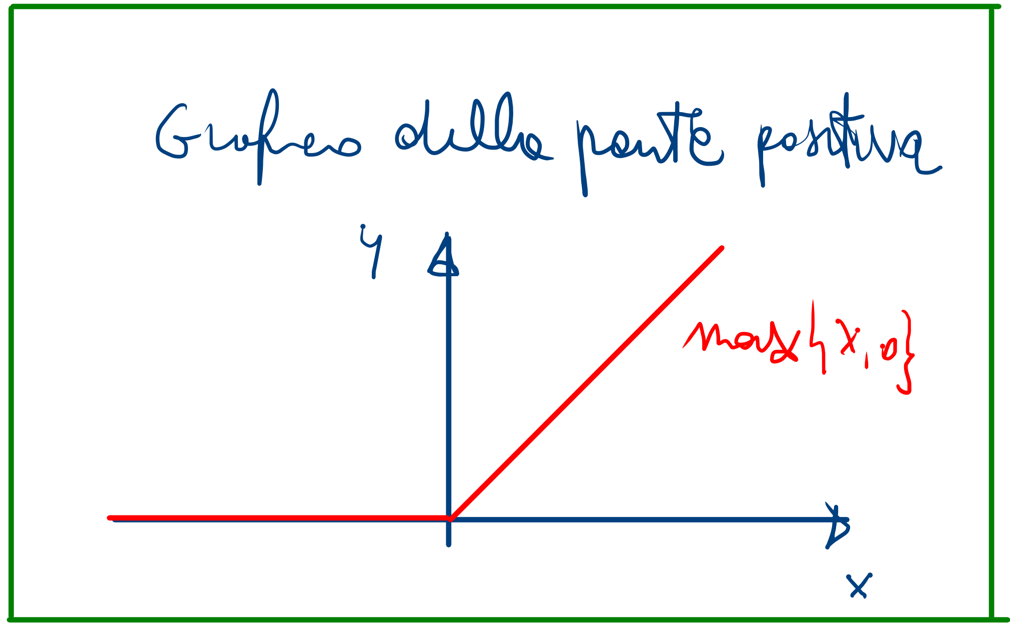
$$= \frac{5}{3} + \frac{1}{25}\left(-\frac{875}{3}+\frac{750}{2}\right) = \frac{5}{3} - \frac{35}{3} + \frac{\cancel{30}^{\,15}}{\cancel{2}_{\,1}} = -\frac{\cancel{30}^{\,10}}{\cancel{3}} + 15 = -10+15 = 5$$

in accordo con quanto anticipato

ESERCIZIO (esempio di v.a. né discreta, né continua).

Sia $X \sim U(-1,1)$. Poi sia $Y = \max\{X, 0\}$ (quindi Y è la "parte positiva" di X).

Trovare la funzione di distribuzione di Y.



RISPOSTA

Si ha $P(Y \geq 0) = 1$ e quindi $F_Y(y) = \begin{cases} 0 & \text{se } y < 0 \\ (*) & \text{se } y \geq 0 \end{cases}$.

Poi si ha

$P(A) = P(A \cap B) + P(A \cap B^c)$ con $A = \{Y \leq y\}$ e $B = \{X \geq 0\}$

$$(*) = P(Y \leq y) = P(\{Y \leq y\} \cap \{X \geq 0\}) + P(\{Y \leq y\} \cap \{X < 0\}) =$$

$$= P(\{X \leq y\} \cap \{X \geq 0\}) + P(\underbrace{\{0 \leq y\}}_{=\Omega} \cap \{X < 0\}) =$$

perché si ha un evento deterministico sempre verificato

$$= P(0 \leq X \leq y) + P(X < 0) = P(X \leq y).$$

In effetti, per $y \geq 0$, si ha (unione disgiunta) $(-\infty, 0) \cup [0, y] = (-\infty, y]$

Quindi

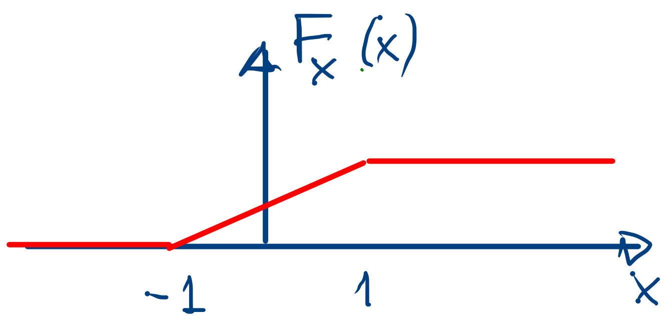
$$F_Y(y) = \begin{cases} 0 & \text{per } y<0 \\ P(X\leq y) & \text{per } y \geq 0 \end{cases} = \begin{cases} 0 & \text{per } y<0 \\ F_X(y) & \text{per } y\geq 0 \end{cases}$$

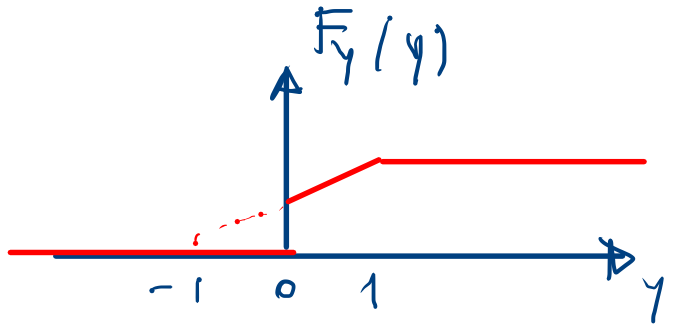
da cui segue

$$F_Y(y) = \begin{cases} 0 & \text{per } y<0 \\ \frac{y-(-1)}{1-(-1)} = \frac{y+1}{2} & \text{per } 0\leq y\leq 1 \\ 1 & \text{per } y>1. \end{cases}$$

OSS.

Graficamente si ha la seguente situazione





OSS.

- La v.a. Y <u>non</u> è continua perché F_Y non è una funzione continua (discontinuità in zero)
- La v.a. Y <u>non</u> è discreta perché ha il tratto lineare che corrisponde ad una "componente uniforme continua" su quell'intervallo.

ESERCIZIO

Una fabbrica ha 2 linee di produzione:

la 1ª produce elementi con tempo di funzionamento che ha distribuzione $Exp(\lambda_1)$;

la 2ª produce elementi con tempo di funzionamento che ha distribuzione $Exp(\lambda_2)$.

In proporzione gli elementi di 1ª e 2ª linea disponibili sono p e $1-p$.

Si sceglie un elemento a caso e lo si mette in funzione; in corrispondenza sia T la v.a. che indica il tempo di funzionamento di questo elemento.

1) Trovare la densità continua di T.

2) Calcolare $\mathbb{E}[T]$.

3) Calcolare $P(E_1 \mid T > s)$ (per $s > 0$) dove $E_1 = \{$l'elemento scelto proviene dalla 1ª linea$\}$.

SVOLGIMENTO

Sia $E_2 = E_1^c$. Allora

FORMULA PROB. TOTALI

1) $F_T(t) = P(T \le t) = P(T \le t \mid E_1) P(E_1) + P(T \le t \mid E_2) P(E_2)$ dove

$$P(T \le t \mid E_1) = \begin{cases} 1 - e^{-\lambda_1 t} & \text{per } t > 0 \\ 0 & \text{per } t \le 0 \end{cases}$$

$$P(T \le t \mid E_2) = \begin{cases} 1 - e^{-\lambda_2 t} & \text{per } t > 0 \\ 0 & \text{per } t \le 0 \end{cases}$$

$P(E_1) = p$

$P(E_2) = 1 - p$

Quindi

$$F_T(t) = \begin{cases} (1 - e^{-\lambda_1 t}) p + (1 - e^{-\lambda_2 t})(1-p) & \text{per } t > 0 \\ 0 \cdot p + 0 \cdot (1-p) = 0 & \text{per } t \le 0 \end{cases}$$

Quindi

$$f_T(t) = \left(\lambda_1 e^{-\lambda_1 t} p + \lambda_2 e^{-\lambda_2 t} (1-p)\right) 1_{(0,\infty)}(t)$$

OSS

Mistura delle densità delle due esponenziali con pesi p e $1-p$.

2) $$\mathbb{E}[T] = \int_0^{\infty} t \left\{ \lambda_1 e^{-\lambda_1 t} p + \lambda_2 e^{-\lambda_2 t} (1-p) \right\} dt =$$

$$= p \underbrace{\int_0^{\infty} t \lambda_1 e^{-\lambda_1 t} dt}_{= 1/\lambda_1} + (1-p) \underbrace{\int_0^{\infty} t \lambda_2 e^{-\lambda_2 t} dt}_{= 1/\lambda_2} = p \cdot \frac{1}{\lambda_1} + (1-p) \frac{1}{\lambda_2}$$

oss. Medie pesate delle medie delle due esponenziali.

3) $$P(E_1 | T > s) = \frac{P(T>s | E_1) P(E_1)}{P(T>s)} = \frac{(1-(1-e^{-\lambda_1 s})) \cdot p}{1 - F_T(s)} = \frac{p e^{-\lambda_1 s}}{1 - F_T(s)}$$

BAYES

dove $$1 - F_T(s) = 1 - \left((1-e^{-\lambda_1 s}) p + (1-e^{-\lambda_2 s})(1-p) \right) = \cancel{1} - \left(\cancel{p} - p e^{-\lambda_1 s} + \cancel{1-p} - (1-p) e^{-\lambda_2 s} \right)$$

$$= p e^{-\lambda_1 s} + (1-p) e^{-\lambda_2 s}$$

Quindi $P(E_1|T>s) = \dfrac{pe^{-\lambda_1 s}}{pe^{-\lambda_1 s} + (1-p)e^{-\lambda_2 s}} = \dfrac{1}{1 + \frac{1-p}{p} e^{(\lambda_1-\lambda_2)s}}$.

OSS

- Se $\lambda_1 = \lambda_2$ (linee di produzione della stessa qualità) si ha

$$P(E_1|T>s) = \frac{1}{1+\frac{1-p}{p}} = \frac{p}{\cancel{p}+1-\cancel{p}} = p = P(E_1) \quad \forall s>0,$$

OSS. E_1 e $\{T>s\}$ sono indipendenti

- Se $\lambda_1 < \lambda_2$ (cioè $\frac{1}{\lambda_1} > \frac{1}{\lambda_2}$, 1ª linea di qualità migliore), e quindi $\lambda_1 - \lambda_2 < 0$, si ha $\lim_{s\to\infty} P(E_1|T>s) = \frac{1}{1+0} = 1$,
- Se $\lambda_1 > \lambda_2$ (cioè $\frac{1}{\lambda_1} < \frac{1}{\lambda_2}$, 2ª linea di qualità migliore), e quindi $\lambda_1 - \lambda_2 > 0$, si ha $\lim_{s\to\infty} P(E_1|T>s) = \frac{1}{1+\infty} = 0$.

In conclusione, se consideriamo anche $E_2 = E_1^c$, abbiamo la seguente situazione:

$$P(E_1 \mid T>s) + P(E_2 \mid T>s) = 1 \qquad \text{per ogni } s>0;$$

le considerazioni fatte per $P(E_1 \mid T>s)$ possono essere fatte anche per $P(E_2 \mid T>s)$.

Poi abbiamo due casi:

1) se $\lambda_1 = \lambda_2$, si ha $P(E_1 \mid T>s) = p$ $\quad P(E_2 \mid T>s) = 1-p$ costanti rispetto a s

e quindi E_1 e $\{T>s\}$ sono indipendenti (anche E_2 e $\{T>s\}$ sono indipendenti).

2) se $\lambda_1 \neq \lambda_2$, $P(E_1 \mid T>s)$ e $P(E_2 \mid T>s)$ non sono costanti rispetto a s, e non si ha l'indipendenza citata sopra. In generale, per $s \to \infty$, una delle due prob. condizionate tende a 1 e l'altra tende a zero. Quella che tende a 1 corrisponde alla linea di produzione di qualità migliore (il minimo tra λ_1 e λ_2).