Problema 1.1 (11 punti). Un server riceve dati a ritmo irregolare. Ogni ora la quantità di dati ricevuti è una variabile aleatoria. Si suppone che tutte queste variabili aleatorie siano i.i.d. (indipendenti e identicamente distribuite).

(7 punti) Supponendo che queste variabili aleatorie abbiano (in Mb) distribuzione binomiale di parametri $n = 150$ e $p = 10\%$, quanto vale approssimativamente la probabilità che in 24 ore i dati ricevuti siano più di 320 Mb?

(2 punti) Si risponda nuovamente nell'ipotesi che le variabili aleatorie abbiano invece distribuzione uniforme continua tra 10 e 20 Mb.

(2 punti) Supponiamo infine che le variabili aleatorie abbiano distribuzione esponenziale di media 15, e sia T il tempo prima che la quantità di dati ricevuti raggiunga i 1000 Mb. Si dica quanto vale (anche approssimativamente) la probabilità che T sia inferiore a 3 giorni.

a) $X_1, X_2, \dots$ dati ricevuti ora per ora

$X_i \sim bin(n,p)$ $\quad n = 150 \quad p = 0{,}1$

indipendenti

$P(X_1+X_2+\dots+X_{24} > 320)$

$$X_i \,\dot\sim\, \mathcal{N}\big(\underset{15}{np},\ \underset{13{,}5}{np(1-p)}\big)$$

$$X_1+X_2+\dots+X_{24} \,\dot\sim\, \mathcal{N}\big(\underset{360}{24\cdot 15};\ \underset{324}{24\cdot 13{,}5}\big)$$

in alternativa, per riproducibilità

$$X_1+\dots+X_{24} \sim bin(24n, p) \,\dot\sim\, \mathcal{N}\big(\underset{360}{24np};\ \underset{324}{24np(1-p)}\big)$$

$$P(X_1+\dots+X_{24} > 320) \cong 1-\Phi\left(\frac{320-360}{\sqrt{324}}\right) = 1-\Phi(-2{,}22) \approx 0{,}9868 \approx 98\%$$

b) $X_i \sim unif(10;20) \qquad E(X_i) = 15 \qquad Var(X_i) = \frac{10^2}{12} \approx 8{,}33$

$$X_1+X_2+\dots+X_{24} \,\dot\sim\, \mathcal{N}(360;\ \underbrace{24\cdot 8{,}33}_{200})$$

$$P(X_1+\dots+X_{24} > 326) \approx 1-\Phi\left(\frac{320-360}{\sqrt{200}}\right) \approx 1-\Phi(-2{,}83) \approx 99\%$$

c) $X_i \sim expo(\lambda) \qquad \lambda = \frac{1}{15}$

$$P(T<3) = P(\underbrace{X_1+X_2+\dots+X_{72}} > 1000) \approx 1-\Phi\left(\frac{1000-1080}{127{,}3}\right) \approx 1-\Phi(-0{,}629) \approx 74\%$$

$\sim gamma(\alpha,\beta) \quad \alpha = 72 \quad \beta = 15$

$\dot\sim\, \mathcal{N}(\alpha\beta;\ \alpha\beta^2) \,\dot\sim\, \mathcal{N}(1080;\ (127{,}3)^2)$

Problema 1.2 (12 punti). Sia X una variabile aleatoria continua con densità

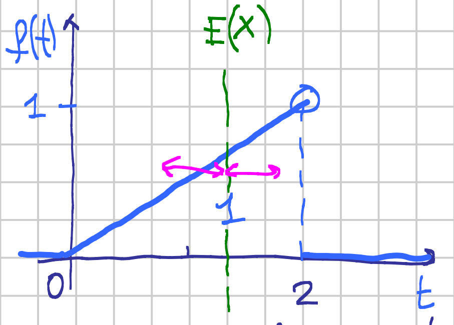
$$f(t) = ct, \qquad 0 \le t \le 2$$

(7 punti) Si determinino c, la media, la varianza e la mediana di X. Si determini la densità di $Z := X/2$. Si tratta di una distribuzione appartenente ad una classe standard?

(3 punti) Sia Y una variabile aleatoria indipendente da X e con la stessa distribuzione. Si determini la probabilità che

$$\frac{1}{2} \le \frac{X}{Y} \le 2$$

(2 punti) Si determini la legge di $X + Y$.



$c = \frac{1}{2}$ graficamente

$$1 = \int_0^2 ct\,dt = c\left[\frac{t^2}{2}\right]_0^2 = c \cdot 2$$

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

$$f(t) = \frac{t}{2} \quad 0 \le t \le 2$$

$$E(X) = \int_0^2 t f(t)\,dt = c\int_0^2 t^2\,dt = c\left[\frac{t^3}{3}\right]_0^2 = \frac{4}{3}$$

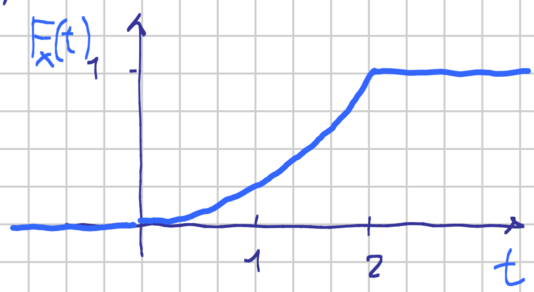
$$E(X^2) = c\int_0^2 t^3\,dt = \dots = c \cdot 4 = 2$$

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{2}{9}$$

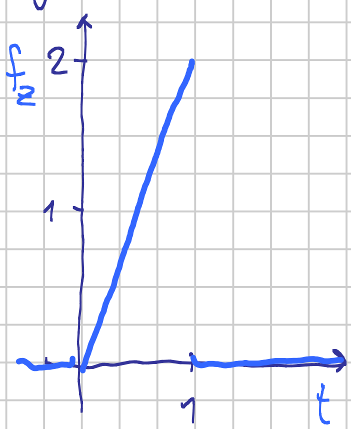
$$\sqrt{\mathrm{Var}(X)} \approx 0{,}47$$

$$F_X(t) := \int_0^t f(s)\,ds = c\,\frac{t^2}{2} = \begin{cases} \frac{t^2}{4} & 0 \le t \le 2 \\ 0 & t < 0 \\ 1 & t > 2 \end{cases}$$

$F_X(2) = 1$ ✓



graficamente $\quad Z = X/2$



$$F_Z(t) = P(Z \le t) = P\left(\frac{X}{2} \le t\right) = P(X \le 2t) = F_X(2t)$$

derivo rispetto a t

$$f_Z(t) = \frac{d}{dt}F_Z(t) = F_X'(2t)\frac{d}{dt}(2t) = f_X(2t) \cdot 2 = c \cdot (2t) \cdot 2$$

$0 \le 2t \le 2$

$0 \le t \le 1$

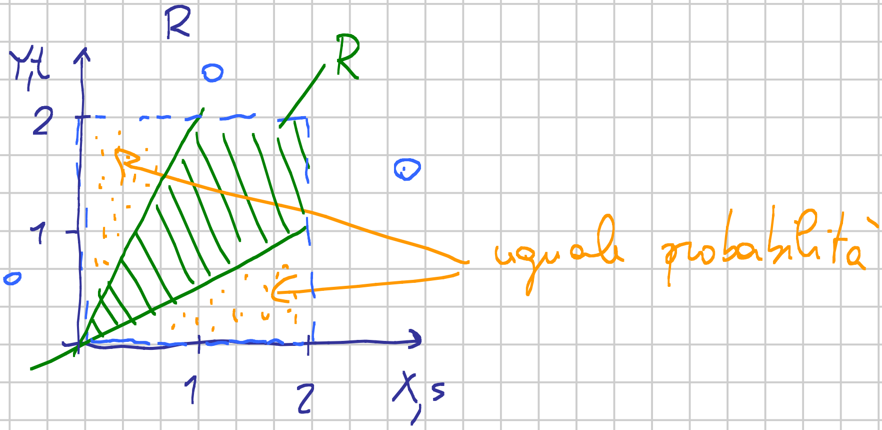
$$= \begin{cases} 2t & 0 \le t \le 1 \\ 0 & \text{altrimenti} \end{cases}$$

$\star$ è una beta $\quad \alpha = 2 \quad \beta = 1$

b) $P\left(\frac{1}{2} \le \frac{X}{Y} \le 2\right) = \iint_R f_{X,Y}(s,t)\,ds\,dt = \iint_R f(s)f(t)\,ds\,dt$

$\frac{1}{2} \le \frac{X}{Y} \iff Y \le 2X$

$\frac{X}{Y} \le 2 \iff X \le 2Y$



$$P(X \le 2Y) = \int_0^2\int_0^{s/2} c^2 st\,dt\,ds = c^2\int_0^2 s\int_0^{s/2} t\,dt\,ds = c^2\int_0^2 s\cdot\frac{s^2}{8}\,ds = c^2\cdot\frac{1}{8}\cdot\left[\frac{s^4}{4}\right]_0^2 = \frac{c^2}{2} = \frac{1}{8}$$

$P(Y \le 2X) = \frac{1}{8}$ per simmetria

$$P\left(\frac{1}{2} \le \frac{X}{Y} \le 2\right) = 1 - 2\cdot\frac{1}{8} = \frac{3}{4}$$

c) $S = X + Y$

$$f_S(t) = \int_{\mathbb{R}} f_X(s) f_Y(t-s)\,ds = \int_{0 \vee t-2}^{2 \wedge t} c\cdot s\cdot c\cdot(t-s)\,ds = \begin{cases} \int_0^t c^2 s(t-s)\,ds & 0 \le t \le 2 \\ \int_{t-2}^2 c^2 s(t-s)\,ds & 2 \le t \le 4 \end{cases}$$

$0 \le s \le 2$ $\qquad$ $0 \le t-s \le 2$

$t-2 \le s \le t$

$\cancel{\ge 0}$ $\qquad$ ≤ 2

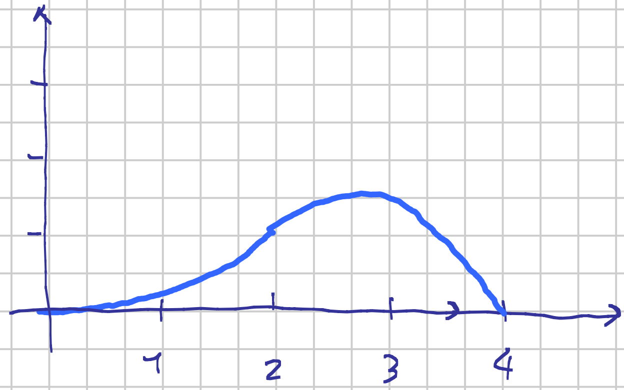
$$= \begin{cases} c^2 t\,\frac{t^2}{2} - c^2\frac{t^3}{3} = c^2\cdot\frac{t^3}{6} = \frac{t^3}{24} \\ c^2 t\left(2 - \frac{(t-2)^2}{2}\right) - c^2\left(\frac{8}{3} - \frac{(t-2)^3}{3}\right) = c^2 t\cdot\frac{1}{2}(4t - t^2) - c^2\cdot\frac{1}{3}(16 - 12t + 6t^2 - t^3) \end{cases}$$

$$= \frac{1}{2}t^2 - \frac{1}{8}t^3 - \frac{4}{3} + t - \frac{1}{2}t^2 + \frac{1}{12}t^3 = -\frac{1}{24}t^3 + t - \frac{4}{3}$$

$f_S(0) = 0$ ✓ $\qquad$ $f_S(2) = \frac{1}{3} \overset{?}{=} \frac{1}{3}$ ✓

$f_S(4) = 0$ ✓

$$f_S(t) = \begin{cases} \frac{1}{24} t^3 & 0 \le t \le 2 \\ t - \frac{4}{3} - \frac{1}{24} t^3 & 2 \le t \le 4 \end{cases}$$



Problema 1.3 (11 punti)**.** Si confrontano due programmi di intelligenza artificiale che si scontrano in un qualche gioco che non prevede il pareggio.

(7 punti) Se si effettuano 400 scontri, quanti deve vincerne come minimo uno dei due perché si possa affermare che sia più forte dell'altro al 5% di significatività? Se finisce 230 a 170, quanto vale il p-dei-dati?

(2 punti) Qual è la potenza del test se la probabilità di vittoria del primo è del 55%?

(2 punti) Quanti scontri bisogna prevedere come minimo se si vuole per lo stesso valore una potenza dell'80%?

a) A, B due programmi

$\alpha = 5\%$ signif

A è meglio di B $\to H_1$

p la prob che vinca A in uno scontro $\quad p_0 = \frac{1}{2}$

$H_0 : p \le \frac{1}{2} \qquad H_1 : p > \frac{1}{2}$

X : # di vincite di A

$X \sim \text{bin}(n, p) \qquad n = 400$

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \stackrel{\sim}{\sim} \mathcal{N}(0,1) \qquad Z := \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \qquad RA_Z = (-\infty ; q)$$

$$q = \Phi^{-1}(1-\alpha) \approx 1{,}645$$

$RA_X = ?$ $\quad Z \le q \iff \hat{p} \le p_0 + q\sqrt{p_0(1-p_0)/n} \iff X \le np_0 + q\sqrt{np_0(1-p_0)}$

$\iff X \le 200 + 16{,}45 \qquad RA_X = [0, 1, \ldots, 216]$

Deve essere $X \ge 217$ per dire H_1

$\alpha^* = ?$ $\quad Z \le \Phi^{-1}(1-\alpha) \iff \Phi(Z) \le 1 - \alpha \iff \alpha \le 1 - \Phi(Z) =: \alpha^*$

$X = 230 \qquad \hat{p} = 0{,}575 \qquad Z = 3 \qquad \Phi(Z) = 0{,}9987 \qquad \alpha^* \le 1\%$

b) $p = 0{,}55 \qquad$ potenza $= P(\text{dire } H_1) = P(X \ge 217) \approx 1 - \Phi\left(\frac{216{,}5 - 220}{9{,}95}\right)$

$np = 220 \quad \sqrt{np(1-p)} \approx 9{,}95 \qquad \text{bin}(n,p) \qquad \approx 1 - \Phi(-0{,}352) \approx 60\%$

c) potenza = 80% $p = 0,55$

$$0,8 = P(Z \geq q) = P\left(\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}} \geq q\right) = P\left(\underbrace{\frac{\hat{p}-p}{\sqrt{p_0(1-p_0)/n}}}_{\mathcal{N}(0,1)} + \frac{p-p_0}{\sqrt{p_0(1-p_0)/n}} \geq q\right)$$

q: 1,645

$$= 1 - \Phi\left(q - \frac{p-p_0}{\sqrt{p_0(1-p_0)/n}}\right)$$

$$\Phi^{-1}(0,8) = 0,8416$$
$$\Phi^{-1}(0,2) = -0,8416$$

$$q - \frac{p-p_0}{\sqrt{p_0(1-p_0)/n}} \approx -0,8416 = -q_1$$

$$\frac{p-p_0}{\sqrt{p_0(1-p_0)}} \cdot \sqrt{n} \approx q + q_1 \qquad \sqrt{n} \approx (q+q_1)\frac{\sqrt{p_0(1-p_0)}}{p-p_0} \approx 24,87$$

$$n \approx 618$$

Problema 1.4 (12 punti)**.** Si analizzano i consumi di un'auto, misurando esattamente il contenuto del serbatoio ogni 20 km.

km	0	20	40	60	80	100	120
litri	30	28.9	28.3	26.65	24.8	23.85	22.8

(7 punti) Si stimi puntualmente e al 95% di confidenza il consumo medio: 1) in litri ogni 20 km, 2) in litri per 100 km e 3) in km/litro.

(2 punti) Si stimi la deviazione standard dei litri consumati ogni 20 km al 90% di confidenza con un intervallo del tipo $[z, \infty)$.

(3 punti) Si ripeta il punto precedente per il consumo in litri per 100 km.

$n = 6$

$X_1 = 1,1$
$X_2 = 0,6$
$X_3 = 1,65$
$X_4 = 1,85$
$X_5 = 0,95$
$X_6 = 1,05$

$X_1 + \ldots + X_6 = 7,2 \qquad \bar{X} = 1,2$ l/20km

$5 \cdot 1,2 = 6$ l/100km

$\frac{100}{6} \approx 16,7$ km/l

$$S^2 = \frac{1}{n-1}\left(\sum X_i^2 - n\bar{X}^2\right) = 0,216 \qquad S \approx 0,465$$

$\sum X_i^2$: 9,72

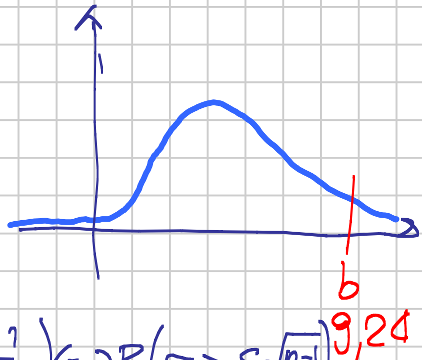
$$X_i \sim \mathcal{N}(\mu, \sigma^2)$$

$$\mu \in \bar{X} \pm q\underbrace{\frac{S}{\sqrt{n}}}_{0,488} \qquad q = F^{-1}_{t(5)}\left(1 - \frac{\alpha}{2}\right) \cong 2,57$$

$$\mu \in 1,2 \pm 0,5 = [0,7 ; 1,7]$$

Per 100 km : $[3,5 \,;\, 8,5]$

Per i km/l : $[11,8 \,;\, 28,6]$



b) $S \approx \sigma$ $\qquad \frac{S^2}{\sigma^2}(n-1) \sim \chi^2(n-1)$

$b = F^{-1}_{\chi^2(5)}(1-\alpha)$

$1-\alpha = P\left(\frac{S^2}{\sigma^2}(n-1) \le b\right) \iff P\left(\sigma^2 \ge S^2 \frac{n-1}{b}\right) \iff P\left(\sigma \ge S\sqrt{\frac{n-1}{b}}\right)$

$\sigma \in [z, +\infty)$ $\qquad z = S\sqrt{\frac{n-1}{b}} \approx 0,342$

$\sigma \approx 0,465$ $\qquad \sigma \ge 0,342$ al 90% di conf

c) 100 km : consumo $= Y_1 + Y_2 + Y_3 + Y_4 + Y_5 = S$

Y_i ciascuna $\mathcal{N}(\mu, \sigma^2)$ $\qquad S \sim \mathcal{N}(5\mu, 5\sigma^2)$

$\sqrt{\mathrm{Var}\, S} = \sqrt{5}\,\sigma$

$\ge 0,765$ al 90% di confidenza