INTRODUZIONE ALLA STATISTICA — scritto 2

Titolo nota — 02/09/2018

Problema 2.1 (11 punti). Un ascensore ha una portata massima di 8 persone o 600 kg.

(6 punti) Supponendo che i pesi in kg delle persone che salgono sull'ascensore siano variabili aleatorie indipendenti ed identicamente distribuite con distribuzione normale $\mathcal{N}(70, 250)$, quanto vale approssimativamente la probabilità che 8 persone superino la portata dell'ascensore? Quanto vale la probabilità che 9 persone non la superino?

(3 punti) Si risponda nuovamente nell'ipotesi che le variabili aleatorie abbiano invece distribuzione come dalla tabella seguente:

kg	50	60	70	80	90
probabilità	15%	30%	25%	20%	10%

(2 punti) Il criterio di 600 kg per 8 persone è a volte generalizzato come: "massimo n persone e $75n$ kg". Esistono valori di n per cui questo criterio non è raccomandabile, nel senso che la probabilità che n persone superino la portata massima è superiore al 25%? (Rispondere per le variabili normali del primo punto.)

$X_1, X_2, \dots$ pesi persone $\mathcal{N}(\mu, \sigma^2)$

$\mu = 70 \quad \sigma^2 = 250$ indip.

$$P\Big(\underbrace{\sum_1^8 X_i}_{\mathcal{N}(8\mu,\, 8\sigma^2)} > 600\Big) = 1 - \Phi\left(\frac{600-8\mu}{\sqrt{8\sigma^2}}\right) \overset{!}{=} 1 - \Phi\left(\frac{600-560}{\sqrt{2000}}\right)$$

$$\overset{!}{=} 1 - \Phi\left(\frac{40}{20\sqrt{5}}\right) = 1 - \Phi\left(\frac{2}{\sqrt{5}}\right) \overset{!}{\approx} 1 - \Phi(0{,}894) \approx 1 - 0{,}8146$$

$$\approx 18{,}5\%$$

$$P\Big(\sum_1^9 X_i \le 600\Big) = \Phi\left(\frac{600-9\mu}{\sqrt{9\sigma^2}}\right) \overset{!}{=} \Phi\left(-\frac{30}{15\sqrt{10}}\right) = \Phi(-0{,}632)$$

$$\approx 1 - 0{,}737 \approx 26\%$$

(b) $X_i \sim$ tabella

TLC $\Rightarrow \sum_i X_i \,\dot{\sim}\, \mathcal{N}(?;?)$

$$\mu := E(X_i) = \sum_k k\,\varphi_{X_i}(k) = 50 \cdot 0{,}15 + 60 \cdot 0{,}3 + 70 \cdot 0{,}25 + 80 \cdot 0{,}2 + 90 \cdot 0{,}1$$

$$= 7{,}5 + 18 + 17{,}5 + 16 + 9 = 64$$

$$\sigma^2 = E(X_i^2) - \underbrace{E(X_i)}_{\mu}{}^2 \qquad E(X_i^2) = \sum_k k^2 \varphi_{X_i}(k) = 2500 \cdot 0{,}15 + 3600 \cdot 0{,}3 + 4900 \cdot 0{,}25 + 6400 \cdot 0{,}2 + 8100 \cdot 0{,}1$$

$$= 375 + 1080 + 1225 + 1280 + 810 = 4770$$

$$\sigma^2 = 4770 - 4096 = 674$$

$$P\Big(\sum_1^8 X_i > 600\Big) \overset{TLC}{\approx} 1 - \Phi\left(\frac{600 - 8\cdot 64}{\sqrt{8 \cdot 674}}\right) \overset{!}{\approx} 1 - \Phi(1{,}20) \approx 11{,}5\%$$

$$P\Big(\sum_1^9 X_i \le 600\Big) \approx \Phi\left(\frac{600 - 9\cdot 64}{\sqrt{9 \cdot 674}}\right) \approx \Phi(0{,}308) \approx 62{,}1\%$$

(c) $$P\Big(\sum_1^n X_i > 75n\Big) > 0{,}25 \iff 1 - \Phi\left(\frac{75n - n\mu}{\sqrt{n\sigma^2}}\right) > 0{,}25$$

$n = 1,2,3,4$ male

$$\iff \Phi(\dots) < 0{,}75 \iff \frac{75-\mu}{\sigma}\sqrt{n} < \Phi^{-1}(0{,}75) \approx \overbrace{0{,}6745}^{q} \iff n < \frac{\sigma^2 q^2}{(75-\mu)^2} \approx 4{,}5$$

Problema 2.2 (13 punti)**.** Sia X una variabile aleatoria continua con densità

$$f_X(t) = c(1-t^2), \qquad -1 \le t \le 1$$

(7 punti) Si determinino c, la media, la varianza e la mediana di X. Si determini la densità di $Z := \frac{X+1}{2}$. Si tratta di una distribuzione appartenente ad una classe standard?

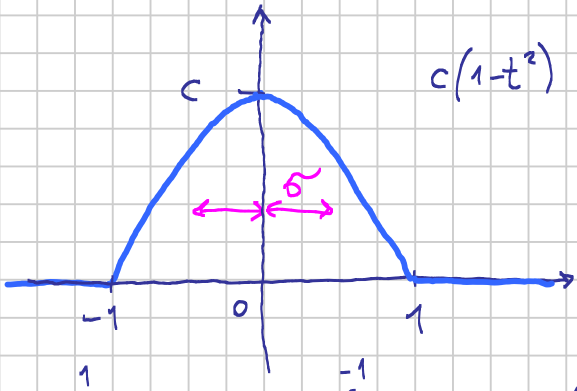
(3 punti) Si determinino densità media e ~~varianza~~ di una variabile aleatoria continua W con funzione di ripartizione

$$F_W(u) = \frac{1}{1+e^{-u}}$$

(3 punti) Sia Y una variabile aleatoria uniforme su $[0,1]$ e indipendente dalla X del primo punto. Si determini la probabilità che

$$0 \le X + Y \le 1$$

$c(1-t^2) \quad |t| \le 1$



$$1 = \int_{-1}^{1} c(1-t^2)\,dt = c\int_{-1}^{1}(1-t^2)\,dt = c\left(2 - \left[\frac{t^3}{3}\right]_{-1}^{1}\right)$$

$$= c\left(2-\frac{2}{3}\right) = c\cdot\frac{4}{3} \iff c = \frac{3}{4}$$

Simmetria → media, mediana $= 0$

$$E(X^2) = Var(X) = \int_{-1}^{1} c t^2(1-t^2)\,dt = c\left(\left[\frac{t^3}{3}\right]_{-1}^{1} - \left[\frac{t^5}{5}\right]_{-1}^{1}\right)$$

$$= c\left(\frac{2}{3}-\frac{2}{5}\right) = \frac{3}{4}\cdot\frac{10-6}{15} = \frac{1}{5}$$

$Z := \frac{X+1}{2}$

$$F_Z(t) = P(Z\le t) = P\left(\frac{X+1}{2}\le t\right) = P(X \le 2t-1) = F_X(2t-1)$$

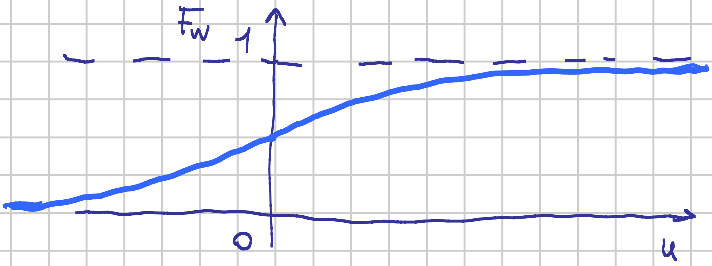
$$f_Z(t) = \frac{d}{dt}F_Z(t) = f_X(2t-1)\cdot 2 = 2c\left(1-(2t-1)^2\right), \quad -1\le 2t-1\le 1$$

$$\overset{!}{=} \frac{3}{2}(1-4t^2+4t-1), \quad 0\le t\le 1$$

$$\overset{!}{=} 6t(1-t), \quad 0\le t\le 1 \qquad \text{beta}$$

(b) $F_W(u) = \frac{1}{1+e^{-u}}$

$$F_W(-u) = \frac{1}{1+e^{u}} = \frac{e^{-u}}{e^{-u}+1} = 1 - F_W(u)$$

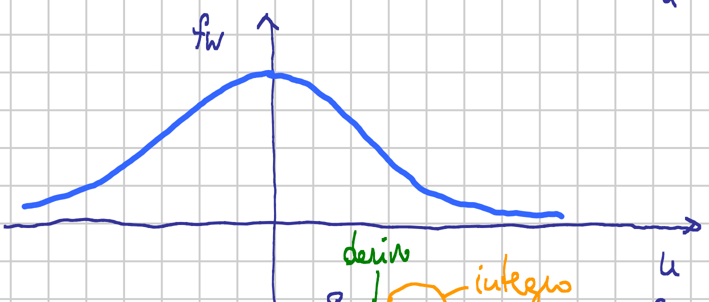


$$f_W(u) = F_W'(u) = \frac{+e^{-u}}{(1+e^{-u})^2}$$

$$f_W(-u) = \frac{e^{u}}{(1+e^{u})^2} = \frac{e^{-2u}}{e^{-2u}}\,\frac{e^{u}}{(1+e^{u})^2} = \frac{e^{-u}}{(e^{-u}+1)^2} = f_W(u)$$

$E(W) = 0$

$$E(W) = \int_{R} u f_W(u)\,du = \int_{R}\frac{ue^{-u}}{(1+e^{-u})^2}\,du$$



deriv

integro

$$E(W) = \lim_{a\to\infty}\int_{-a}^{a}\frac{ue^{-u}}{(1+e^{-u})^2}\,du = \lim_{a\to\infty}\left\{\left[\frac{u}{1+e^{-u}}\right]_{-a}^{a} - \int_{-a}^{a}\frac{1}{1+e^{-u}}\,du\right\} = \lim_{a\to\infty} 0 = 0$$

$$\frac{a}{1+e^{-a}} + \frac{a}{1+e^{a}} - \int_0^a \frac{du}{1+e^{-u}} - \underbrace{\int_{-a}^0 \frac{du}{1+e^{-u}}}_{v=-u} = \frac{a}{1+e^{-a}} + \frac{a e^{-a}}{e^{-a}+1} - \int_0^a du \left\{ \frac{1}{1+e^{-u}} + \frac{1}{1+e^{u}} \right\}$$

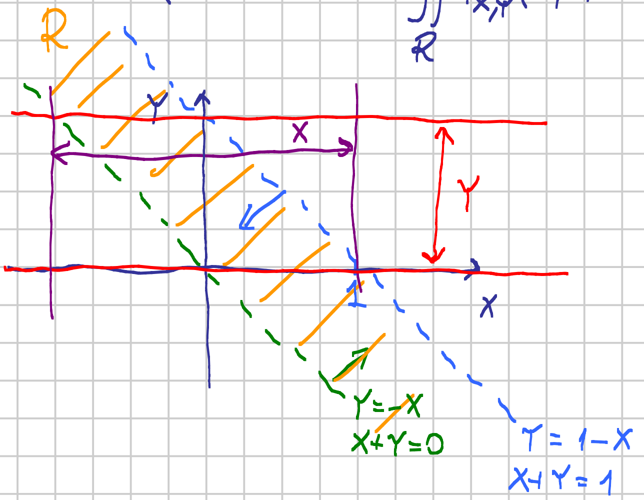
$$-\int_a^0 \frac{-dv}{1+e^{v}} = -\int_0^a \frac{dv}{1+e^{v}}$$

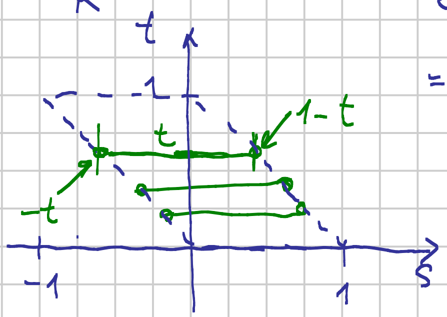
$$\underbrace{\frac{1}{1+e^{-u}} + \frac{e^{-u}}{e^{-u}+1}}_{1}$$

$$= a \cdot 1 - \int_0^a du \cdot 1 = a - a = 0$$

(c) $Y \sim unif(0,1)$ indip da X

$$P(0 \le X+Y \le 1) = \iint_R f_{X,Y}(s,t)\,ds\,dt = \iint_R f_X(s) f_Y(t)\,ds\,dt = \int_0^1 \int_{-t}^{1-t} c(1-s^2)\,ds\,dt$$





$$= c\int_0^1 \left(1 - \left[\frac{s^3}{3}\right]_{-t}^{1-t}\right) dt$$

$$= c\int_0^1 \left(1 - \frac{(1-t)^3}{3} + \frac{(-t)^3}{3}\right) dt$$

$$= c\left\{1 - \frac{1}{3}\int_0^1 \left(1 - 3t + 3t^2 - \cancel{t^3} + \cancel{t^3}\right) dt\right\}$$

$$= c\left\{1 - \frac{1}{3}\underbrace{\left[1 - \frac{3}{2} + \frac{3}{3}\right]}_{\frac{1}{2}}\right\} = \frac{\cancel{3}}{4} \cdot \frac{5}{\cancel{6}_2} = \frac{5}{8}$$

Problema 2.3 (12 punti)**.** Una azienda che produce zampironi testa la durata di una nuova formulazione. Un campione di 10 esemplari mostra le durate seguenti (in minuti):

459	495	461	494	477
485	446	485	458	488

(6 punti) Questi dati sono compatibili al 5% di significatività con l'ipotesi che la durata media sia di almeno 8 ore?

(3 punti) Si stimi al 95% di confidenza la deviazione standard della durata degli zampironi e si testi tramite il calcolo del p-dei-dati l'ipotesi che la deviazione standard sia di 10 minuti.

(3 punti) Supponendo che la durata degli zampironi sia normale con deviazione standard 10 e media incognita, si stimi al 95% di confidenza la frazione di singoli zampironi con durata inferiore alle 8 ore.

(a)

$n=10$ $\quad X_1, X_2, \dots, X_n$ $\quad \mu$ media σ dev. st.

$X_i \sim \mathcal{N}$

$H_0: \mu \geq \mu_0 \quad H_1: \mu < \mu_0 \quad \mu_0 = 480, \alpha = 5\%$

test μ dati gaussiani σ incognita

$$T := \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \overset{\mu=\mu_0}{\sim} t(n-1)$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = 474{,}8 \text{ min}$$

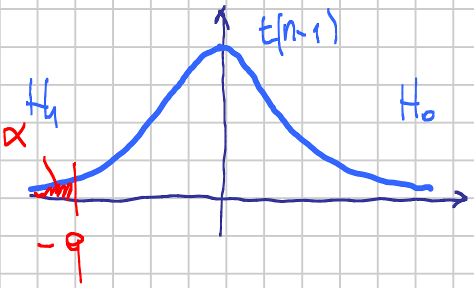
$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \approx 17{,}4 \text{ min}$$

$$T \approx -0{,}945$$

$$-q = F^{-1}_{t(n-1)}(\alpha) \qquad q = F^{-1}_{t(n-1)}(1-\alpha) \approx 1{,}83$$

$$RA_T = [-q; +\infty) = [-1{,}83; +\infty) \ni T$$

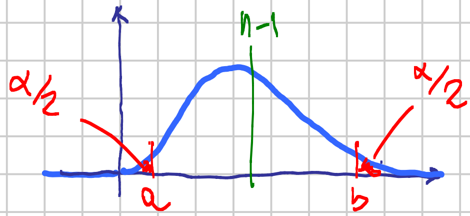
accetto H_0



(b) Intervallo di confidenza σ $\qquad 1-\alpha = 95\%$

$$\frac{S^2}{\sigma^2}(n-1) \sim \chi^2(n-1) \qquad a = F^{-1}_{\chi^2(n-1)}\left(\frac{\alpha}{2}\right) \approx 2{,}70$$

$$b \approx 19{,}02$$



$$1-\alpha = P\left(a \leq \frac{S^2}{\sigma^2}(n-1) \leq b\right) = \dots = P\left(\sqrt{\frac{n-1}{b}}\,S \leq \sigma \leq \sqrt{\frac{n-1}{a}}\,S\right)$$

$$\sigma \in \left[\sqrt{\frac{n-1}{b}}\,S ; \sqrt{\frac{n-1}{a}}\,S\right] = [11{,}97 ; 31{,}77] \approx [12 ; 32]$$

$H_0: \sigma = \sigma_0 \qquad H_1: \sigma \neq \sigma_0 \qquad \sigma_0 = 10 \qquad p$ dei dati

$$W = \frac{S^2}{\sigma_0^2}(n-1) \overset{H_0}{\sim} \chi^2(n-1) \qquad W \approx 27{,}25$$

$$RA_W = [a, b] \qquad a = F^{-1}_{\chi^2(n-1)}\left(\frac{\alpha}{2}\right) \qquad b = F^{-1}_{\chi^2(n-1)}\left(1-\frac{\alpha}{2}\right)$$

accetto H_0 α $\quad a \leq W \leq b \iff \frac{\alpha}{2} \leq F_{\chi^2(n-1)}(W) \leq 1 - \frac{\alpha}{2}$

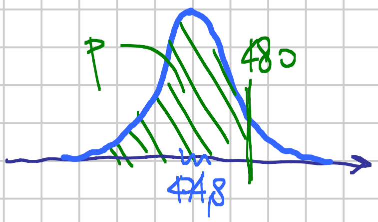
$$\Leftrightarrow \begin{cases} \alpha \le 2\, F_{\chi^2}(w) \\ \alpha \le 2\left(1 - F_{\chi^2}(w)\right) \end{cases} \quad \text{entrambe} \qquad \alpha \le 2 \min\left\{ F_{\chi^2}(w),\ \underline{1 - F_{\chi^2}(w)} \right\} =: \alpha^*$$

$$F_{\chi^2}(w) \simeq F_{\chi^2}(27{,}25) > 0{,}995 \qquad \alpha^* = 2\left(1 - F_{\chi^2(n-1)}(w)\right) \lesssim 1\%$$

dico H_1 : i dati dimostrano che $\sigma \neq 10$ min

(c) $X_i \sim \mathcal{N}(\mu, \sigma^2)$ $\quad \sigma = 10 \qquad \mu \simeq \bar{X} = 474{,}8$

p : frazione unità con durata < 480



stima puntuale $\quad p \simeq F_{\mathcal{N}(\bar{X}, \sigma^2)}(480) \simeq 70\%$

valore vero $\quad p = F_{\mathcal{N}(\mu, \sigma^2)}(480)$

intervallo di confidenza $\quad r_1 \le p \le r_2 \quad$ al 95% di conf

$$r_1 \le F_{\mathcal{N}(\mu,\sigma^2)}(480) \le r_2 \quad \Leftrightarrow \quad r_1 \le \Phi\left(\frac{480-\mu}{\sigma}\right) \le r_2$$

$$\Leftrightarrow \quad \Phi^{-1}(r_1) \le \frac{480-\mu}{\sigma} \le \Phi^{-1}(r_2)$$

$$\Leftrightarrow \quad 480 - \sigma\Phi^{-1}(r_2) \le \mu \le 480 - \sigma\Phi^{-1}(r_1)$$

ma l'intervallo di conf. per μ è $\quad \mu \in \bar{X} \pm q\frac{\sigma}{\sqrt{n}} \qquad q = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$

$$480 - \sigma\Phi^{-1}(r_2) = \bar{X} - q\frac{\sigma}{\sqrt{n}} \quad \Leftrightarrow \quad \sigma\Phi^{-1}(r_2) = 480 - \bar{X} + q\frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow \quad r_2 = \Phi\left(\frac{480-\bar{X}}{\sigma} + \frac{q}{\sqrt{n}}\right) \approx \Phi(1{,}14) = 0{,}8729$$

$$r_1 = \Phi\left(\frac{480-\bar{X}}{\sigma} - \frac{q}{\sqrt{n}}\right) \approx \Phi(-0{,}1) = 1 - \Phi(0{,}1) = 0{,}4602$$

$$\frac{480-\bar{X}}{\sigma} \approx 0{,}52 \qquad \frac{q}{\sqrt{n}} \approx 0{,}62$$

Al 95% di conf $\quad p \in [46\% \,;\, 87\%]$

Problema 2.4 (11 punti). Un tipo va a funghi. Il tempo che passa prima di trovare il primo porcino è ipotizzato avere distribuzione esponenziale, e così i tempi per ciascuno dei porcini successivi. Al 15-esimo porcino trovato, sono passate in totale 3 ore e mezza e il fungaiolo decide di tornare a casa.

(7 punti) Si stimi puntualmente e al 95% di confidenza il parametro della distribuzione esponenziale introdotta. Qual è l'unità di misura di tale parametro?

(2 punti) Consideriamo un test per verificare al 10% di significatività se la media di questa distribuzione esponenziale sia pari a 10 minuti. Si determini la regione di accettazione relativa alla media campionaria di 15 tempi di raccolta.

(2 punti) Si determini la potenza del test del punto precedente per una distribuzione esponenziale di media 20 minuti.

(a)

$X_1, X_2, \ldots$ tempi successivi di ricerca

$X_i \sim \text{expo}(\lambda)$ $\quad E(X_i) = \beta = \frac{1}{\lambda}$

$X_1 + X_2 + \ldots + X_{15} = 3{,}5\ h = 210$ min

$\beta = E(X_i) \approx \bar{X} = \frac{1}{n}\sum_i X_i \approx 0{,}2\bar{3}\ h$

$= 14$ min

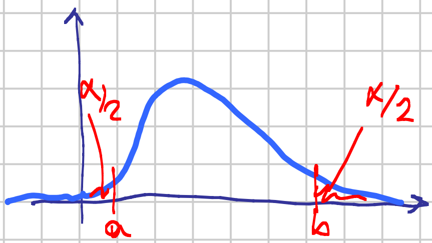
$\lambda \approx \frac{1}{\bar{X}} = \frac{1}{14} \approx 0{,}07$ min^{-1}

$\approx 4{,}28\ h^{-1}$

Stimo di trovare circa 4,28 funghi all'ora, 0,07 funghi al minuto

[funghi / tempo] è l'unità di misura

$2n\lambda\bar{X} \sim \chi^2(2n)$



$1-\alpha = 95\%$

$a \cong 16{,}8$

$b \approx 47$

$1-\alpha = P(a \le 2n\lambda\bar{X} \le b) \ldots$

al livello di conf $1-\alpha$, $\lambda \in \left[\frac{a}{2n\underbrace{\bar{X}}_{3,5}} ; \frac{b}{2n\underbrace{\bar{X}}_{3,5}}\right] \cong [2{,}4 ; 6{,}71]$

(b) $H_0: \beta = \beta_0$ $\quad H_1: \beta \neq \beta_0$ $\quad \beta_0 = 10$ min $\quad \alpha = 10\%$ lvl sign.

$W := \frac{2n\bar{X}}{\beta_0} \overset{H_0}{\sim} \chi^2(2n)$ $\quad a = F^{-1}_{\chi^2(2n)}\left(\frac{\alpha}{2}\right) \approx 18{,}5$

$b = F^{-1}_{\chi^2(2n)}\left(1-\frac{\alpha}{2}\right) \approx 43{,}8$

$RA_W = [a, b]$

dico H_0 se $a \le \frac{2n\bar{X}}{\beta_0} \le b \iff a\frac{\beta_0}{2n} \le \bar{X} \le b\frac{\beta_0}{2n}$

$RA_{\bar{X}} = \left[a\frac{\beta_0}{2n} ; b\frac{\beta_0}{2n}\right] = [6{,}17 ; 14{,}6]$

(c) Potenza per $\beta = 20$ min

$$P(\text{dire } H_1 \mid \beta = 20) = 1 - P(a \le W \le b \mid \beta = 20) = 1 - P\left(a \le \frac{2n\bar{X}}{\beta_0} \le b \,\middle|\, \beta = 20\right)$$

$$= 1 - P\Big(a\frac{\beta_0}{\beta} \le \underbrace{\frac{2n\bar{X}}{\beta}}_{\chi^2(2n)} \le b\frac{\beta_0}{\beta}\Big) = 1 - \Big[\underbrace{F_{\chi^2(2n)}\big(\underbrace{b\tfrac{\beta_0}{\beta}}_{21{,}9}\big)}_{0{,}14} - \underbrace{F_{\chi^2(2n)}\big(\underbrace{a\tfrac{\beta_0}{\beta}}_{9{,}25}\big)}_{\text{piccolo } \ll 0{,}005}\Big] \approx 86\%$$