

33 Uniforme, di Cauchy, esponenziale, quantili

Come abbiamo visto, una v.a. è una funzione $X : \Omega \rightarrow \mathbb{R}$ che ad eventi associa numeri e si dice continua se i valori riempiono almeno un intervallo.

Per esempio il peso del primo neonato del prossimo anno, misurato con "infiniti decimali" (numero reale, con precisione infinita), ovviamente da considerarsi prima dell'evento.

Ogni singolo valore ha probabilità nulla: $P(X = t) = 0, \forall t$.

In questa trattazione elementare, le v.a. continue che considereremo sono **molto regolari**: la **funzione di ripartizione**

$$F_X(z) := P(X \leq z)$$

- è continua su tutto \mathbb{R}
- e derivabile salvo al più in 1 o 2 punti.

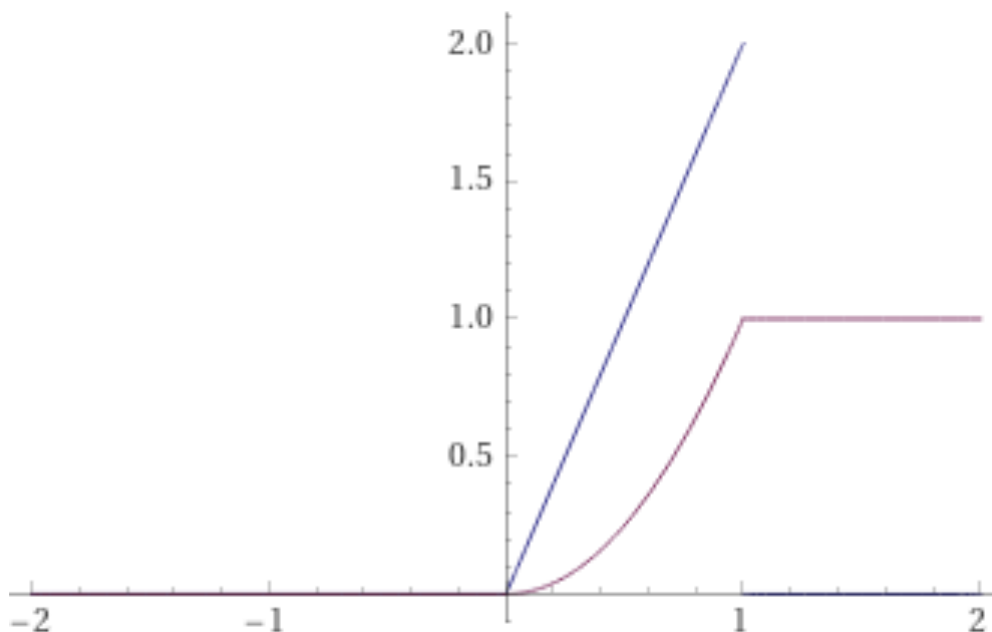


Figure 48: Blu: densità con 1 punto di discontinuità. Magenta: sua funzione di ripartizione con 1 punto di non derivabilità (ma comunque continua). Screenshot da WolframAlpha.

La derivata della f.r., completata ad arbitrio in quegli eventuali punti particolari ovvero *singolari*,

$$f(z) := F'_X(z) = DP(X \leq z) \quad \text{spesso denotata } f_X, \text{ con argomento } z, t, u, x \dots$$

si chiama **densità** della variabile aleatoria.

Per fissare le idee nella mente, si immagini

la funzione di ripartizione come una curva sigmoide che cresce da 0 a 1

la densità come una funzione a campana nel senso più lato possibile con asintoto l'asse delle ascisse.

Variazioni sono possibili nei singoli casi ma intanto fissiamo nella mente qualcosa da immaginare per la generalità dei casi:

f.r. \leftrightarrow sigmoide *tipo arcotangente ma riscalata fra 0 e 1*

densità \leftrightarrow campana

Una funzione f è densità di una v.a. se e solo se

$$f(t) \geq 0 \quad \wedge \quad \int_{-\infty}^{+\infty} f(t) dt = 1.$$

Se f e F_X sono rispettivamente densità e f.r. di una variabile aleatoria continua X , allora (teorema)

$$F_X(x) = \int_{-\infty}^x f(t) dt \quad \forall x$$

$$P(X = x) = 0 \quad \forall x$$

$$P(a \leq X \leq b) = \int_a^b f(t) dt \quad \forall a, b, \text{ ogni } \leq \text{ sostituibile con } <$$

e in quest'ultima a e b possono essere infiniti (col $<$, ovvio).

33.1 Variabile aleatoria uniforme continua

Considereremo qua la v.a. $\sim \mathbb{U}[a, b]$ oppure $\mathbb{U}(a, b)$, uniforme su $[a, b]$.

$$f(t) := \begin{cases} \frac{1}{b-a} & \text{se } a \leq t \leq b \\ 0 & \text{altrimenti} \end{cases} \quad F_X(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{se } a \leq x \leq b \\ 1 & \text{se } x > b \end{cases} .$$

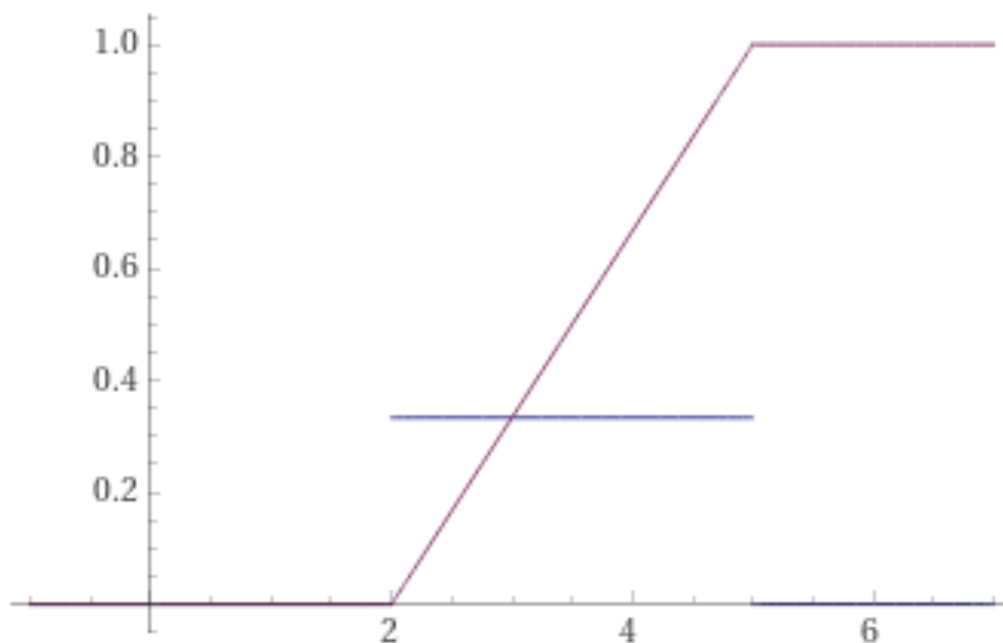


Figure 49: Blu: densità uniforme continua di parametri 2 e 5. Magenta: sua f.r.. Screenshot da WolframAlpha.

La densità ha 2 punti di discontinuità e corrispondentemente la f.r. ha 2 punti di non derivabilità.

33.2 Variabile aleatoria di Cauchy.

$$f(t) := \frac{1}{\pi} \cdot \frac{1}{1+t^2}$$

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x$$

La distribuzione di Cauchy ha in Farmacia almeno 5 utilità:

- fornisce un modello di densità a campana di cui si possono calcolare i valori con le 4 operazioni
- fornisce un modello di densità a campana di cui si possono calcolare gli integrali con funzioni elementari (l'arcotangente)
- fornisce un esempio di variabile aleatoria priva di speranza matematica
- permette di chiarire, nel confronto con la v.a. normale, la questione delle *code pesanti*
- è usata in Spettroscopia: [LINK->](#)

33.3 Variabile aleatoria esponenziale

Queste sono densità e funzione di ripartizione di una v.a. X esponenziale di parametro $\lambda > 0$:

$$f(t) := \begin{cases} \lambda e^{-\lambda t} & \text{se } t \geq 0 \\ 0 & \text{altrimenti} \end{cases} \quad F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

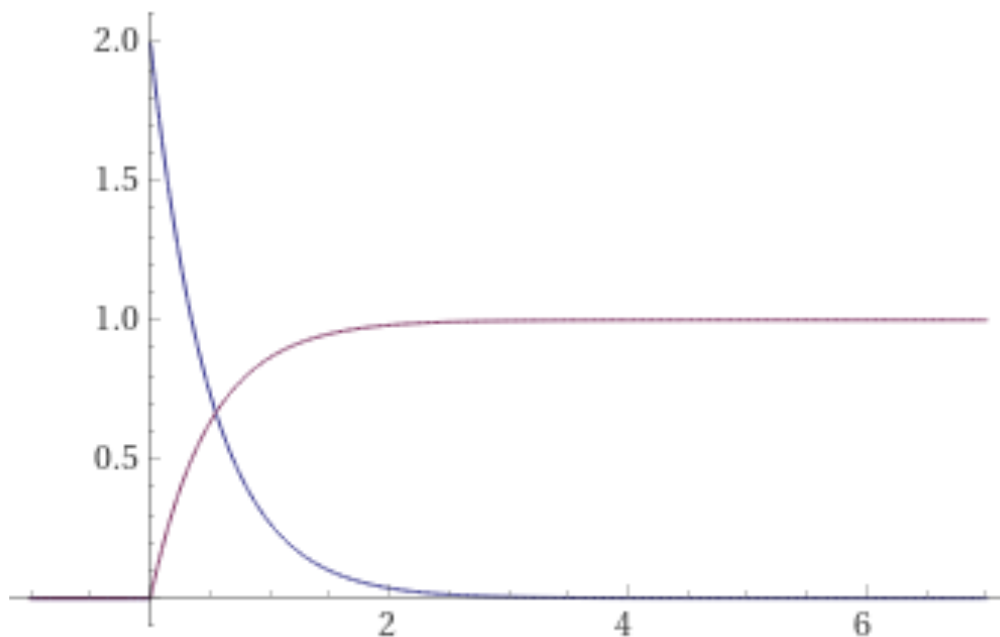


Figure 50: Blu: densità esponenziale di parametro 2. Magenta: sua f.r.. Screenshot da WolframAlpha.

La densità esponenziale modella molte cose fra cui l'*intertempo* fra 2 chiamate telefoniche a una farmacia o a un qualunque servizio in un tempo del giorno in cui esse arrivano in ogni minuto con la stessa probabilità: Leggiamo sulla Wikipedia in inglese alla voce *Exponential distribution*:

if we focus on a time interval during which the rate is roughly constant, such as from 2 to 4 p.m. during work days, the exponential distribution can be used as a good approximate model for the time until the next phone call arrives.

Ecco 11 valori (di una simulazione informatica) della variabile aleatoria esponenziale di parametro 3.5, con 3 cifre decimali:

0.163 0.190 0.695 0.132 0.979 0.157 0.023 0.236 0.291 0.113 0.015

Si notino i valori piccoli e i valori grandi, distribuiti in un modo tipico, caratterizzante appunto la variabile aleatoria esponenziale.

Ecco online su WolframAlpha un po' di determinazioni, ad ogni reload nuove, (di una simulazione informatica) della variabile aleatoria esponenziale di parametro 3.5: [link->](#)

La distribuzione esponenziale ha un'ampia ricorrenza nelle Scienze Applicate. Traiamo da Wikipedia in inglese, l'enciclopedia libera, alla voce

The exponential distribution occurs naturally when describing the lengths of the inter-arrival times in a homogeneous Poisson process.

The exponential distribution may be viewed as a continuous counterpart of the geometric distribution, which describes the number of Bernoulli trials necessary for a

discrete process to change state. In contrast, the exponential distribution describes the time for a continuous process to change state.

In real-world scenarios, the assumption of a constant rate (or probability per unit time) is rarely satisfied. For example, the rate of incoming phone calls differs according to the time of day. But if we focus on a time interval during which the rate is roughly constant, such as from 2 to 4 p.m. during work days, the exponential distribution can be used as a good approximate model for the time until the next phone call arrives. Similar caveats apply to the following examples which yield approximately exponentially distributed variables:

- The time until a radioactive particle decays, or the time between clicks of a Geiger counter
- The time it takes before your next telephone call
- The time until default (on payment to company debt holders) in reduced form credit risk modeling

Exponential variables can also be used to model situations where certain events occur with a constant probability per unit length, such as the distance between mutations on a DNA strand, or between roadkills on a given road.

In queuing theory, the service times of agents in a system (e.g. how long it takes for a bank teller etc. to serve a customer) are often modeled as exponentially distributed variables. (...)

Per ciascuno di quei fenomeni abbiamo qualcosa del tipo

$$P(\text{dura meno di } x) = 1 - e^{-\lambda x}$$

con un parametro λ caratteristico di quel fenomeno.

In particolare

$$P(\text{dura meno di } 0) = 0 \text{ cioè almeno un istante dura.}$$

Abbiamo introdotto la speranza di vita con la variabile aleatoria geometrica, col robottino che può “morire” solo all’ultimo giorno dell’anno. La variabile aleatoria esponenziale è in qualche modo il corrispondente continuo della variabile aleatoria geometrica, e tiene conto della possibilità di morte – o di guasto per gli apparati tecnologici, dai motori alle lampadine ai fusibili – in ogni momento. Entrambe le variabili aleatorie godono dell’assenza di memoria: resta da vivere sempre la stessa quantità di tempo, mediamente. Esattamente come per la variabile aleatoria geometrica, c’è però da dire che sia per gli esseri viventi che per le macchine, la mortalità non è costante: è molto maggiore presso la nascita, e nella vecchiaia. Entrambi i modelli allora sono molto approssimativi, seppure danno un’idea della questione.

33.4 Definizione semplificata dei quantili

Per ogni variabile aleatoria **continua** X di interesse pratico in Statistica Medica e nelle Scienze della Farmacia

il quantile $q_{0.95}$ di ordine 0.95
 ovvero ovvero 95%
 di X ovvero della sua densità f
 è l’unico numero tale che

$$P(X \leq q_{0.95}) = 0.95$$

e similmente a 0.95 vale per
 ogni numero fra 0 e 1 esclusi

(107)

ed è allora soluzione dell’equazione nell’incognita $q_{0.95}$

$$F_X(q_{0.95}) = 0.95$$

cioè in definitiva ha il valore (teorico, spesso difficile da calcolare)

$$q_{0.95} = F_X^{-1}(0.95)$$

(e similmente a 0.95 vale per ogni numero fra 0 e 1 esclusi).
(Esistono anche i quantili delle variabili aleatorie discrete ma non ce ne occupiamo).

33.5 Mediana di variabile aleatoria continua

La mediana è un particolare quantile, quello del 50%:

$$\text{mediana di v.a. continua} := q_{0.5} \quad (108)$$

e allora

$$P(\text{variabile aleatoria} < \text{mediana}) := 50\% \quad (109)$$

(173)

Per esempio la mediana dell'esponenziale di parametro λ è $\frac{1}{\lambda} \ln 2$.

E abbastanza banalmente la mediana dell'uniforme continua di parametri a e b è la media di a e b .

33.6 Anticipazione di Statistica Inferenziale

Accenniamo a una classica densità (quella detta⁽¹⁷⁴⁾ chi-quadrato) che in Statistica Inferenziale consente il più importante calcolo della Statistica dei farmaci.

Fra farmaco e placebo, e fra guariti e non guariti, si formano 4 classi, ma dai 4 numeri possiamo ritenere statisticamente significativo, cioè oltre la normale fluttuazione casuale, l'effetto curativo del farmaco? Dai 4 numeri con una formula (che vedremo in seguito) si produce

¹⁷³In pratica la mediana di una v.a. continua ha il 50% di lasciare alla sua sinistra un valore casuale tratto dalla variabile aleatoria considerata (e ovviamente 0.5=50% ha sostituito nella (107) il valore 0.95).

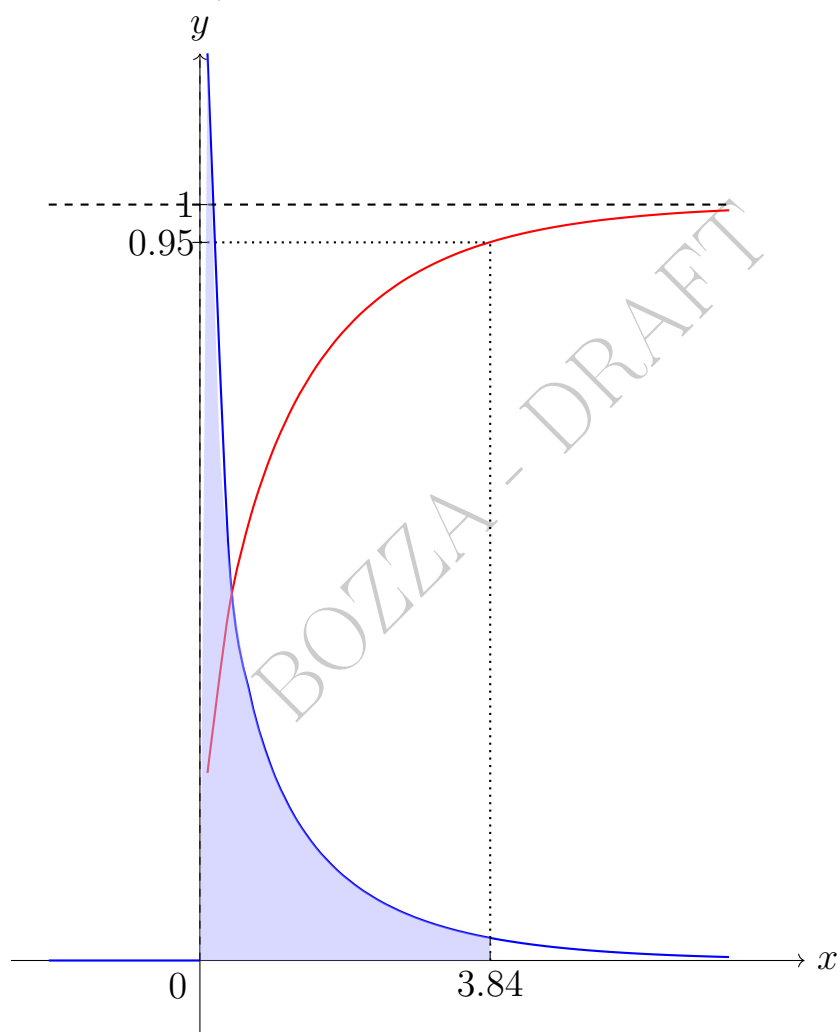
¹⁷⁴In effetti è la densità del chi-quadrato a 1 grado di libertà:

$$f(x) := \begin{cases} \frac{e^{-\frac{x}{2}}}{\sqrt{2\pi}\sqrt{x}} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases} \quad (110)$$

1 numero, e se supera 3.84 si ritiene statisticamente significativo l'effetto del farmaco.

Quantile del chi-quadrato al 95% per la tabella $2 \times 2 = 3.84$

(A rigore ≈ 3.84).



In figura è ombreggiata (purtroppo con una lacuna in alto) l'area del sottografico della densità da 0 a 3.84, il quale è il quantile di ordine 0.95, e prima di 3.84 c'è il 95% dell'area 1 del sottografico.

In qualche modo, molto approssimativamente e a rigore inesattamente, si intende che solo col 5% di probabilità potevano venire solo per caso numeri come quei 4, senza che il farmaco avesse causato il

numero di guariti (questione molto sottile, che un po' verrà approfondita in Statistica Inferenziale).

Il 5% ovvero il 95% viene considerato classicamente la soglia della "ragionevole significatività".

33.7 Proprietà di densità e funzioni di ripartizione

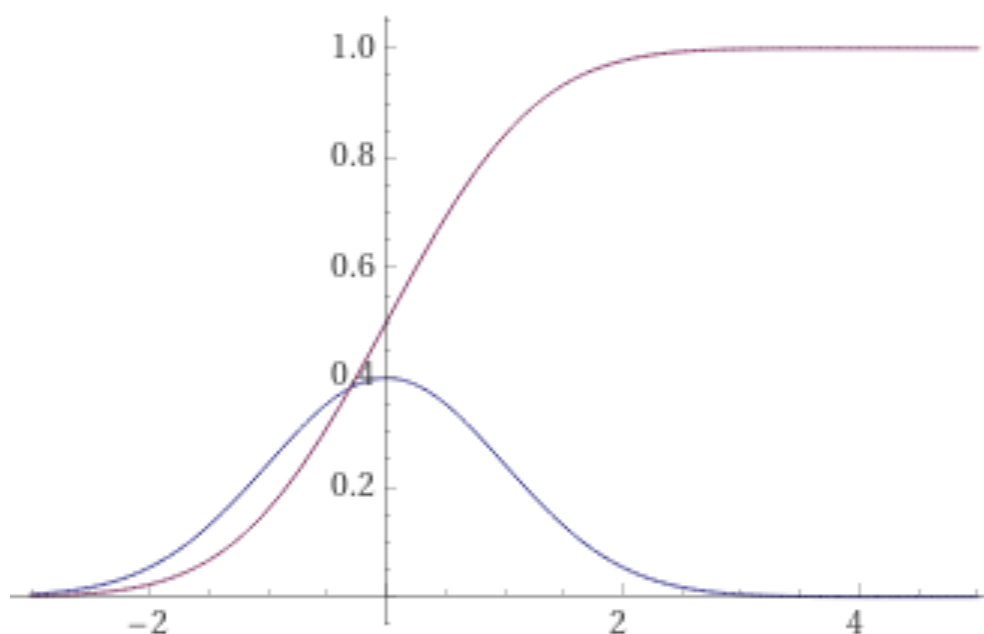


Figure 51: Blu: densità normale standard. Magenta: sua f.r.. Screenshot da WolframAlpha.

Proprietà di ogni f.r. $F(x)$ di v.a. continua:

- 1) ha valori fra 0 e 1, cioè $0 \leq F(x) \leq 1$;
- 2) è non decrescente: $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$;
- 3) tende a 0 per $x \rightarrow -\infty$;
- 4) tende a 1 per $x \rightarrow +\infty$;
- 5) è continua (in trattazioni di livello superiore continua a destra);
- 6) dove la densità $f = F'$ è positiva⁽¹⁷⁵⁾ la f.r. è crescente puntualmente.

¹⁷⁵Dove f vale 0 la f.r. può essere puntualmente crescente oppure no.

Proprietà di ogni densità $f(x)$ di v.a. continua:

- 1) ha valori ≥ 0 , sappiamo;
- 2) ha integrale 1 su tutto \mathbb{R} , sappiamo; e poi si dimostra che
- 3) è continua salvo al più 1 o 2 punti (in questa trattazione).

3 bis) Tutte le densità che considereremo in questa trattazione tendono a 0 per $x \rightarrow -\infty$ e $x \rightarrow +\infty$.

Ma densità particolarmente "capricciose", considerabili a un livello superiore di questo testo elementare, potrebbero non avere uno o entrambi quei limiti, con infinite oscillazioni con "campate" sempre più strette, come nella figura sottostante.

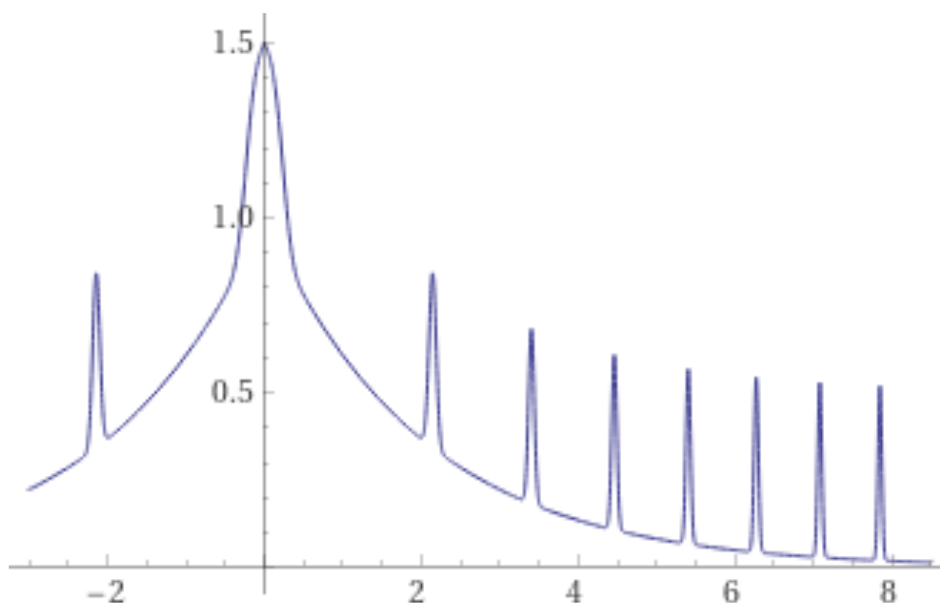


Figure 52: Densità "capricciosa" del tipo che ci proponiamo di non considerare: non tende a 0 per $x \rightarrow +\infty$. Screenshot da WolframAlpha.

33.8 Complementi – Definizione più completa di quantile

Se l'insieme dove f è positiva è un intervallo, anche se illimitato, allora (si dimostra) per ogni $\alpha \in]0, 1[$ esiste un'unica soluzione dell'equazione

$$F_X(t) = \alpha$$

che si chiama *quantile di ordine* α , indicato talvolta con q_α :

$$P(X \leq q_\alpha) = \alpha.$$

Esempio. Per una legge esponenziale di parametro 3 calcolare $P\left(X \geq \frac{2}{3} \ln 2\right)$.

La densità è

$$f(x) := \begin{cases} 3e^{-3x} & \text{se } x \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

e allora

$$P\left(X \geq \frac{2}{3} \ln 2\right) = \int_{\frac{2}{3} \ln 2}^{+\infty} f(t) dt = \int_{\frac{2}{3} \ln 2}^{+\infty} 3e^{-3t} dt = \star$$

calcoliamo l'integrale indefinito con la (88) con $z := -3t + 0$

$$\int 3e^{-3t} dt = \frac{1}{-3} \left(\int 3e^z dz \right)_{z=-3t} = -\frac{1}{3} (3e^z)_{z=-3t} = -e^{-3t}$$

$$\text{riprendiamo } \star = \left[-e^{-3t} \right]_{\frac{2}{3} \ln 2}^{+\infty} = -e^{-\infty} - \left(-e^{-3 \cdot \frac{2}{3} \ln 2} \right) =$$

$$\begin{aligned} \text{intendendo } e^{-\infty} \text{ nel senso del limite e allora } 0 &= 0 + e^{-2 \ln 2} = e^{\ln 2^{-2}} = 2^{-2} = \\ &= \frac{1}{4} = 0.25 = 25\% \end{aligned}$$

(e allora $\frac{2}{3} \ln 2$ è il quantile di ordine $\frac{1}{4}$ ossia 0.25 ossia 25%).

Calcolando la probabilità p dell'evento complementare $X < \frac{2}{3} \ln 2$, e poi facendo $1 - p$, si potevano evitare gli infiniti.

Esercizi.

- I principali quantili considerati in statistica sono:
 - i *quartili* $q_{0.25}$, $q_{0.5}$, $q_{0.75}$, e il secondo è la mediana;
 - i *decili* $q_{0.1}$, $q_{0.2}, \dots, q_{0.9}$, corrispondenti a 10%, 20%, ..., 90%;
 - i *centili* o *percentili* $q_{0.01}$, $q_{0.02}, \dots, q_{0.99}$, corrispondenti a 1%, 2%, ..., 99%;
 - due *ventili*, precisamente $q_{0.05}$ e $q_{0.95}$.

Si calcolino relativamente alla densità esponenziale.

Per esempio per i decili si troverà

$$q_{k/10} = \frac{1}{\lambda} \ln \frac{10}{10-k} \quad k = 1, \dots, 9$$

(Si noti che per $k = 5$ si ottiene la mediana).

- Per una v.a. X di densità esponenziale di parametro 5 calcolare $P(X < 2)$, $P(X^3 > 2)$, $P(X^2 \geq 4)$.
- Si consideri una v.a. Z di densità

$$f(z) := \begin{cases} 6z(1-z) & \text{se } 0 \leq z \leq 1 \\ 0 & \text{altrimenti} \end{cases} .$$

Senza fare calcoli ma col disegno del grafico stabilire quanto vale la mediana e mettere in ordine crescente

$$P(Z < 0.4), \quad P(Z > 0.9), \quad P(Z \leq 0.2), \quad P(Z \geq 0.7).$$