Funzione di densità di probabilità

Riccardo Rimondi
26 gennaio 2024

Premessa

Quando parliamo di variabili casuali discrete, ovvero variabili che possono assumere un numero di valori finito o numerabile, possiamo parlare della probabilità che questa variabile assuma un singolo valore.

La funzione che prende in input i valori discreti che può assumere la variabile e fa uscire in output le probabilità ad essi associati si chiama **funzione di massa di probabilità**:

$$p(x_i) = P(X = x_i)$$

Inoltre questa funzione deve soddisfare sempre la seguente equazione, ovvero la **condizione di normalizzazione**:

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Questo perché la somma delle probabilità che X assuma almeno uno dei valori possibili è uguale ad 1.

Una conseguenza diretta di ciò è che la probabilità associata a un singolo valore deve sempre essere compresa tra 0 e 1.

$$0 \leq p(x_i) \leq 1 \; \forall x_i$$

Problemi nel continuo

Parlando di variabili casuali continue, la funzione di massa di probabilità non funziona più. Questo perché, potendo la variabile assumere valori in un intervallo continuo (a, b), quando proviamo a calcolare la probabilità che esca un preciso numero usando la probabilità classica avremmo:

$$\frac{casi\ favorevoli}{casi\ totali} = \frac{1}{\infty} = 0$$

Quando consideriamo variabili casuali continue parliamo quindi di probabilità che il valore che assume la variabile sia in un certo intervallo.

Esempio

Scegliendo un numero reale casuale tra 0 e 1, qual è la probabilità che questo numero sia esattamente uguale a 0.5?

La probabilità che questo accada è nulla, poiché il numero può assumere valori nel continuo.

Qual è invece la probabilità che questo numero sia compreso tra 0.5 e 0.7?

Questa probabilità non è nulla, si può calcolare e vedremo come.

Funzione di densità di probabilità (PDF)

La funzione di densità di probabilità è l'analogo per le variabili continue della massa di probabilità delle variabili discrete.

Descrive la "densità" di probabilità nell'insieme che rappresenta i valori assumibili da una variabile casuale continua.

$$p(x)=\frac{\lim_{dx\to 0} P(x<X\leq(x+dx))}{dx}$$

Come già abbiamo detto, non è molto interessante la probabilità associata a un singolo valore, bensì la probabilità che questo valore cada in un certo intervallo:

$$P(x\in A)=\int_A p(x)\,dx$$

L'integrale è l'analogo per il continuo della sommatoria.

Caratteristiche necessarie della funzione

Anche la densità di probabilità deve seguire le stesse regole della massa di probabilità, quindi:

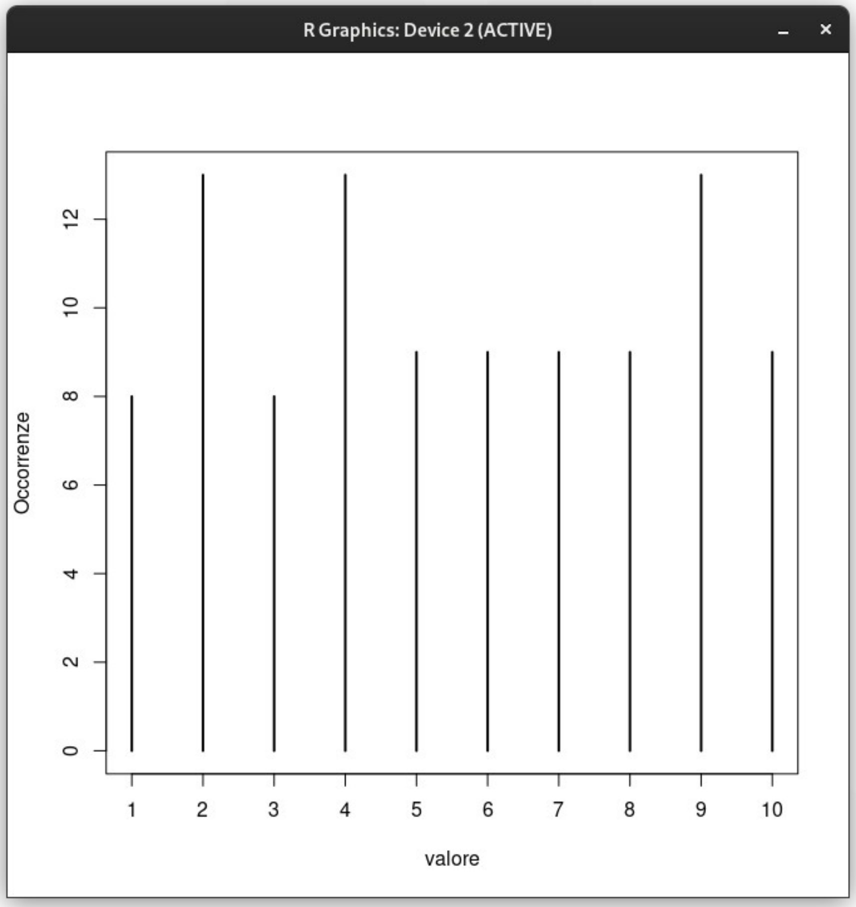
1. La probabilità non può mai essere negativa

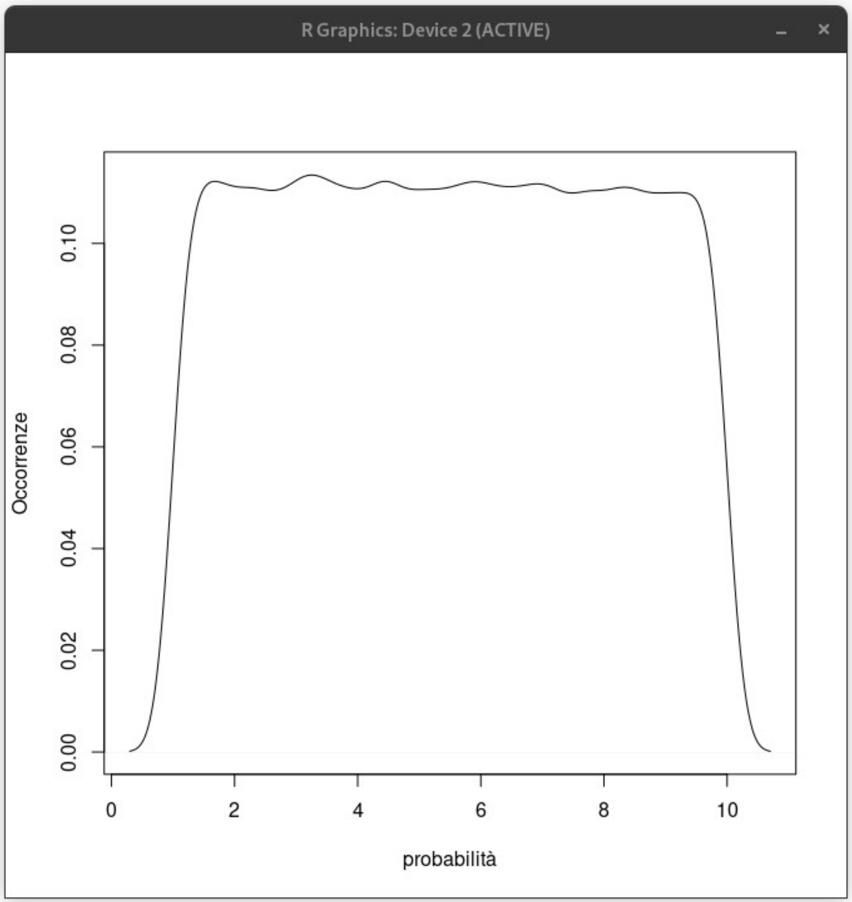
   $p(x) \geq 0 \, \forall \, x$

2. L'integrale sull'insieme contenente tutti i valori possibili deve essere uguale a 1

   $\int_{\mathbb{R}} f(x) \, dx = 1$

Esempi PMF e PDF





Funzione di ripartizione (CDF)

La funzione di ripartizione associata a una variabile continua X è la funzione:

$$F(x)=P(X\leq x)=\int_{-\infty}^{x} f(t)dt$$

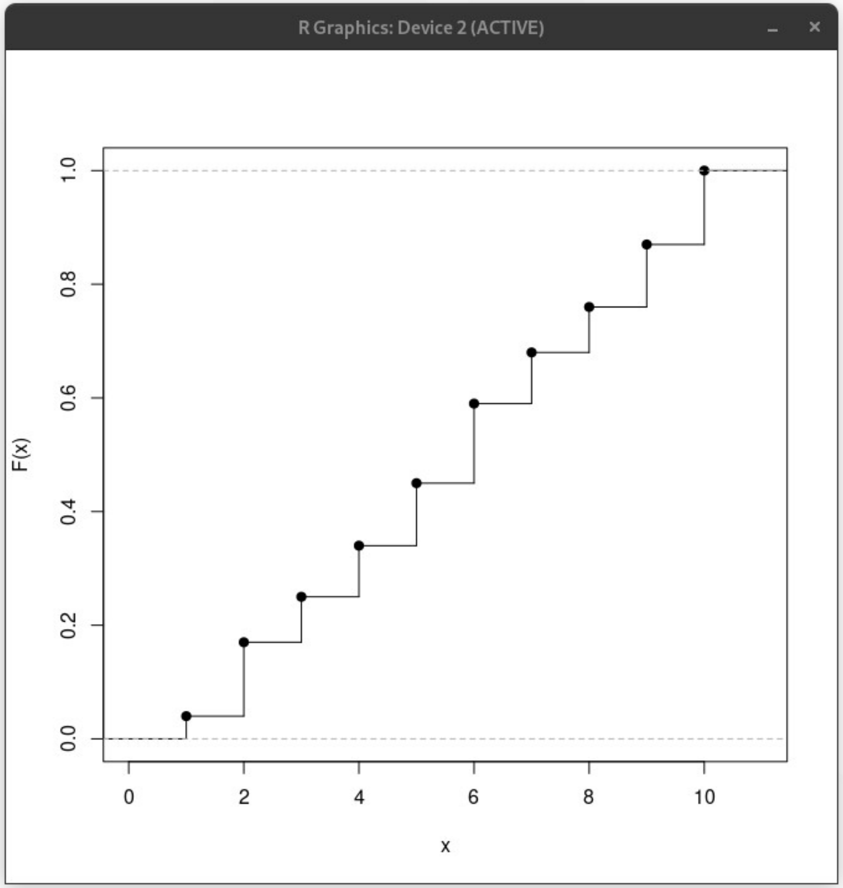
Questa funzione descrive la probabilità che la variabile assuma valori inferiori a un dato valore x

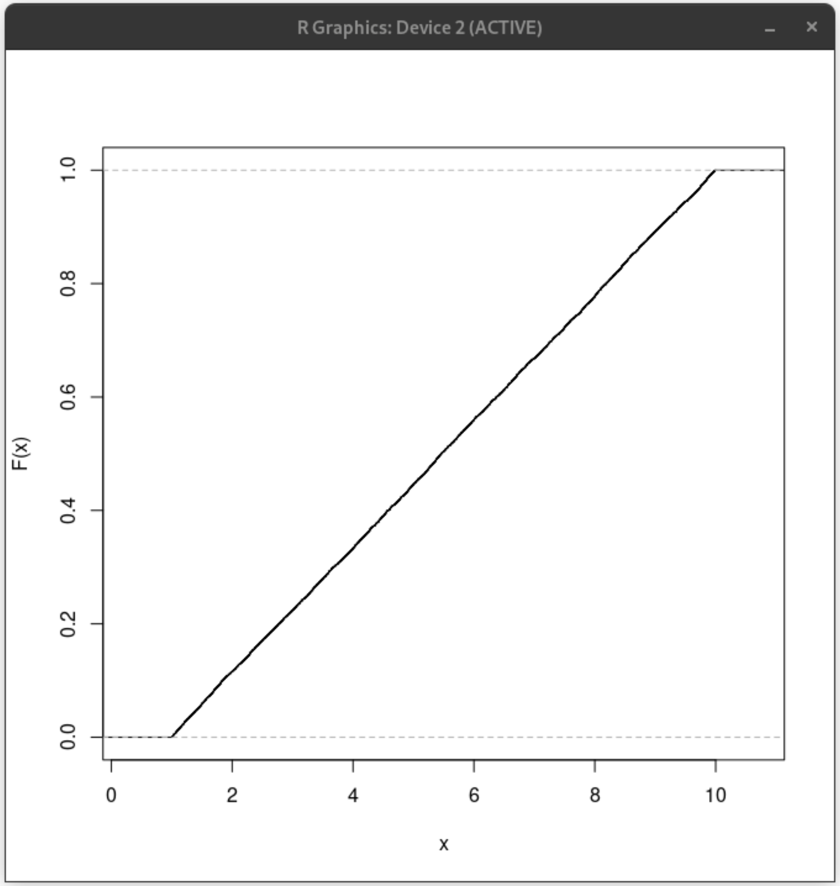
La funzione di ripartizione si può estendere alle variabili discrete sostituendo l'integrale con la sommatoria.

$$F(x)=P(X\leq x)=\sum_{n=-\infty}^{x} p(n)$$

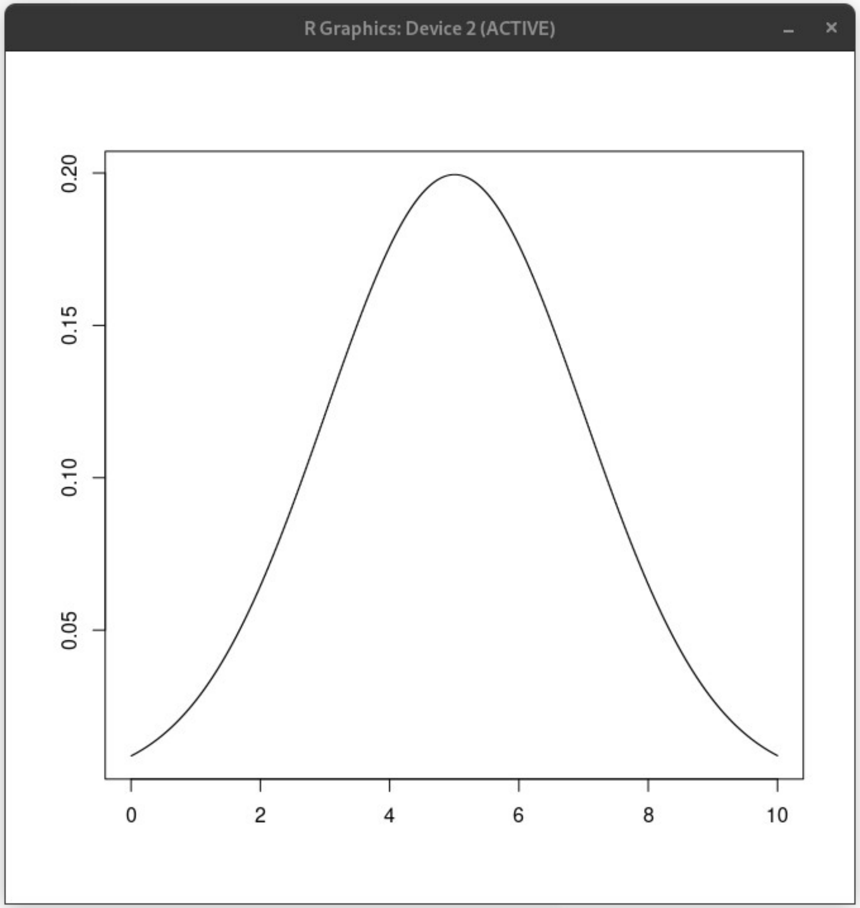
Dove p(n) è la massa di probabilità della variabile discreta.

Esempi CDF variabili discrete e continue





Distribuzione normale



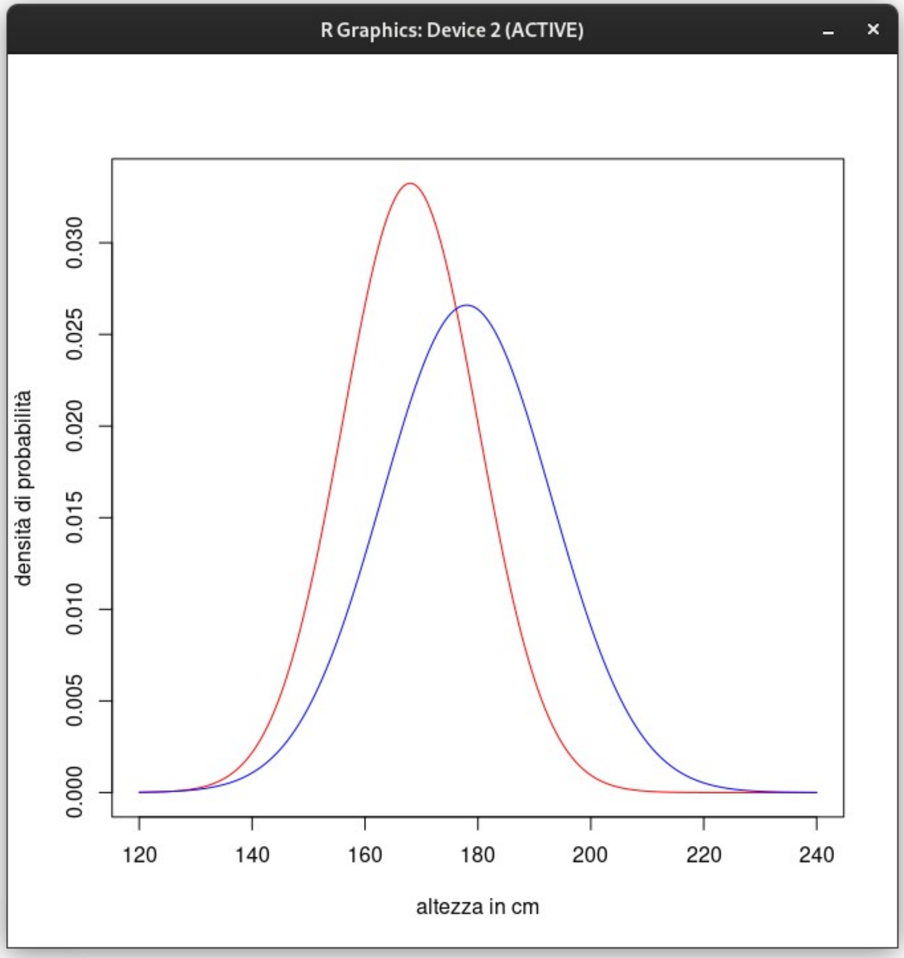
L’altezza degli Italiani

La distribuzione normale ha come PDF:

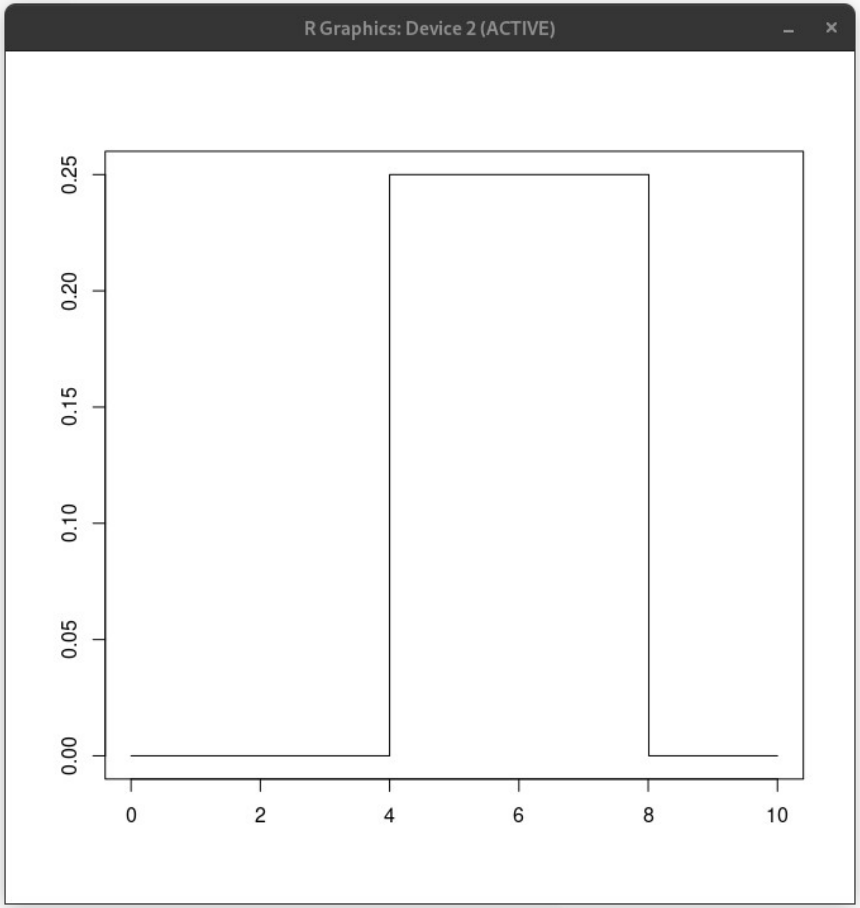
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{(2\sigma)^2}}$$

Questa distribuzione è importantissima poiché descrive tantissimi fenomeni naturali come l’altezza, il peso, la larghezza delle foglie di una pianta, il qi ecc.

L’altezza media per una donna in Italia è di 168cm con deviazione standard di 12cm, invece per gli uomini la media è di 178cm e la deviazione standard di 15cm.



Distribuzione uniforme



La funzione random()

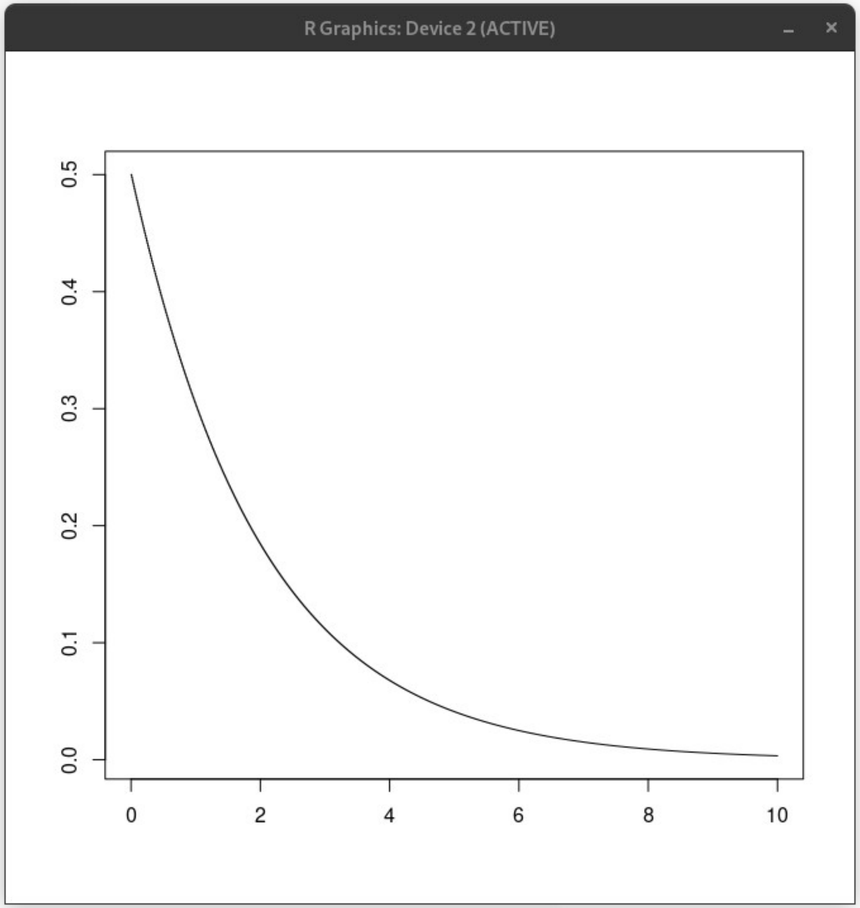
La distribuzione uniforme descrive la funzione che genera numeri casuali in tutti i linguaggi di programmazione.

In realtà però un computer non è in grado di generare numeri effettivamente casuali, infatti tutte le funzioni random() generano numeri pseudo-casuali, ovvero si basano su sequenze di numeri che sembrano casuali.

L'unico modo per ottenere numeri effettivamente casuali è basarsi su eventi fisici.

Alcuni generatori di numeri casuali per esempio si basano sul tempo al quale viene richiesto un numero casuale, prendendo l'ultima di una lunga serie di cifre significative.

Distribuzione esponenziale



Decadimento radioattivo

La distribuzione esponenziale ha come PDF:

$$f(x;\lambda)=\lambda e^{-\lambda x}$$

Questa distribuzione descrive il decadimento delle particelle radioattive.

$$N(t)=N_0 e^{-\lambda t}$$

Dove N(t) è la quantità di isotopi radioattivi ad un tempo t.

Si può utilizzare la funzione inversa per calcolare l’età degli oggetti antichi.

Il parametro lambda viene chiamato in questo caso “costante di decadimento” e la “speranza di vita” di una particella è descritta dal valor medio.

$$E(X)=\frac{1}{\lambda}$$

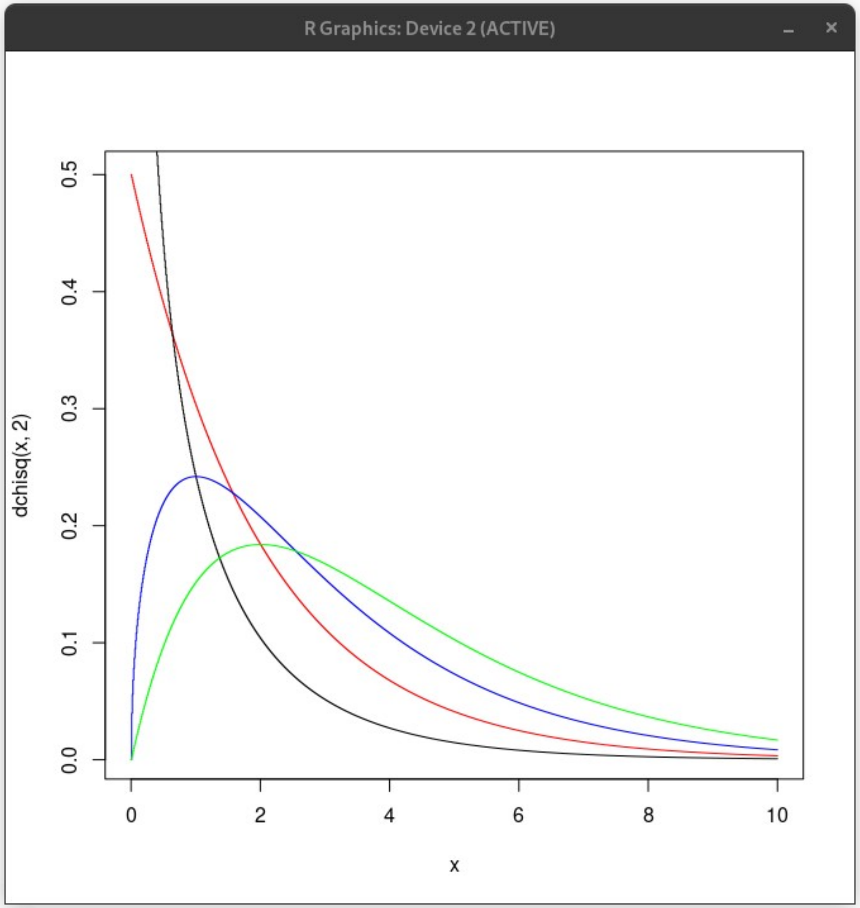
Distribuzione χ^2

La distribuzione chi-quadro con “k” gradi di libertà ha come PDF:

$$\chi_k^2 = \sum_{i=1}^{k} x_i^2 = x_1^2 + \ldots + \ldots x_k^2$$

Dove “$x_1, \ldots, x_k$“ sono variabili aleatorie indipendenti con distribuzione normale standard (media=0 e varianza = 1).

Gradi di libertà
nero = 1
rosso = 2
blu = 3
verde = 4



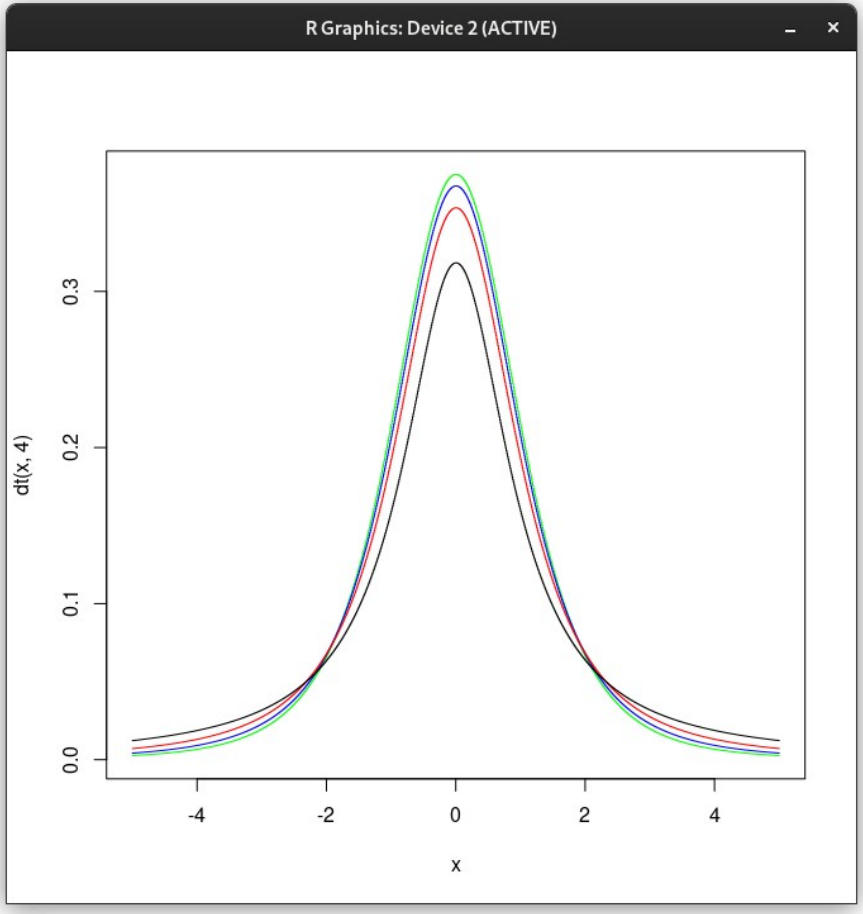
Distribuzione *t*

La distribuzione t con “n” gradi di libertà ha come PDF:

$$T_n = \frac{Z}{\sqrt{K/n}}$$

Dove “Z” e “K” sono due variabili aleatorie indipendenti che seguono rispettivamente la distribuzione normale standard (media=0 e varianza = 1) e la distribuzione chi-quadro con “n” gradi di libertà.

Gradi di libertà
nero = 1
rosso = 2
blu = 3
verde = 4

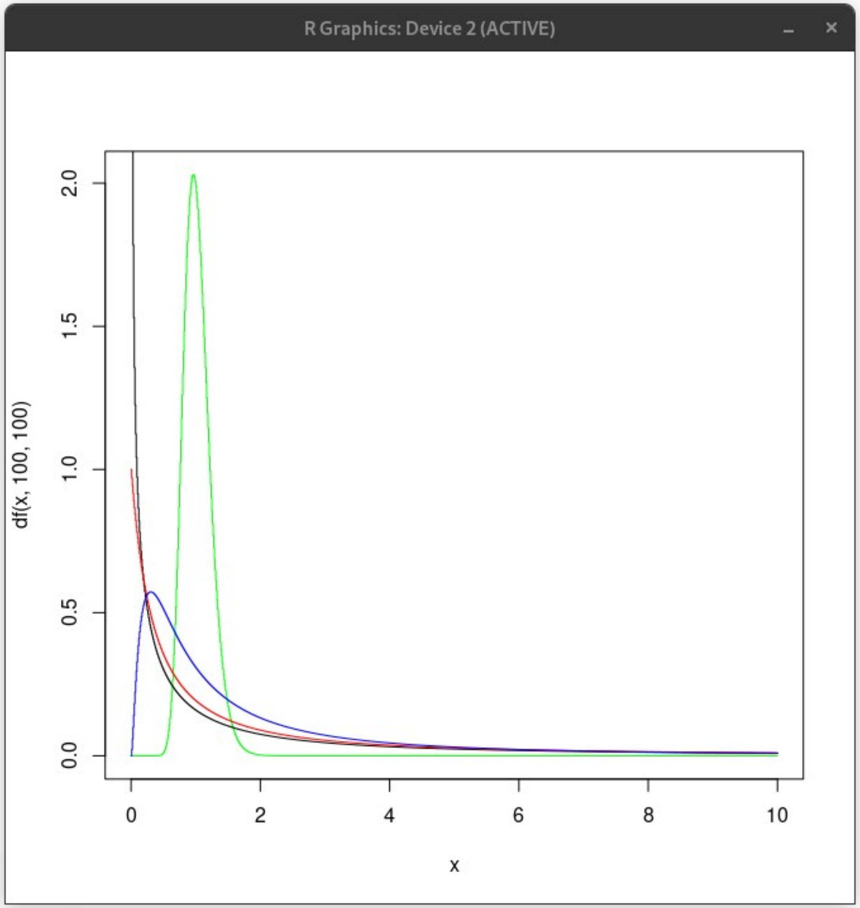


Distribuzione *F*

La distribuzione F con parametri naturali ("m", "n") ha come PDF:

$$F = \frac{X/m}{Y/n}$$

Dove "X" e "Y" sono due variabili aleatorie indipendenti che seguono la distribuzione chi-quadro con rispettivamente "m" e "n" gradi di libertà.



(m, n)
Nero = (1, 1)
Rosso = (2, 1)
Blu = (5, 2)
Verde = (100, 100)