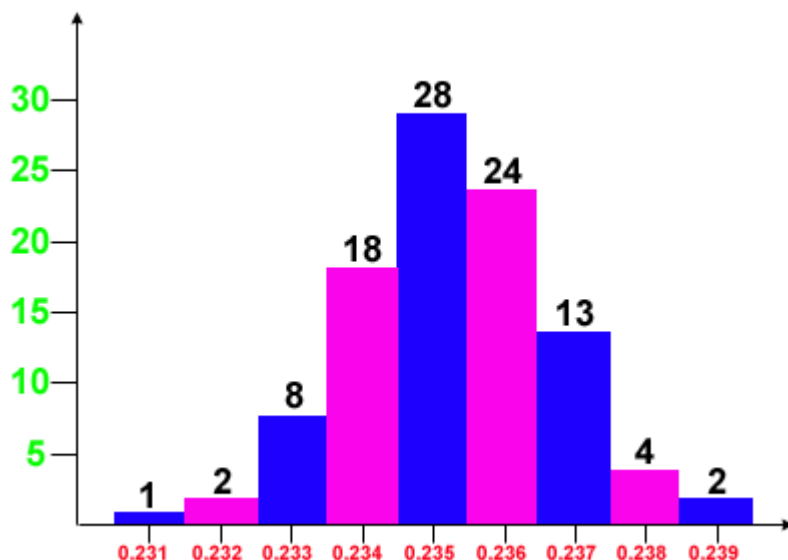


Si fa' osservare che il termine "**frequenza**" indica di solito il rapporto fra il numero di volte che si è presentato uno specifico valore osservato della variabile  $X$  e il numero totale di valori osservati, però, essa viene anche usata per indicare il numeratore di questo rapporto. Qualora nel seguito vi fossero dubbi sul significato del termine "frequenza" faremo uso rispettivamente dei termini **frequenza relativa** e **frequenza assoluta** nella tabella 1 sono riportate le frequenze assolute.

Per rappresentare graficamente i risultati della tabella 1 si usa un grafico chiamato "**istogramma**" che ora passiamo a costruire:

fig 1)



Un istogramma è un grafico come quello illustrato in figura 1 in cui si usano aree di rettangoli per rappresentare le frequenze osservate, in particolare le frequenze relative, così l'area del rettangolo centrato, ad esempio, in  $x=0.234$  dovrebbe essere uguale alla frequenza relativa 0.18. Tuttavia spesso in pratica si usa scegliere una conveniente unità di misura sull'asse y con il risultato che le aree dei rettangoli possono essere soltanto proporzionali, anziché uguali, alle corrispondenti frequenze relative. L'istogramma mostrato nella figura 1 è stato costruito con una tale conveniente scelta di unità di misura, quindi come si può osservare le aree sono soltanto proporzionali alle frequenze relative. Se l'istogramma si deve costruire in modo tale che le aree siano uguali alle frequenze relative, allora l'area totale dell'istogramma deve essere uguale a 1, perché la somma delle frequenze relative deve essere uguale a 1.

Se  $h$  indica la distanza fra valori di  $x$  consecutivi, l'altezza del rettangolo centrato, ad esempio, in  $x_i$  sarà uguale a  $\frac{f_i}{N \cdot h}$  dove  $f_i$  indica la frequenza assoluta di  $x_i$  e  $N$  rappresenta il numero totale di osservazioni. Questo risultato è ovvio quando si pensa che questa ordinata moltiplicata per la base  $h$  deve essere uguale alla frequenza relativa  $\frac{f_i}{N}$ .

L'istogramma di figura 1 indica la frequenza con cui si ottengono i vari valori della variabile  $X$  per 100 esecuzioni dell'esperimento. Se si fossero fatte 200 esecuzioni l'istogramma risultante sarebbe stato grande il doppio di quello basato su 100 esecuzioni.

Per confrontare istogrammi basati su numeri di esecuzioni dell'esperimento diversi fra loro, è necessario scegliere unità di misura sull'asse y in modo tale che l'area totale dell'istogramma sia sempre uguale a 1. Con questa scelta di unità di misura ci si dovrebbe aspettare che l'istogramma tendesse ad un istogramma fisso cioè che non cambia più all'aumentare del numero di esecuzioni dell'esperimento.

Inoltre se si fa l'ipotesi che la variabile  $X$  si possa misurare tanto accuratamente quanto si vuole così che l'unità  $h$  sull'asse  $x$  si possa rendere tanto piccola quanto si vuole allora ci si dovrebbe aspettare



che l'istogramma si smussasse e tendesse ad una curva continua quando il numero di esecuzioni aumenta indefinitamente ed  $h$  si seglie sempre più piccolo cioè tende a 0.

Quando l'area dell'istogramma è uguale a 1, segue da quanto detto, che la somma delle aree di rettangoli vicini fra loro è uguale alla frequenza relativa con cui il valore della variabile  $X$  cade nell'intervallo costituito dalle basi di quei rettangoli. Poiché questa proprietà continuerà a valere all'aumentare del numero di esecuzioni dell'esperimento, l'area sottesa dalla curva fra due qualsiasi valori di  $X$  dovrebbe essere uguale alla frequenza relativa con la quale ci si aspetterebbe che il valore della variabile  $X$  cadesse nell'intervallo determinato da quei valori. *La funzione  $f(x)$  il cui grafico è concepito come forma limite dell'istogramma, viene presa come modello matematico per la variabile casuale continua  $X$  e viene chiamata la **funzione di densità di probabilità della variabile**.*

Poiché la frequenza relativa nel caso di un istogramma è sostituita dalla probabilità, nel caso di un modello matematico la definizione di funzione di densità di probabilità di una variabile continua si può stabilire nella forma seguente:

### Definizione

**Una funzione di densità di probabilità di una variabile casuale continua  $X$  è una funzione  $f$  che soddisfa le seguenti proprietà:**

1.  $f(x) \geq 0$
2.  $\int_{-\infty}^{+\infty} f(x)dx = 1$  **che si può leggere anche come  $P\{-\infty\} < X < +\infty$**
3.  $\int_a^b f(x)dx = P\{a < X < b\}$  **dove  $a$  e  $b$  sono 2 valori qualsiasi di  $X$  con  $a < b$**

La prima proprietà è ovviamente necessaria perché la probabilità deve essere non negativa, la seconda proprietà corrisponde a richiedere che la probabilità di un evento che è certo di verificarsi deve essere uguale a 1.

Infatti certamente la variabile  $X$  assumerà un qualche valore reale quando viene fatta una sua osservazione. Nel calcolare poi le probabilità di una variabile casuale continua si richiede soltanto che la variabile giaccia in qualche intervallo. Come risultato si ha che *le probabilità delle variabili continue sono sempre fornite da integrali, mentre quelle delle variabili discrete da somme*.

Se il dominio della variabile  $X$  non è l'intera retta reale si fa l'ipotesi che la funzione  $f(x)$  sia definita nulla per i valori esterni allo specificato dominio della variabile.

### Esempio

Come esempio consideriamo la possibilità di usare la funzione  $f(x) = ke^{-x}$  dove  $k$  è una costante, come funzione di densità della variabile  $X$ . Per la prima proprietà è chiaro che la costante  $k$  deve essere positiva  $k > 0$ .

Poiché l'integrale tra  $\int_{-\infty}^{+\infty} e^{-x} dx = -e^{-x} \Big|_{-\infty}^{+\infty} = 0 + e^{+\infty}$  segue che la variabile  $X$  deve essere limitata, quindi facciamo l'ipotesi per esempio che  $X$  possa assumere soltanto valori non negativi, cioè  $x \geq 0$ . Allora la  $f(x)$  sarà definita *nulla per valori negativi*, e *data dalla formula suddetta per valori non negativi*. Per la seconda proprietà segue allora che  $k$  deve essere uguale a 1, poiché

$$\int_0^{+\infty} e^{-x} dx = -e^{-x} \Big|_0^{+\infty} = 0 + 1 = 1$$

Si conclude quindi può essere una funzione di densità della variabile  $x$  la funzione

$$\begin{cases} f(x) = e^{-x}, x \geq 0 \\ f(x) = 0, x < 0 \end{cases}$$

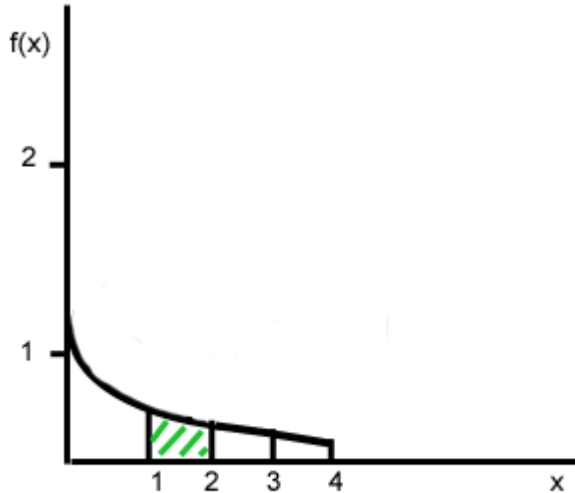


Il calcolo ad esempio della probabilità che  $X$  sia compreso tra 1 e 2 diventa

$$P\{1 < X < 2\} = \int_1^2 e^{-x} dx = -e^{-x} \Big|_1^2 = -e^{-2} + e^{-1} \approx 0.23$$

Il grafico di questa funzione di densità e la rappresentazione come area della probabilità che  $X$  sia compreso tra 1 e 2 sono mostrati nella figura seguente

fig 2)



È importante sottolineare che sebbene in qualsiasi dato problema la funzione  $f(x)$  si possa scegliere a piacimento, naturalmente secondo le conoscenze di cui dispone lo sperimentatore, una scelta per cui le probabilità risultanti non approssimano bene le frequenze relative osservate non è certamente una scelta utile.

