STATISTICA E ANALISI DEI DATI

Seconda Parte

Dott. Stefano Cirillo
Dott. Luigi Di Biasi

a.a. 2025-2026

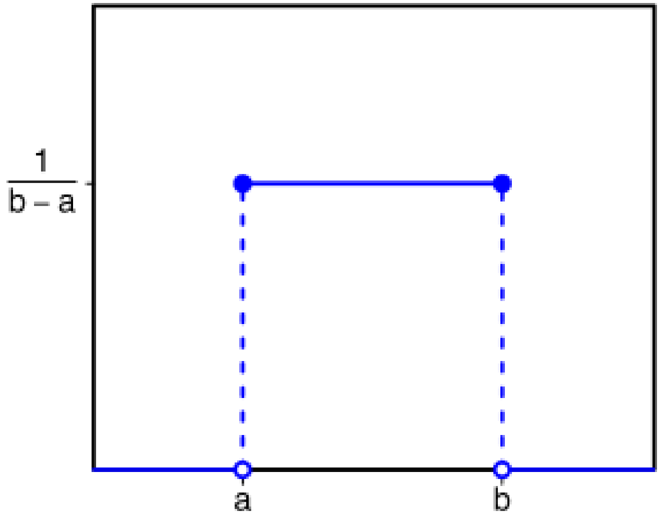
Densità di Probabilità

- Le **variabili aleatorie continue** sono variabili il cui valore può assumere qualsiasi numero reale all'interno di un intervallo (limitato o illimitato)
- Una variabile casuale continua X è caratterizzata dalla sua **funzione di densità di probabilità**, $f(x)$ (o $p(x)$) tale che l'area sottesa alla funzione in un intervallo $[a; b]$, è pari alla probabilità che X assuma un valore in quell'intervallo:

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

oppure $\quad P(a \leq X \leq b) = \int_a^b p(x)\, dx$

dove $p(x)$ è la funzione di probabilità



- La funzione di densità deve verificare le proprietà:

1) $f(x) \geq 0$ $\qquad$ 2) $\int_{-\infty}^{+\infty} f(x)\, dx = 1$

Funzione di Distribuzione Cumulativa

- La funzione di distribuzione cumulativa, indicata con $F(x)$ descrive il comportamento probabilistico di una variabile aleatoria, sia essa discreta o continua.
- $F(x)$ fornisce la probabilità cumulativa fino a un certo valore x ed è definita come:

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

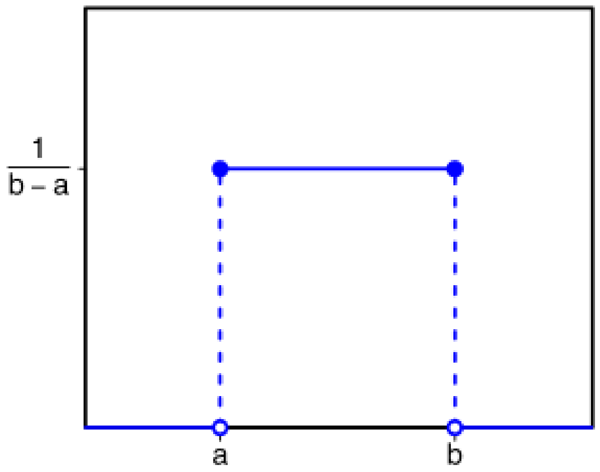
dove $f(x)$ è la funzione di probabilità di X

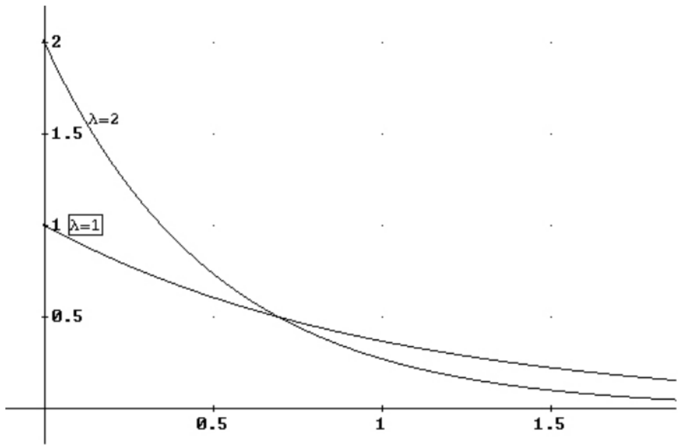
- **Proprietà:**
  1. $F(x)$ è una funzione monotona non decrescente, ovvero: $F(x_1) \leq F(x_2)$ per ogni $x_1 \leq x_2$
  2. I valori di $F(x)$ variano tra 0 e 1, significa che $\lim_{x \to -\infty} F(x) = 0$ oppure $\lim_{x \to \infty} F(x) = 1$
  3. La probabilità che X assuma un valore in un intervallo $[a; b]$ si ottiene come differenza tra i valori di $F(x)$ ai limiti dell'intervallo:
     - Inoltre poiché $F(X) = P(X \leq x)$ e $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$ si ha che:

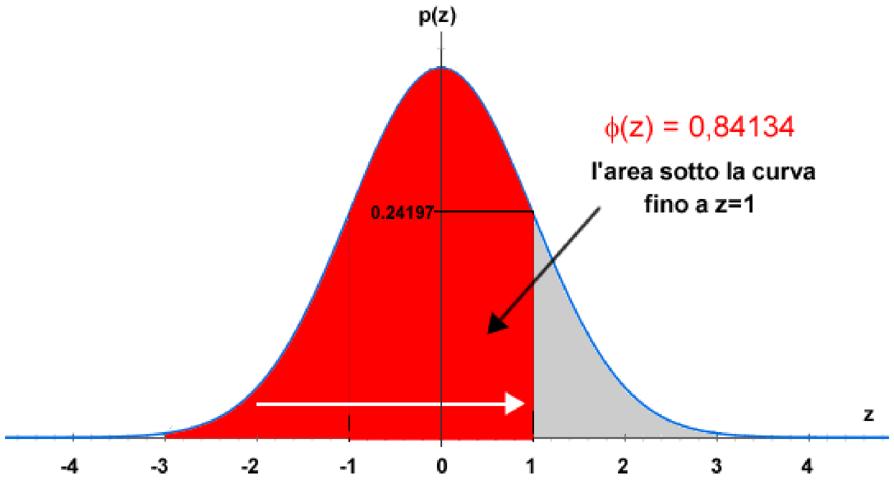
$$P(a \leq X \leq b) = F(b) - F(a)$$

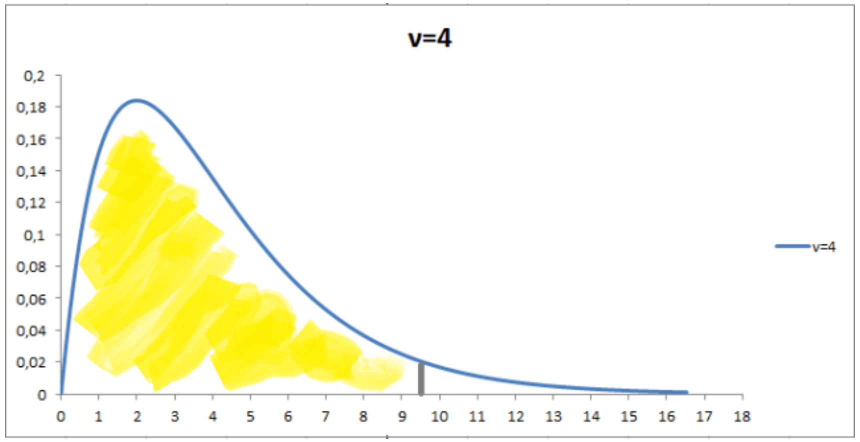
Variabili Aleatorie Continue

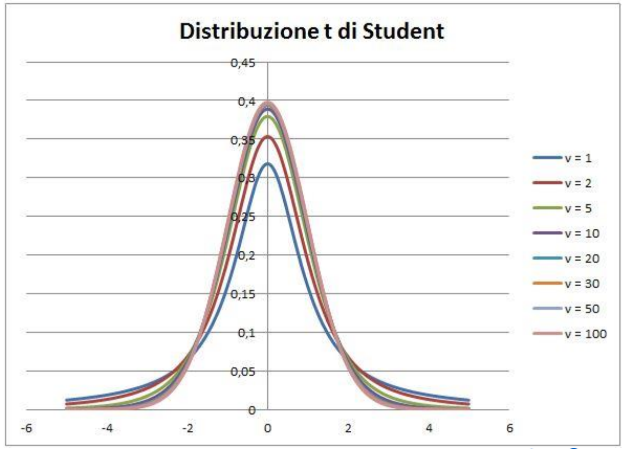
- Nelle prossime lezioni discuteremo delle seguenti distribuzioni discrete:
  - Distribuzione uniforme;
  - Distribuzione esponenziale;
  - Distribuzione normale;
  - Distribuzione chi–quadrato;
  - Distribuzione di Student.









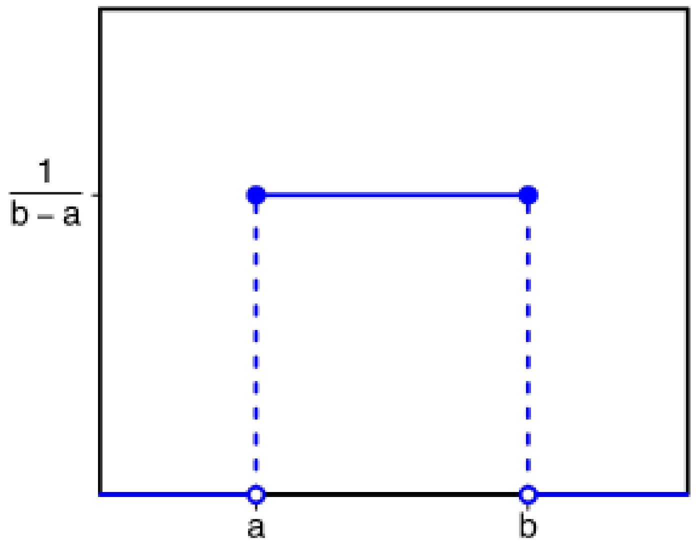


STATISTICA E ANALISI DEI DATI

Distribuzione Uniforme

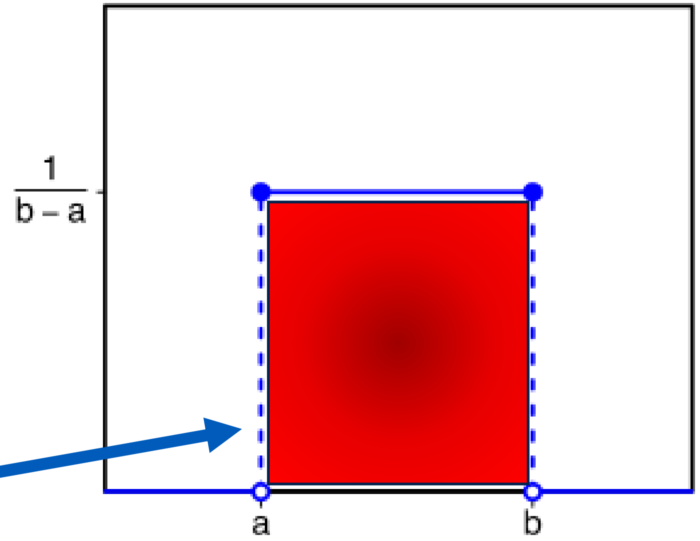
Densità di Probabilità Uniforme

- Si consideri la variabile aleatoria X che può assumere tutti i valori dell'intervallo reale $[a; b]$ con $a < b$ in modo che tutti i sottointervalli di uguale ampiezza abbiano la stessa probabilità
  - La **densità di probabilità** è: $f(x) = \begin{cases} \frac{1}{b-a} & se\ x \in [a; b] \\ 0 & altrimenti \end{cases}$



Densità di Probabilità Uniforme

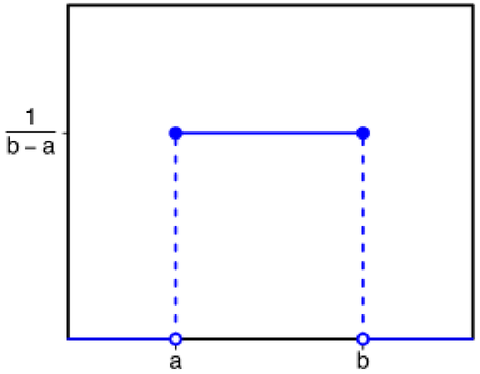
- Si consideri la variabile aleatoria X che può assumere tutti i valori dell'intervallo reale $[a; b]$ con $a < b$ in modo che tutti i sottointervalli di uguale ampiezza abbiano la stessa probabilità
  - La **densità di probabilità** è: $f(x) = \begin{cases} \frac{1}{b-a} & se\ x \in [a; b] \\ 0 & altrimenti \end{cases}$



- Perché $\frac{1}{b-a}$?
  - La probabilità totale per una variabile aleatoria deve essere uguale a 1
    - Ciò significa che l'area sotto l'intervallo $[a; b]$ è 1
  - Base = $b - a$ (la lunghezza dell'intervallo)
  - Altezza = $f(x)$
    - $f(x) * (b - a) = 1 \qquad \Rightarrow f(x) = \frac{1}{(b-a)}$

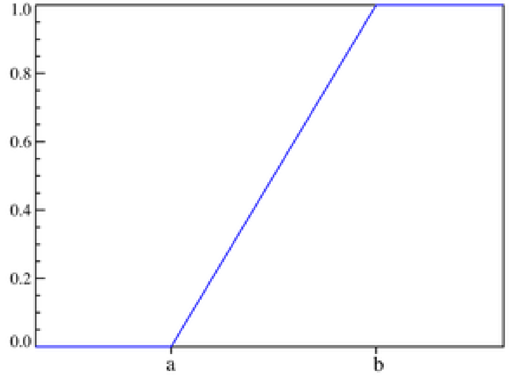
Distribuzione Uniforme

- Si consideri la variabile aleatoria X che può assumere tutti i valori dell'intervallo reale $[a; b]$ con $a < b$ in modo che tutti i sottointervalli di uguale ampiezza abbiano la stessa probabilità



  - La **funzione di distribuzione** è: $F_X(x) = P(X \leq x) = \begin{cases} 0 & se\ x < a \\ \frac{x-a}{b-a} & se\ x \in [a; b] \\ 1 & se\ x > b \end{cases}$

si dice uniformemente distribuita nell'intervallo $[a; b]$



$$F_X(x) = \frac{distanza\ di\ x\ dall'inizio\ dell'intervallo}{lunghezza\ totale\ intervallo} = \frac{x-a}{b-a}$$

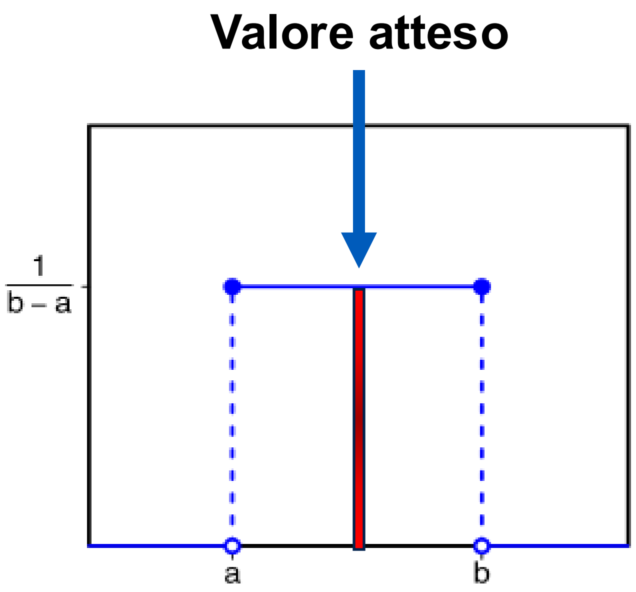
La densità uniforme continua traduce l'idea di equiprobabilità del caso discreto distribuendo la probabilità in modo uniforme sull'intero intervallo $[a, b]$, rendendo ogni sottointervallo di uguale ampiezza equiprobabile, anche se la probabilità puntuale $P(X = x)$ per ogni singolo valore è zero

Valore Atteso

- La distribuzione uniforme è utilizzata in qualsiasi situazione in cui si sceglie un valore "a caso" in un fissato intervallo, senza alcuna preferenza per valori inferiori, superiori o medi nell'intervallo
- **Notazione**:
  - $X \sim U(a, b)$ intenderemo che X è una variabile aleatoria avente distribuzione uniforme nell'intervallo (a, b)
- Il valore medio della distribuzione uniforme risulta essere:

$$E(X) = \int_a^b x * f(x)\, dx \Rightarrow \int_a^b x * \frac{1}{b-a} dx \Rightarrow \frac{1}{b-a} \int_a^b x \; dx \Rightarrow \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b$$

$$= \frac{1}{b-a} \left(\frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{a+b}{2}$$

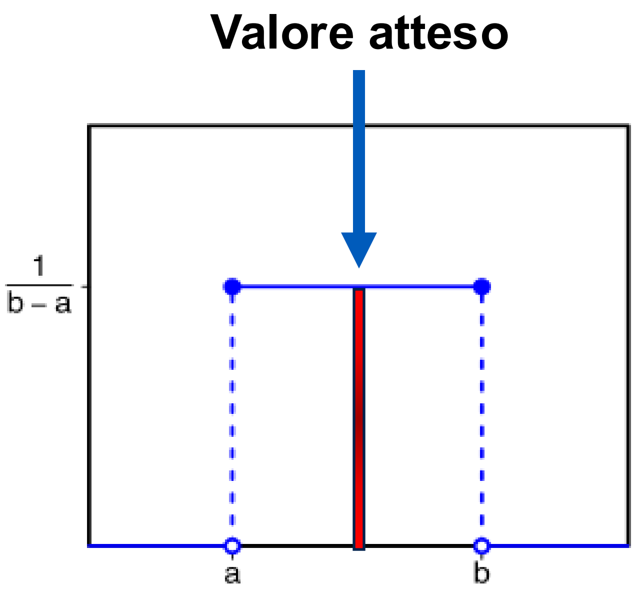


Valore Atteso

- La distribuzione uniforme è utilizzata in qualsiasi situazione in cui si sceglie un valore "a caso" in un fissato intervallo, senza alcuna preferenza per valori inferiori, superiori o medi nell'intervallo
- **Notazione**:
  - $X \sim U(a, b)$ intenderemo che X è una variabile aleatoria avente distribuzione uniforme nell'intervallo (a, b)
- Il valore medio della distribuzione uniforme risulta essere:

$$E(X) = \frac{a+b}{2}$$

$$E(X^2) = \int_a^b x^2 * f(x)\, dx \Rightarrow \frac{1}{b-a}\int_a^b x^2\, dx = \frac{a^2+ab+b^2}{3}$$



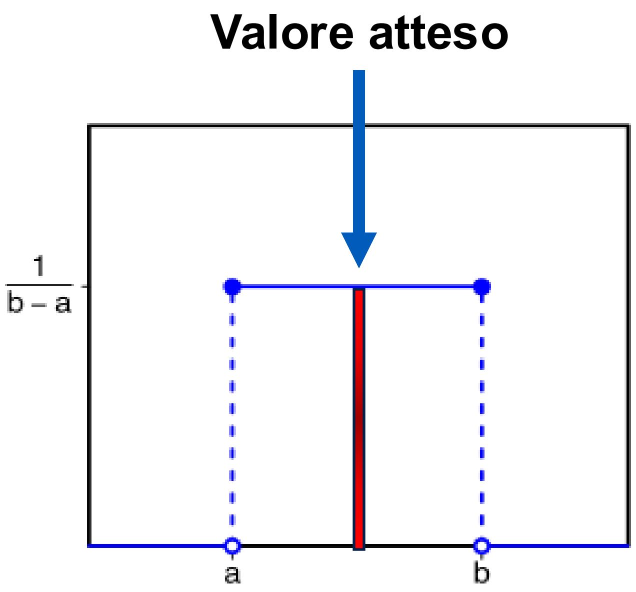
Varianza

- La distribuzione uniforme è utilizzata in qualsiasi situazione in cui si sceglie un valore “a caso” in un fissato intervallo, senza alcuna preferenza per valori inferiori, superiori o medi nell’intervallo
- **Notazione**:
  - $X \sim U(a, b)$ intenderemo che X è una variabile aleatoria avente distribuzione uniforme nell’intervallo (a, b)
- Il valore medio e la **varianza** della distribuzione uniforme risultano essere:

$$E(X) = \frac{a+b}{2}$$

$$E(X^2) = \frac{a^2 + ab + b^2}{3}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$



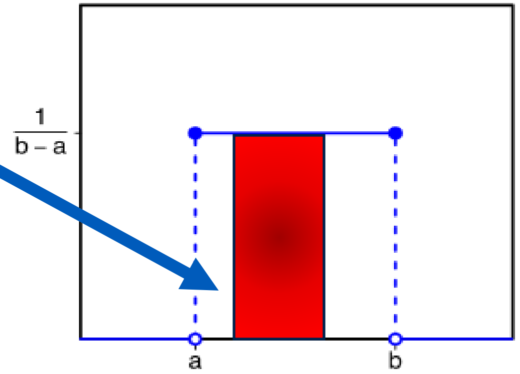
Esempio

- Immaginiamo di avere una variabile casuale X che rappresenta la posizione di un punto casuale su un segmento di lunghezza 5, dove il punto può trovarsi ovunque tra $[0;5]$
- La densità di probabilità uniforme in questo caso sarebbe:

$$f(x) = \begin{cases} \frac{1}{b-a} & se\ x \in [a;b] \\ 0 & altrimenti \end{cases} = \begin{cases} \frac{1}{5-0} & se\ x \in [0;5] \\ 0 & altrimenti \end{cases}$$

  - Questo significa che ogni punto all'interno dell'intervallo $[0;5]$ ha la stessa probabilità di essere scelto
- Se volessimo calcolare la probabilità che X cada nell'intervallo $[1;3]$, <u>potremmo calcolare l'area sotto la curva</u> della densità di probabilità in quell'intervallo, che sarebbe semplicemente l'ampiezza dell'intervallo moltiplicata per l'altezza costante della densità di probabilità:

$$P(a \leq X \leq b) = P(1 \leq X \leq 3) = \int_1^3 f(x)\,dx = \frac{1}{5} * (3-1) = \frac{2}{5} = 0.4$$



Distribuzione Uniforme

- Per il calcolo della <u>densità uniforme</u> si utilizza la funzione:

```
dunif(x, min=a, max=b)
```

dove

- **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria uniforme;
- **min** e **max** sono il minimo e il massimo dell'intervallo in cui la densità uniforme è positiva; se il minimo ed il massimo non sono specificati essi per default assumono i valori 0 e 1

Distribuzione Uniforme

- Per il calcolo della densità uniforme si utilizza la funzione:

```
dunif(x, min=a, max=b)
```

dove

- **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria uniforme;
- **min** e **max** sono il minimo e il massimo dell'intervallo in cui la densità uniforme è positiva; se il minimo ed il massimo non sono specificati essi per default assumono i valori 0 e 1

- Per calcolare la funzione di distribuzione uniforme invece utilizziamo la funzione:

```
punif(x, min=a, max=b, lower.tail = TRUE)
```

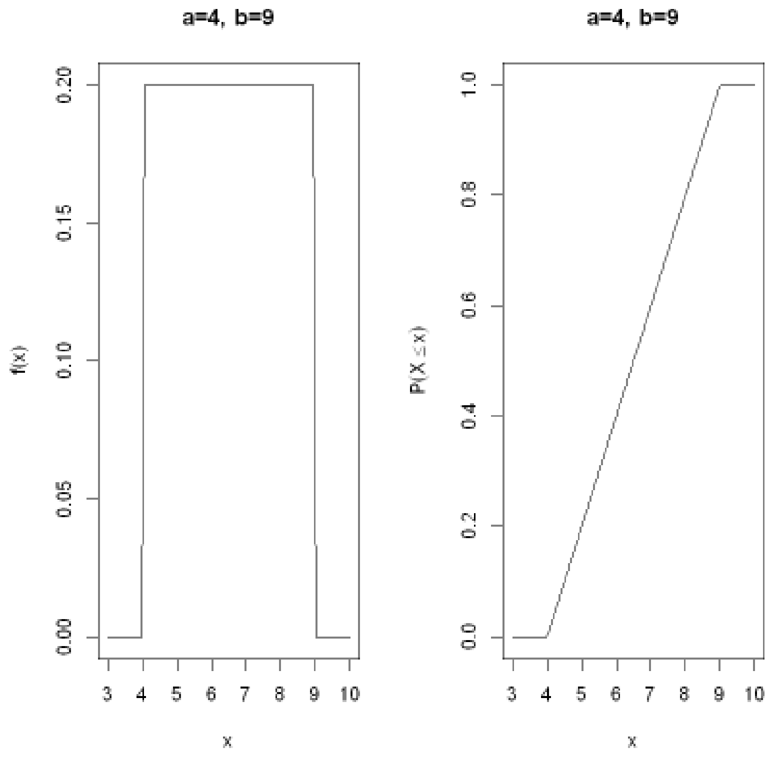
dove

- **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria uniforme;
- **min** e **max** sono il minimo e il massimo dell'intervallo in cui la densità uniforme è positiva;
- **lower.tail** se tale parametro è TRUE (caso di default) calcola $P(X \leq x)$, mentre se tale parametro è FALSE calcola $P(X > x)$

Esempio (R)

- **Esempio**:
  - Calcoliamo la densità di probabilità e la funzione di distribuzione di una variabile aleatoria $X \sim U(4,9)$

```
>par(mfrow=c(1,2))
>curve(dunif(x,min=4,max=9),from=3, to=10,xlab="x",
+ylab="f(x)",main="a=4,b=9")
>
>curve(punif(x,min=4,max=9),from=3, to=10,xlab="x",
+ylab=expression(P(X<=x)),main="a=4,b=9")
```



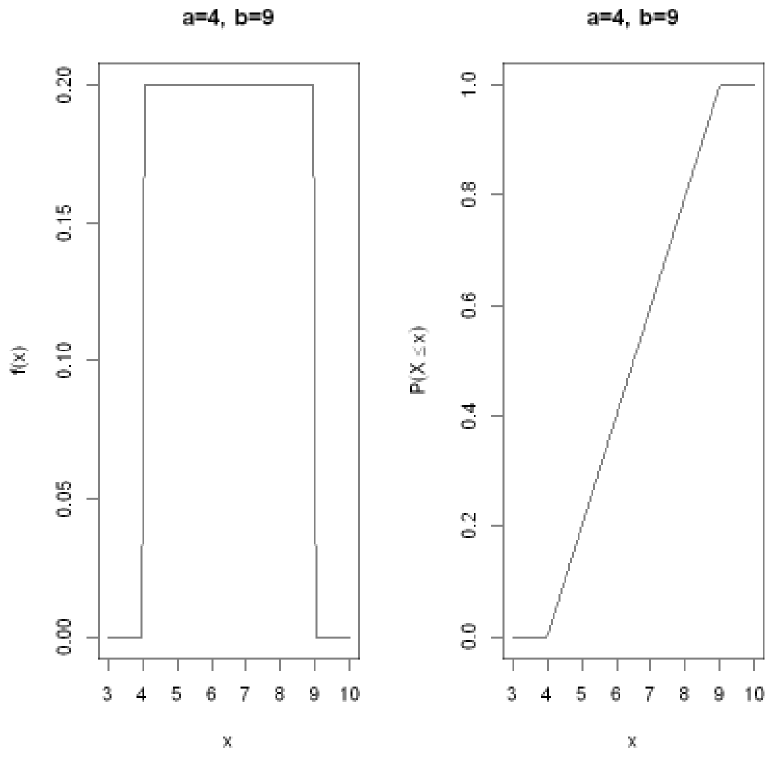
Distribuzione Uniforme

- **Esempio**:
  - Calcoliamo la densità di probabilità e la funzione di distribuzione di una variabile aleatoria $X \sim U(4,9)$

```
>par(mfrow=c(1,2))
>curve(dunif(x,min=4,max=9),from=3, to=10,xlab="x",
+ylab="f(x)",main="a=4,b=9")
>
>curve(punif(x,min=4,max=9),from=3, to=10,xlab="x",
+ylab=expression(P(X<=x)),main="a=4,b=9")
```

x non è definito

x non è definito



Distribuzione Uniforme

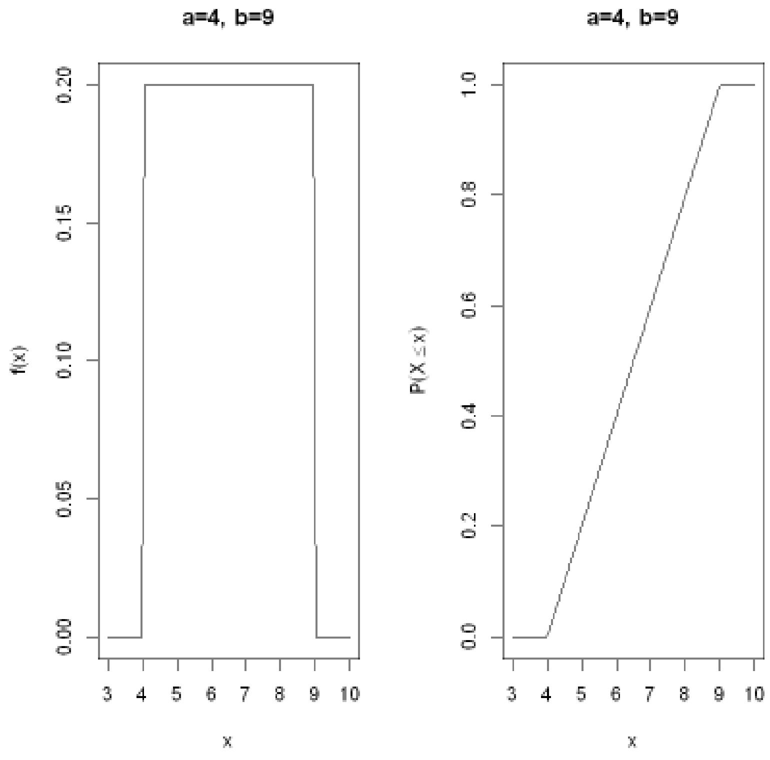
- **Esempio**:
  - Calcoliamo la densità di probabilità e la funzione di distribuzione di una variabile aleatoria $X \sim U(4,9)$

```
>par(mfrow=c(1,2))
>curve(dunif(x,min=4,max=9),from=3, to=10 xlab="x",
+ylab="f(x)",main="a=4,b=9")
>
>curve(punif(x,min=4,max=9),from=3, to=10 xlab="x",
+ylab=expression(P(X<=x)),main="a=4,b=9")
```

x non è definito

Assume i valori nel range considerato

x non è definito



Distribuzione Uniforme

- **Esempio**:
  - Calcoliamo la densità di probabilità e la funzione di distribuzione di una variabile aleatoria $X \sim U(4,9)$

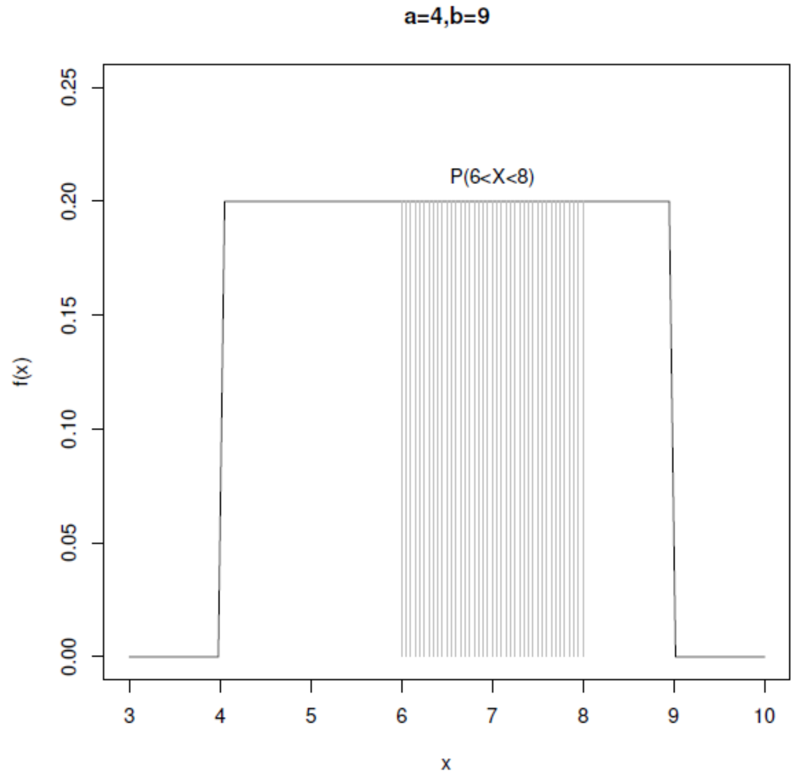
```
>par(mfrow=c(1,2))
>curve(dunif(x,min=4,max=9),from=3, to=10,xlab="x",
+ylab="f(x)",main="a=4,b=9")
>
>curve(punif(x,min=4,max=9),from=3, to=10,xlab="x",
+ylab=expression(P(X<=x)),main="a=4,b=9")
```

  - La probabilità che la variabile aleatoria $X \sim U(4,9)$ assuma valori nell'intervallo $(6,8)$ è:

$$P(6 \leq X \leq 8) = \frac{8-6}{9-4} = \frac{2}{5} = 0.4$$

```
>curve(dunif(x,min=4,max=9),from=3, to=10,xlab="x",
+ylab="f(x)",main="a=4,b=9",ylim=c(0,0.25))
>x<-seq(6,8,0.05)
>lines(x,dunif(x,min=4,max=9),type="h",col="grey")
>text(7.0,0.21,"P(6<X<8)")
>
> punif(8,min=4,max=9)-punif(6,min=4,max=9)
[1] 0.4
```

Si calcola la differenza tra le probabilità di 8 e 6



Quantili

- Per calcolare i quantili (percentili) della distribuzione uniforme

```
qunif(z, min=a, max=b)
```

dove

  - ***z*** è il valore assunto (o i valori assunti) dalle probabilità relative al percentile $z \cdot 100$–esimo, cioè il vettore delle probabilità;
  - **min** e **max** sono il minimo e il massimo dell'intervallo in cui la densità uniforme è positiva

- Il risultato della funzione è il percentile $z \cdot 100$–esimo, ossia il più piccolo numero x assunto dalla variabile aleatoria uniforme X tale che

$$F_X(x) = P(X \leq x) \geq z \qquad (a \leq x \leq b)$$

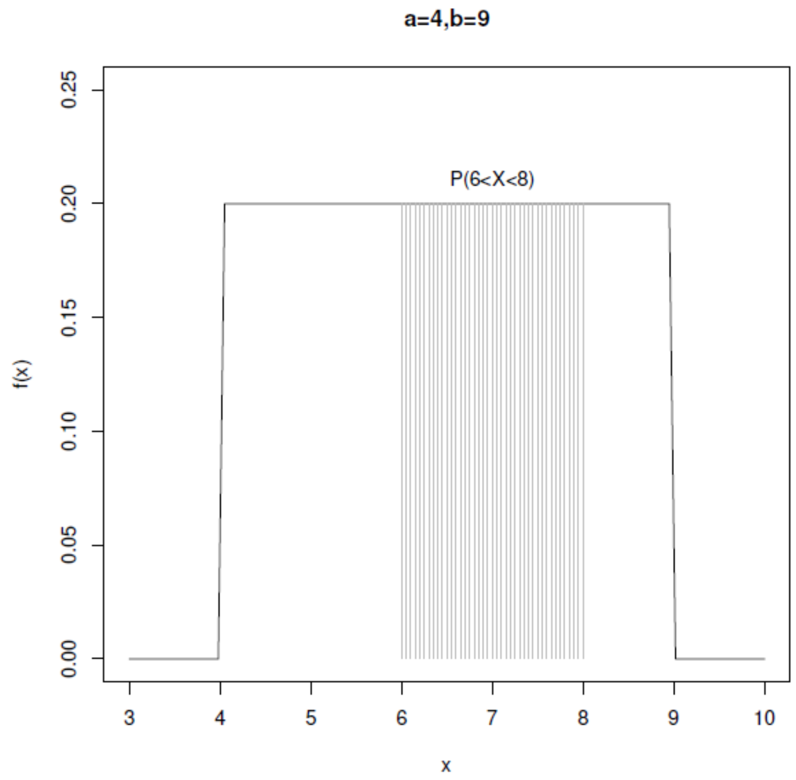
Quantili

- **Esempio**: Calcoliamo i quartili Q_0, Q_1, Q_2, Q_3, Q_4 di $X \sim U(4,9)$:

```
>z<-c(0,0.25,0.5,0.75,1)
>qunif(z, min=4, max=9)
[1] 4.00 5.25 6.50 7.75 9.00
```

- Ricordando che $P(X \leq x) = \frac{(x-a)}{(b-a)} = \frac{(x-4)}{(9-4)}$ se $4 \leq x < 9$ si ha:

$$P(X \leq x) = \frac{x-4}{5} \geq 0.25 \iff x \geq 4 + 5 \cdot 0.25 = 5.25,$$
$$P(X \leq x) = \frac{x-4}{5} \geq 0.50 \iff x \geq 4 + 5 \cdot 0.50 = 6.50,$$
$$P(X \leq x) = \frac{x-4}{5} \geq 0.75 \iff x \geq 4 + 5 \cdot 0.75 = 7.75.$$



Generazione Randomica di Numeri

- È possibile simulare la variabile aleatoria uniforme nell'intervallo (a, b) generando una sequenza di numeri pseudocasuali mediante la funzione

```
runif(N, min=a, max=b)
```

dove

  - **N è** lunghezza della sequenza da generare;
  - **min** e **max** sono il minimo e il massimo dell'intervallo in cui la densità uniforme è positiva

- **Esempio**: Generiamo una sequenza di 10000 numeri pseudocasuali simulando una variabile aleatoria uniforme $X \sim U(4, 9)$:

```
> sim<-runif(10000,min=4,max=9)
> mean(sim)
[1] 6.482359
> var(sim)
[1] 2.062938
```

- **Valore atteso**: $E(X) = \frac{a+b}{2} = \frac{9+4}{2} = 6.5$ **Varianza**: $Var(X) = \frac{(b-a)^2}{12} = \frac{(9-4)^2}{12} = 2.083$

Generazione Randomica di Numeri

- È possibile simulare la variabile aleatoria uniforme nell'intervallo (a, b) generando una sequenza di numeri pseudocasuali mediante la funzione

```
runif(N, min=a, max=b)
```

dove

- **N è** lunghezza della sequenza da generare;
- **min** e **max** sono il minimo e il massimo dell'intervallo in cui la densità uniforme è positiva

- **Esempio**: Generiamo una sequenza di 10000 numeri pseudocasuali simulando una variabile aleatoria uniforme $X \sim U(4,9)$:

```
> sim<-runif(10000,min=4,max=9)
> mean(sim)
[1] 6.482359
> var(sim)
[1] 2.062938
```

Valori Simulati

Valori teorici

- **Valore atteso**: $E(X) = \frac{a+b}{2} = \frac{9+4}{2} = \boxed{6.5}$

Varianza: $Var(X) = \frac{(b-a)^2}{12} = \frac{(9-4)^2}{12} = \boxed{2.083}$

Esempio (i)

- **Esempio**:
  - Due clienti A e B entrano contemporaneamente in un supermercato.
  - Supponendo che il tempo impiegato per fare la spesa sia una variabile aleatoria uniformemente distribuita nell'intervallo (10,20) per A e nell'intervallo (15,25) per B e che siano tra loro indipendenti, **si desidera determinare la probabilità che B finisca la spesa dopo di A**

Esempio (i)

- **Esempio**:
  - Due clienti A e B entrano contemporaneamente in un supermercato.
  - Supponendo che il tempo impiegato per fare la spesa sia una variabile aleatoria uniformemente distribuita nell'intervallo (10,20) per A e nell'intervallo (15,25) per B e che siano tra loro indipendenti, **si desidera determinare la probabilità che B finisca la spesa dopo di A**
  - Sia $X \sim U(10, 20)$ variabile aleatoria che descrive il tempo per fare la spesa di A
  - Sia $Y \sim U(15, 25)$ variabile aleatoria che descrive il tempo per fare la spesa di B

Esempio (i)

- **Esempio**:
  - Due clienti A e B entrano contemporaneamente in un supermercato.
  - Supponendo che il tempo impiegato per fare la spesa sia una variabile aleatoria uniformemente distribuita nell’intervallo (10,20) per A e nell’intervallo (15,25) per B e che siano tra loro indipendenti, **si desidera determinare la probabilità che B finisca la spesa dopo di A**
  - Sia $X \sim U(10, 20)$ variabile aleatoria che descrive il tempo per fare la spesa di A
  - Sia $Y \sim U(15, 25)$ variabile aleatoria che descrive il tempo per fare la spesa di B
  - **Condizione** dell'evento $Y > X$:
    - Dobbiamo determinare la regione nel piano dove $15 \leq Y \leq 25$ e $10 \leq X \leq 20$, e $Y > X$

Esempio (i)

- **Esempio**:
  - Siano $X \sim U(10,20)$ e $Y \sim U(15,25)$ variabili aleatorie che descrivono il tempo per fare la spesa di A e B
  - Per trovare l'intersezione tra $Y > X$, dobbiamo considerare che:

$$Y > X \text{ è valido solo per } x \leq y$$

  - L'intervallo di sovrapposizione tra X e Y è [15,20]
  - La probabilità richiesta può essere calcolata come:

$$P(Y > X) = \int\int_{\mathcal{D}} \boxed{f_{XY}(x,y)}\, dx\, dy = \int\int_{\mathcal{D}} \boxed{f_X(x)\, f_Y(y)}\, dx\, dy$$

Densità Congiunta

dove il dominio D è definito:

$$\mathcal{D} = \{(x,y) : 10 < x < 20, 15 < y < 25, y > x\}$$

Densità Congiunta

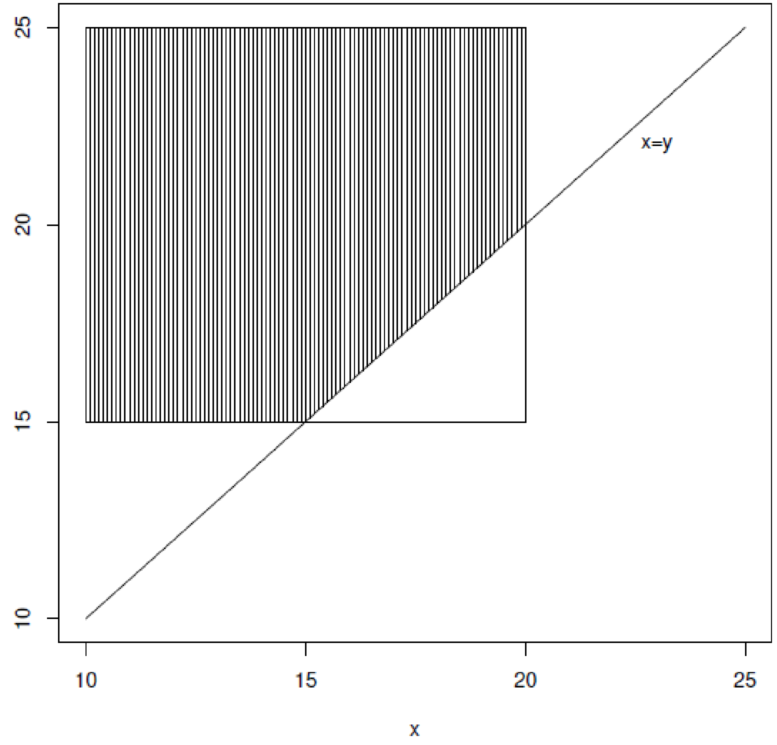
- La **densità congiunta** di due o più variabili aleatorie è una funzione che descrive la probabilità congiunta di ottenere specifici valori per tutte le variabili coinvolte
- Nel caso di due variabili aleatorie continue la densità congiunta è spesso denotata come $f_{X,Y}(x, y)$, dove X e Y sono le variabili aleatorie in questione
- Se consideriamo due variabili aleatorie continue indipendenti e ciascuna segue una distribuzione uniforme
  - Supponiamo che X sia uniformemente distribuita nell'intervallo $[a, b]$ e Y sia uniformemente distribuita nell'intervallo $[c, d]$, dove $a < b$ e $c < d$
  - La **densità di probabilità** uniforme per ciascuna variabile è **costante** all'interno del proprio intervallo e zero al di fuori
  - La densità congiunta di X e Y sarà data dalla moltiplicazione delle densità di probabilità delle due variabili, poiché sono indipendenti:

$$f_{X,Y}(x, y) = f_X(x) * f_Y(y)$$

Esempio (i)

- **Esempio**:
  - Rappresentiamo il dominio D:

```
>plot(10:25,10:25,xlab="x",ylab="y",type="l")
>text(23,22,"x=y")
>rect(10,15,20,25)
>x1<-seq(10,15,0.1)
>segments(x1,15,x1,25)
>x2<-seq(15,20,0.1)
>segments(x2,x2,x2,25)
```

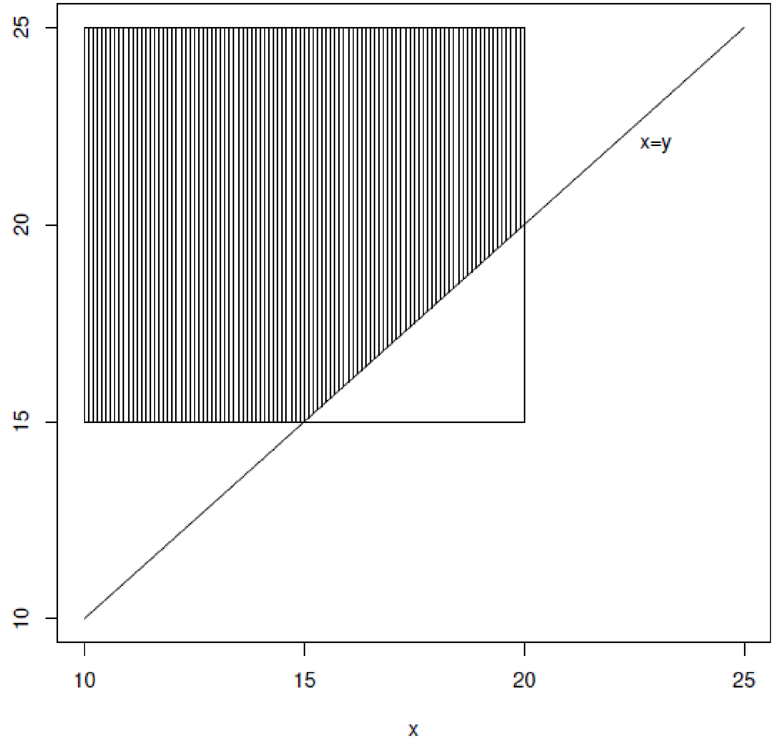


  - Pertanto, tale probabilità può essere calcolata moltiplicando la densità congiunta per l’area tratteggiata rappresentata

Esempio

- Sappiamo che:
  - X sia uniformemente distribuita nell'intervallo $[10, 20]$
  - Y sia uniformemente distribuita nell'intervallo $[15, 25]$
- La densità di probabilità uniforme:
  - $f_X(x) = \frac{1}{20-10} = \frac{1}{10}. \quad x \in [10; 20]$
  - $f_Y(y) = \frac{1}{25-15} = \frac{1}{10}. \quad y \in [15; 25]$
- La **densità congiunta** di X e Y è:

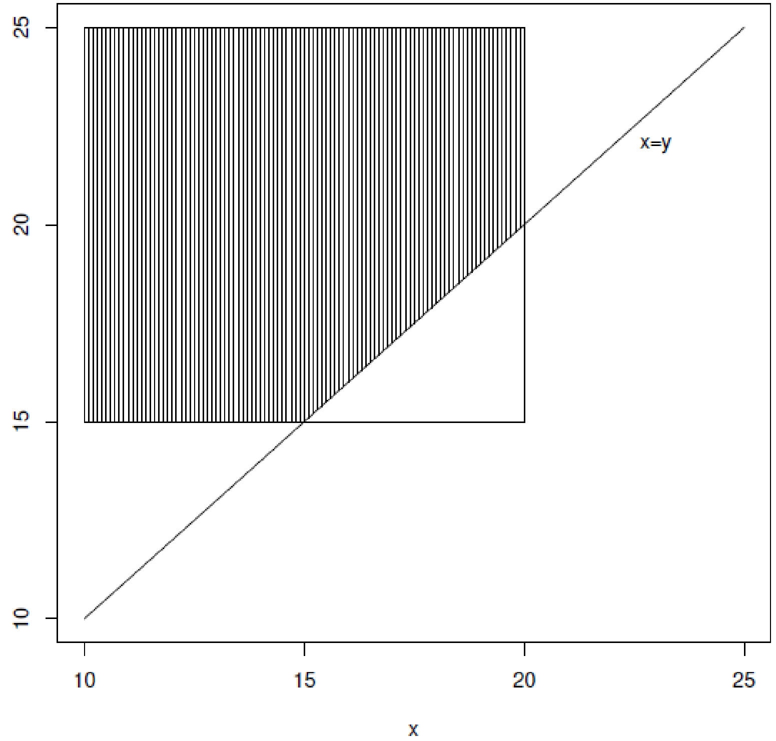
$$f_{X,Y}(x, y) = f_X(x) * f_Y(y) = \frac{1}{10} * \frac{1}{10} = \frac{1}{100}$$



Esempio

- **Esempio**:
  - Pertanto, tale probabilità può essere calcolata moltiplicando la densità congiunta per l’area tratteggiata rappresentata:

$$P(Y > X) = \frac{1}{100}\left[100 - \frac{25}{2}\right] = \frac{7}{8} = 0.875.$$



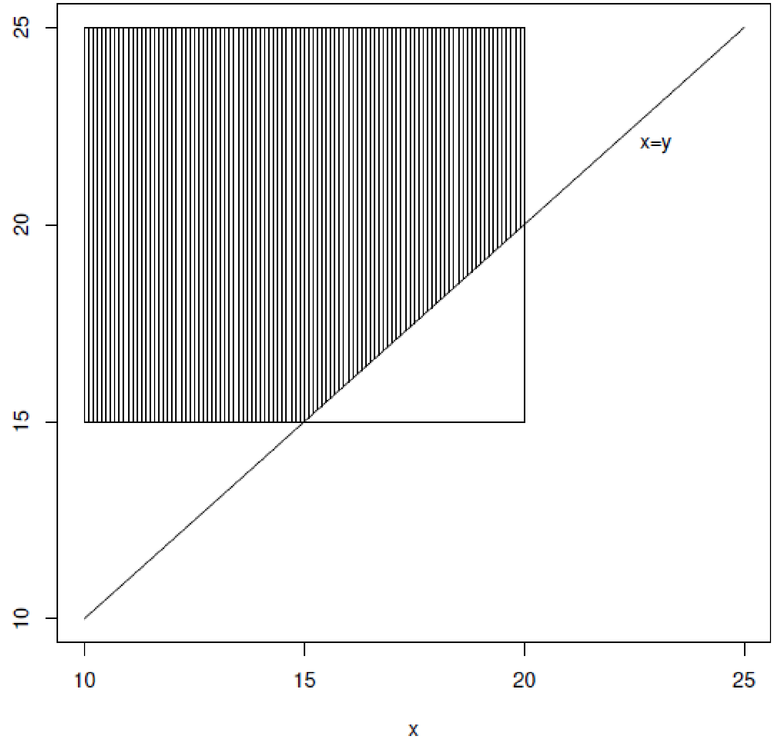
Esempio

- **Esempio**:
  - Pertanto, tale probabilità può essere calcolata moltiplicando la densità congiunta per l'area tratteggiata rappresentata:

$$P(Y > X) = \frac{1}{100}\left[100 - \frac{25}{2}\right] = \frac{7}{8} = 0.875.$$

  - Calcoliamo ora tale probabilità utilizzando la simulazione delle variabili aleatorie A e B effettuando 1.000.000 di simulazioni:

```
> N<-1000000
> x<-runif(N,min=10,max=20)
> y<-runif(N,min=15,max=25)
> diff<-y-x
> sum(diff>0)/length(diff)
[1] 0.875204
```



Esempio

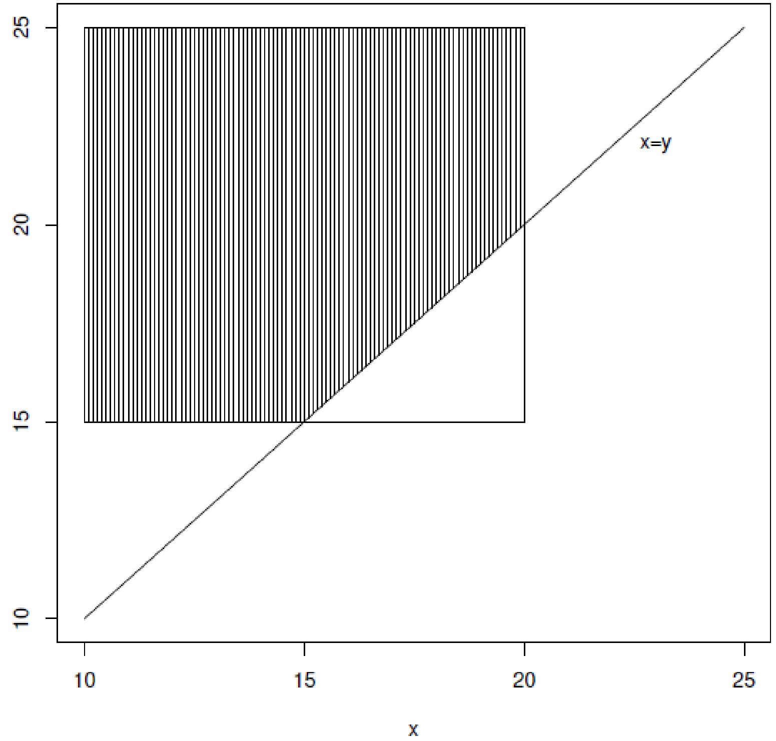
- **Esempio**:
  - Pertanto, tale probabilità può essere calcolata moltiplicando la densità congiunta per l'area tratteggiata rappresentata:

$$P(Y > X) = \frac{1}{100}\left[100 - \frac{25}{2}\right] = \frac{7}{8} = 0.875.$$

  - Calcoliamo ora tale probabilità utilizzando la simulazione delle variabili aleatorie A e B effettuando 1.000.000 di simulazioni:

```
> N<-1000000
> x<-runif(N,min=10,max=20)
> y<-runif(N,min=15,max=25)
> diff<-y-x
> sum(diff>0)/length(diff)
[1] 0.875204
```

Differenza valore per valore sui vettori x e y



  - Quindi affinché B completi la spesa dopo di A occorre considerare gli elementi del vettore *diff* positivi

Esempio

- **Esempio**:
  - Pertanto, tale probabilità può essere calcolata moltiplicando la densità congiunta per l'area tratteggiata rappresentata:

$$P(Y > X) = \frac{1}{100}\left[100 - \frac{25}{2}\right] = \frac{7}{8} = 0.875.$$

  - Calcoliamo ora tale probabilità utilizzando la simulazione delle variabili aleatorie A e B effettuando 1.000.000 di simulazioni:

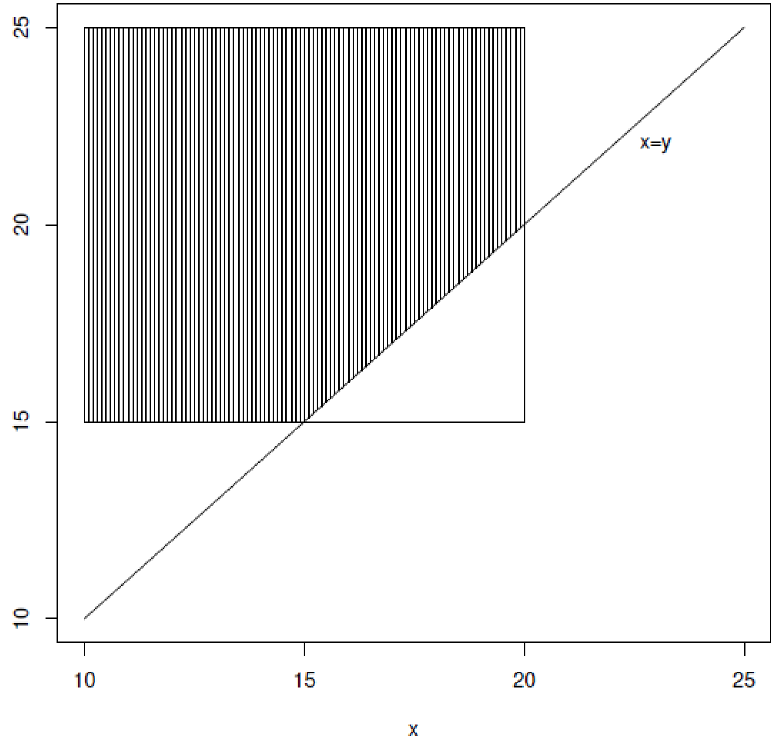
```
> N<-1000000
> x<-runif(N,min=10,max=20)
> y<-runif(N,min=15,max=25)
> diff<-y-x
> sum(diff>0)/length(diff)
[1] 0.875204
```

Differenza valore per valore sui vettori x e y

Casi Possibili

Casi Favorevoli

  - Quindi affinché B completi la spesa dopo di A occorre considerare gli elementi del vettore *diff* positivi



Esempio

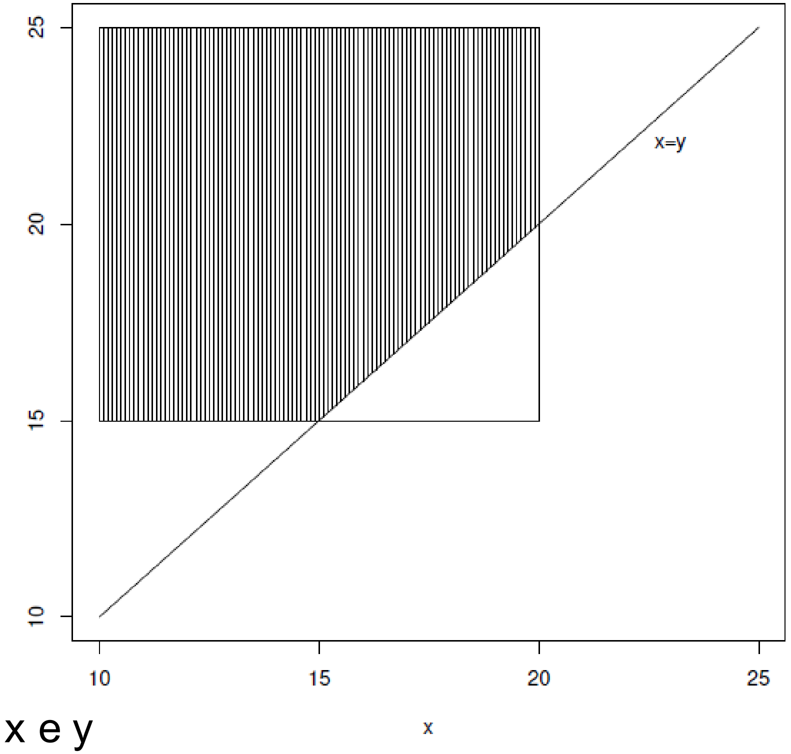
- **Esempio**:
  - Pertanto, tale probabilità può essere calcolata moltiplicando la densità congiunta per l'area tratteggiata rappresentata:

$$P(Y > X) = \frac{1}{100}\left[100 - \frac{25}{2}\right] = \frac{7}{8} = \boxed{0.875.}$$

  - Calcoliamo ora tale probabilità utilizzando la simulazione delle variabili aleatorie A e B effettuando 1.000.000 di simulazioni:

```
> N<-1000000
> x<-runif(N,min=10,max=20)
> y<-runif(N,min=15,max=25)
> diff<-y-x
> sum(diff>0)/length(diff)
[1] 0.875204
```

Differenza valore per valore sui vettori x e y



  - Quindi affinché B completi la spesa dopo di A occorre considerare gli elementi del vettore *diff* positivi
  - Per ottenere la probabilità che B completi la spesa dopo di A occorre considerare il rapporto tra i casi favorevoli e i casi possibili

STATISTICA E ANALISI DEI DATI

Distribuzione Esponenziale

Distribuzione Esponenziale

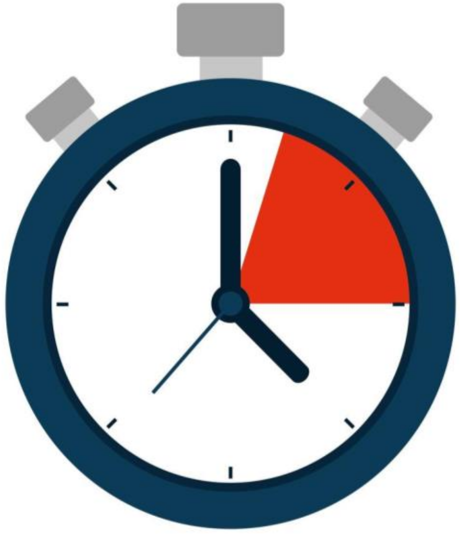
- La distribuzione esponenziale è una distribuzione di probabilità continua che descrive **il tempo che intercorre tra eventi successivi** in un **processo di Poisson**, dove gli eventi avvengono in media a una velocità costante
- La densità di probabilità esponenziale si può interpretare come l'analogo nel continuo della **funzione di probabilità geometrica**
- La distribuzione esponenziale viene usata spesso quando:
  - Si studiano sistemi di servizio in cui è ragionevole assumere che i tempi di interarrivo degli utenti oppure i tempi di espletamento dei servizi siano distribuiti esponenzialmente
  - In contesti come la durata di vita di oggetti o il tempo di attesa tra eventi casuali.
  - Si considera la durata di funzionamento di componenti elettronici
  - Si valutano dispositivi di varia natura che si guastano per cause accidentali (corti circuiti, fulmini, scariche elettriche, imprevedibili sollecitazioni meccaniche,…)

Parametro λ

- Il significato del parametro λ della distribuzione esponenziale può essere chiarito osservando il valore atteso:

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

Se X descrive un tempo, misurato in minuti, allora λ è una frequenza, misurata in $\frac{1}{minuti}$



- **Esempio**:
  - Se un call center riceve in media 3 chiamate al minuto, allora $\lambda = \mathbf{3}$ **chiamate/minuto**.
  - Il **tempo medio atteso tra un evento e l'altro** è l'inverso di lambda: $E(X) = \frac{1}{\lambda}$
    - Nell'esempio del call center, il tempo medio tra una chiamata e la successiva è **1/3 di minuto**, cioè **20 secondi**
- Il parametro λ ha lo stesso significato del parametro λ della distribuzione di Poisson, che descrive invece il numero di arrivi

Da Poisson ad Esponenziale

- In un processo di Poisson la probabilità di osservare $\boldsymbol{k}$ eventi in un intervallo di tempo $\boldsymbol{t}$ è:

$$P(k\ eventi\ in\ tempo\ t) = \frac{(\lambda t)^x}{k!} e^{(-\lambda t)}$$

- Consideriamo ora il **tempo di attesa T** fino al **primo evento**
- È più facile partire dalla domanda:

Qual è la probabilità che il primo evento accada **dopo** il tempo $\boldsymbol{t}$?

- $\boldsymbol{P(T > t)}$ = Probabilità che il primo evento accada dopo il tempo t

  = Probabilità che **nessun evento** sia accaduto entro il tempo t

  $= P(0\ eventi\ in\ tempo\ t) = \frac{(\lambda t)^0}{0!} e^{(-\lambda t)} = e^{(-\lambda t)}$

- Ora, la **funzione di distribuzione $\boldsymbol{F(t)}$** è la probabilità che il primo evento accada **entro** il tempo t:

$$F(t) = P(T \leq t) = 1 - P(T > t) = 1 - e^{(-\lambda t)}$$

- La **funzione di densità $\boldsymbol{f(t)}$** è la **derivata** della funzione di ripartizione

$$f(t) = \frac{d}{dt} F(t) = \frac{d}{dt} 1 - e^{(-\lambda t)} = \lambda e^{-\lambda t}$$

Distribuzione Esponenziale

- Considerato $\lambda > 0$, la variabile aleatoria X con:
  - **Funzione di distribuzione** è: $F_X(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$
  - **Densità di probabilità** è: $f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & altrimenti \end{cases}$

  si dice esponenzialmente distribuita con parametro λ
- La $F_X(x)$ risponde alla domanda:
  - Qual è la probabilità che il tempo fino al prossimo evento sia minore o uguale a un certo valore x?
- <u>Notazione</u>:
  - $X \sim \varepsilon(1, \lambda)$ variabile aleatoria esponenziale di parametro λ
- Per una variabile aleatoria esponenziale si ha:

Valore atteso: $E(X) = \frac{1}{\lambda}$ **Varianza**: $Var(X) = \frac{1}{\lambda^2}$

Assenza di Memoria

- La distribuzione esponenziale, così come la geometrica e Poisson, gode della proprietà di **assenza di memoria**
- Proprietà assenza di memoria:
  - Una variabile aleatoria esponenziale X con parametro $\lambda > 0$ gode della proprietà di mancanza di memoria se per ogni s ed t interi non negativi:

$$P(X > s + t \mid X > s) = P(X > t) \qquad per \quad s, t \geq 0$$

Cioè la probabilità che l'evento si verifichi in un intervallo futuro di lunghezza *t* non dipende dal tempo che è già trascorso s

- **Dimostrazione**:

$$P(X > s + t \mid X > s) = \frac{P(X > s + t \mid X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)}$$

Dato che: $P(X \leq x) = 1 - e^{-\lambda x}$ si ha che $P(X > x) = e^{-\lambda x}$ da cui:

$$\frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \frac{e^{-\lambda s}e^{-\lambda t}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t)$$

- Se si interpreta X come un tempo di attesa, la probabilità condizionata che il tempo di attesa X sia maggiore di $t + s$ non dipende da quanto si è gia atteso, ossia da s

Distribuzione Esponenziale

- Per il calcolo della densità esponenziale si utilizza la funzione:

```
dexp(x, rate = lambda)
```

dove

- **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria;
- **rate** è la frequenza λ della densità esponenziale

- Per calcolare la funzione di distribuzione esponenziale invece utilizziamo la funzione:

```
pexp(x, rate = lambda, lower.tail = TRUE)
```

dove

- **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria;
- **rate** è la frequenza λ della densità esponenziale
- **lower.tail** se tale parametro è TRUE (caso di default) calcola $P(X \leq x)$, mentre se tale parametro è FALSE calcola $P(X > x)$

Esempio (i)

- Supponiamo di avere un processo in cui gli eventi si verificano in media ogni 2 giorni, e vogliamo modellare il tempo tra gli eventi con una distribuzione esponenziale
- Quindi, il parametro λ della distribuzione esponenziale è $\frac{1}{2}$ (poiché la media è l'inverso di λ)
- La funzione di densità di probabilità diventa:

$$f_X(x) = \begin{cases} \frac{1}{2} e^{-\frac{1}{2}x}, & x > 0 \\ 0, & altrimenti \end{cases}$$

- Qual è la probabilità che passino almeno 3 giorni tra due eventi successivi?

$$P(X \geq 3) = \int_3^\infty \frac{1}{2} e^{-\frac{1}{2}x}\, dx$$

Esempio (i)

- **Esempio**:
  - La probabilità che la variabile aleatoria esponenziale di valore medio $E(X) = \frac{1}{2}$ assuma valori nell'intervallo (0.5,1.5):

$$P(0.5 < X < 1.5) = [1 - e^{-2\cdot 1.5}] - [1 - e^{-2\cdot 0.5}] = 0.3180923$$

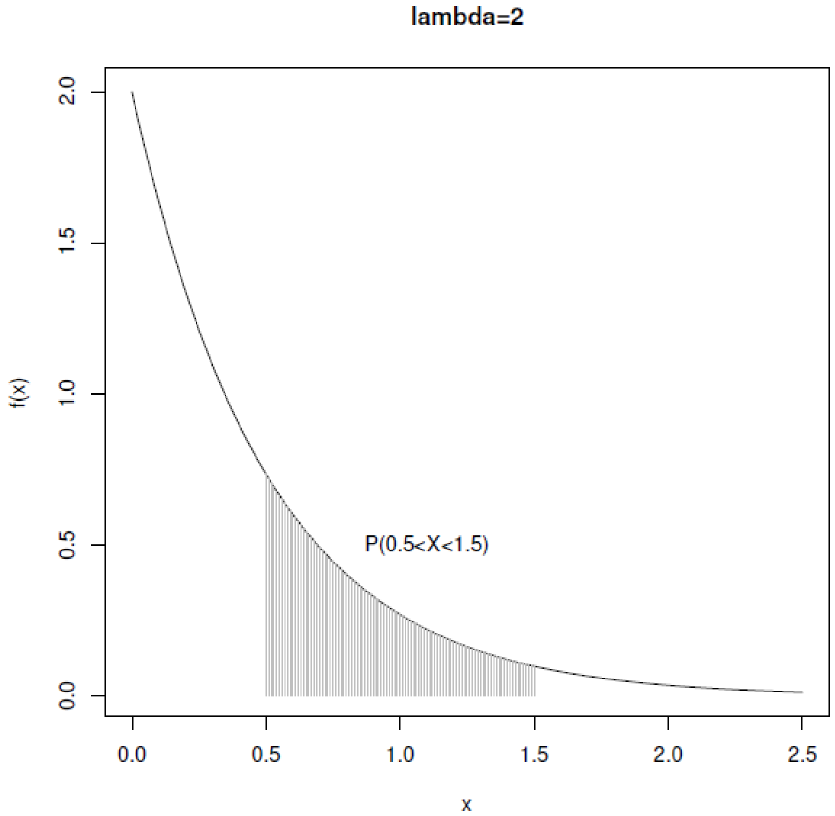
Esempio (i)

- **Esempio**:
  - La probabilità che la variabile aleatoria esponenziale di valore medio $E(X) = \frac{1}{2}$ assuma valori nell'intervallo (0.5,1.5):

$$P(0.5 < X < 1.5) = [1 - e^{-2\cdot 1.5}] - [1 - e^{-2\cdot 0.5}] = 0.3180923$$

  - In R si ha che:

```
>curve(dexp(x,rate=2),from=0,to=2.5,xlab="x",ylab="f(x)")
>x<-seq(0.5,1.5,0.01)
>lines(x,dexp(x,rate=2),type="h",col="grey")
>text(1.1,0.5,"P(0.5<X<1.5)")
> pexp(1.5,2)-pexp(0.5,2)
[1] 0.3180924
```



  - corrisponde all'area sottesa dalla densità esponenziale

Esempio (ii)

- Un macchinario industriale si guasta, in media, **una volta ogni 8 ore**. Si assume che i guasti seguano un processo di Poisson, e quindi il tempo tra un guasto e l'altro segue una distribuzione esponenziale. Si vuole calcolare:
    1. La probabilità che il prossimo guasto si verifichi entro **2 ore**.
    2. La probabilità che il macchinario funzioni per almeno **10 ore** senza guastarsi.

Esempio (ii)

- Un macchinario industriale si guasta, in media, **una volta ogni 8 ore**. Si assume che i guasti seguano un processo di Poisson, e quindi il tempo tra un guasto e l'altro segue una distribuzione esponenziale. Si vuole calcolare:
  1. La probabilità che il prossimo guasto si verifichi entro **2 ore**.
  2. La probabilità che il macchinario funzioni per almeno **10 ore** senza guastarsi.
- **Soluzione 1**:
  - La **frequenza media** dei guasti è $\lambda = \frac{1}{8}$ per ora
  - La probabilità che il prossimo guasto si verifichi entro 2 ore è: $F_X(x) = P(X \leq 2) = 1 - e^{-\frac{1}{8}2} = 0.221$

Esempio (ii)

- Un macchinario industriale si guasta, in media, **una volta ogni 8 ore**. Si assume che i guasti seguano un processo di Poisson, e quindi il tempo tra un guasto e l'altro segue una distribuzione esponenziale. Si vuole calcolare:
  1. La probabilità che il prossimo guasto si verifichi entro **2 ore**.
  2. La probabilità che il macchinario funzioni per almeno **10 ore** senza guastarsi.
- **Soluzione 1**:
  - La **frequenza media** dei guasti è $\lambda = \frac{1}{8}$ per ora
  - La probabilità che il prossimo guasto si verifichi entro 2 ore è: $F_X(x) = P(X \leq 2) = 1 - e^{-\frac{1}{8}2} = 0.221$
- **Soluzione 2**:
  - La probabilità che il tempo tra gli eventi superi x è complementare alla CDF:

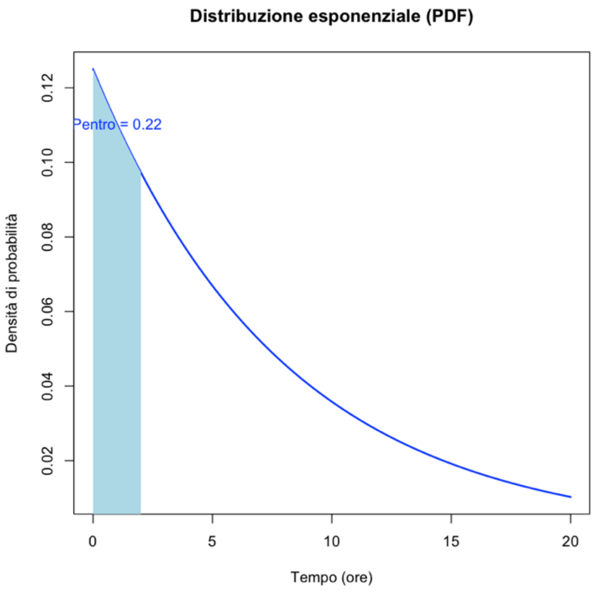
$$P(X > 10) = 1 - \left(1 - e^{-\lambda x}\right) = e^{-\lambda x} = e^{-\frac{10}{8}} = 0.2865$$

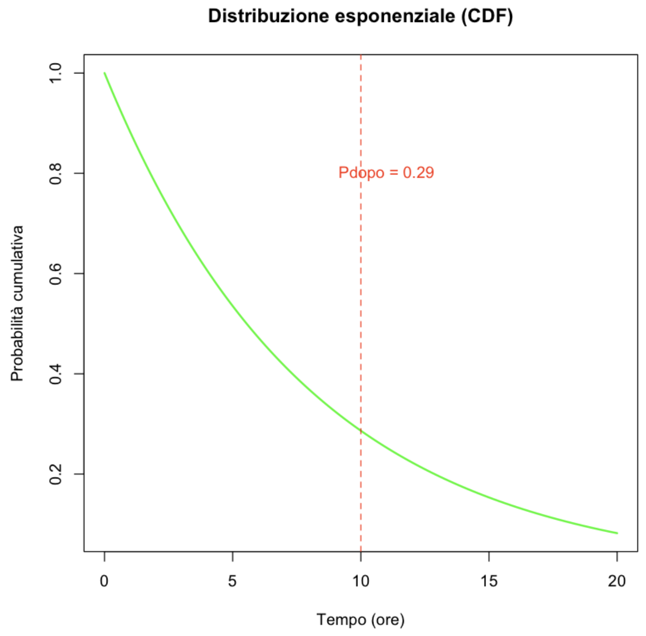
  - La probabilità che il macchinario funzioni per almeno 10 ore senza guasti è **28.6%**.

Esempio (ii) R

```r
# Parametri
lambda <- 1 / 8  # Tasso medio (guasti per ora)

# Valori per il grafico
x_vals <- seq(0, 20, by = 0.1)  # Intervallo di tempo da 0 a 20 ore

# Calcoli per i punti specifici
x1 <- 2  # Tempo entro cui calcolare la probabilità

prob_guasto_entro_2 <- pexp(x1, rate = lambda)  # Probabilità entro 2 ore

# 1. Grafico della PDF con evidenziamento del prossimo guasto entro 2 ore
pdf_vals <- dexp(x_vals, rate = lambda)  # Valori della densità di probabilità
plot(x_vals, pdf_vals, type = "l", col = "blue", lwd = 2,
     xlab = "Tempo (ore)", ylab = "Densità di probabilità",
     main = "Distribuzione esponenziale (PDF)")
polygon(c(0, seq(0, x1, by = 0.1), x1),
        c(0, dexp(seq(0, x1, by = 0.1), rate = lambda), 0),
        col = "lightblue", border = NA)
text(x1 / 2, dexp(x1 / 2, rate = lambda),
     labels = sprintf("Pentro = %.2f", prob_guasto_entro_2), col = "blue")
```

```r
# 2. Grafico della CDF con evidenziamento di almeno 10 ore senza guasti
x2 <- 10  # Tempo minimo di funzionamento
prob_guasto_dopo_10 <- 1 - pexp(x2, rate = lambda)  # Probabilità oltre 10 ore

# Valori della funzione di distribuzione cumulativa
cdf_vals <- pexp(x_vals, rate = lambda, lower.tail=FALSE)
plot(x_vals, cdf_vals, type = "l", col = "green", lwd = 2,
     xlab = "Tempo (ore)", ylab = "Probabilità cumulativa",
     main = "Distribuzione esponenziale (CDF)")
abline(v = x2, col = "red", lty = 2)  # Linea verticale per 10 ore
text(x2 + 1, 0.8, labels = sprintf("Pdopo = %.2f", prob_guasto_dopo_10), col = "red")
```





Quantili

- Per calcolare i quantili (percentili) della distribuzione esponenziale

```
qexp(z, rate = lambda)
```

dove

  - **z** è il valore assunto (o i valori assunti) dalle probabilità relative al percentile z · 100–esimo, cioè il vettore delle probabilità;
  - **rate** è la frequenza λ della densità esponenziale

- Il risultato della funzione è il percentile z ·100–esimo, ossia il più piccolo numero *x* assunto dalla variabile aleatoria X tale che

$$F_X(x) = P(X \leq x) \geq z \qquad (a \leq x \leq b)$$

Quantili

- **Esempio**: Calcoliamo i quartili Q_0, Q_1, Q_2, Q_3, Q_4 della variabile aleatoria esponenziale di valore medio $E(X) = \frac{1}{2}$

```
>z<-c(0,0.25,0.5,0.75,1)
>qexp(z, rate=2)
[1] 0.0000000 0.1438410 0.3465736 0.6931472       Inf
```

- Ricordando che $P(X \le x) = 1 - e^{-\lambda x}$ si ha:

$$1 - e^{-\lambda x} \geq z \quad \Longleftrightarrow \quad x \geq -\frac{\log(1-z)}{\lambda}$$

$$P(X \leq x) = 1 - e^{-2x} \geq 0.25 \quad \Longleftrightarrow \quad x \geq -\frac{\log(1-0.25)}{\lambda} = 0.1438410,$$

$$P(X \leq x) = 1 - e^{-2x} \geq 0.50 \quad \Longleftrightarrow \quad x \geq -\frac{\log(1-0.50)}{\lambda} = 0.3465736,$$

$$P(X \leq x) = 1 - e^{-2x} \geq 0.75 \quad \Longleftrightarrow \quad x \geq -\frac{\log(1-0.75)}{\lambda} = 0.6931472.$$

Generazione Randomica di Numeri

- È possibile simulare la variabile aleatoria esponenziale generando una sequenza di numeri pseudocasuali mediante la funzione

```
rexp(N, rate=lambda)
```

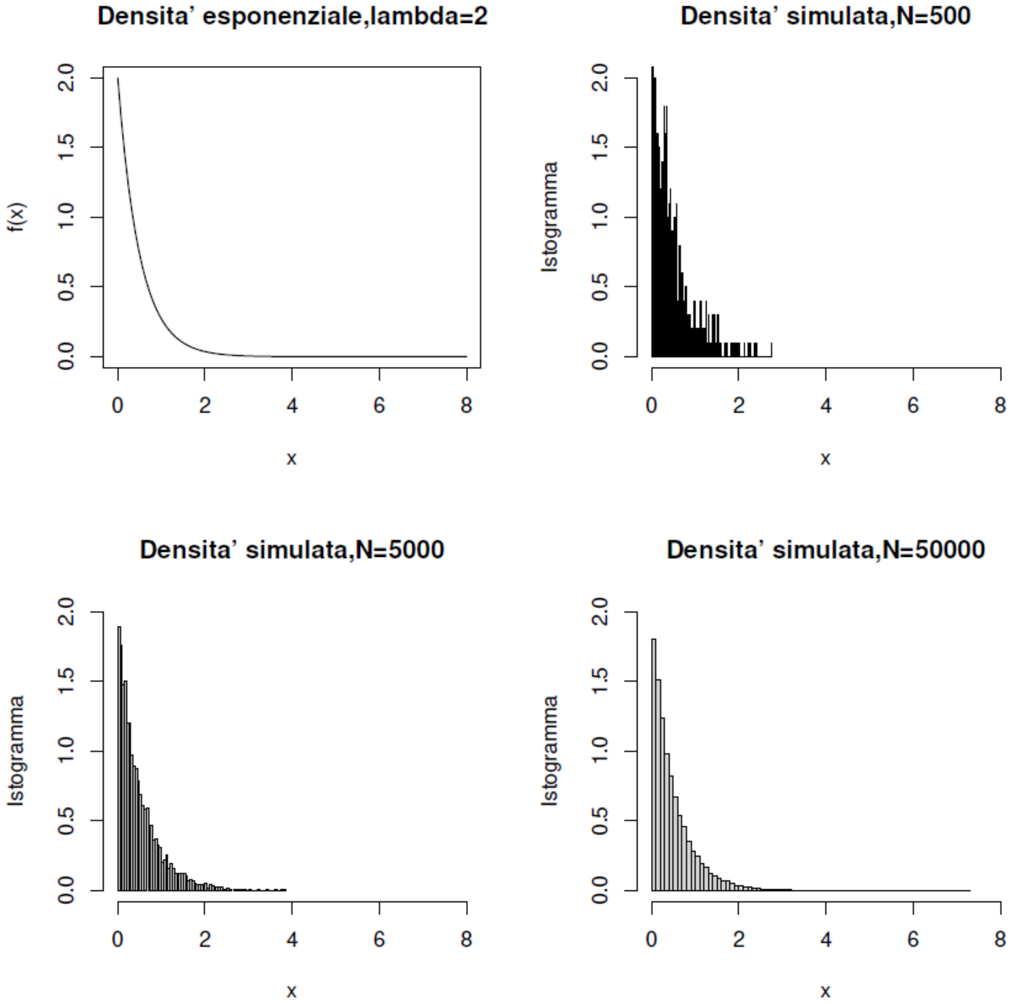
dove

- **N è** lunghezza della sequenza da generare;
- **rate** è la frequenza λ della densità esponenziale

Generazione Randomica di Numeri

- **Esempio**: Consideriamo una variabile aleatoria esponenziale di valore medio $E(X) = \frac{1}{2}$ e generiamo tre sequenze di $N = 500, 5000, 50000$ numeri pseudocasuali:

```
>par(mfrow=c(2,2))
>curve(dexp(x,rate=2),from=0, to=8,xlab="x",ylab="f(x)",
+ylim=c(0,2),main="Densita' esponenziale,lambda=2")
>
>sim1<-rexp(500,rate=2)
>hist(sim1,freq=F,xlim=c(0,8),ylim=c(0,2),breaks=100,xlab="x",
+ylab="Istogramma",main="Densita' simulata,N=500")
>
>sim2<-rexp(5000,rate=2)
>hist(sim2,freq=F,xlim=c(0,8),ylim=c(0,2),breaks=100,xlab="x",
+ylab="Istogramma",main="Densita' simulata,N=5000")
>
>sim3<-rexp(50000,rate=2)
>hist(sim3,freq=F,xlim=c(0,8),ylim=c(0,2),breaks=100,xlab="x",
+ylab="Istogramma",main="Densita' simulata,N=50000")
```

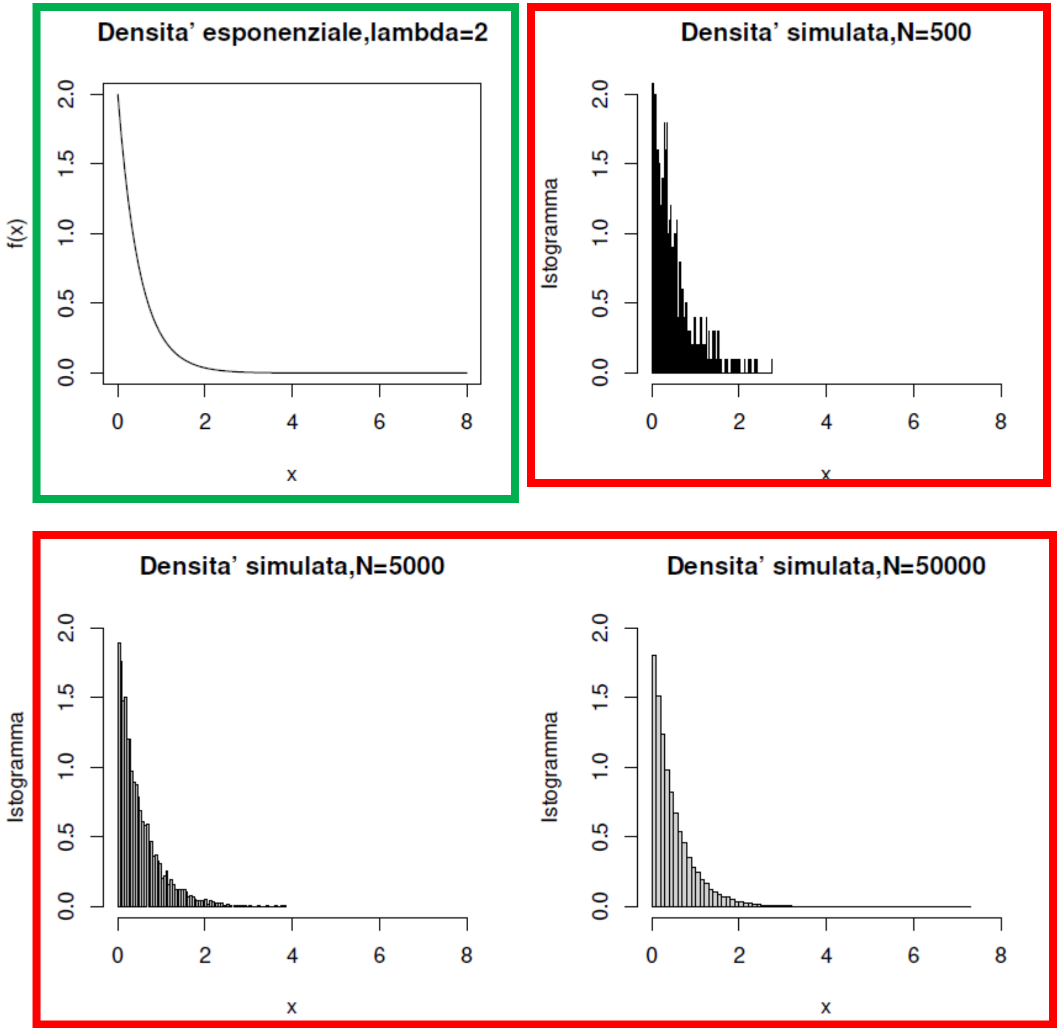


- Notiamo che al crescere di N la la densità simulata si avvicina alla densità esponenziale teorica di valore medio $E(X) = \frac{1}{2}$

Generazione Randomica di Numeri

- **Esempio**: Consideriamo una variabile aleatoria esponenziale di valore medio $E(X) = \frac{1}{2}$ e generiamo tre sequenze di $N = 500, 5000, 50000$ numeri pseudocasuali:

```
>par(mfrow=c(2,2))
>curve(dexp(x,rate=2),from=0, to=8,xlab="x",ylab="f(x)",
+ylim=c(0,2),main="Densita' esponenziale,lambda=2")
>
>sim1<-rexp(500,rate=2)
>hist(sim1,freq=F,xlim=c(0,8),ylim=c(0,2),breaks=100,xlab="x",
+ylab="Istogramma",main="Densita' simulata,N=500")
>
>sim2<-rexp(5000,rate=2)
>hist(sim2,freq=F,xlim=c(0,8),ylim=c(0,2),breaks=100,xlab="x",
+ylab="Istogramma",main="Densita' simulata,N=5000")
>
>sim3<-rexp(50000,rate=2)
>hist(sim3,freq=F,xlim=c(0,8),ylim=c(0,2),breaks=100,xlab="x",
+ylab="Istogramma",main="Densita' simulata,N=50000")
```



- Notiamo che al crescere di N la la densità simulata si avvicina alla densità esponenziale teorica di valore medio $E(X) = \frac{1}{2}$