
4 – LA PROBABILITÀ

TECNICHE DI ANALISI DI DATI I



"Probabilities are very slippery things"

K. Pearson

"As to what probability is and how it is connected with statistics, there has seldom been such complete disagreement and breakdown of communication since the Tower of Babel"

Savage, 1954

Variabili ed esperimenti aleatori

Un altro modo per analizzare una distribuzione X è chiedersi quanto sia **verosimile**, cioè **probabile**, che un caso X_i della distribuzione si verifichi. Ogni modalità di X ha una certa probabilità di manifestarsi, ovvero **a ogni modalità è associabile la rispettiva probabilità**, come in una distribuzione X a ogni modalità è associata la rispettiva frequenza.

Il collegamento tra probabilità e statistica descrittiva si realizza nelle **variabili casuali/aleatorie/stocastiche** → **quantità** il cui valore dipende **dall'esito di un esperimento casuale**

Un **esperimento casuale** è ogni **atto** o **processo** la cui singola esecuzione (**prova**) dà luogo ad un **risultato non prevedibile**: tirare una moneta, lanciare due dadi...;

Invece, negli **esperimenti deterministici** ogni **singola prova dà lo stesso risultato**, a parità di condizioni sperimentali e in assenza di **covariate**.

Anche se il risultato della **singola prova** è **imprevedibile**, se **ripetiamo molte volte** un **esperimento casuale** è possibile rilevare **regolarità** nei risultati, ovvero **nella distribuzione** di un certo fenomeno statistico (per esempio, di quante volte esce Testa)



Le frequenze relative o **proporzioni** con cui si manifestano i valori delle distribuzioni di frequenza sono assimilabili alle loro **probabilità** di realizzarsi.



L'insieme dei valori assumibili da un fenomeno statistico e delle probabilità che gli attribuiamo costituisce un modello,

cioè una descrizione di quello che può verificarsi e delle probabilità con cui ci aspettiamo che questi eventi si verifichino.

Definizioni di probabilità

Definizione classica (Pascal, 1623-1662) “La **probabilità** di un evento è il **rapporto tra il numero dei casi favorevoli all’evento e il numero dei casi possibili**, purché questi siano **tutti ugualmente probabili**”.)

*La definizione è **tautologica**: per definire la probabilità bisogna sapere cosa significa: “ugualmente probabili”, cioè sapere cos’è la probabilità. Inoltre, si applica solo a eventi equiprobabili.*

Definizione frequentista di probabilità o **legge empirica del caso**: in una successione di prove nelle stesse condizioni, la **frequenza di un evento si avvicina alla sua probabilità**, e l’approssimazione tende a migliorare con l’aumentare del numero delle prove → la **probabilità è il valore costante intorno al quale tende a stabilizzarsi la frequenza relativa di un evento al crescere del numero di prove di un dato esperimento**, ovvero il **limite** cui tende la **proporzione** di successi man mano che il numero di prove cresce a infinito.

Ma: le frequenze non sono una successione numerica data mediante legge, ma numeri rilevati sperimentalmente, per i quali il concetto di limite non è chiaro.

Inoltre, va considerata una successione di prove fatte nelle stesse condizioni e ripetute indefinitamente, restringendo l’applicabilità della definizione a situazioni ben delimitate (i lanci di un dado).

Legge dei grandi numeri

Pascal: la probabilità di ottenere Testa T lanciando una moneta è **1 (casi attesi) su 2 (casi equiprobabili: T o C)**

Buffon (1708-1788): ha tirato **4040** volte una moneta ottenendo **2048** volte Testa

Pearson (1857-1936) ha tirato per **24000** volte una moneta ottenendo **12012** volte Testa:

$$P(T) = \frac{1}{2} = .500$$

$$P(T) = \frac{2048}{4040} = .5063907$$

$$P(T) = \frac{12012}{24000} = .5005$$

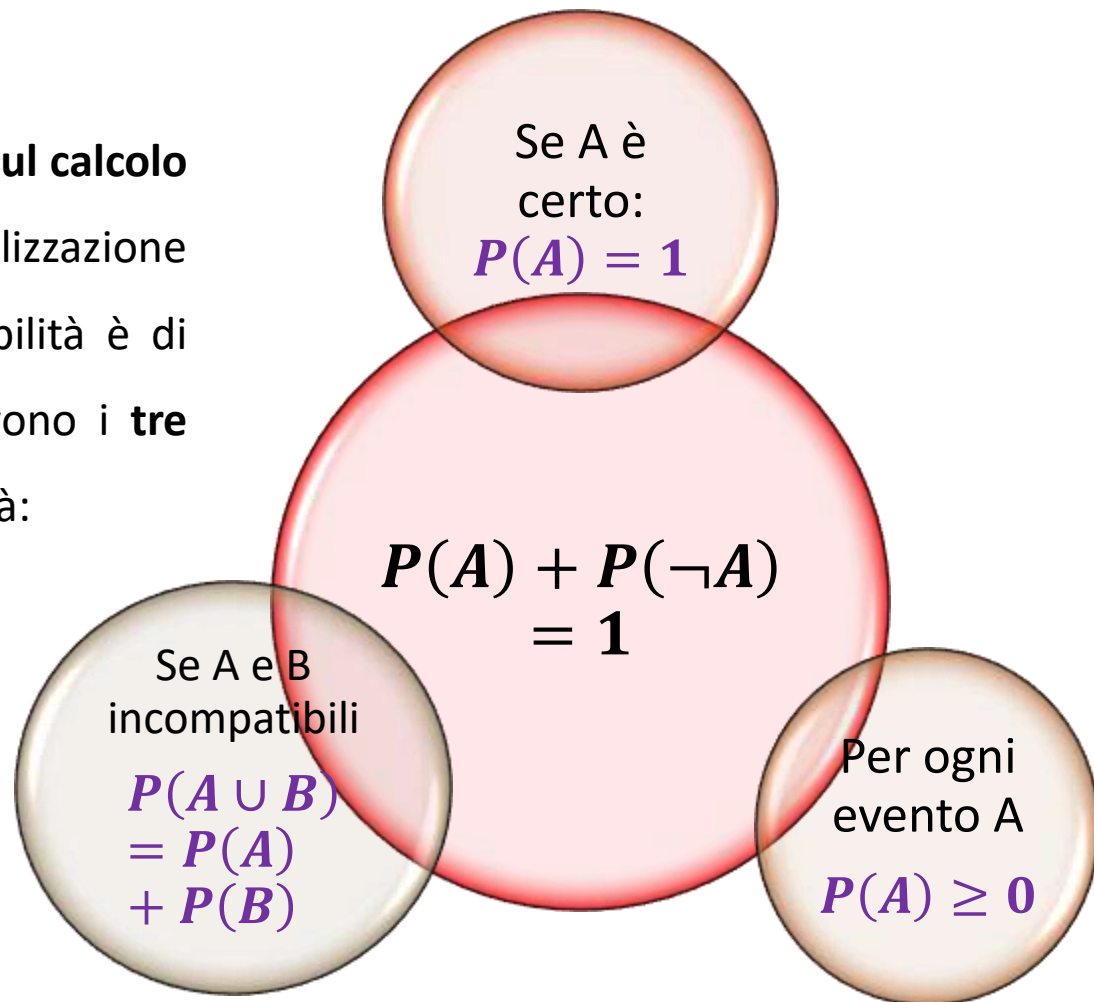
Legge dei grandi numeri (Bernoulli): per un numero tendente **all'infinito** di ripetizioni **identiche** di un esperimento **aleatorio**,
la probabilità di un evento tende a coincidere con la sua frequenza relativa.

La media calcolata di una sequenza di prove indipendenti ed equiprobabili di una variabile aleatoria, ripetute per un numero di volte tendente a infinito, è sufficientemente vicina alla media vera, quella calcolabile teoricamente.

Calcolo della probabilità

La probabilità di un evento **varia da 0** (l'evento non può in alcun caso verificarsi) **a 1** (evento **certo**); se vi è più semplice, potete **interpretarla** (non calcolarla!) in termini **percentuali**.

Al di là delle definizioni, c'è **accordo sul calcolo delle probabilità**: la prima formalizzazione della teoria del calcolo delle probabilità è di **Kolmogorov** (1903-1987), cui si devono i **tre assiomi** fondamentali della probabilità:



Probabilità di più eventi

Due eventi sono **mutualmente escludentisi (incompatibili)** se il verificarsi dell'uno **impedisce il verificarsi dell'altro** (tirando un solo dado, il verificarsi di "cinque" impedisce le altre facce).

La probabilità del verificarsi di due eventi mutualmente escludentisi è uguale **alla somma della probabilità di verificarsi dei singoli eventi: regola della somma**

$$P(A \text{ o } B) = P(A) + P(B) = 1$$

La probabilità che si verifichi almeno uno tra A e B è =1, cioè è certo che uno tra A o B si verifichi.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1$$

La probabilità che si verifichi uno tra A e B è data dalla somma della probabilità di A e di B meno la **probabilità che A e B si verifichino insieme** ($P(A \cap B)$): poiché sono incompatibili, $P(A \cap B) = 0$.

Tirando **un** solo dado, il verificarsi **dell'evento "faccia dispari"** ("1" o "3" o "5") è uguale alla somma della probabilità di "1", "3" e "5", meno la probabilità che si verifichino assieme

$$P(\text{dispari}) = P(1 \cup 3 \cup 5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} - 0 = \frac{3}{6} = .50$$

Se gli eventi sono **compatibili**, il verificarsi di uno non esclude il verificarsi dell'altro: in questo caso la regola della somma funziona in maniera del tutto simile, ma poiché $P(A \cap B) \neq 0$, non potremo ignorarla.

Tirando **un solo dado**, qual è la probabilità di ottenere **“faccia dispari”** (A: “1”, “3”, “5”) **o un multiplo di 3** (B: “3”)?

La faccia “tre” soddisfa contemporaneamente A – dispari e B – multiplo di 3. Quindi,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1$$

$$P(\text{dispari} \cup \text{multiplo}_3) = \frac{3}{6} + \frac{1}{6} - \frac{1}{6} = .50$$

Gli eventi compatibili possono essere **indipendenti** oppure **dipendenti**: vediamo le caratteristiche.

Due eventi A e B sono **indipendenti** se il **verificarsi di A non influenza il verificarsi di B** (tirando due dadi, il fatto che uno segni “cinque” non influenza il risultato dell’altro dado).

La probabilità del verificarsi **contemporaneo o in successione** di due eventi **indipendenti** è uguale al **prodotto** della probabilità di verificarsi dei singoli eventi: **regola del prodotto (o delle probabilità composte)**

$$P(A \text{ e } B) = P(A) \times P(B)$$

$$P(A \cap B) = P(A) \times P(B)$$

Tirando due dadi, il verificarsi **dell’evento “somma = 2”** (dado 1: “1” e dado 2: “1”), è uguale al prodotto della probabilità di avere “uno” con il primo **e** “uno” con il secondo dado)

$$P\left(\sum 2\right) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} = .028$$

L’evento “somma 2” è 1 sui 36 eventi **possibili che possono verificarsi tirando due dadi**

Due eventi A e B sono **dipendenti** se il **verificarsi di A influenza il verificarsi di B**: in una estrazione **senza rimpiazzo** (tombola, lotto), l'evento "estrazione di un numero pari" ha probabilità **45/90** alla prima estrazione, ha probabilità **45/89** alla seconda se il primo estratto è dispari, e **44/89** se il primo estratto è pari, e così via.

Quando gli eventi sono dipendenti, il **principio del prodotto cambia**:

probabilità condizionata o composta di B una volta che si sia **verificato A**: **$B|A$**

$$P(A \text{ e } B) = P(A) \times P(B|A)$$

$$P(A \text{ e } B) = P(B) \times P(A|B)$$

La probabilità che i primi due numeri in una estrazione del lotto siano entrambi pari è:

$$P(\text{primi 2 numeri pari}) = \frac{45}{90} \times \frac{44}{89} = \frac{1980}{8010} = .247$$

Ora vedremo la probabilità in esperimenti
un po' più complessi.

A seconda che nella variabile aleatoria si
manifestino modalità discrete o
continue, avremo distribuzioni di
probabilità discrete o continue.

Distribuzioni di probabilità discrete

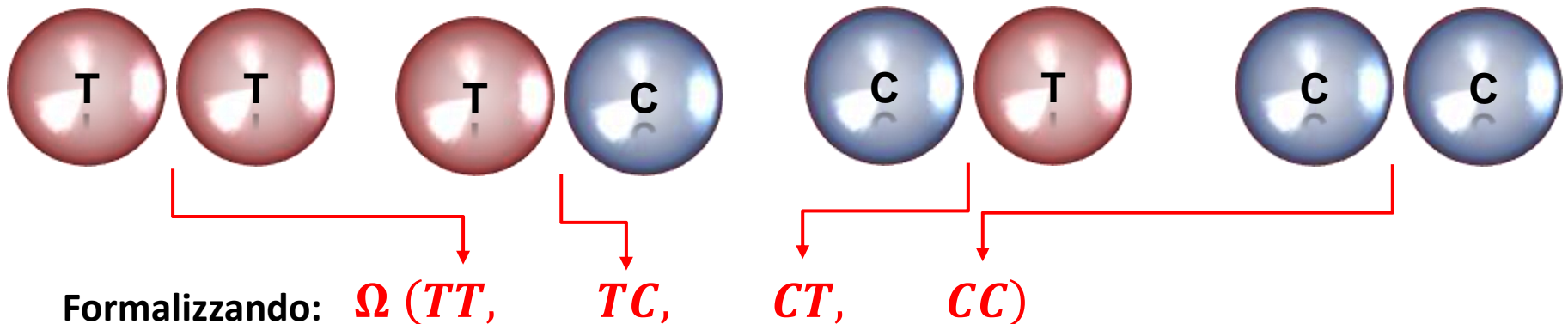
Funzione di probabilità o massa di probabilità o probabilità discreta

Se una **variabile X** può assumere **solo valori interi e positivi** $x_1, x_2, x_3, \dots, x_k$, con **probabilità** rispettivamente $P_1, P_2, P_3, \dots, P_k$ (la cui somma è = 1), la **variabile X** ha una **distribuzione di probabilità discreta**.

Facciamo l'esperimento di **due lanci di una moneta**, in cui ogni faccia (T, C) ha la **stessa probabilità** di manifestarsi in ogni lancio:

$$P(T) = P(C) = \frac{1}{2} = .50$$

L'insieme dei risultati possibili, cioè lo **spazio campionario Ω** dell'esperimento, è :



Consideriamo T come **evento atteso** e **contiamo il numero di T** possibili nello spazio campionario, cioè associamo a ogni **prova possibile di Ω** un valore $X(\omega) = x$:

ω	CC	TC	CT	TT
$X(\omega) = x$	0	1	1	2
$P(\omega)$	1/4	1/4	1/4	1/4

$$X(0): f = 1: f_r = 1/4$$

$$P(X_{0T}) = 1/4 = .25$$

$$X(1_T): f = 2: f_r = 2/4$$

$$P(X_{1T}) = 1/4 \times 2 = .50$$

$$X(2_T): f = 1: f_r = 1/4$$

$$P(X_{2T}) = 1/4 = .25$$

$X = x$	0	1	2	Totale
$P(X = x)$.25	.50	.25	$\Sigma = 1$

La probabilità che nei due lanci T non esca mai è pari a 0.25, la probabilità che T esca una volta è pari a .50, la probabilità che T esca due volte è pari a .25

Abbiamo creato un **modello matematico** che descrive i possibili risultati numerici di un **esperimento casuale** con l'attribuzione delle probabilità di verificarsi dei singoli eventi.

La **funzione che assegna a ogni possibile valore di X (variabile aleatoria discreta) la probabilità del possibile esito** dell'esperimento casuale è la **FUNZIONE DI PROBABILITÀ $P(X=x)$** (o **funzione di massa di probabilità** o **densità discreta**)

Funzione di ripartizione

$$P(X \leq x_i) = \sum_{x_i \leq x} P(X = x_i)$$

Una volta nota la distribuzione di probabilità di una variabile aleatoria, si può calcolare la **probabilità di un evento minore o uguale a x_i** , cioè **$P(X \leq x_i)$** , con le **probabilità cumulate**: si sommano le probabilità degli eventi inferiori o uguali a x_i .

La funzione che calcola la probabilità $P(X \leq x_i)$ si chiama **funzione di ripartizione**, perché ripartisce la distribuzione delle probabilità in due parti: fino a x_i – oltre x_i).

Come esperimento casuale, **tiriamo due dadi** (trial= 2; eventi possibili = 6x6) e calcoliamo la distribuzione di probabilità della **somma dei dadi**, da evento 2 a evento 12:

$X = x$	$\Sigma_{ab}2$	$\Sigma_{ab}3$	$\Sigma_{ab}4$	$\Sigma_{ab}5$	$\Sigma_{ab}6$	$\Sigma_{ab}7$	$\Sigma_{ab}8$	$\Sigma_{ab}9$	$\Sigma_{ab}10$	$\Sigma_{ab}11$	$\Sigma_{ab}12$	Tot
	1-1	1-2,2-1	1-3,3-1, 2-2	1-4,4-1, 2-3,3-2	1-5,5-1, 2-4,4-2, 3-3	1-6,6-1 ,2-5,5-2, 3-4,4-3	2-6,6-2, 3-5,5-3, 4-4	3-6,6-3, 4-5,5-4	4-6,6-4,5-5	5-6,6-5	6-6	
f	1	2	3	4	5	6	5	4	3	2	1	36
	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	
$P(X = x)$.028	.083	.111	.139	.167	.194	.167	.139	.111	.083	.028	1

→ Qual è la probabilità di ottenere una somma pari o inferiori a 6?

Calcoliamola, usando le frazioni per usare tutti i decimali:

$X = x$	$\Sigma_{ab}2$	$\Sigma_{ab}3$	$\Sigma_{ab}4$	$\Sigma_{ab}5$	$\Sigma_{ab}6$
$P(X = x)$	1/36	2/36	3/36	4/36	5/36

```
cumulata_sei <- 1/36 + 2/36 + 3/36 + 4/36 + 5/36  
cumulata_sei  
[1] 0.4166667
```

La probabilità di ottenere una **somma uguale o minore a 6 è =.42.**

Per conoscere la **probabilità di ottenere un dato maggiore di x_i** , cioè $P(X > x_i)$, si **sottrae alla probabilità totale la probabilità cumulata per i valori inferiori o uguali a $x-1$.**

La probabilità di ottenere una somma uguale o maggiore di 7 è :

```
cumulata_sette_più <- 1 - (cumulata_sei)  
cumulata_sette_più  
[1] 0.5833333
```

$X = x$	$\Sigma_{ab}7$	$\Sigma_{ab}8$	$\Sigma_{ab}9$	$\Sigma_{ab}10$	$\Sigma_{ab}11$	$\Sigma_{ab}12$
$P(X = x)$	6/36	5/36	4/36	3/36	2/36	1/36

Distribuzione di probabilità binomiale


Lanciare una moneta N volte è un esempio di **distribuzione di probabilità binomiale**: è **discreta** e calcola la **probabilità di ottenere X successi** in **N prove indipendenti** di un esperimento aleatorio i cui risultati siano **dicotomici, mutualmente escludentisi** ed **equiprobabili** in ogni prova.

Per **calcolare la probabilità di elementi** di una **distribuzione binomiale**, si usa la **funzione di probabilità binomiale**:

$$P(X) = C_k^N \times p^k \times (1 - p)^{N-k}$$

N = numero di prove;

p = probabilità dell'evento atteso



C: combinazioni semplici di N oggetti di classe k (il numero di elementi in ogni combinazione): tutti i gruppi che si possono formare con gli N oggetti a k a k .

La funzione di R corrispondente alla formula è **dbinom(x, size, prob)**: **x**= vettore aleatorio o singola modalità, **size**= numero di prove, **prob**= probabilità dell'evento atteso in ogni prova. La **d** del nome sta per **density** (*densità* o massa di probabilità).

Qual è la probabilità di ottenere T in due lanci di una moneta , con **vettore aleatorio da 0 a 2**, **due lanci** e una **probabilità** di ottenere T in **ogni lancio** pari a $\frac{1}{2}$?

```
dbinom(0:2, size = 2, prob = .50)  
[1] 0.25 0.50 0.25
```

Probabilità di nessuna T

Probabilità di una T

Probabilità di due T

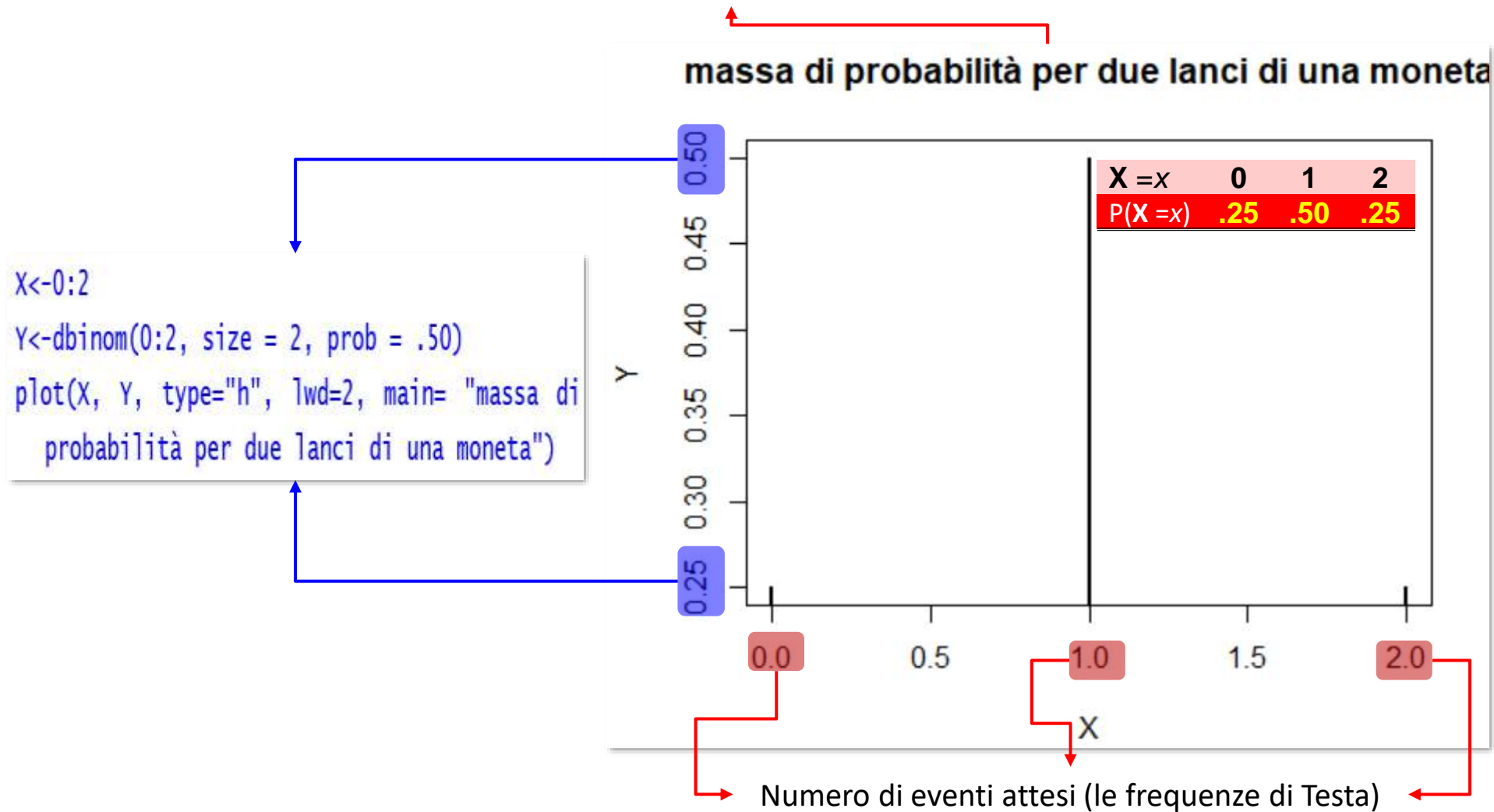
X =x	0	1	2
P(X =x)	.25	.50	.25

La probabilità di **un solo figlio maschio** (singolo evento atteso) **su tre nascite** (prove=3, probabilità di M in ogni nascita = .50)

```
dbinom(1,size = 3,prob = .50)  
[1] 0.375
```

La **rappresentazione grafica** della **distribuzione** di probabilità binomiale è **un grafico a barre**. Vediamo quella dei due lanci di una moneta

Massa di probabilità, ovvero la **probabilità di ciascuno degli eventi in ascissa**.



Per **calcolare la funzione di ripartizione** di una **distribuzione** binomiale si usa `pbinom(q, size, prob, lower.tail=TRUE/FALSE)`; **q**= quantile della distribuzione che identifica la partizione, **size**= prove, **prob**= probabilità dell'evento atteso in ogni trial.

Se **lower.tail=TRUE (default)** si considera la parte inferiore della distribuzione, quindi è riportata la probabilità cumulata **FINO al quantile q**; se **FALSE**, si ignora la parte inferiore ed è riportata la probabilità cumulata **DAL quantile q in su**.

*Tirando due volte una moneta, la probabilità di ottenere T **0 o 1 volta** è uguale a*

```
pbinom(1,size = 2,prob = .5, lower.tail=TRUE)
[1] 0.75
```

```
dbinom(0:2, size = 2, prob = .50)
[1] 0.25 0.50 0.25
```

```
pbinom(1,size = 2,prob = .5, lower.tail = FALSE)
[1] 0.25
```

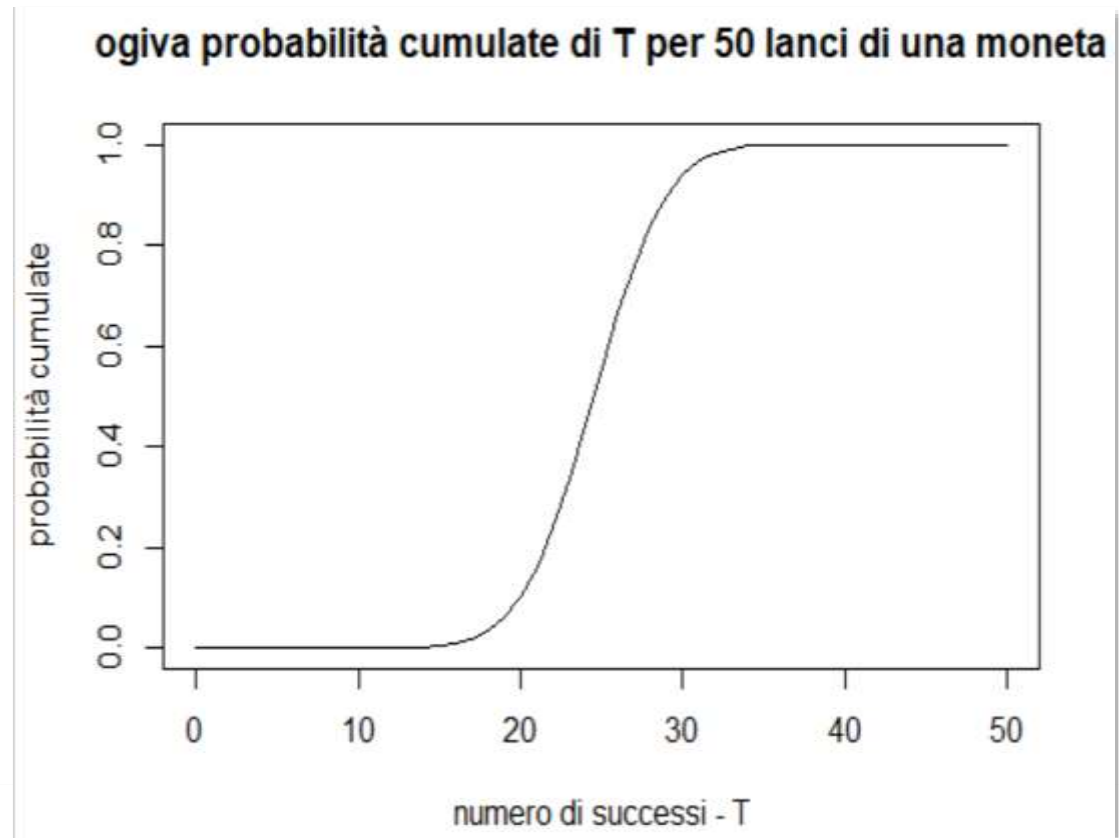
```
dbinom(0:2, size = 2, prob = .50)
[1] 0.25 0.50 0.25
```

La probabilità cumulata di ottenere il quantile 1 o un quantile <1 è = .75; quella di ottenere un quantile > 1 è = 0.25

pbinom: p sta per *partition*

Graficamente, **la funzione di ripartizione è un'ogiva di probabilità cumulate**. Possiamo rappresentarla con `plot`, specificando in X la variabile aleatoria e in Y la funzione di ripartizione; `type= "l"` (`line`) indica che i punti del grafico sono uniti da una linea.

*Applichiamola a un esperimento un po' più corposo: vediamo le probabilità di ottenere **T** tirando una moneta 50 volte.*



```
X<-0:50
Y<-pbinom(q = X, size = 50, prob = .50)
plot(X, Y, type = "l", xlab="numero di successi - T", ylab="probabilità cumulate", main = "ogiva
probabilità cumulate di T per 50 lanci di una moneta")
```

L'inverso di `pbinom` è `qbinom(p, size, prob, lower.tail= TRUE/FALSE)`: dà il quantile corrispondente alla probabilità cumulata `p=` ; `size`: numero di trial, `prob`: probabilità dell'evento atteso in ogni trial. `lower.tail`: se `TRUE` (default) dà il quantile corrispondente a $P(X \leq x)$; se `FALSE`, quello corrispondente a $P(X > x)$

```
qbinom(p = .75, size =2, prob = .50, lower.tail = TRUE)  
[1] 1
```

```
dbinom(0:2, size = 2, prob = .50)  
[1] 0.25 0.50 0.25
```

```
qbinom(p = .75, size =2, prob = .50, lower.tail = FALSE)  
[1] 0
```

```
dbinom(0:2, size = 2, prob = .50)  
[1] 0.25 0.50 0.25
```

Il quantile corrispondente a una probabilità cumulata $\leq .75$ è = 1; quello corrispondente a una probabilità cumulata $>.75$ è 0.

Infine, è possibile **generare una variabile aleatoria** composta da numeri **casuali** che assumono una distribuzione binomiale con `rbinom (n, size, prob)`: `n=` numero di osservazioni, `size=` prove, `prob=` probabilità dell'evento atteso in ogni prova.

Riassumiamo con un esempio: una prova d'esame è composta da **30 domande** a scelta multipla: **un'alternativa corretta e due errate**. Si assegna un punto alla risposta **corretta (evento atteso)** e zero punti a quella sbagliata (evento non atteso) → distribuzione **binomiale**. La **probabilità** dell'evento atteso è, per ogni prova, $1/3 = 0.33$; supponiamo che le alternative di risposta siano tutte realmente **equiprobabili** per chi risponde completamente a caso.

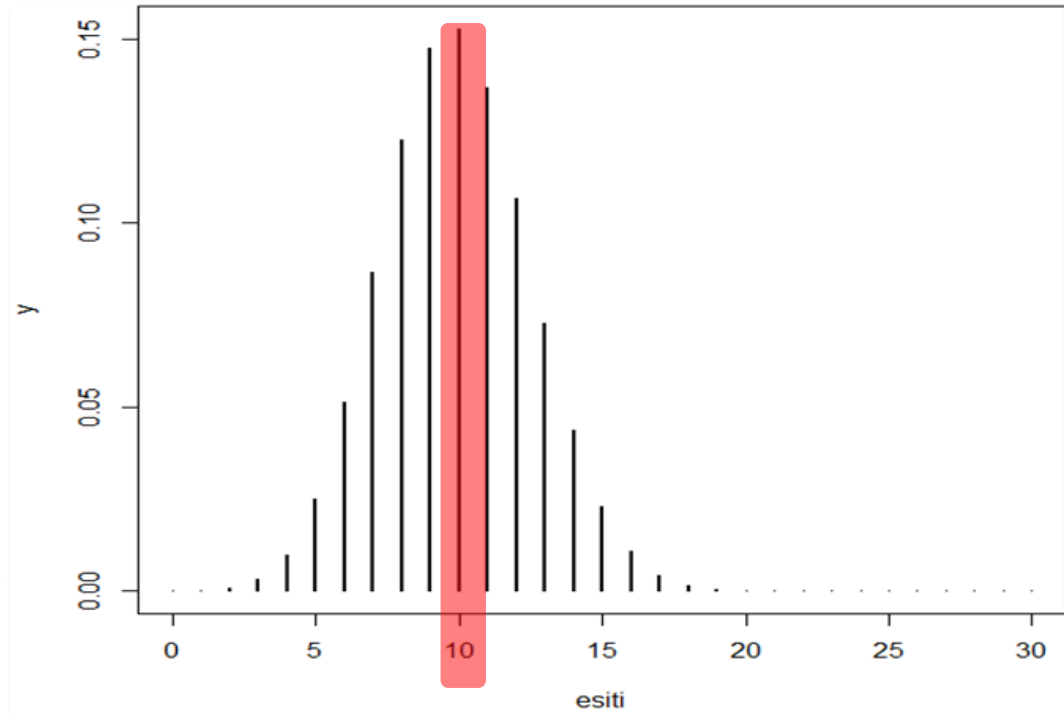
Quanto è probabile che uno studente, rispondendo **del tutto** a caso, ottenga 18?

```
dbinom(x = 18, size = 30, prob = .333)  
[1] 0.001700154
```

Molto poco probabile...-

Costruiamo il plot della massa di probabilità per vedere qual è il risultato più probabile e quali quelli praticamente impossibili: in *X* il vettore degli esiti possibili, in *Y* la massa di probabilità

```
esiti<-c(0:30)  
y<-dbinom(esiti, 30, .33)  
plot(esiti, y, type="h", lwd=2)
```



Quale sia la probabilità massima riscontrata nella distribuzione binomiale?

```
max(dbinom(esiti, 30, .33))  
[1] 0.1529001
```

La probabilità di rispondere correttamente per caso **al massimo a 12** domande (funzione di ripartizione) è uguale a:

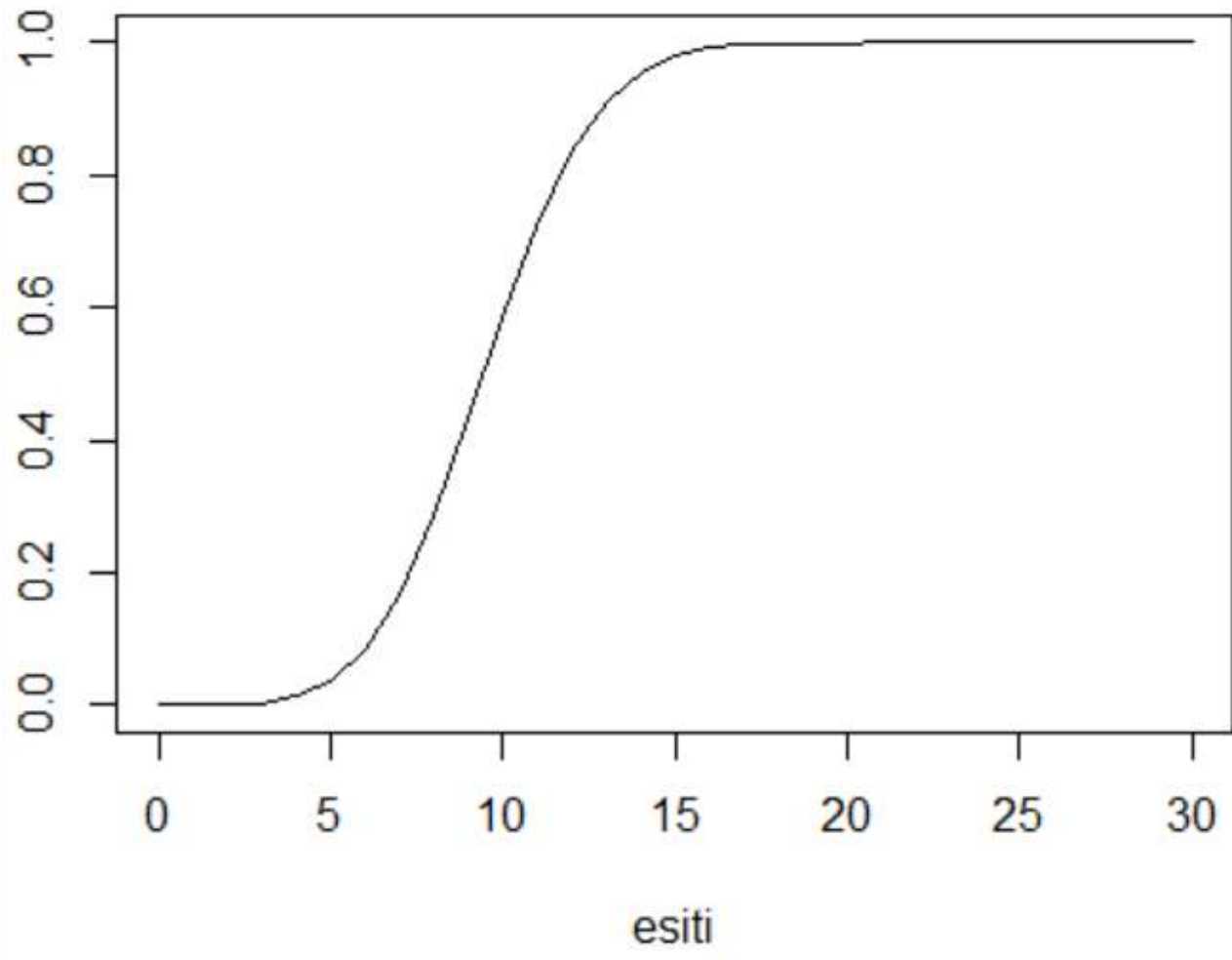
```
pbinom(q = 12, size = 30, prob = .33, lower.tail = TRUE)  
[1] 0.1562739
```

Invece, la probabilità di rispondere correttamente per caso a **10 o più** domande è uguale

```
pbinom(q = 9, size = 30, prob = .33, lower.tail = FALSE)  
[1] 0.447119
```

Costruiamo il plot della funzione di ripartizione.

```
esiti<-c(0:30)  
y<-dbinom(esiti, 30, .33)  
plot(esiti, y, type="h", lwd=2)
```



Altre distribuzioni di probabilità discrete

Se il campione è estratto da una popolazione **finita, senza possibilità di reinserimento**, la distribuzione binomiale non è applicabile, dato che **gli eventi non sono indipendenti** (l'estrazione di un elemento del campione condiziona le probabilità dei successivi).

Si utilizza la **distribuzione ipergeometrica**, che misura la **probabilità dei successi / eventi attesi X** in un campione di N elementi, **presi a caso e senza reinserimento, da una popolazione di N elementi.**

$$P(X = k) = \frac{\binom{a}{k} \binom{N-a}{n-k}}{\binom{N}{n}}$$

Se il campione estratto è molto piccolo rispetto alla popolazione (meno del 5%), il mancato reinserimento ha scarso effetto nella probabilità di successo di ogni prova e allora la distribuzione binomiale è una buona approssimazione di quella ipergeometrica.

Per stimare la probabilità di un evento da una distribuzione ipergeometrica usiamo **dhyper**, i cui argomenti sono: **x= evento atteso, m= numero successi in popolazione, n= numero insuccessi in popolazione, k= numerosità campione**)

Probabilità di selezionare casualmente **2 palline bianche (x)** da un'urna che contiene **5 palline bianche (m)** e **cinque palline nere (n)**, estraendo senza reinserimento **6 palline (k)**:

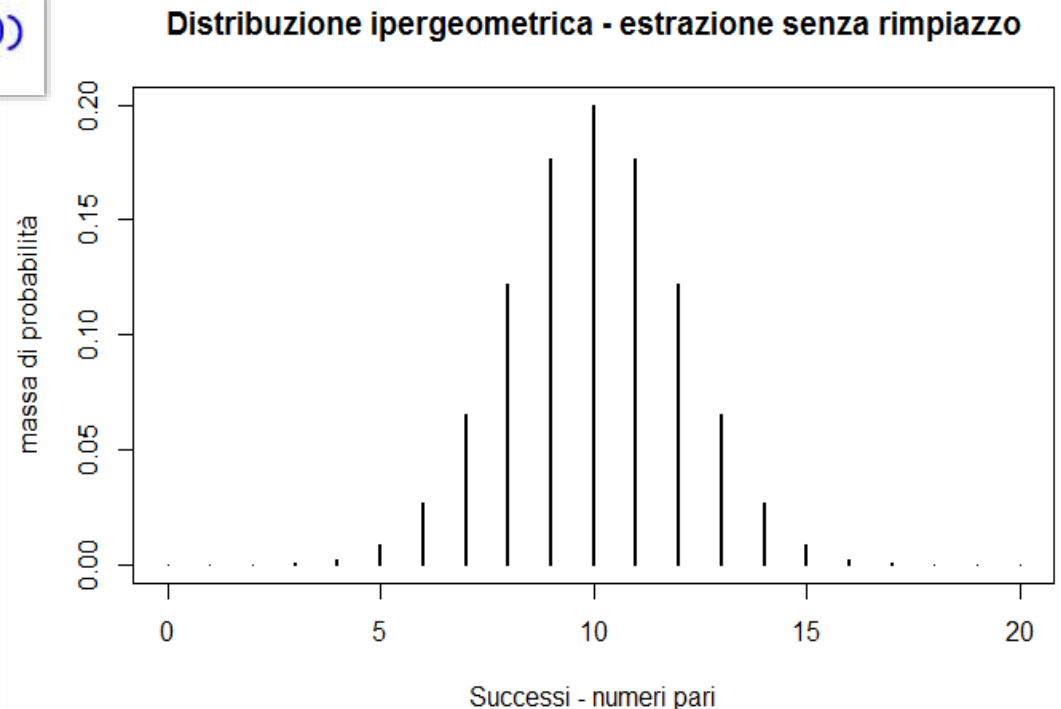
```
dhyper(x = 2, m = 5, n = 5, k = 6)  
[1] 0.2380952
```

Probabilità di estrarre **7 numeri pari (x)** dal bussolotto della tombola, che contiene **45 numeri pari (m)** e **45 dispari (n)**, nelle prime **20 estrazioni (k)**:

```
dhyper(x = 7, m = 45, n = 45, k = 20)  
[1] 0.06498521
```

Possiamo plottare la massa di probabilità ipergeometrica come prima:

```
x<-0:20  
y<-dhyper(x, m = 45, n = 45, k = 20)
```



```
> plot(x, y, type="h", main="Distribuzione ipergeometrica - estrazione senza rimpiazzo", ylab="massa di  
probabilità", xlab="Successi - numeri pari", lwd=2)
```

Come per la binomiale, possiamo calcolare la **funzione di ripartizione** con `phyper(q= , m= , n= , k= , lower.tail=TRUE/FALSE)`: `q` è il quantile in corrispondenza del quale effettuare la ripartizione; `m`, `n` e `k` danno le stesse indicazioni di `dhyper`.

L'inverso della funzione di ripartizione si ottiene con `qhyper(p= , m= , n= , k= , lower.tail=TRUE/FALSE)`

Una **distribuzione casuale** di numeri che segue la distribuzione ipergeometrica si ottiene con `rhyper(nn= , m= , n= , k=)`: `nn` indica il numero di osservazioni.

Ritroveremo la distribuzione ipergeometrica nel **test della probabilità esatta di Fisher**.

Se l'evento atteso è molto raro, o per conoscere la probabilità di un dato numero di successi per unità di tempo, se gli eventi sono indipendenti e il numero medio di successi per unità di tempo si mantiene costante, si usa la **distribuzione di Poisson**.

La distribuzione di Poisson si usa quando il numero di prove N è grande e la probabilità p che si verifichi l'evento è piccola (**evento raro: legge dei piccoli numeri**). Un evento è raro se il **numero delle prove N** è almeno pari a **50**, mentre $N \times p$ è < 5 .

Questa distribuzione ha un gran numero di applicazioni: controllo qualità, pazienti sofferenti di malattie rare, terremoti, telefonate a un destinatario sbagliato, bersagli colpiti a Londra dai bombardamenti durante la WWII, errori di stampa nei libri...

La formula della massa di probabilità di Poisson non è friendly: $P(X) = \lambda^X \times e^{-\lambda} / X!$

X : numero di successi; λ : numero medio di successi per unità di tempo; e : base del sistema di logaritmi naturali = 2.71828.

... Ma R ci toglie da ogni impaccio.

Per calcolare la probabilità di ricevere solo 2 mail di lavoro in un'ora a Ferragosto, sapendo che la media oraria negli altri giorni dell'anno è =5, usiamo `dpois (x=, lambda =)`: `x` numero di successi, `lambda` media per unità di tempo.

```
dpois(x = 2, lambda = 5)
[1] 0.08422434
```

Possiamo calcolare la **funzione di ripartizione** con `ppois(q= , lambda= , lower.tail=TRUE/FALSE)`: `q` è il quantile in corrispondenza del quale effettuare la ripartizione; `lambda` come in `dpois`.

L'inverso della funzione di ripartizione si ottiene con `qpois(p= , lambda= , lower.tail=TRUE/FALSE)`.

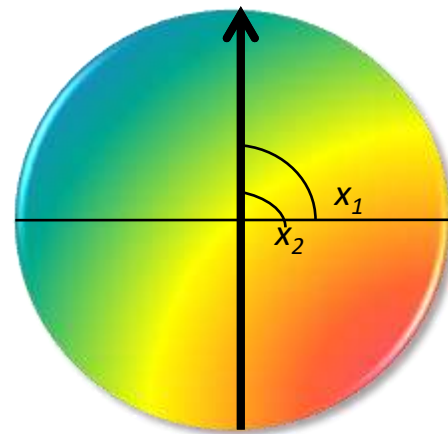
Una **distribuzione casuale** di numeri che segue la distribuzione di Poisson si ottiene con `rpois (n= , lambda=)`: `n` numero di osservazioni.

Distribuzioni di probabilità continue

Funzione di densità di probabilità

Alcuni esperimenti casuali non generano valori discreti: quando la variabile aleatoria X assume un **insieme continuo** di valori, è una **variabile casuale (o aleatoria) continua**.

Pensiamo a una “ruota della fortuna”, in cui l’ago è fatto ruotare finché terminerà casualmente la sua corsa in una certa posizione, che registreremo come **numero di gradi x_1** da una linea di riferimento. Se si fa ruotare l’ago una seconda volta, si otterrà un nuovo valore in gradi **x_2** , e così via.



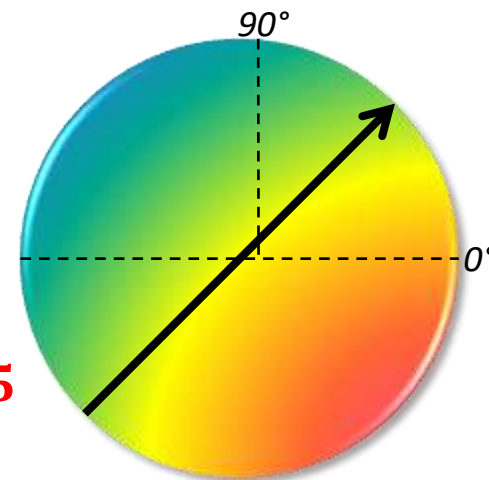
Possiamo **ottenere un numero infinito di diversi valori x_i** → **spazio campionario infinito Ω** , corrispondente a tutte le possibili posizioni assumibili. Poiché la rotazione è casuale, proviamo a calcolare la probabilità con cui la variabile aleatoria X assume un dato valore x_i usando la formula per eventi equiprobabili di una variabile discreta:

$$P(X = x) = \frac{\text{casi favorevoli}}{\text{casi possibili}} = \frac{1}{\infty} = \mathbf{0}$$

La formula comporta che **ciascuno dei possibili valori di X ha probabilità nulla di verificarsi!** Quindi, è un **problema** concettualizzare la probabilità di una variabile aleatoria continua come se fosse una variabile discreta. Ma non è impossibile calcolare le probabilità relative a eventi che coinvolgono variabili aleatorie continue: basta cambiare approccio...

Per esempio, **qual è la probabilità** che l'ago si fermi in un settore della ruota **compreso tra 0 e 90 gradi**? Dato che il settore corrisponde a **un quarto** dell'intera area, potremo stimare:

$$P(0 \leq x_i \leq 90) = P(x_i \leq 90) = \frac{1}{4} = .25$$



Dovremo valutare quanto **cambia la probabilità da un valore all'altro di un intervallo**, anche **infinitesimale**: $(x, x + dx)$, in cui dx esprime l'incremento di X ; se preferite, dovremo valutare la probabilità che **X assuma determinati valori nell'intervallo di riferimento** (rapporto incrementale).

Le variabili aleatorie continue sono **definite sempre a partire dall'insieme dei valori assumibili** e dalle **probabilità del tipo $P(X \leq x)$** : queste probabilità possono essere espresse attraverso la funzione **$f(x)$** , detta **funzione di densità di probabilità**.

La **funzione di densità di probabilità** associa a ogni \mathbb{R} il limite, per dx tendente a 0, del rapporto tra la probabilità che X assuma valori nell'intervallo $(x, x + dx)$ e l'ampiezza dx . Consente, quindi, di calcolare quanto valga la **probabilità attorno a X_i in rapporto all'ampiezza dell'intervallo** – ed ecco perché si definisce “**densità**”.

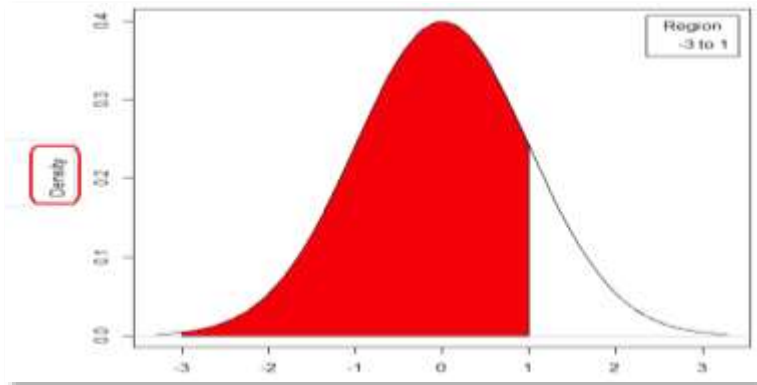
L'aspetto di $f(x)$ non è rassicurante:
$$F(x) = P(X \leq x) = \int_{-\infty}^{+\infty} f(u) du$$

... ma non la useremo mai direttamente. Consideriamola un'estensione della formula che abbiamo visto applicata alle variabili aleatorie discrete, in cui l'integrale sostituisce la sommatoria e la funzione di densità $f(x)$ sostituisce i termini $P(X = x_i)$.

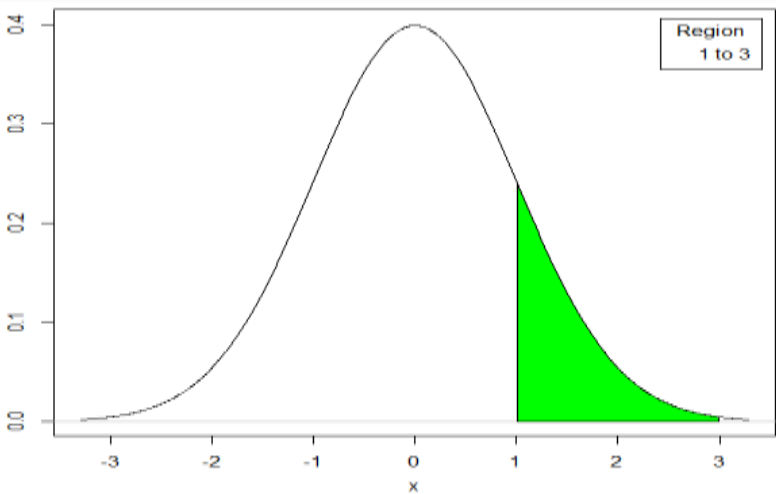
Calcolare l'integrale della funzione di densità corrisponde a calcolare l'area sotto la curva $f(x)$ stessa. La **somma delle probabilità** in corrispondenza di tutti i valori assumibili dalla variabile aleatoria continua è $= 1 \rightarrow$ **l'area totale sotto la curva è $f(x) = 1$.**

È utile usare un'interpretazione grafica della densità.

Con Rcommander, disegniamo tre distribuzioni aleatorie continue, di forma **normale**: Distribuzioni → distribuzioni continue → distribuzione normale → Disegna distribuzione normale).

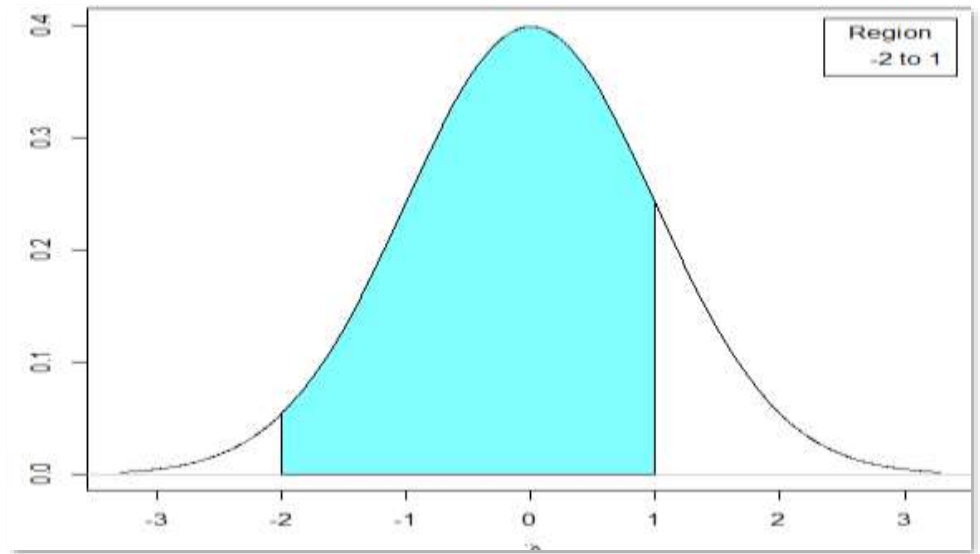


Abbiamo calcolato la funzione di densità da $-\infty$ a $+1$



Abbiamo rappresentato l'area a destra del punto $+1$, cioè $P(> 1)$

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du = (a)$$



Abbiamo calcolato il valore dell'area sotto $f(x)$ compresa tra -2 e $+1$, cioè $P(-2 < X \leq 1)$

Distribuzione di probabilità normale

Le funzioni applicabili in R a distribuzioni discrete si **applicano anche a variabili continue**, con qualche cambiamento negli argomenti.

Per calcolare la **funzione di densità** in una distribuzione **normale** si usa `dnorm(x, m, s)`, che restituisce **l'altezza della curva (la densità di probabilità) normale**, con media = `m` e deviazione standard = `s`, per ogni x_i . Se `m` e `s` non sono specificate, R assume che la distribuzione normale sia standardizzata ($\mu = 0, \sigma = 1$).

La densità di probabilità di **+2** in una distribuzione normale con `m= 0` e `s= 1` è:

```
dnorm(x = 2, mean = 0, sd = 1)
[1] 0.05399097
```

Dato che la curva normale è simmetrica, la densità di probabilità di **-2** è identica:

```
dnorm(x = -2, mean = 0, sd = 1)
[1] 0.05399097
```

La curva di densità normale si disegna con `curve(expr=funzione, da=, a=)`: la funzione di `expr=` sarà, naturalmente, `dnorm`.

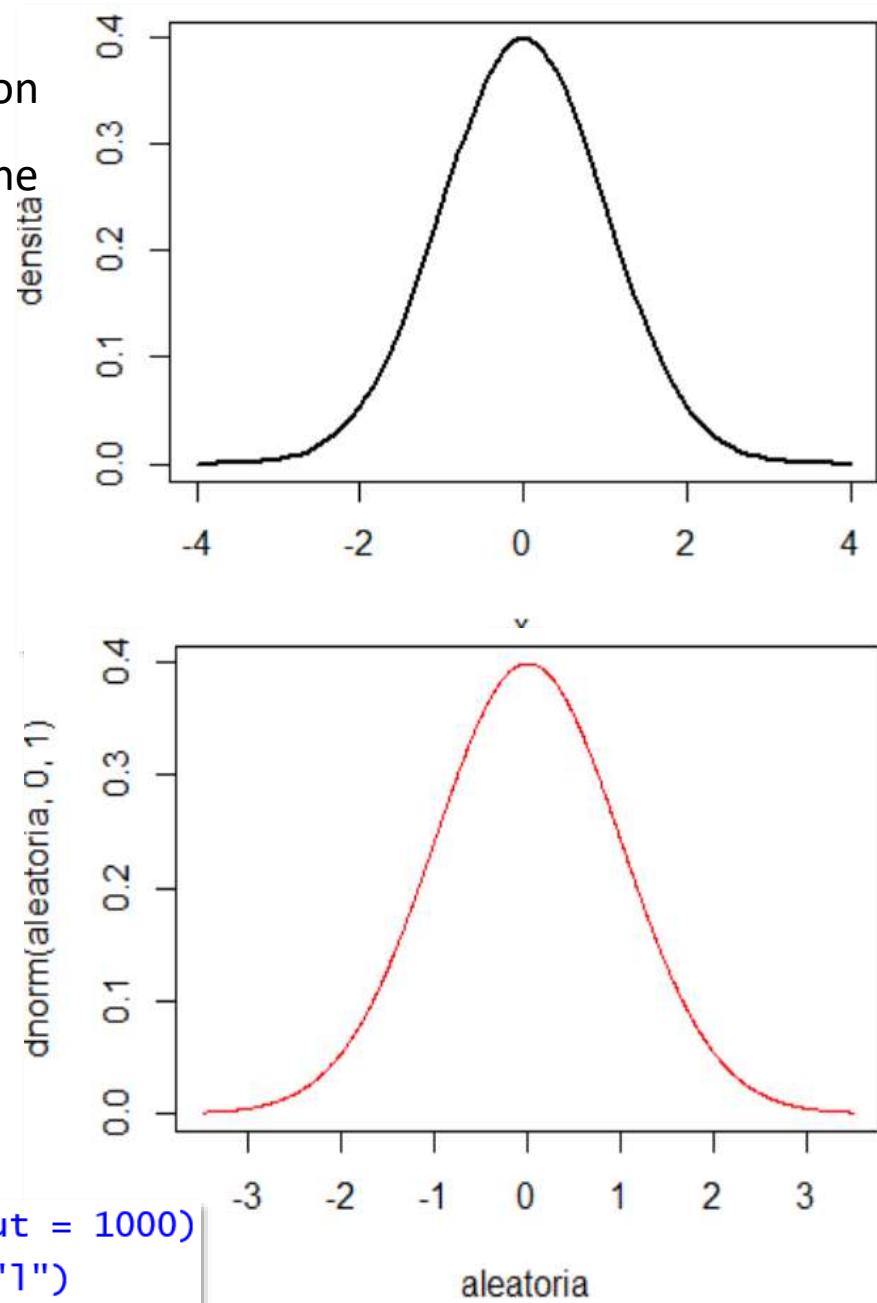
```
curve(dnorm, from = -4, to = 4, lwd=2,  
      ylab="densità", xlab="x")
```

Possiamo anche rappresentare `dnorm` con `plot`.

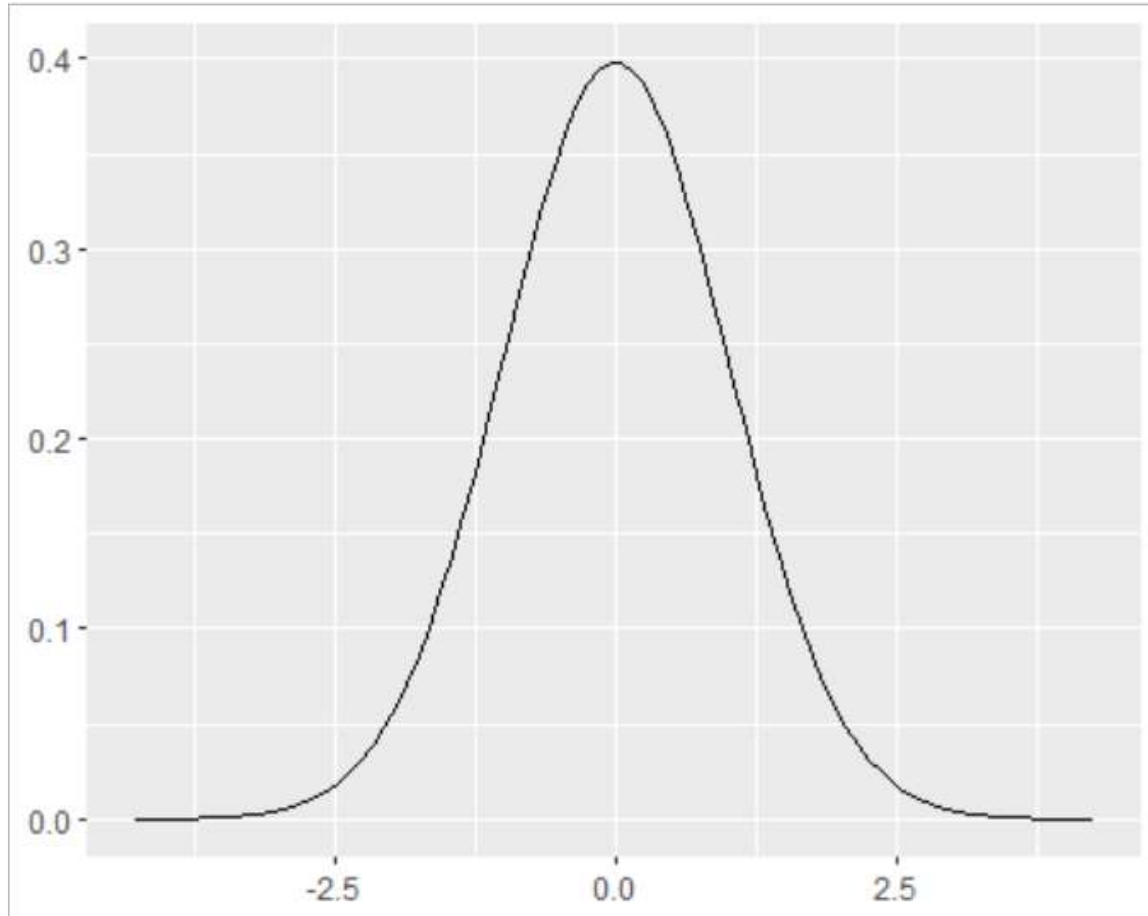
Il grafico sarà di `type= "l"` (`lines`) per visualizzare la curva come linea continua (provate a omettere `type` ...).

Si crea una variabile aleatoria e la si plotta:

```
aleatoria<-seq(from = -3.5,to = 3.5, length.out = 1000)  
plot(aleatoria, dnorm(aleatoria, 0,1), type= "l")
```



Infine, si può usare anche `dist_norm(mean= , sd=)` di `sjPlot`. Per una normale standardizzata, usiamo `mean=0` e `sd=1` :



La **funzione di ripartizione** `pnorm(q= quantile, m, s, lower.tail=TRUE/FALSE)` calcola la probabilità cumulata: **da $-\infty$ fino a x** (`lower.tail=TRUE`) o **da x a $+\infty$** (`lower.tail=FALSE`).

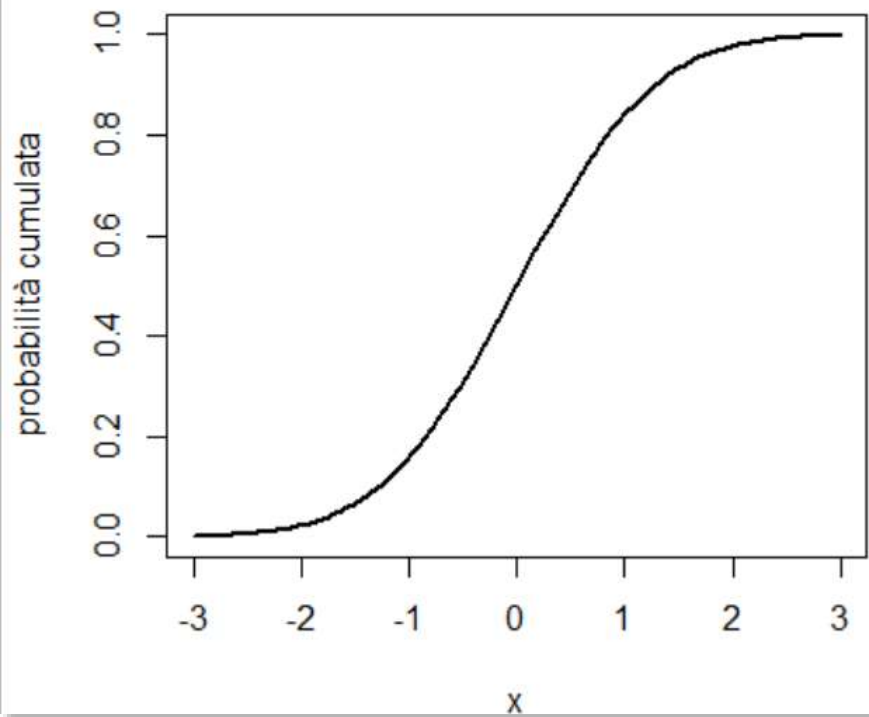
Usiamo lo “speciale” quantile della distribuzione normale standardizzata **$z= 1.96$** : la **probabilità cumulata** di ottenere un quantile **z uguale o inferiore a 1.96** è:

```
pnorm(q = 1.96, mean = 0, sd = 1, lower.tail = TRUE)  
[1] 0.9750021
```

La **probabilità cumulata** di ottenere un quantile **z maggiore di 1.96** è:

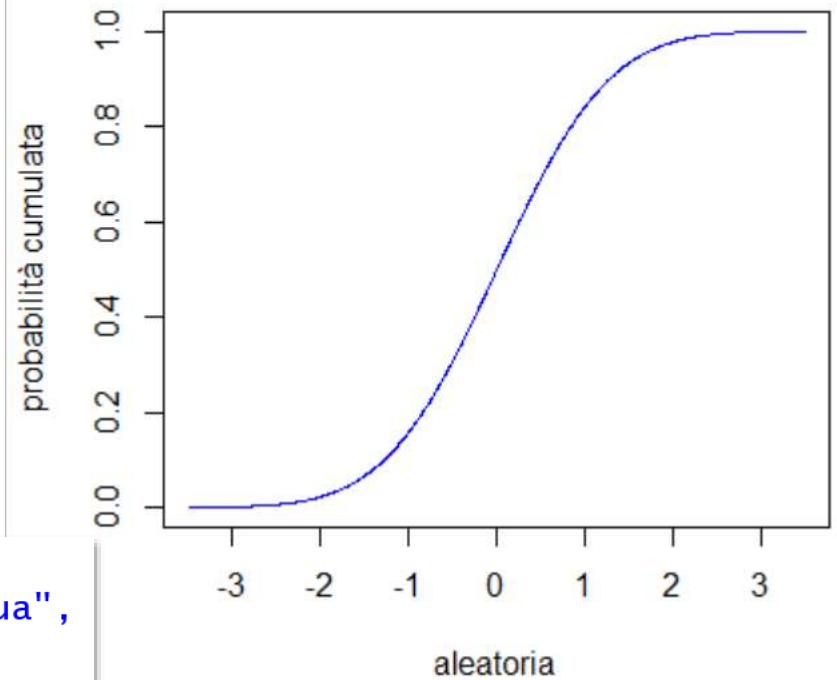
```
pnorm(q = 1.96, mean = 0, sd = 1, lower.tail = FALSE)  
[1] 0.0249979
```

Disegniamo l'**ogiva** delle probabilità cumulate con `curve`, applicata a `pnorm`, oppure applicando `plot` a una distribuzione aleatoria normale:



```
curve(pnorm, from = -3,to = 3, lwd=2,
      ylab= "probabilità cumulata",
      xlab="x")
```

probabilità cumulata variabile continua



```
plot(aleatoria, pnorm(aleatoria, 0,1),
     main="probabilità cumulata variabile continua",
     type="l", ylab="probabilità cumulata",
     col="blue")
```

L'inverso della funzione di ripartizione è `qnorm(q=quantile, m, s, Lower.tail=TRUE/FALSE)`, che restituisce il **quantile** della **variabile aleatoria normale** corrispondente alla probabilità cumulata indicata nella funzione.

Il quantile che corrisponde a una **probabilità cumulata =.50** è:

```
qnorm(p = .50, mean = 0, sd =1)
[1] 0
```

Un altro quantile “magico”, che **lascia alla sua sinistra** (fino a $-\infty$) il **95%** della distribuzione e che ritroveremo nell’inferenza, è:

```
qnorm(p = .95, mean = 0, sd =1)
[1] 1.644854
```

```
qnorm(p = .05, mean = 0, sd =1)
[1] -1.644854
```



Per **generare** una **variabile aleatoria di numeri casuali** con distribuzione **normale** usiamo `rnorm(n=, m, s)`; `n=` è il numero di elementi della distribuzione.

```
random_normale<-rnorm(n = 1000, mean = 10, sd = 5)
```

Altre distribuzioni di probabilità continue

La distribuzione **chi quadrato** (χ^2 , K. Pearson) è una distribuzione di **valori standard al quadrato** (da 0 a $+\infty$), **indipendenti**, estratti da una variabile standardizzata distribuita normalmente, con μ e σ noti. : χ^2

Il parametro ν (ni) indica i **gradi di libertà** (df) della distribuzione χ^2 : **determina la sua forma.**

$$\chi^2_{\nu} = \sum_1^{\nu} \frac{(x_i - \mu)^2}{\sigma^2}$$

Useremo la distribuzione χ^2 per stimare probabilità nel test chi quadrato, nella verifica della normalità multivariata e, nell'esame di Tecniche di analisi di dati II, nella regressione logistica.

Vediamo **tre esempi di distribuzioni** χ^2 per **diversi gradi di libertà** ($\nu = 6$; $\nu = 30$; $\nu = 60$); creiamo la variabile **chi** e usiamo **dchisq(variabile, df)** per la funzione di densità; plottiamole variando l'argomento **df= gradi di libertà**, per verificare cosa succede alla curva **man mano che i df aumentano**, e impariamo a **sovrapporre in un solo grafico.**

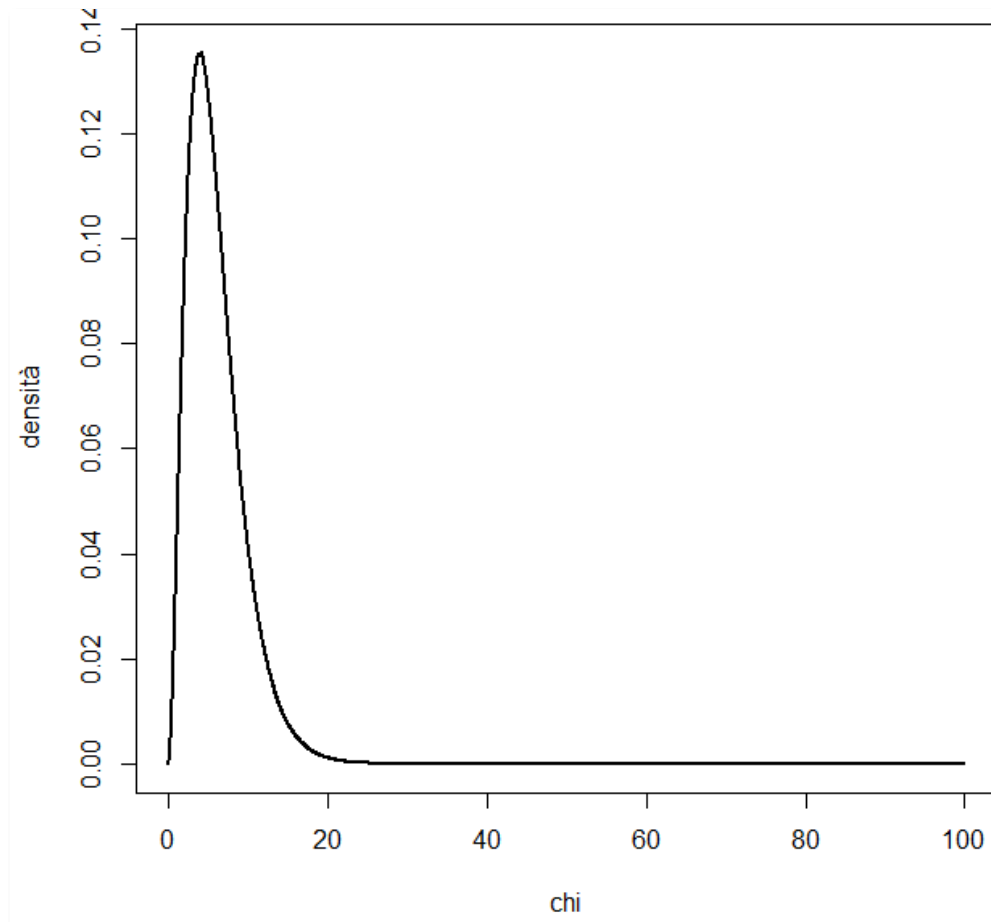
```
chi<-seq(0, 100, by=.05)
```

Assegniamo a `chi_6`, `chi_30` e `chi_60` la
funzione di densità per $\nu = 6, \nu = 30, \nu = 60$

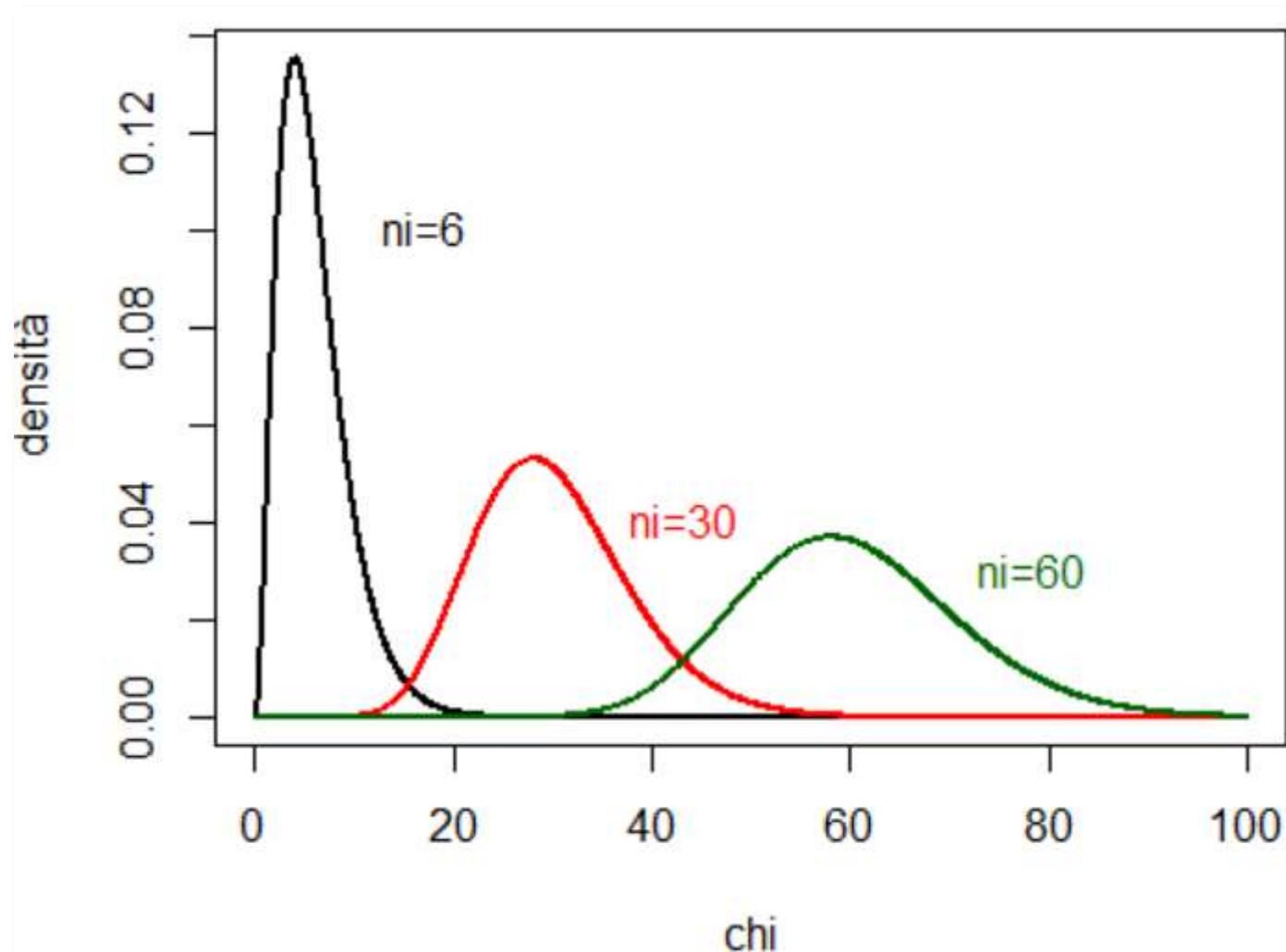
```
chi_6<-dchisq(x = chi, df = 6)  
chi_30<-dchisq(x = chi, df = 30)  
chi_60<-dchisq(x = chi, df = 60)
```

```
plot(chi, chi_6, type = "l", lwd=2, xlab="chi", ylab="densità")
```

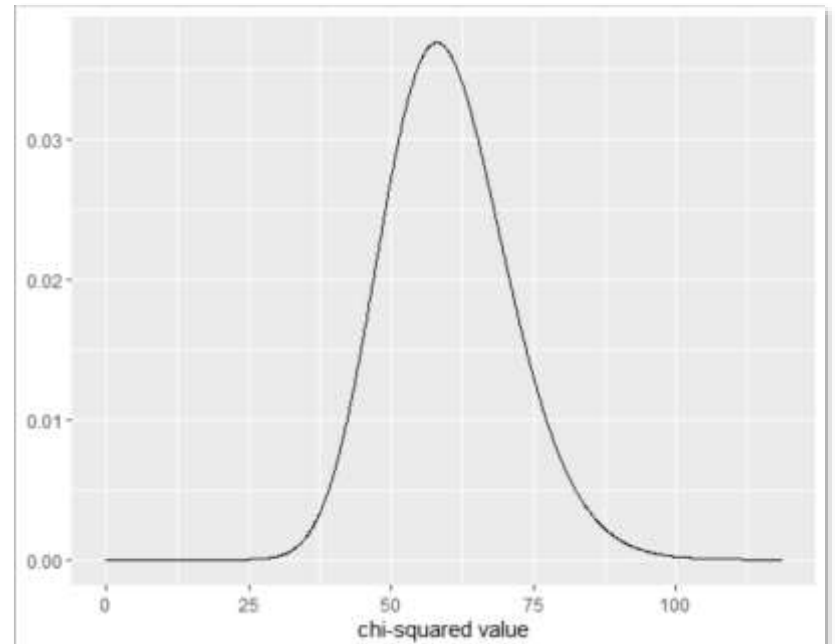
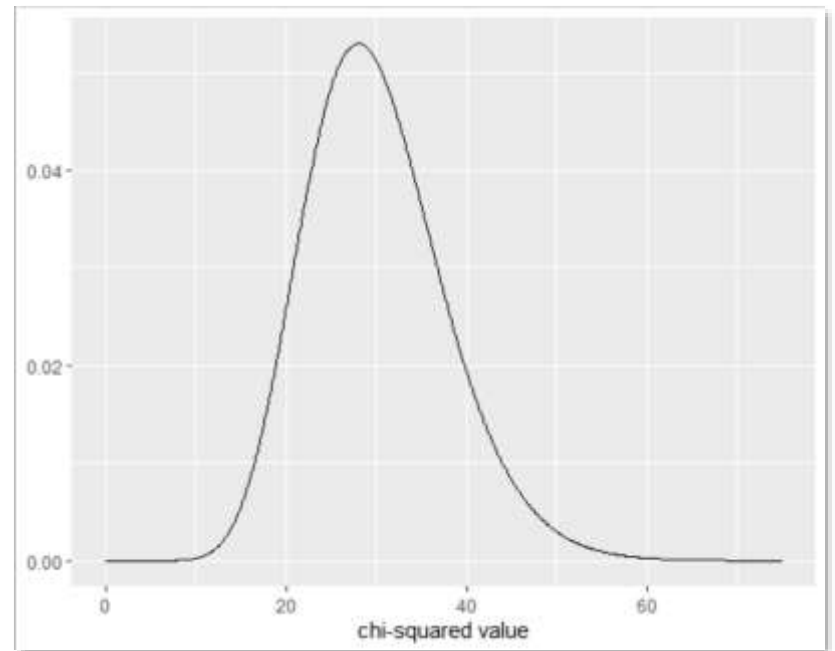
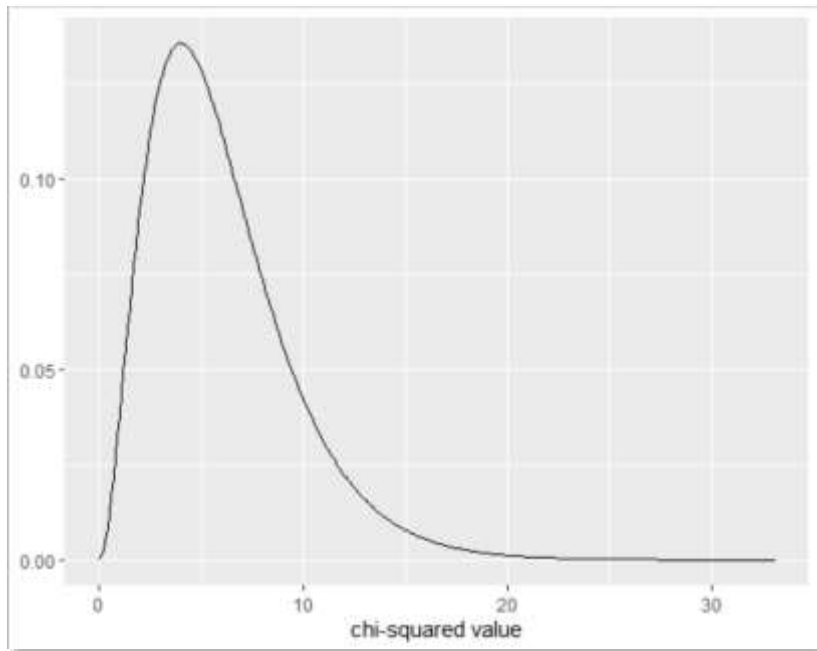
Invece di creare **altri** due plot,
disegniamo le distribuzione di
densità **in questo stesso plot** con
lines(x, y): `x` è ancora `chi`, `y` è
prima `chi_30`, poi `chi_60`.



Notate cosa succede alle distribuzioni man mano che i df aumentano?



Provate a creare i grafici delle tre distribuzioni, aggiungendo con [text](#) anche il testo nel grafico




Si può usare anche `dist_chisq(df=)` di `sjPlot`, specificando solo i diversi gradi di libertà:

Usiamo `rchisq(numero di osservazioni, df)` per creare un vettore di numeri casuali con distribuzione χ^2

`pchisq(quantile, df, lower.tail = TRUE / FALSE)` per la probabilità cumulata e

`qchisq(probabilità cumulata, df, lower.tail=TRUE/FALSE)` come inverso della precedente.

La distribuzione **F di Snedecor** (così da lui battezzata in onore di Fisher) è data dal **rapporto tra due distribuzioni χ^2 indipendenti**, quindi assume solo valori positivi.

$$F_{v1,v2} = \frac{\chi_{v1}^2}{\chi_{v2}^2} \times \frac{v1}{v2}$$


I **gradi di libertà** che definiscono la distribuzione F sono **due**: uno per la distribuzione al **numeratore**, l'altro per quella al **denominatore**.

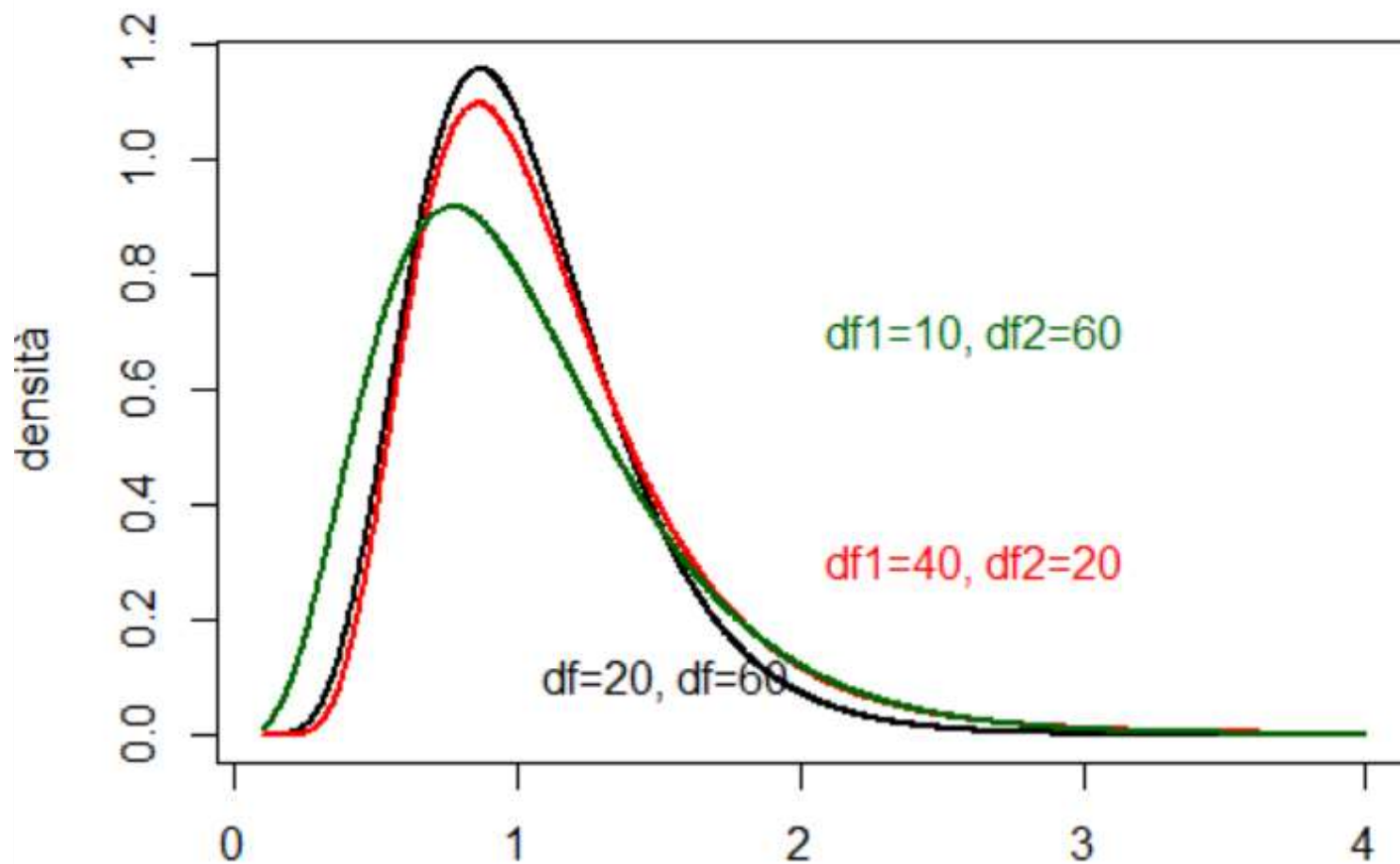
Useremo la distribuzione F nei modelli lineari: regressione multipla e analisi della varianza.

df(variabile, df1, df2) stima la densità,

rf(numero di osservazioni, df1, df2) crea una distribuzione di numeri casuali con distribuzione F,

pf(quantile, df1, df2 lower.tail=TRUE/FALSE) ricava la probabilità cumulata e

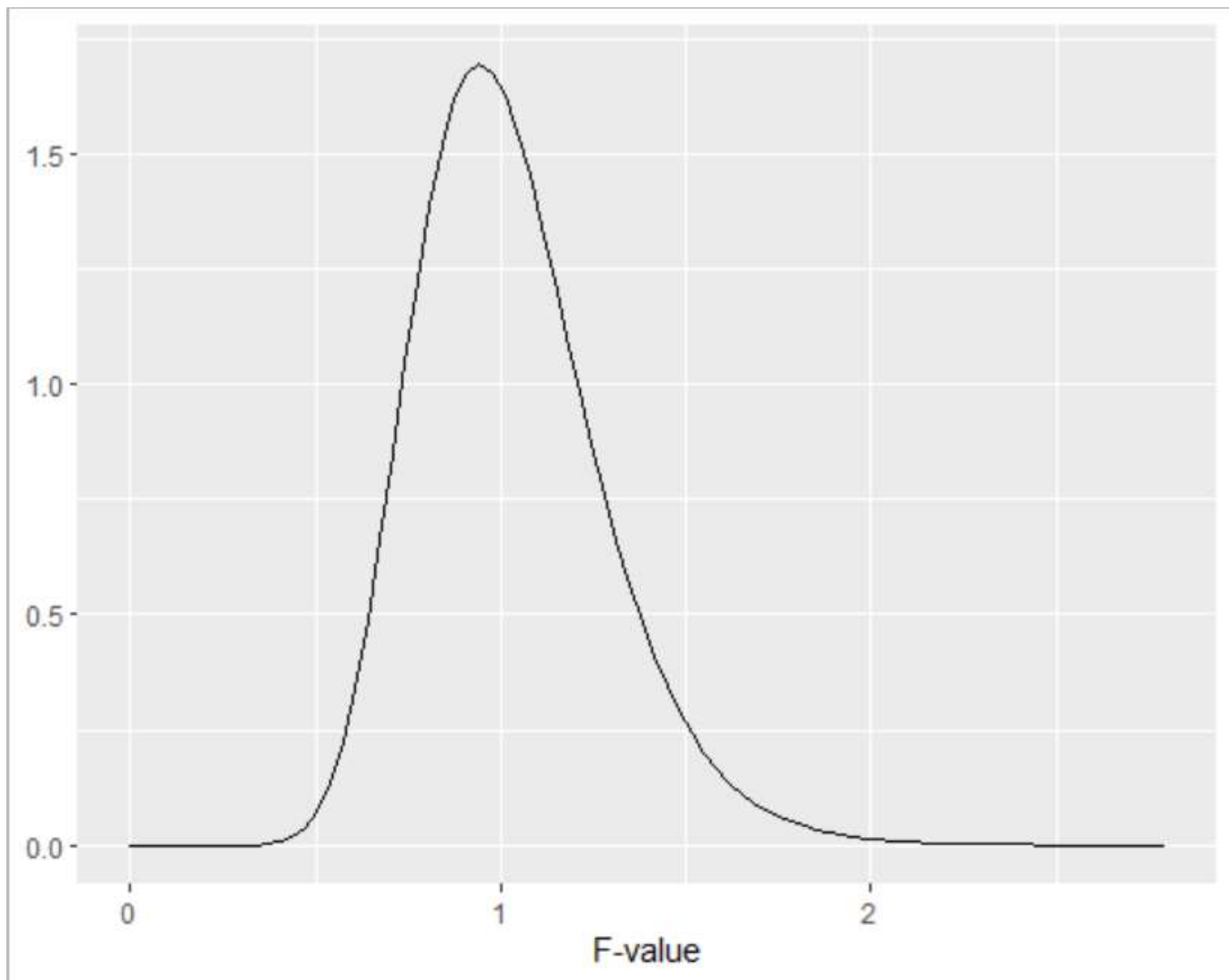
qf(probabilità cumulata, df1, df2, lower.tail=TRUE/FALSE) come inverso della precedente.



```


effe<-seq(.1, 4, length.out = 1000)
plot(effe, df(x = effe, df1 = 20, df2 = 60), type="l", lwd=2, ylab="densità")
lines(effe, df(effe,df1 = 40, df2 = 20), type="l", col="red", lwd=2)
lines(effe, df(effe,df1 = 10, df2 = 60), type="l", col="dark green", lwd=2)
text(x=c(1, 2,2), y=c(.1, .3, .7), labels=c("df=20, df=60", "df1=40, df2=20", "df1=10,
df2=60"),col=c("black", "red", "dark green"), pos=4)

```



Questo è il grafico ottenuto con `dist_f(deg.f=)` di `sjPlot`: $\nu_1 = \text{dfg.f}=100$, $\nu_2 = \text{deg.f2}=80$.

La distribuzione **t di Student** (pseudonimo di Gosset) è data dalla **radice quadrata** della distribuzione **F**; anche in questo caso il parametro è $\nu = N - 1$.

$$t_\nu = \sqrt{F_{1,\nu}}$$


Per **df** tendenti a infinito (cioè sufficientemente **grandi**), si **approssima alla distribuzione normale**.

È **simmetrica intorno a $t = 0$** : assume quindi valori negativi e positivi.

Troveremo la distribuzione t in molti test: test per un campione, correlazione, t test (naturalmente), modelli lineari (regressione multipla e contrasti a priori e pianificati nell'analisi della varianza).

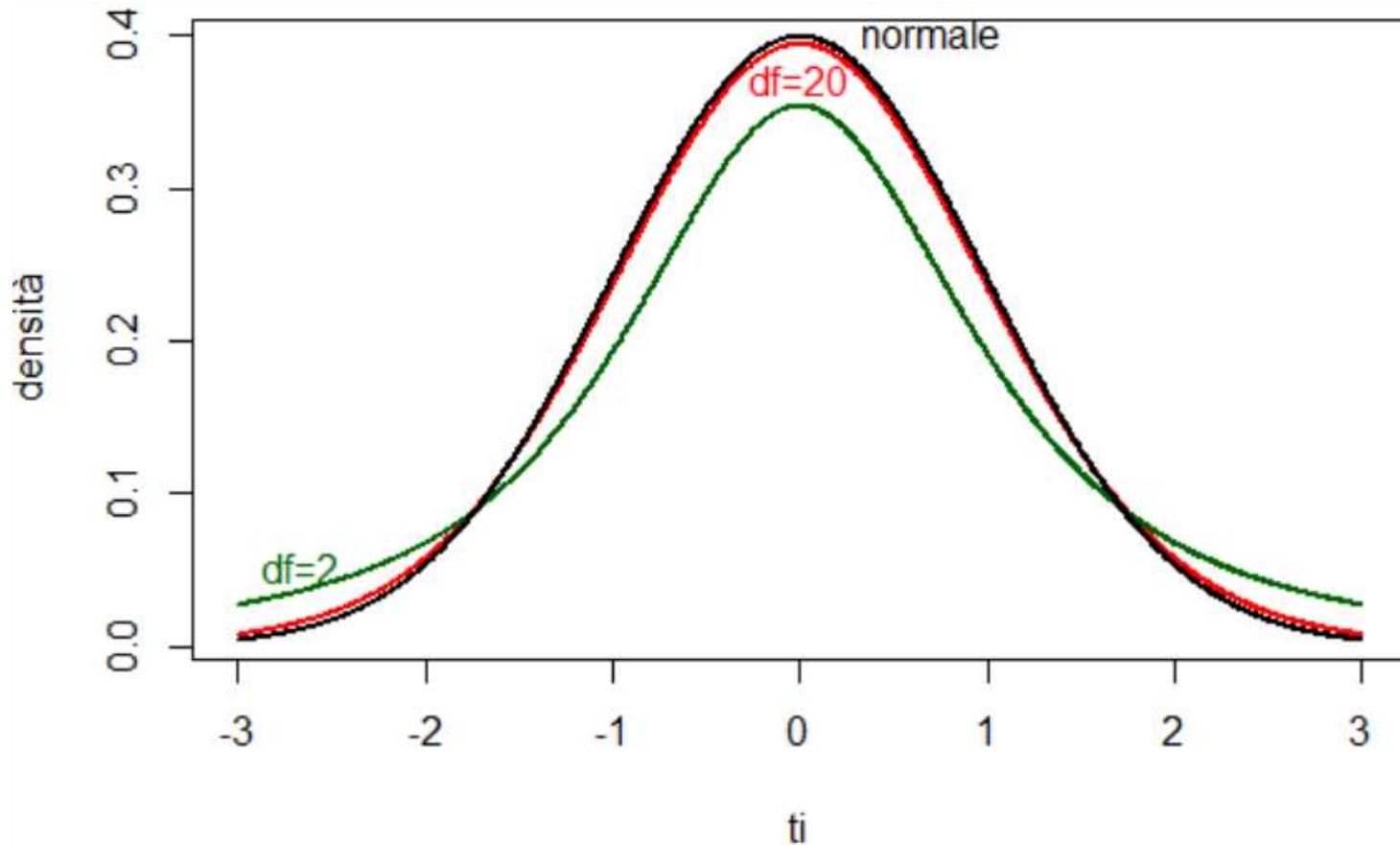
dt(variabile, df) stima la densità,

rt(numero di osservazioni, df) crea una distribuzione di numeri casuali con distribuzione t,

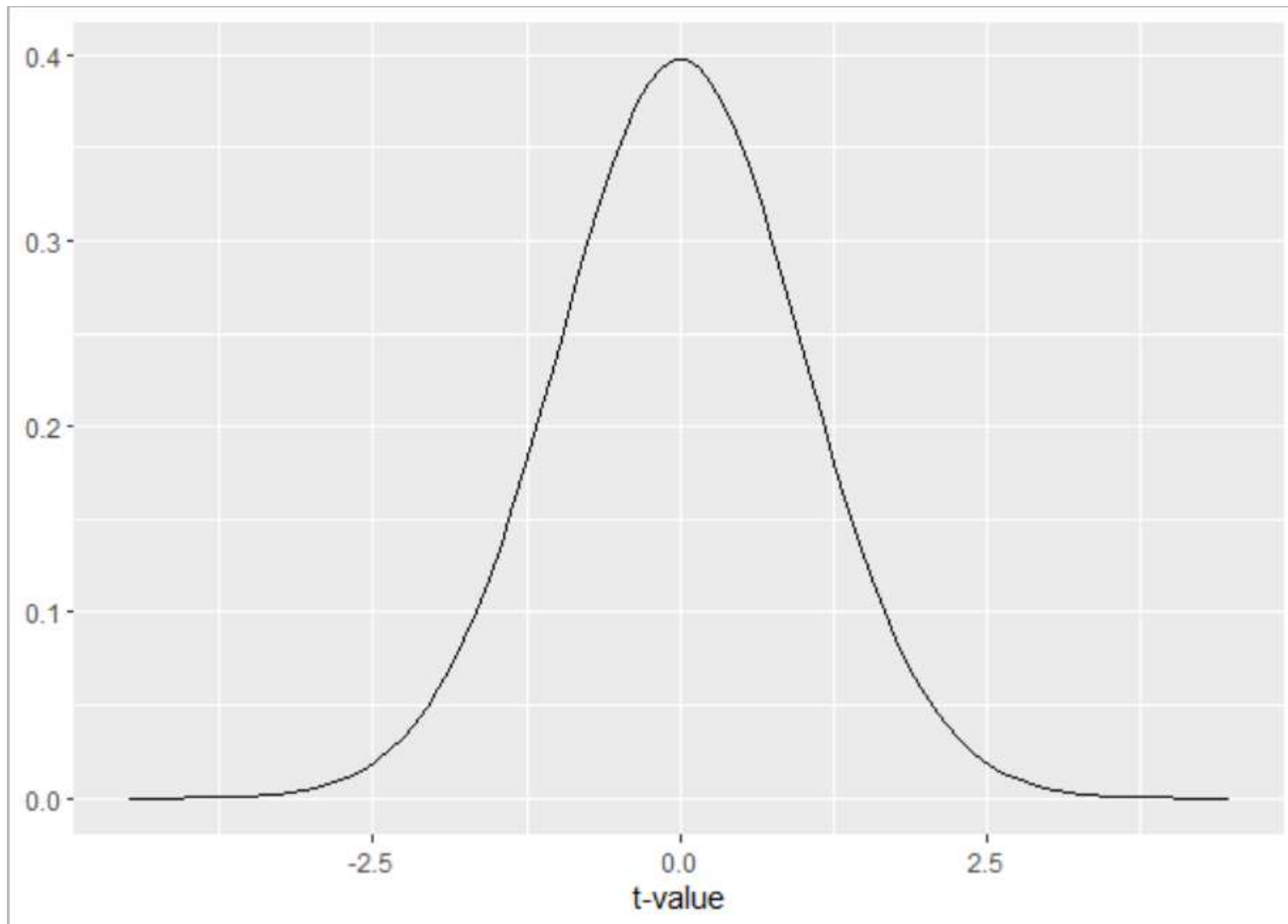
pf(quantile, df, lower.tail=TRUE/FALSE) ricava la probabilità cumulata e

qt(probabilità cumulata, df, lower.tail=TRUE/FALSE) è l'inverso della precedente.

```
ti<-seq(-3, 3, length.out = 1000)
normale<-seq(-3, 3, length.out = 1000)
```



```
plot(ti, dt(x = ti, df = 20), type="l", lwd=2, col="red", ylab="densità")
lines(ti, dt(x = ti, df = 2), type="l", lwd=2, col="dark green")
lines(normale, dnorm(normale, 0, 1), type="l", lwd=2, col="black")
text(x=c(-3, -.4, .2), y=c(.05, .37, .4), pos=4, labels=c("df=2", "df=20", "normale"),
     col=c("dark green", "red", "black"))
```



Questo è il grafico ottenuto con `dist_t(deg.f=)` di `sjPlot`: $\nu_1 = \text{dfg.f}=100$

Il quoziente d'intelligenza è distribuito normalmente in popolazione, con $\mu=100$ e $SE=15$.

Tra quali due valori di QI (eliminate tutti i decimali) si trova il 68.2% dei valori di QI della popolazione?

Quale proporzione della popolazione ha un QI < 80? E un QI < 110?

Quale proporzione della popolazione ha un QI compreso tra 95 e 115? E un QI compreso tra 70 e 85?

Quale valore di QI (eliminate tutti i decimali) ha il 10% della popolazione con il QI più alto?

Quale valore di QI (eliminate tutti i decimali) ha il 10% della popolazione con il QI più basso?