Laboratorio di Probabilità e Statistica

lezione 4

Massimo Guerriero – Ettore Benedetti

Indice Lezione

- Prerequisiti dalla lezione scorsa
- Permutazioni e coefficiente binomiale
- Variabili casuali discrete
  - Esperimento di Bernoulli e variabili casuali derivate
  - Variabile casuale di Poisson
- Variabili casuali continue
  - Variabile casuale Normale
  - Approssimazione della Binomiale con la Gaussiana
  - Variabile Chi-quadrato (X^2)
  - Variabile t di Student
  - Variabile F di Fisher

Prerequisiti dalla lezione scorsa

- Probabilità e probabilità condizionate
- Creazione e gestione di tabelle di contingenza
- Analisi di dipendenza
- Capacità di osservazione (evitare paradossi)

Permutazioni e coefficiente binomiale 1/2

Le **permutazioni** sono i modi di sistemare n oggetti differenti scambiandoli di posizione.

$$n! = n \cdot (n-1) \dots \cdot 2 \cdot 1$$

In R si utilizza lo stesso comando introdotto per l'esempio dei compleanni:
prod*(1 : n)*

Es.

prod(1:4) → *effettua l'operazione:* $1 \cdot 2 \cdot 3 \cdot 4 = 4!$

$= 24$

Permutazioni e coefficiente binomiale 2/2

Se invece vogliamo calcolare in quanti modi possibili possiamo estrarre da un gruppo di n, un sottogruppo di k oggetti,
utilizziamo il **coefficiente binomiale**:

$$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$$

In R si calcola con il comando ***choose(****n,k****)***

Es.

choose(4,3)
[1] 4

Variabili casuali discrete

Una variabile casuale è una grandezza che trasforma gli eventi elementari di un esperimento casuale in numeri della retta reale.
Si dice **discreta** quando questa variabile assume una quantità finita o numerabile di valori.

Es.

- *Il lancio di un dado può essere modellato da una variabile casuale che può assumere i valori 1, 2, 3, 4, 5, 6, tutti con probabilità 1/6*

- *L'esperimento di due lanci di una moneta può essere modellato da una variabile casuale che può assumere come valori il numero di volte in cui esce testa o croce.*

$\Omega = \{TT, TC, CT, CC\}$ $\quad P(w \in \Omega) = \{¼, ¼, ¼, ¼\}$

$X(w)=x$	$(2, 1, 1, 0)$	: *numero di teste*
$X=x$	$(2, 1, 0)$	: *variabile aleatoria X*
$P(X=x)$	$(¼, ½, ¼)$	: *distribuzione di probabilità di X*

Esperimento di Bernoulli e variabili casuali derivate 1/4

Un **esperimento Bernoulliano** consiste in un insieme di prove ripetute dove:

- Ogni singola prova può avere solo 2 esiti possibili: «successo» o «insuccesso»
- La probabilità che da origine al «successo» è costante
- I risultati delle prove sono indipendenti

Es.

L'esito di un sondaggio referendario:

- *supponiamo sia del tipo sì/no*
- *se prendiamo una scheda a caso, la probabilità di estrarre un «sì» è pari alla frequenza relativa.*
- *ogni elettore vota in modo indipendente l'uno dall'altro*

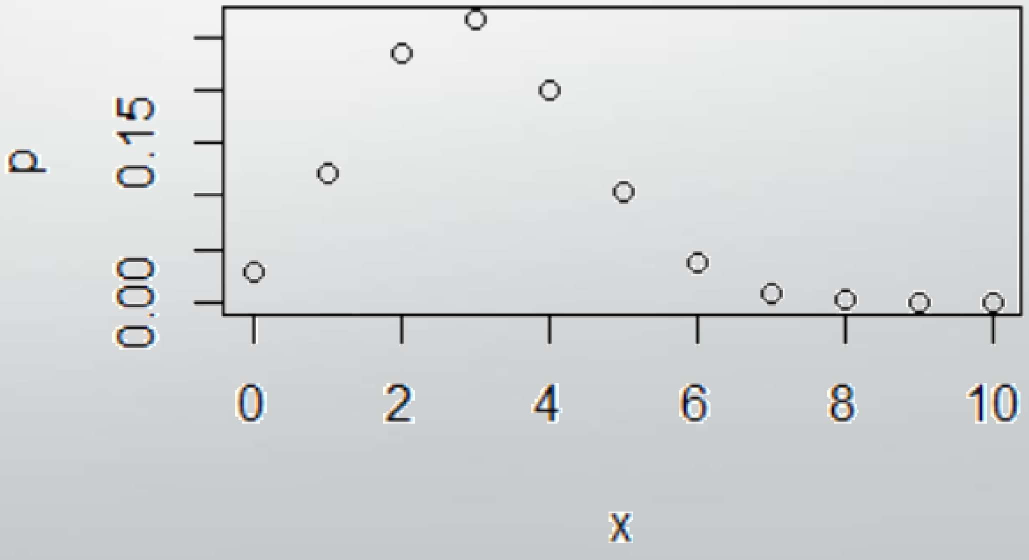
Esperimento di Bernoulli e variabili casuali derivate 2/4

La **variabile casuale di Bernoulli** è una variabile casuale discreta che può assumere solo due valori: *0 e 1* con probabilità *p* ed *1-p*.

E' un caso particolare della variabile casuale **binomiale,** la quale descrive in generale il numero di successi che si possono ottenere in *n* prove di Bernoulli.

Es.

Probabilità di successo per 10 esperimenti bernulliani con p=0.3



media = np   varianza = np(1-p)

Esperimento di Bernoulli e variabili casuali derivate 3/4

In R, per calcolare probabilità relative ad eventi che coinvolgono variabili casuali di tipo Binomiale o Bernoulli, si può usare il comando ***pbinom(****x,n,p****)***

Es. Ci chiediamo la probabilità di avere al più tre successi (P(X <= 3)) ripetendo 10 volte un esperimento bernoulliano in cui la probabilità di successo p è 0.3:

```
pbinom(3,10,0.3)
[1]  0.6496107
```

la probabilità complementare si calcola con il parametro lower.tail impostato a FALSE

```
pbinom(3,10,0.3, lower.tail=FALSE)
[1]  0.3503893
```

Se volessimo invece calcolare la densità di probabilità di X in un punto x dovremmo utilizzare il comando ***dbinom(****x,n,p****)***

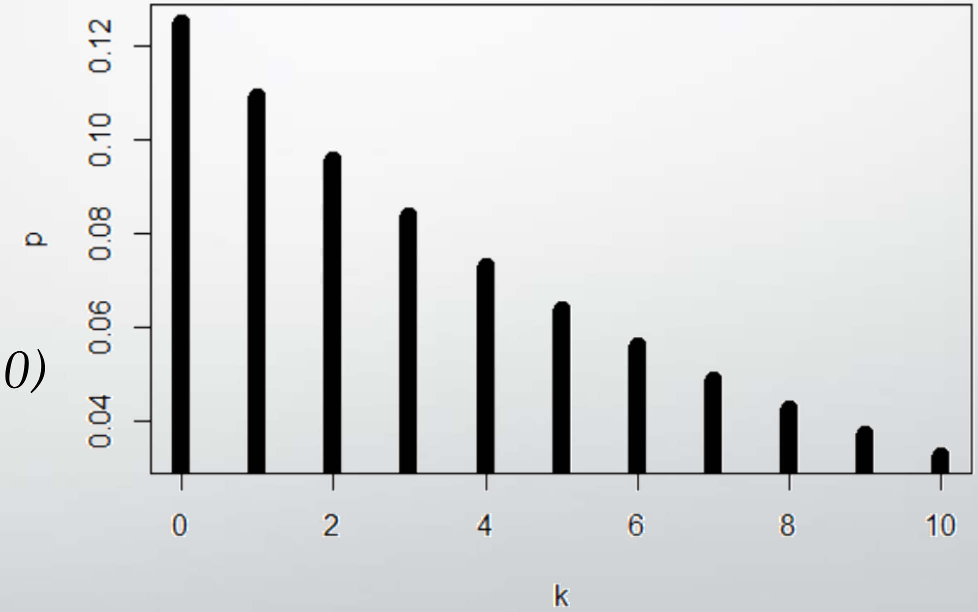
N.B Per Bernoulli basta ricordare che n = 1.

Esperimento di Bernoulli e variabili casuali derivate 4/4

In un esperimento Bernoulliano ci si può chiedere quanto tempo si deve aspettare per avere il primo successo.
La variabile casuale **geometrica** descrive la probabilità di avere il primo successo dopo k prove.

Es.

k <- 0:10
*p <- **dgeom(**k,1/8**)***
plot(k, p, type="h", lwd=10)



media= (1-p)/p
varianza= $(1\text{-}p)/p^2$

Il numero di insuccessi k che si devono avere prima di ottenere l'n-esimo successo è contato dalla variabile casuale **binomiale negativa**, una generalizzazione della geometrica (n=1).

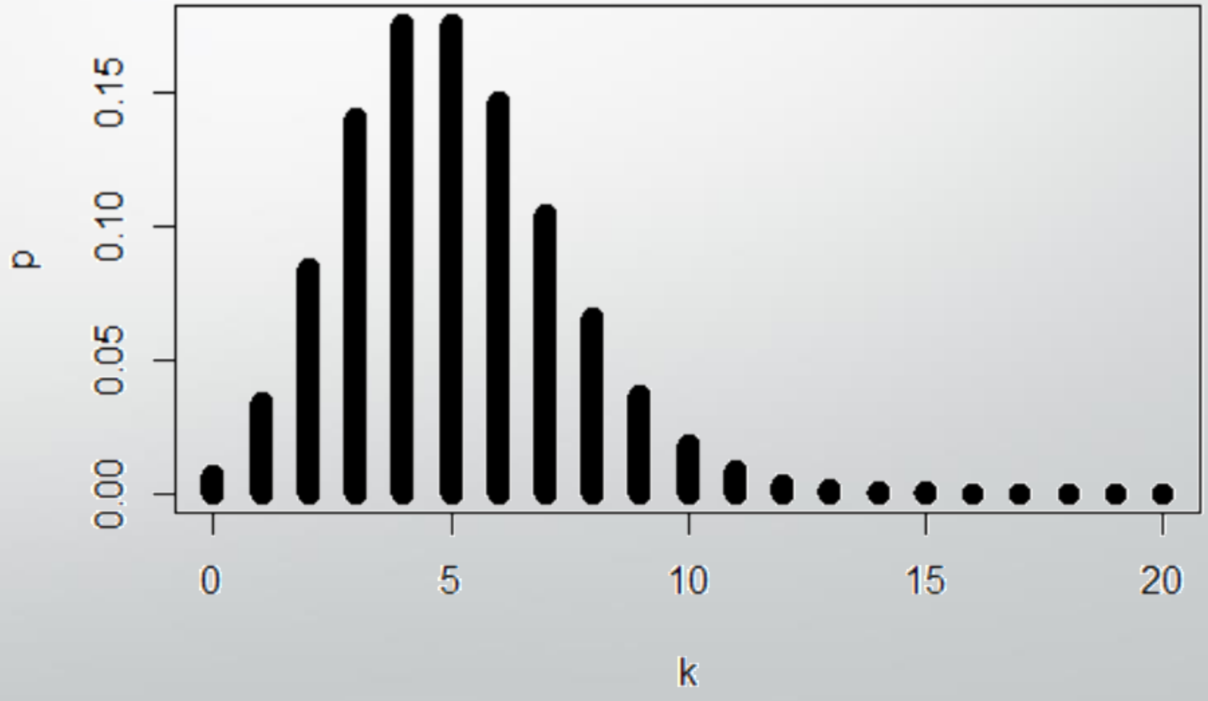
Variabile casuale di Poisson 1/2

Quando n è molto grande e p molto piccola (evento raro), la distribuzione binomiale viene approssimata correttamente dalla distribuzione di Poisson.

Es.

k <- 0:20
*p <- **dpois(**k, 5**)***
plot(k, p, type="h", lwd=10)

dove 5 è λ costante



media = λ = np = varianza

Variabile casuale di Poisson 2/2

Es.

mediamente un centralino riceve 20 chiamate in un'ora.
Qual è la probabilità che in 5 minuti non arrivano chiamate?
E che in 10 minuti si abbiano al più 10 chiamate?

ogni minuto arrivano mediamente 20/60 chiamate
*in 5 minuti arrivano mediamente 20/60*5 chiamate*
*in 10 minuti arrivano mediamente 20/60*10 chiamate*

ppois(**0,20/60*5**)	*ppois(10,20/60*10)*
[1] 0.1888756	*[1] 0.9993085*

Schema riassuntivo

Variabile Casuale	Bin(n,p)	Bin Neg(n,p)
Densità	dbinom(k, n, p)	dnbinom(k, n, p)
Ripartizione	pbinom(x, n, p)	pnbinom(x, n, p)
Quantili	qbinom(α, n, p)	qnbinom(α, n, p)

Variabile Casuale	Geom(p)	Poisson(λ)
Densità	dgeom(k, p)	dpois(k, λ)
Ripartizione	pgeom(x, p)	ppois(x, λ)
Quantili	qgeom(α, p)	qpois(α, λ)

Consegna

1. Un casello autostradale è frequentato mediamente da 120 macchine in 60 minuti di servizio.
   Ci si chiede la probabilità che in 15 minuti si abbiano al più 20 macchine.

2. Rappresentare il fenomeno con un grafico.

3. Ripetendo 20 volte un esperimento bernoulliano con p=0.25,
   qual è la probabilità che l'esperimento sia verificato
   più di 4 volte?

Indice Lezione

- Prerequisiti dalla lezione scorsa
- Permutazioni e coefficiente binomiale
- Variabili casuali discrete
  - Esperimento di Bernoulli e variabili casuali derivate
  - Variabile casuale di Poisson
- Variabili casuali continue
  - Variabile casuale Normale
  - Approssimazione della Binomiale con la Gaussiana
  - Variabile Chi-quadrato (X^2)
  - Variabile t di Student
  - Variabile F di Fisher

Variabili casuali continue

Una variabile casuale è una grandezza che trasforma gli eventi elementari di un esperimento casuale in numeri della retta reale.
Si dice **continua** quando questa variabile assume infiniti valori.

La probabilità che una variabile casuale X assuma infiniti valori, come un intervallo, non può essere calcolata semplicemente sommando la probabilità dei singoli valori, occorre integrare la funzione **densità di probabilità**.

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

Variabile casuale Normale 1/3

Si tratta di una variabile casuale continua molto utilizzata in diversi ambiti.

Normale standard:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad -\infty < x < \infty \qquad \text{media 0, varianza 1}$$

In generale:

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

La normale standard è fondamentale per il calcolo delle probabilità, di essa si conoscono tutte le probabilità, risolte con l'integrale, fornite da algoritmi standard come quelli utilizzati da R

Variabile casuale Normale 2/3

Ci si può ricondurre alla normale standard tramite l'operazione di **standardizzazione**

$$Z = \frac{X - \mu}{\sigma}$$

In R la standardizzazione viene effettuata automaticamente se passiamo come parametri media e deviazione standard.

Es.

calcoliamo P(X< 3) con X ~ N(5,3)

***pnorm(**3,5,sqrt(3)**)** #3 parametri: standardizzazione automatica*

[1] 0.1241065

pnorm((3-5)/sqrt(3))#1 parametro: standardizzazione manuale

[1] 0.1241065

Variabile casuale Normale 3/3

Vediamo ora come disegnarne il grafico con R:

curve(dnorm(x,-4,1), -10, 12, ylab=" ",axes=F)

curve(dnorm(x,7,2), -10, 12, ylab=" ",add=T)

Intervalli notevoli

$P(\mu - \sigma < X < \mu + \sigma)$

pnorm(1) – pnorm(-1)

[1] 0.6826895

$P(\mu - 2\sigma < X < \mu + 2\sigma)$

pnorm(2) – pnorm(-2)

[1] 0.9544997

$P(\mu - 3\sigma < X < \mu + 3\sigma)$

pnorm(3) – pnorm(-3)

[1] 0.9973002

Approssimazione della Binomiale con la Gaussiana

Un altro modo di ricavare la Normale è come limite della variabile casuale Binomiale.

De Moivre si accorse che con p=0.5 la distribuzione della binomiale è simmetrica rispetto alla media np.

Standardizzando:

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \cong N(0,1)$$

Per grandi campioni (n>30), l'approssimazione si può ritenere valida.

Variabile Chi-quadrato (X^2)

Se Z ~ N(0,1), la variabile $Y = Z^2$ si distribuisce con una legge nota col nome di distribuzione di Chi-quadrato con g=1 grado di libertà.

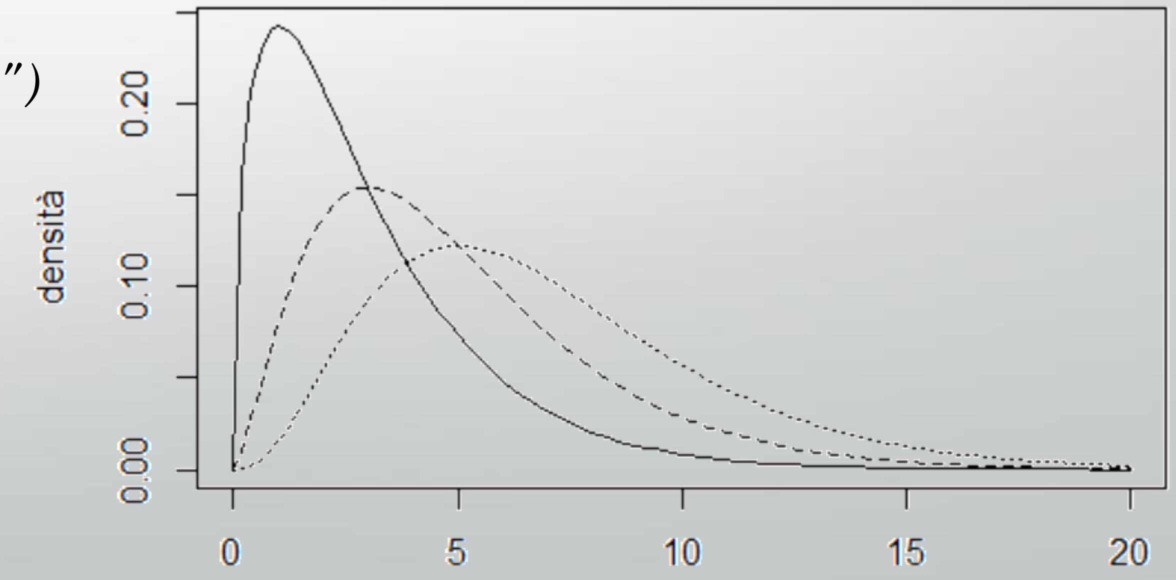
Questa distribuzione assume solo valori positivi ed è fortemente asimmetrica quando *g* è piccolo. Per *n* abbastanza grande, (*n>30*), la distribuzione chi-quadrato converge alla Normale.

Es.

*curve(***dchisq(***x, df=3***)***, 0., 20, ylab="densità")*

curve(dchisq(x, df=5), 0., 20,add=T,lty=2)

curve(dchisq(x, df=7), 0., 20,add=T,lty=3)



Variabile t di Student

Questa variabile si ottiene tramite la trasformazione:

$$t = \frac{Z}{\sqrt{\frac{X^2_g}{g}}} \qquad \textit{dove } Z \sim N(0,1) \textit{ indipendente da } X^2_g$$

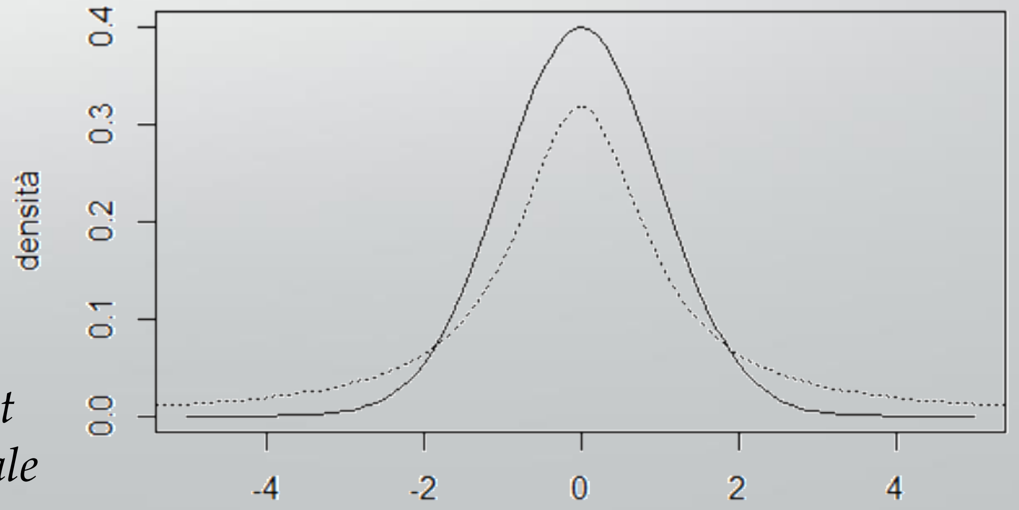
Anche questa distribuzione è caratterizzata dai gradi di libertà, è simile alla normale con la differenza che le code sono leggermente più alte.
Per *n* abbastanza grande, (*n>30*), la distribuzione t di Student converge alla Normale.

Es.

curve(dnorm(x), -5, 5, ylab ="densità")

*curve(**dt(**x, df=1**)**, -6, +6, lty=3, add=T)*

in tratteggio la t di Student
in linea continua la Normale



Variabile F di Fisher

Date due variabili $X \sim X^2_m$ e $Y \sim X^2_n$ indipendenti, il rapporto

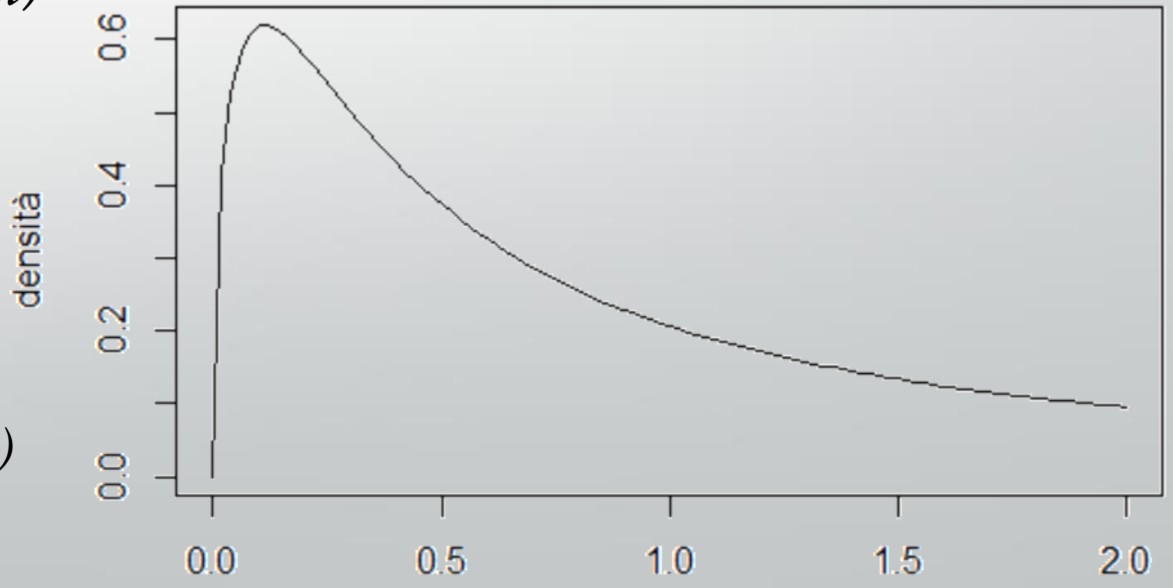
$$F = \frac{\frac{X^2_m}{m}}{\frac{X^2_n}{n}}$$

ha una distribuzione nota come F di Fisher con *m* ed *n* gradi di libertà.

Per *n (inteso come numero di campioni)* abbastanza grande, (*n*>*30*), la distribuzione F di Fisher converge alla Normale.

Es.

*curve(**df(**x,df1=3,df2=1**)**, 0, 2, ylab ="densità")*



Schema riassuntivo

Variabile Casuale	Normale(μ, σ)	t di Student (t_g)
Densità	dnorm(x, μ, σ)	dt(x, g)
Ripartizione	pnorm(x, μ, σ)	pt(x, g)
Quantili	qnorm(α, μ, σ)	qt(α, g)

Variabile Casuale	Chi-quadrato X^2_g	F(g_1, g_2)
Densità	dchisq(x, g)	df(k, g_1 , g_2)
Ripartizione	pchisq(x, g)	pf(x, g_1 , g_2)
Quantili	qchisq(α, g)	qf(α, g_1 , g_2)

Consegna

1. Calcolare gli intervalli notevoli della gaussiana su normali di media e varianza a piacere.

2. Costruire un grafico di una binomiale con p=0.5 ed n=100 sovrapposta alla normale standard

3. Calcolare i quantili c(25, 50, 75, 100) delle distribuzioni di probabilità delle variabili continue viste a lezione.