Laboratorio di ST1 - Lezione 2

Antonietta di Salvatore

Dipartimento di Matematica
Universitá degli Studi Roma Tre

Outline

- variabili discrete
  - Binomiale
  - Poisson
- variabili continue
  - Normale
  - Esponenziale
  - Gamma
  - Chi quadro
- teorema del limite centrale
- quantili

Comandi principali

dNOME() Funzione di densitá o distribuzione di probabilitá per variabili discrete
pNOME() Funzione di ripartizione
qNOME() Quantili
rNOME() Generatore di numeri pseudo-casuali
Ogni distribuzione ha dei suoi parametri: richiamare l'help per una corretta definizione di questi

ESEMPIO: distribuzione Binomiale

```
dbinom(x, size, prob)
pbinom(q, size, prob)
qbinom(p, size, prob)
rbinom(n, size, prob)
```

Densitá, funzione di distribuzione, quantili e funzione generatrice di numeri casuali per una funzione di distribuzione binomiale di parametri numerositá (size) e probabilitá (prob)

Le Variabili Discrete

Binomiale: $X \backsim Bin(p, n)$

Dato un esperimento dicotomico $S = (e_s, e_i)$, tale variabile rappresenta il numero dei successi e_i che si verificano in una sequenza di n sottoprove indipendenti nelle quali é costante la probabilit p di successo.

$$Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ per } x = 0, 1, 2, \ldots, n$$

Valore Atteso: $E(X) = np$
Varianza: $V(X) = np(1-p)$

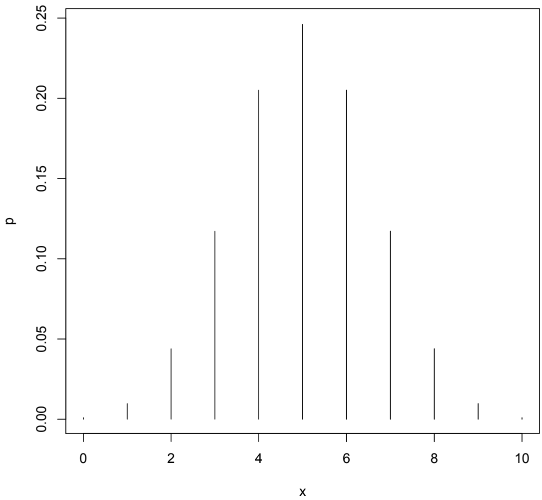
Consideriamo la probabilitá associata alla sequenza da 0 a 10 da una $Bin(10, 1/2)$ e disegnamo lo Spike plot

```
BINO = dbinom(0:10, size = 10, prob = 1/2)
plot(0:10, BINO, type = 'h', main= 'Spike plot di una
distribuzione Binomiale Bin(10, 1/2)', xlab= 'x', ylab = 'p')
```

Somma delle frequenze relative

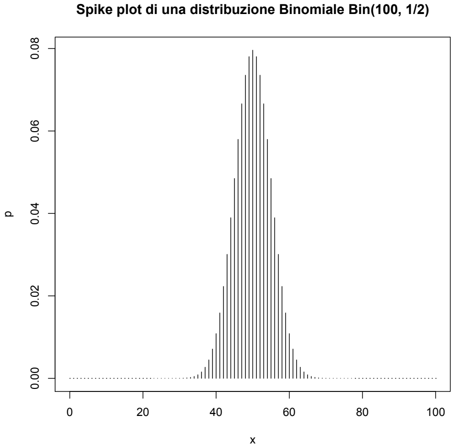
```
> sum(BINO)
[1] 1
```

Spike plot di una distribuzione Binomiale Bin(10, 1/2)

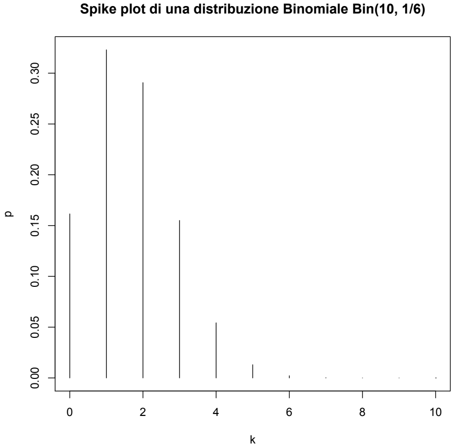


Cambiamo i parametri n e p

```
BINO = dbinom(0:100, size = 100, prob = 1/2)
plot(0:100, BINO, type = 'h', main= 'Spike plot di una
distribuzione Binomiale Bin(100, 1/2)', xlab= 'x', ylab = 'p')
```



```
BINO = dbinom(0:10, size = 10, prob = 1/6)
plot(0:10, BINO, type = 'h', main= 'Spike plot di una
distribuzione Binomiale Bin(10, 1/6)', xlab= 'k', ylab = 'p')
```



Poisson: $X \backsim Po(\lambda)$

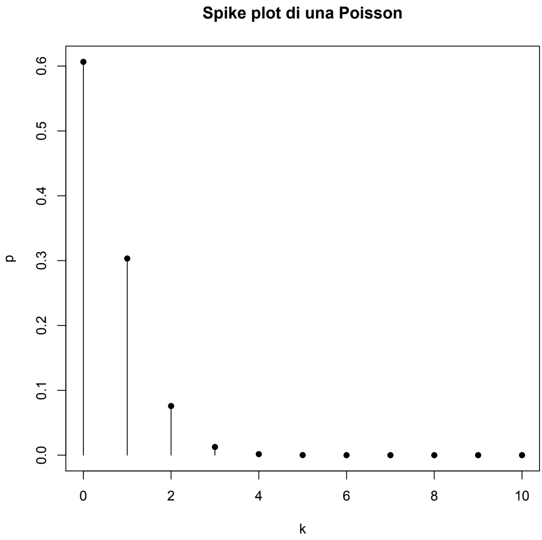
o anche 'legge degli eventi rari', riguarda il numero di eventi registrati in un ambito circoscritto, di tipo temporale, spaziale o concettuale

$$p(X = x) = \lambda^x \frac{exp(-\lambda)}{x!} \text{ per } x = 0, 1, 2. \ldots$$

$E(X) = \lambda$, $V(X) = \lambda$

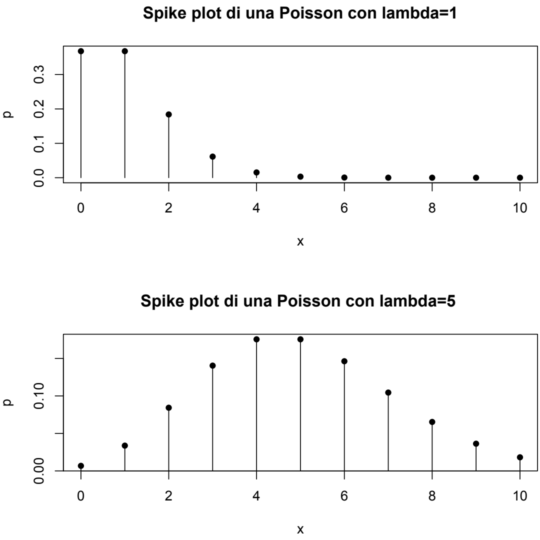
Ora si usa dpois(x,λ):

```
POI = dpois(0:10, lambda=1/2)
plot(0:10, POI, type = 'h', main= 'Spike plot di una Poisson',
xlab='k', ylab = 'p')
points(0:10,POI,pch = 16, cex = 1)
```


Spike plot di una Poisson
p
0.0
0.1
0.2
0.3
0.4
0.5
0.6
k
0
2
4
6
8
10

Cambiamo il parametro λ

```
par(mfrow = c(2, 1))

POI2 = dpois(0:10, lambda=1)

plot(0:10, POI2, type = 'h', main='Spike plot di una Poisson con
lambda=1', xlab='x', ylab = 'p')

points(0:10,POI2,pch = 16, cex = 1)

POI3 = dpois(0:10, lambda=5)

plot(0:10, POI3, type = 'h', main= 'Spike plot di una Poisson
con lambda=5 ', xlab='x', ylab = 'p')

points(0:10,POI3,pch = 16, cex = 1)
```

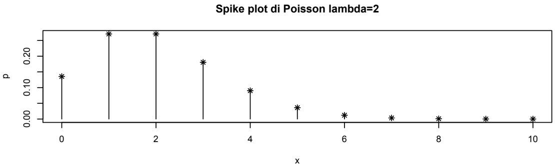


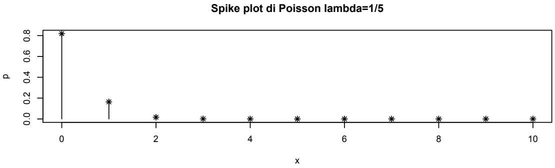
Spike plot di una Poisson con lambda=1
p
0.0
0.1
0.2
0.3
0
2
4
6
8
10
x
Spike plot di una Poisson con lambda=5
p
0.00
0.10
0
2
4
6
8
10
x

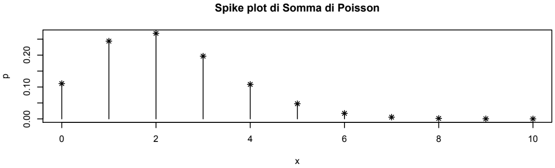
Somma di Poisson

La somma di variabili aleatorie *indipendenti* con distribuzioni di Poisson di parametri λ_1 e λ_2 é una variabile aleatoria con distribuzione di Poisson di parametro $\lambda = \lambda_1 + \lambda_2$ (proprietá riproduttiva)

```
par(mfrow = c(3, 1))

POI = dpois(0:10, lambda=2)
plot(0:10,POI,type = 'h', main= 'Spike plot di Poisson
lambda=2', xlab='x', ylab ='p')

points(0:10,POI,pch = 8,cex =1)

POI2 = dpois(0:10, lambda=1/5)
plot(0:10, POI2, type='h',main='Spike plot di Poisson
lambda=1/5', xlab='x', ylab = 'p')

points(0:10,POI2,pch=8,cex= 1)

POI3=dpois(0:10, lambda = 11/5)
plot(0:10, POI3,type = 'h', main='Spike plot di Somma di
Poisson', xlab='x', ylab ='p')

points(0:10,POI3,pch = 8,cex = 1)
```


Spike plot di Poisson lambda=2
p
0.00
0.10
0.20
x
0
2
4
6
8
10


Spike plot di Poisson lambda=1/5
p
0.0
0.2
0.4
0.6
0.8
x
0
2
4
6
8
10


Spike plot di Somma di Poisson
p
0.00
0.10
0.20
x
0
2
4
6
8
10

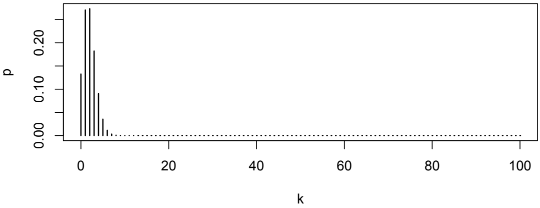
Binomiale → Poisson

Se n é molto grande (≥ 50) e p molto piccolo, tale che $np < 10$ e $p(1-p) \backsim p$, allora la binomiale puó essere approssimata con una distribuzione di Poisson con parametro $\lambda = np$.

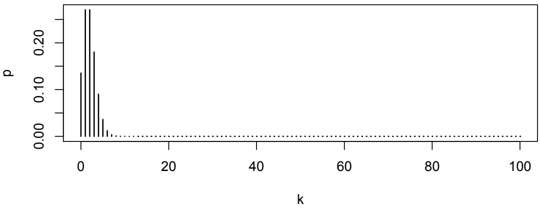
ESEMPIO: $n = 100, p = 1/50, np = 2$

```
par(mfrow = c(2, 1))

BINO = dbinom(0:100, size = 100, prob = 1/50)
plot(0:100, BINO, type= 'h', main= 'Spike plot di una
distribuzione Binomiale Bin(100,1/50)', xlab= 'x', ylab = 'p')

POI = dpois(0:100, lambda=2)
plot(0:100,POI,type = 'h', main= 'Spike plot di Poisson Poi(2)',
xlab='x', ylab ='p')
```

Spike plot di una distribuzione Binomiale Bin(100,1/50)



Spike plot di Poisson Poi(2)



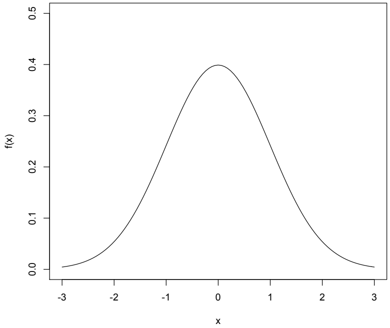
Le Variabili Continue

Normale: $X \backsim N(\mu, \sigma)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
$$-\infty < x < \infty, \mu \in R, \sigma > 0$$

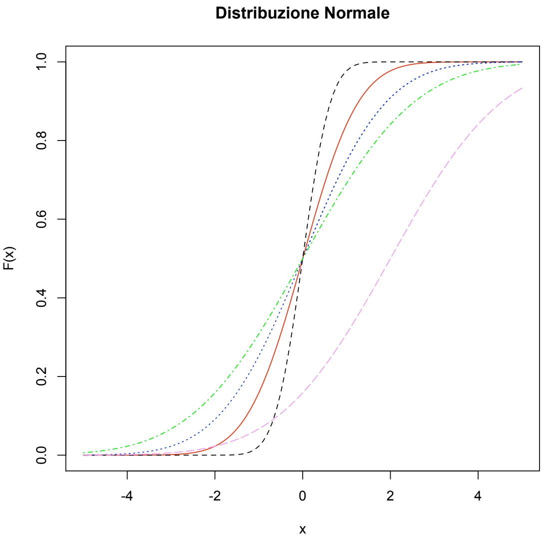
Disegnamo i valori di una curva Normale Standardizzata $Z \backsim N(0,1)$ tra -3 e 3 con il comando ”curve”

```
curve(dnorm(x),-3,3, axes = TRUE, ylab = 'f(x)', xlab = 'x',
ylim =c(0,.5))
```



Funzione di ripartizione

```
curve(pnorm(x), col = 'red', ylab = 'F(x)', from = -5, to = 5,
main ='Distribuzione Normale')

curve(pnorm(x, mean=0, sd=0.5), from = -5, to = 5, add = T, lty
= 2)

curve(pnorm(x, mean=0, sd=1.5), col = 'blue', from = -5, to = 5,
add = T, lty = 3)

curve(pnorm(x, mean=0, sd=2), col = 'green', from = -5, to = 5,
add = T, lty = 4)

curve(pnorm(x, mean=2, sd=2), col = 'violet', from = -5, to = 5,
add = T, lty = 5)
```



Distribuzione Normale
F(x)
1.0
0.8
0.6
0.4
0.2
0.0
-4
-2
0
2
4
x

Evidenziare le code

Evidenziamo la coda di SINISTRA

```
vals = seq(-3,-1, length = 100)
x =c(-3, vals, -1, -3)
y = c(0, dnorm(vals),0, 0)
polygon(x,y, density= 20, angle = 45)
```

Evidenziamo la coda di DESTRA

```
vals = seq(1,3,length = 100)
x = c(1, vals, 3, 1)
y = c(min(y), dnorm(vals),min(y), min(y))
polygon(x,y,density = 20,angle = 45)
```

Un piccolo esercizio sui valori medi: campioniamo dati da una Normale $N(5, 2)$ e vediamo come varia la media campionaria all'aumentare della numerositá campionaria. Usiamo il comando `rnorm(n, mean = , sd = )`

```
campione1=rnorm(10,5,2)
m1=mean(campione1)
m1
```

```
campione2=rnorm(100,5,2)
m2=mean(campione2)
m2
```

```
campione3=rnorm(1000,5,2)
m3=mean(campione3)
m3
```

e ora controlliamo cosa é successo alla varianza campionaria

```
v1=var(campione1)
v2=var(campione2)
v3=var(campione3)
```

Esponenziale: $X \backsim E(\lambda)$

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

con $x > 0$ e $\lambda > 0$

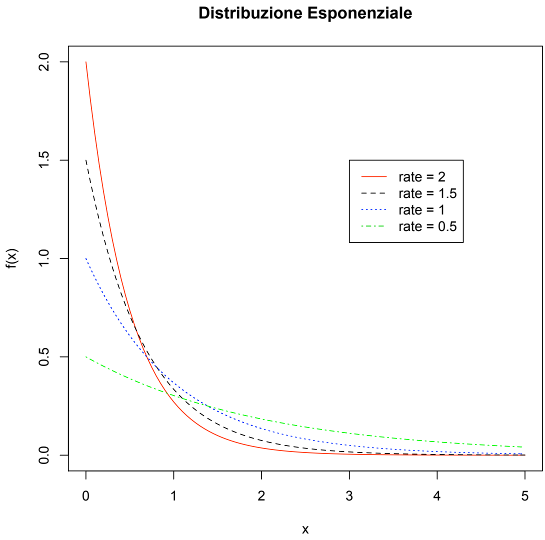
$E(X) = 1/\lambda$
$V(X) = 1/\lambda^2$

I relativi comandi sono:

```
dexp(x, rate = 1, log = FALSE)
pexp(q, rate = 1, lower.tail = TRUE, log.p = FALSE)
qexp(p, rate = 1, lower.tail = TRUE, log.p = FALSE)
rexp(n, rate = 1)
```

Realizzazioni di Esponenziali

```
curve(dexp(x, rate = 2), col = 'red', ylab = 'f(x)', from = 0,
to = 5, main ='Distribuzione Esponenziale')

curve(dexp(x, rate = 1.5), from = 0, to = 5, add = T, lty = 2)

curve(dexp(x, rate = 1), col = 'blue', from = 0, to = 5, add =
T, lty = 3)

curve(dexp(x, rate = 0.5), col = 'green', from = 0, to = 5, add
= T, lty = 4)

legend(1.2,6, c('rate = 2', 'rate = 1.5', 'rate = 1', 'rate =
0.5'), lty = c(1, 2, 3, 4), col=c(2,1,4,3))
```


Distribuzione Esponenziale
rate = 2
rate = 1.5
rate = 1
rate = 0.5
f(x)
x
0.0
0.5
1.0
1.5
2.0
0
1
2
3
4
5

Gamma: $X \backsim G(\alpha, \beta)$

$$f(x) = \frac{\alpha^{\beta}}{\Gamma(\beta)} x^{\beta-1} \exp\left(-\alpha x\right)$$

$$x > 0, \alpha > 0, \beta > 0$$

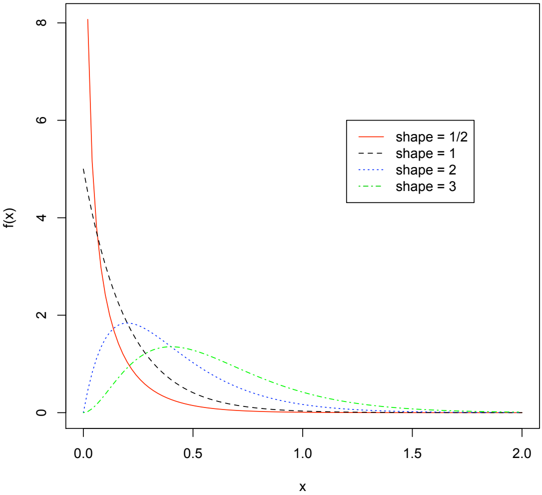
I relativi comandi sono

```
dgamma(x, shape, rate = 1, scale = 1/rate, log = FALSE)
pgamma(q, shape, rate = 1, scale = 1/rate, lower.tail = TRUE,
log.p = FALSE)
qgamma(p, shape, rate = 1, scale = 1/rate, lower.tail = TRUE,
log.p = FALSE)
rgamma(n, shape, rate = 1, scale = 1/rate)
```

Realizzazioni di Gamma: variamo il parametro α

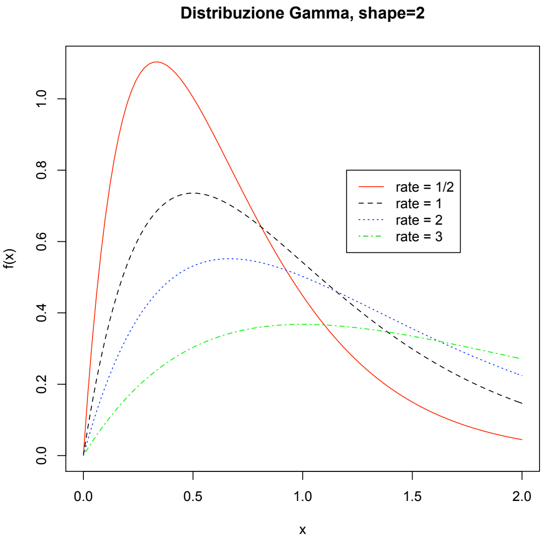
```
curve(dgamma(x, shape = 0.5, rate = 5), col = 'red', ylab =
'f(x) ', from = 0, to = 2, main ='Distribuzione Gamma, rate=5')

curve(dgamma(x, shape = 1, rate = 5), from = 0, to = 2, add = T,
lty = 2)

curve(dgamma(x, shape = 2, rate = 5), col = 'blue', from = 0, to
= 2, add = T, lty = 3)

curve(dgamma(x, shape = 3, rate = 5), col = 'green', from = 0,
to = 2, add = T, lty = 4)

legend(1.2,6, c('shape = 1/2', 'shape = 1', 'shape = 2', 'shape
= 3'), col=c(2,1,4,3), lty = c(1, 2, 3, 4))
```

Distribuzione Gamma, rate=5



Realizzazioni di Gamma: variamo il parametro β

```
curve(dgamma(x, shape = 2, rate = 3), col = 'red', ylab =
'f(x)', from = 0, to = 2, main ='Distribuzione Gamma, shape=2')

curve(dgamma(x, shape = 2, rate = 2), from = 0, to = 2, add = T,
lty = 2)

curve(dgamma(x, shape = 2, rate = 1.5), col = 'blue', from = 0,
to = 2, add = T, lty = 3)

curve(dgamma(x, shape = 2, rate = 1), col = 'green', from = 0,
to = 2, add = T, lty = 4)

legend(1.2,0.8, c('rate = 1/2', 'rate = 1', 'rate = 2', 'rate =
3'), lty = c(1, 2, 3, 4), col=c(2,1,4,3))
```


Distribuzione Gamma, shape=2
rate = 1/2
rate = 1
rate = 2
rate = 3
f(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0

Realizzazioni di Gamma

```
s=c()

for (i in 1:10000) s[i]=mean(rgamma(n=1,1))
hist(s,prob=T,xlim=c(0,4),ylim=c(0,2))

for (i in 1:10000) s[i]=mean(rgamma(n=5,1))
hist(s,prob=T,xlim=c(0,4),ylim=c(0,2))

for (i in 1:10000) s[i]=mean(rgamma(n=10,1))
hist(s,prob=T,xlim=c(0,4),ylim=c(0,2))

for (i in 1:10000) s[i]=mean(rgamma(n=20,1))
hist(s,prob=T,xlim=c(0,4),ylim=c(0,2))
```

Chi Quadro

La distribuzione $\chi^2(k)$ descrive la variabile aleatoria

$$\chi^2(k) = \sum_i^k X_i^2 = X_1^2 + \dots X_k^2$$

dove $X_1, ..., X_k$ sono variabili aleatorie indipendenti con distribuzione normale standard $N(0,1)$. Il parametro k detto numero di gradi di libertá (df).

I relativi comandi sono:

```
dchisq(x, df, ncp=0, log = FALSE)
pchisq(q, df, ncp=0, lower.tail = TRUE, log.p = FALSE)
qchisq(p, df, ncp=0, lower.tail = TRUE, log.p = FALSE)
rchisq(n, df, ncp=0)
```

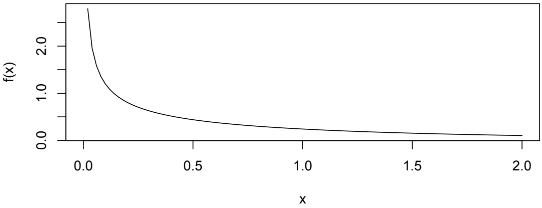
Gamma $\leftrightarrow$ Chi Quadro

Per $k > 0$, la distribuzione gamma con parametro di forma $\alpha = k/2$ e parametro di scala $\beta = 1/2$ (o 2) da luogo alla distribuzione chi-square con k gradi di libertá.

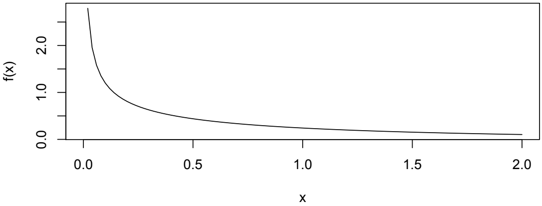
ESEMPIO con k=1

```
par(mfrow = c(2, 1))

curve(dgamma(x, shape = 1/2, rate = 1/2), ylab = 'f(x)', from =
0, to = 2, main ='Distribuzione Gamma, k=1')

curve(dchisq(x, 1), ylab = 'f(x)', from = 0, to = 2, main
='Distribuzione Chi Quadro, k=1')
```

Distribuzione Gamma, k=1



Distribuzione Chi Quadro, k=1



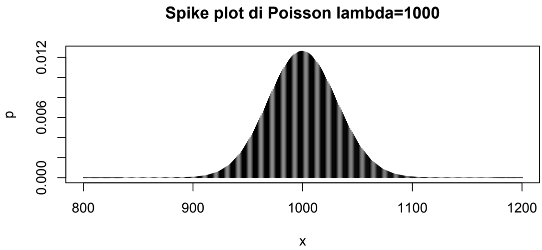
Teorema del limite centrale: una applicazione

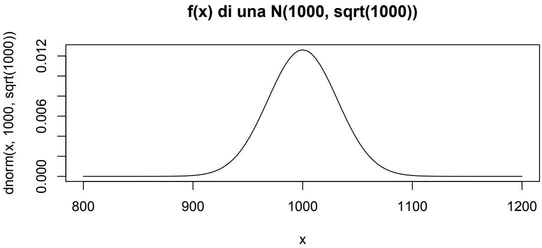
Abbiamo gi visto che la somma di Poisson ancora una Poisson con parametro λ uguale alla somma dei parametri dele variabili coinvolte.
Possiamo inoltre osservare che se il numero delle variabili sommate é elevato tale che la $\lambda > 1000$ la distribuzione che si ottiene puó essere approssimata alla Normale $N(\lambda, \lambda)$

```
par(mfrow = c(2, 1))
```

```
POI = dpois(800:1200, lambda=1000)
plot(800:1200,POI,type = 'h', main= 'Spike plot di Poisson
lambda=1000', xlab='x', ylab ='p')
```

```
curve(dnorm(x, 1000, sqrt(1000)), main= 'Densita N(1000,
sqrt(1000))')
```


Spike plot di Poisson lambda=1000
p
0.000
0.006
0.012
800
900
1000
1100
1200
x


f(x) di una N(1000, sqrt(1000))
dnorm(x, 1000, sqrt(1000))
0.000
0.006
0.012
800
900
1000
1100
1200
x

I quantili

Calcolo e individuazione sul grafico del quantile 0.95 di una distribuzione Normale(0,1).

```
curve(dnorm(x), -3,3, ylab = 'f(x)' , col ='red', main =
Distribuzione N(0,1))
qnorm(0.95)
points(qnorm(0.95), 0, pch='|')
```

Il quantile 0.95 di una $N(0,1)$ é 1.644854. Notare che il quadrato di 1.96 é 3.84, infatti:

$$P(Z^2 \leq 3,84) = 0,95 \iff P(|Z| \leq 1,96) = 0,95$$

La probabilitá che il chi-quadrato con 1 g.l. sia minore di 3.84 é uguale alla probabilitá che la normale standard sia compresa tra -1,96 e 1,96.