

UNIVERSITÀ DEGLI STUDI DI SALERNO



Dipartimento di Informatica

Corso di Laurea Magistrale in Informatica

Inferenza Statistica

Docente:

Prof. A.G. Nobile

Studente:

Gino Farisano

Corso di Metodi e Tecniche per l'Analisi dei Dati

Anno Accademico 2014/2015

Indice

1. Introduzione	3
1.1 Distribuzione normale.....	5
1.2 Distribuzione Chi-Quadrato.....	7
1.3 Distribuzione di Student.....	9
2. Intervalli di confidenza	11
2.1 Popolazione normale.....	13
2.2 Intervallo di confidenza per μ con σ^2 nota	15
2.3 Intervallo di confidenza per μ con σ^2 non nota	20
2.4 Intervallo di confidenza per σ^2 con valore medio μ noto.....	27
2.5 Intervallo di confidenza per σ^2 con valore medio μ non noto.....	32

1. Introduzione

Una VARIABILE ALEATORIA QUANTITATIVA X può essere definita come un fenomeno le cui modalità sono espresse da numeri o, più precisamente, come una funzione definita sullo spazio degli eventi Ω che associa ad ogni elemento ω di Ω un numero reale $X(\omega)=x$. Le variabili aleatorie quantitative sono DISCRETE se producono risposte numeriche che derivano da un processo di conteggio. Ad es. "il numero dei componenti di una famiglia", "il numero delle stanze di un'abitazione", "il numero dei biglietti dell'autobus utilizzati in un mese", ecc. Le variabili aleatorie quantitative sono CONTINUE se generano risposte che derivano da un processo di misurazione. Ad es. "l'altezza", "il reddito", "il fatturato", ecc.

Ad una variabile aleatoria discreta è associata una DISTRIBUZIONE DI PROBABILITÀ (rappresentata dall'elenco delle modalità che la variabile assume, a ciascuna delle quali è associata la relativa probabilità). Chiamiamo la funzione $P(X = x_i) = P(x_i)$ FUNZIONE DI PROBABILITÀ. Poiché le modalità elencate sono mutuamente esclusive e collettivamente esaustive, le probabilità sommano ad 1.

Un'altra funzione che descrive il comportamento di una variabile casuale è la FUNZIONE DI RIPARTIZIONE. La FUNZIONE DI RIPARTIZIONE fornisce la probabilità che la variabile aleatoria X assuma valori \leq ad un particolare valore x_i .

Nel caso si parli di fenomeni continui si può calcolare la probabilità che la v.a. assuma valori compresi in un intervallo, mentre la probabilità che assuma un particolare valore è zero. Questa caratteristica distingue i fenomeni che risultano da un processo di misurazione (come ad es. il tempo) da quelli ottenuti attraverso un processo di conteggio. Il calcolo di probabilità, del valore atteso e dello scarto quadratico medio per un fenomeno continuo coinvolge espressioni matematiche che richiedono pertanto la conoscenza del calcolo integrale. Una variabile casuale X continua è caratterizzata dall'esistenza di una funzione $f(x)$ continua, detta FUNZIONE DI DENSITÀ, che gioca un ruolo analogo alla funzione di probabilità nel caso discreto, con la differenza fondamentale che $f(x)$ non è interpretabile come una probabilità.

Analogamente al caso discreto, possiamo definire la FUNZIONE DI RIPARTIZIONE (a differenza di $f(x)$ è una probabilità), il VALORE ATTESO e la VARIANZA di una variabile casuale continua. Dobbiamo però sostituire l'integrale alla sommatoria...

Esempi di variabili casuali discrete sono:

- Variabile casuale uniforme discreta
- Variabile casuale di Bernoulli
- Variabile casuale binomiale
- Variabile casuale di Poisson
- Variabile casuale geometrica
- Variabile casuale ipergeometrica

Esempi di variabili casuali continue sono:

- Variabile casuale normale o gaussiana
- Variabile casuale Gamma
- Variabile casuale di Student
- Variabile casuale esponenziale
- Variabile casuale di Chi Quadrato
- Variabile casuale beta
- Variabile casuale di Cauchy

In questo documento ci soffermeremo sulla distribuzione normale, quella di Student e di Chi Quadrato, per poi arrivare agli intervalli di confidenza.

1.1 Distribuzione normale

La funzione di distribuzione normale, detta anche di Gauss o gaussiana, ha un ruolo importante nel calcolo delle probabilità e nella statistica in quanto costituisce una distribuzione limite alla quale tendono diverse funzioni di distribuzioni sotto determinate ipotesi. La distribuzione normale è caratterizzata dalla seguente funzione di densità:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\{(x-\mu)^2/2\sigma^2\}} \text{ con } -\infty < x < \infty$$

Dove μ è il valore atteso e σ^2 la varianza.

In R la densità normale si calcola mediante la funzione:

```
> dnorm (x, mean = mu, sd = sigma, log = FALSE)
```

dove *mean* ed *sd* indicano rispettivamente il valore medio e la deviazione standard della densità normale ed *x* è il valore o i valori assunti dalla variabile aleatoria normale. Questi parametri verranno utilizzati anche nelle seguenti funzioni dove *x* può essere sostituito da *q* ma con lo stesso significato.

La funzione di ripartizione di una variabile normale X si calcola mediante la funzione:

```
> pnorm (q, mean = mu, sd = sigma, lower.tail = TRUE, log.p = FALSE)
```

Dove se *lower.tail* = TRUE viene calcolato $P(X \leq q)$ mentre se *lower.tail* = FALSE viene calcolato $P(X > q)$.

I quantili della distribuzione possono essere calcolati mediante la funzione:

```
> qnorm (z, mean = mu, sd = sigma, lower.tail = TRUE, log.z = FALSE)
```

Dove *z* indica il valore o i valori assunti dalle probabilità relative al percentile *z**100-esimo. Infine sempre tramite R è possibile simulare la variabile aleatoria normale generando una sequenza di numeri pseudocasuali mediante la funzione:

```
> rnorm (N, mean = mu, sd = sigma)
```

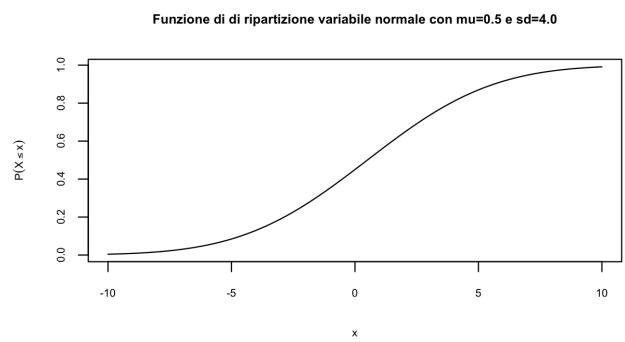
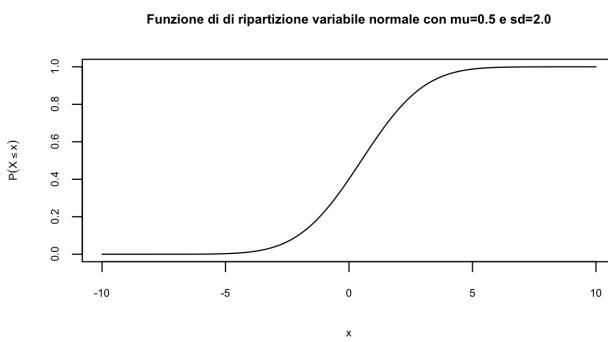
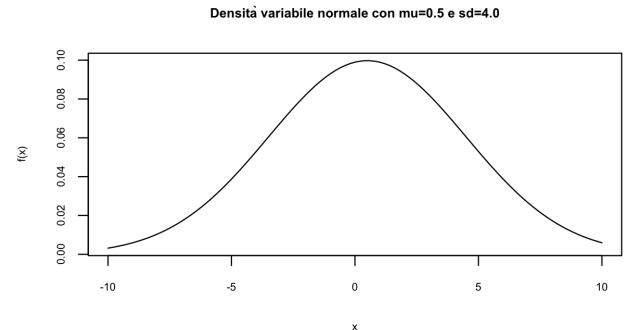
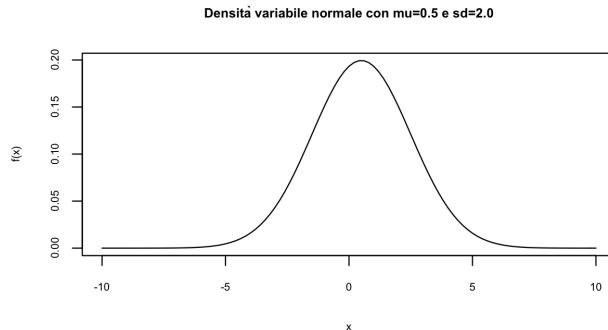
Dove *N* indica la lunghezza della sequenza di numeri da generare.

Riportiamo di seguito un esempio grafico della densità e della funzione di ripartizione di una variabile normale:

```

> par(mfcol=c(2,2), ps = 8)
> curve(dnorm(x, mean=0.5, sd=2.0), from=-10, to=10, xlab="x", ylab="f(x)", main="Densità variabile normale con mu=0.5 e sd=2.0")
> curve(pnorm(x,mean=0.5,sd=2.0),from=-10,to=10,xlab="x", ylab=expression(P(X<=x)),main="Funzione di ripartizione variabile normale con mu=0.5 e sd=2.0")
> curve(dnorm(x, mean=0.5, sd=4.0), from=-10, to=10, xlab="x", ylab="f(x)", main="Densità variabile normale con mu=0.5 e sd=4.0")
> curve(pnorm(x,mean=0.5,sd=4.0),from=-10,to=10,xlab="x", ylab=expression(P(X<=x)),main="Funzione di ripartizione variabile normale con mu=0.5 e sd=4.0")

```



1.2 Distribuzione Chi-Quadrato

In teoria delle probabilità una distribuzione chi quadrato o chi quadro è una distribuzione di probabilità che descrive la somma dei quadrati di alcune variabili aleatorie indipendenti aventi distribuzione normale standard. Una variabile aleatoria X di densità di probabilità

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1} e^{-x/2}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

con n intero positivo ed $\Gamma(v)$ definita nel seguente modo:

$$\Gamma(v) = \int_0^{+\infty} x^{v-1} e^{-x} dx, \quad v > 0$$

si dice distribuzione chi-quadrato con n gradi di libertà. È assai utile in Statistica per la valutazione delle varianze di campioni estratti da una popolazione. Per una sua notevole proprietà asintotica, al crescere di n la distribuzione Chi-quadrato tende ad assumere una forma “a campana” simmetrica rispetto al suo valor medio, e si può dimostrare che per $n \gg 1$ è ben approssimata da una legge normale N (n, 2n) con media n e varianza 2n.

La densità di una distribuzione chi-quadrato può essere calcolata mediante la funzione:

```
dchisq (x, df , log = FALSE )
```

Dove x indica il valore assunto o i valori assunti dalla variabile aleatoria chi-quadrato e df indica il numero di gradi di libertà.

La funzione di ripartizione viene calcolata mediante la funzione:

```
pchisq (q, df , lower.tail = TRUE , log.p = FALSE )
```

dove q indica il valore assunto o i valori assunti dalla variabile aleatoria, df indica il numero dei gradi di libertà ed il parametro lower.tail si comporta come nel caso del calcolo della funzione di distribuzione di una variabile normale.

Inoltre è possibile calcolare anche i quantili e simulare una variabile chi-quadrato mediante le funzioni:

```

qchisq (p, df , lower .tail = TRUE , log .p = FALSE )

rchisq (N, df)

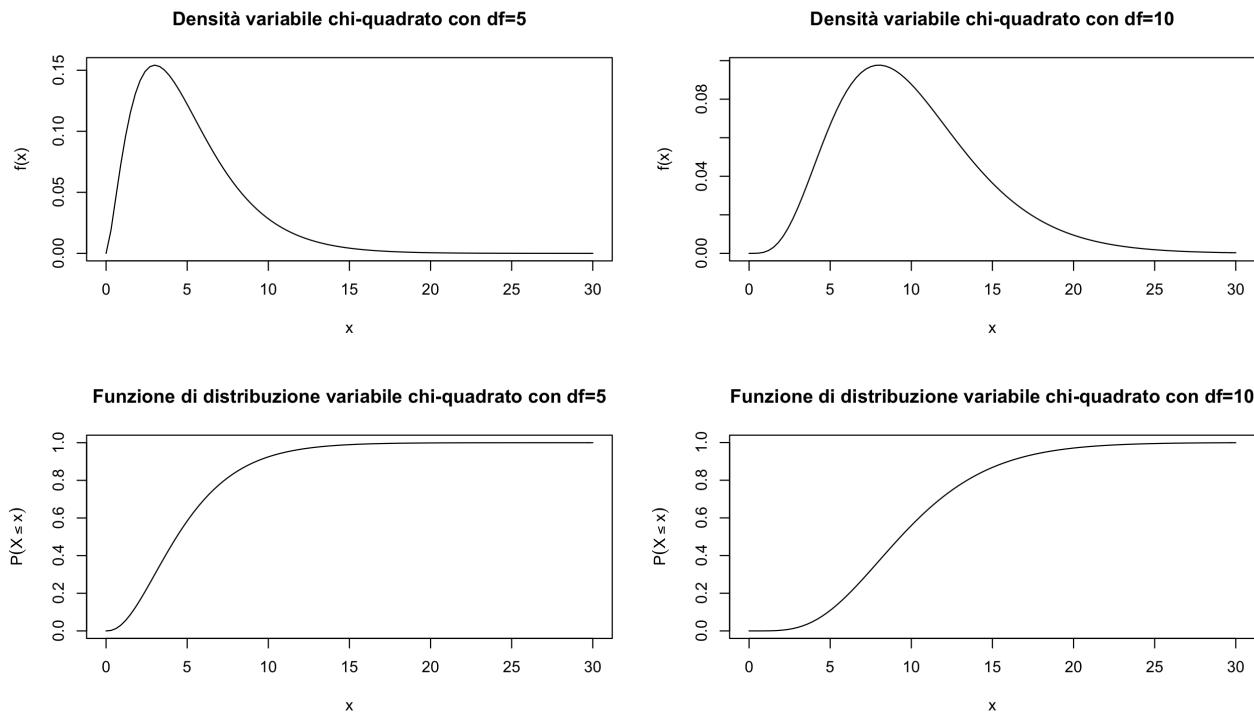
```

dove df indica il numero di gradi di libertà e gli altri parametri sono stati già visti precedentemente.
Mostriamo di seguito un esempio di rappresentazione grafica della densità di una variabile chi-quadrato e della funzione di distribuzione.

```

> par(mfcol=c(2,2))
> curve(dchisq(x,df=5),xlab="x",ylab="f(x)",main="Densità variabile chi-quadrato con df=5",from=0,to=30)
> curve(pchisq(x,df=5),xlab="x",ylab=expression(P(X<=x)),main="Funzione di distribuzione variabile chi-quadrato con df=5", from=0,to=30)
> curve(dchisq(x,df=10),xlab="x",ylab="f(x)",main="Densità variabile chi-quadrato con df=10",from=0,to=30)
> curve(pchisq(x,df=10),xlab="x",ylab=expression(P(X<=x)),main="Funzione di distribuzione variabile chi-quadrato con df=10", from=0,to=30)

```



1.3 Distribuzione di Student

In teoria delle probabilità la distribuzione di Student, o t di Student, è una distribuzione di probabilità continua che governa il rapporto tra due variabili aleatorie, la prima con distribuzione normale e la seconda il cui quadrato ha distribuzione chi quadrato. Viene usata in statistica per stimare il valore medio di una popolazione quando sia disponibile un campione di piccole dimensione (meno di 30 elementi). Se il campione è più numeroso le distribuzioni gaussiana e quella di Student differiscono di poco, pertanto è indifferente usare una o l'altra.

W.S. Gossett il cui pseudonimo era Student mostrò che se la popolazione è distribuita normalmente, la distribuzione delle medie dei campioni di dimensioni p sono distribuite secondo la distribuzione di Student di ordine $p-1$ che è il numero dei gradi di libertà.

Una variabile aleatoria X di densità di probabilità:

$$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{2}\right)^{-(n+1)/2}$$

con n intero positivo e $\Gamma(v)$ definita come per chi-quadrato, si dice distribuzione di Student, o avere “distribuzione t di Student”, con n gradi di libertà. La densità può essere calcolata direttamente mediante la funzione:

```
dt(x, df, log = FALSE )
```

dove i parametri x e df sono stati già definiti in precedenza.

La funzione di distribuzione può essere calcolata mediante la funzione:

```
pt(q, df, lower.tail = TRUE, log.p = FALSE )
```

dove i parametri q e df sono gli stessi delle altre distribuzioni, ed infine è possibile calcolare i quantili e simulare una variabile di Student mediante le funzioni:

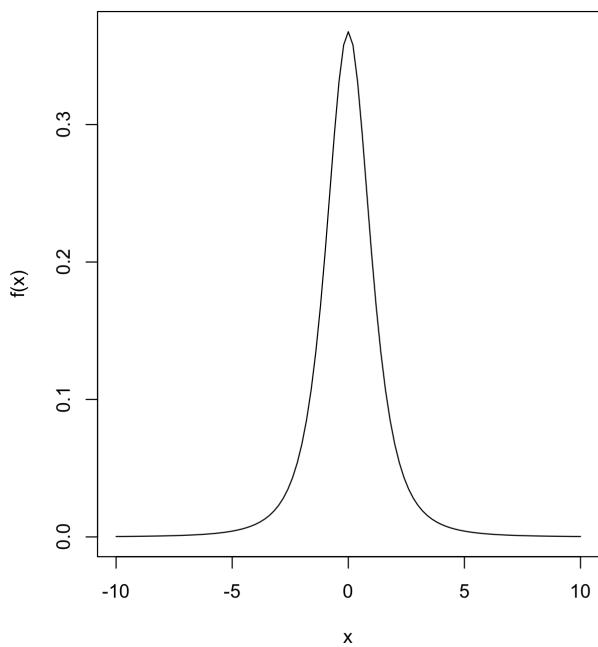
```
qt(p, df, lower.tail = TRUE, log.p = FALSE )
```

```
rt(N, df)
```

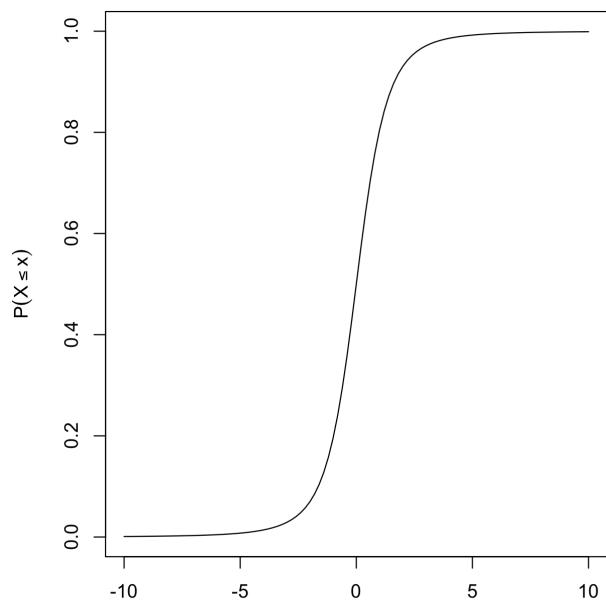
Di seguito mostriamo con un esempio la rappresentazione grafica della densità e della funzione di distribuzione di una variabile di Student:

```
> par(mfcol=c(1,2))
> curve(dt(x,df=3),xlab="x",ylab="f(x)",main="Densità variabile di Student con df=3",from=-10,to=10)
> curve(pt(x,df=3),xlab="",ylab=expression(P(X <=x)),main="Funzione di distribuzione variabile di Student con df=3", from=-10,to=10)
```

Densità variabile di Student con df=3



Funzione di distribuzione variabile di Student con df=3



2. Intervalli di confidenza

Alla stima puntuale di un parametro non noto di una popolazione (costituita da un singolo valore reale) spesso si preferisce sostituire un intervallo di valori, detto *intervallo di confidenza*, ossia si cerca di determinare in base ai dati del campione, due limiti (uno inferiore ed uno superiore) entro i quali sia compreso il parametro non noto con un certo *coefficiente di confidenza* (detto anche *grado di fiducia*). Sia X_1, X_2, \dots, X_n un campione casuale di ampiezza n estratto da una popolazione con funzione di probabilità nel caso discreto, o densità di probabilità nel caso continuo, $f(x; \theta)$, dove θ denota il parametro non noto della popolazione. Indichiamo con $\underline{C}_n = g_1(X_1, X_2, \dots, X_n)$ e con $\overline{C}_n = g_2(X_1, X_2, \dots, X_n)$ due statistiche, ovvero funzioni osservabili del campione, che soddisfino la condizione $\underline{C}_n < \overline{C}_n$, cioè che godono della proprietà che per ogni campione fissato x abbiamo $g_1(x) < g_2(x)$. Fissato un coefficiente di confidenza $1 - \alpha$ ($0 < \alpha < 1$), se è possibile scegliere le statistiche \underline{C}_n e \overline{C}_n in modo tale

$$P(\underline{C}_n < \theta < \overline{C}_n) = 1 - \alpha$$

allora si dice che $(\underline{C}_n, \overline{C}_n)$ è un intervallo di confidenza $1 - \alpha$ per θ e le statistiche $\underline{C}_n, \overline{C}_n$ sono dette limite inferiore e limite superiore dell'intervallo. In questo caso l'intervallo $(g_1(x), g_2(x))$, dove x è il campione utilizzato, è detto stima dell'intervallo di confidenza di grado $1 - \alpha$ per θ ed i punti finali $g_1(x), g_2(x)$, di tale intervallo sono detti rispettivamente stima del limite inferiore e stima del limite superiore dell'intervallo di confidenza.

In generale esistono numerosi intervalli di confidenza dello stesso grado $1 - \alpha$ per un parametro non noto θ della popolazione. La scelta dell'intervallo di confidenza deve essere effettuata in base ad alcune proprietà statistiche. Ad esempio, fissato un coefficiente di confidenza $1 - \alpha$ alcune proprietà desiderabili sono che la lunghezza dell'intervallo di confidenza $L(X_1, X_2, \dots, X_n) = \overline{C}_n - \underline{C}_n$ sia la più piccola possibile oppure che la lunghezza media di tale intervallo sia la più piccola possibile.

Un metodo per costruire intervalli di confidenza è il metodo pivotale. Tale metodo consiste nel determinare una variabile aleatoria di pivot $\gamma(X_1, X_2, \dots, X_n; \theta)$ che dipende dal campione casuale X_1, X_2, \dots, X_n e dal parametro non noto θ e la cui funzione di distribuzione non contiene il parametro da stimare. Tale variabile non è una statistica in quanto non è osservabile. Per ogni fissato coefficiente α ($0 < \alpha < 1$) siano α_1 e α_2 due valori dipendenti solo dal coefficiente fissato tali che per ogni θ si abbia

$$P(\alpha_1 < \gamma(X_1, X_2, \dots, X_n; \theta) < \alpha_2) = 1 - \alpha$$

Se per ogni possibile campione x e per ogni θ si riesce a dimostrare che

$$\alpha_1 < \gamma(x; \theta) < \alpha_2 \Leftrightarrow g_1(x) < \gamma(x; \theta) < g_2(x)$$

Se poniamo $\underline{C}_n = g_1(x)$ e $\overline{C}_n = g_2(x)$ allora $(\underline{C}_n, \overline{C}_n)$ è un intervallo di confidenza $1 - \alpha$ per θ .

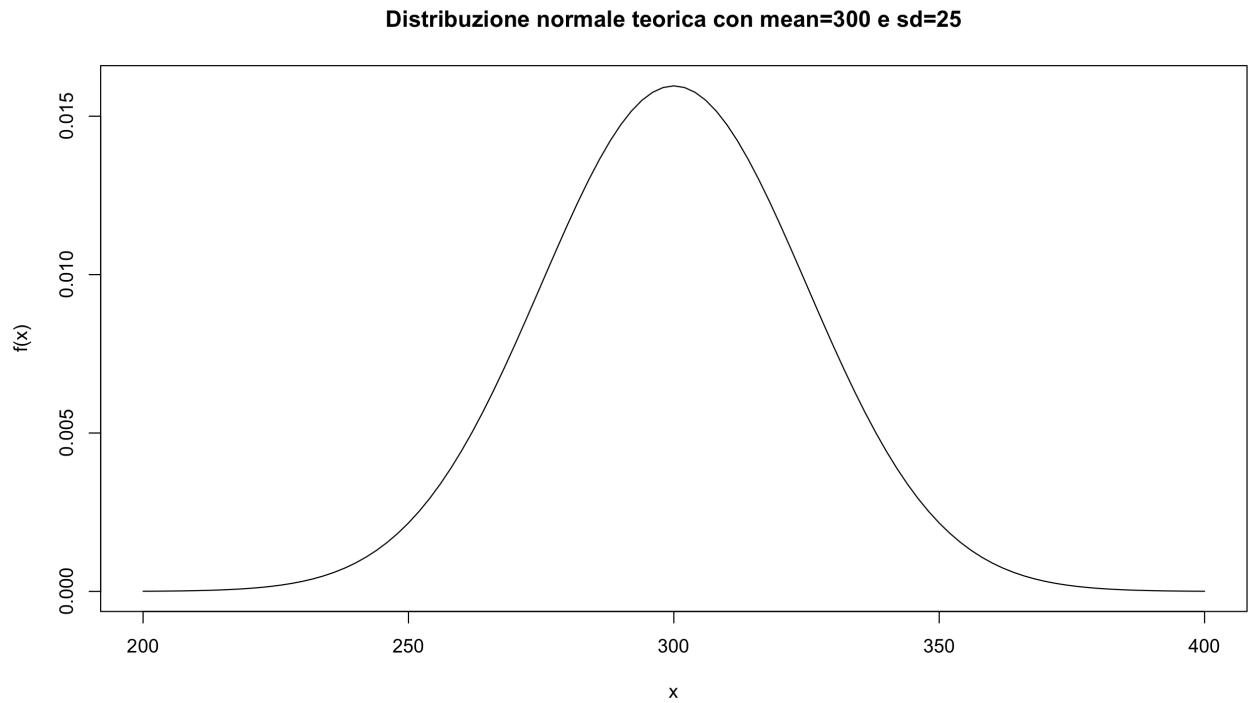
2.1 Popolazione normale

Sia X_1, X_2, \dots, X_n un campione casuale di ampiezza n estratto da una popolazione normale con valore medio μ e varianza σ^2 . Considerando questi due parametri si possono analizzare i seguenti problemi:

- Determinare un intervallo di confidenza di grado $1-\alpha$ per il valore medio μ nel caso in cui la varianza σ^2 della popolazione normale risulti essere nota.
- Determinare un intervallo di confidenza di grado $1-\alpha$ per il valore medio μ nel caso in cui la varianza σ^2 della popolazione normale non è nota.
- Determinare un intervallo di confidenza di grado $1-\alpha$ per la varianza σ^2 nel caso in cui il valore medio μ della popolazione normale è noto.
- Determinare un intervallo di confidenza di grado $1-\alpha$ per la varianza σ^2 nel caso in cui il valore medio μ della popolazione normale non è noto.

Di seguito verranno create delle popolazioni normali in R mediante la funzione `rnorm` vista precedentemente. Di seguito attraverso un grafico mostriamo la differenza tra la densità teorica e quella simulata

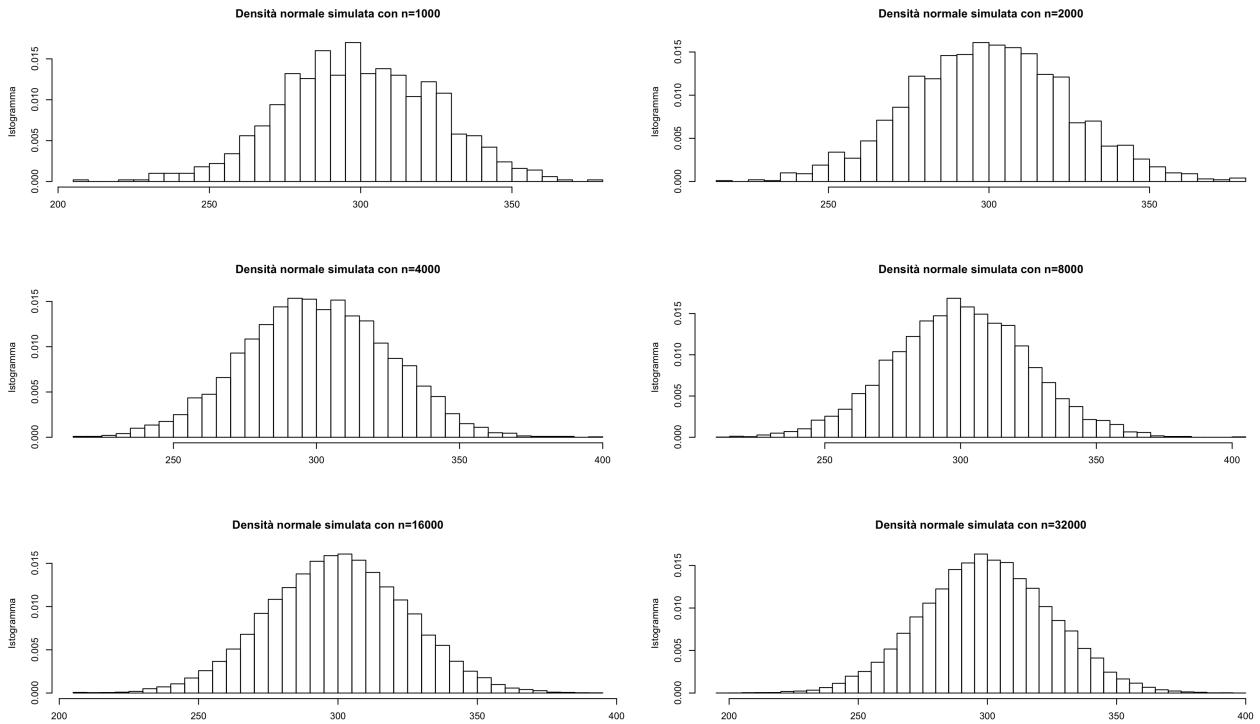
```
> curve(dnorm(x,mean=300,sd=25),from=200, to=400, xlab="x", ylab="f(x)", main="Distribuzione normale teorica con mean=300 e sd=25")
```



```

> par(mfrow=c(3,2))
> popolazione1 <- rnorm(1000,mean=300,sd=25)
> hist(popolazione1,breaks=50,freq=F,xlab="",ylab="Istogramma", main="Densità normale simulata con n=1000")
> popolazione2 <- rnorm(2000,mean=300,sd=25)
> hist(popolazione2,breaks=50,freq=F,xlab="",ylab="Istogramma", main="Densità normale simulata con n=2000")
> popolazione3 <- rnorm(4000,mean=300,sd=25)
> hist(popolazione3,breaks=50,freq=F,xlab="",ylab="Istogramma", main="Densità normale simulata con n=4000")
> popolazione4 <- rnorm(8000,mean=300,sd=25)
> hist(popolazione4,breaks=50,freq=F,xlab="",ylab="Istogramma", main="Densità normale simulata con n=8000")
> popolazione5 <- rnorm(16000,mean=300,sd=25)
> hist(popolazione5,breaks=50,freq=F,xlab="",ylab="Istogramma", main="Densità normale simulata con n=16000")
> popolazione6 <- rnorm(32000,mean=300,sd=25)
> hist(popolazione6,breaks=50,freq=F,xlab="",ylab="Istogramma", main="Densità normale simulata con n=32000")

```



Dal grafico possiamo notare come all'aumentare della popolazione la densità simulata tende a quella teorica.

2.2 Intervallo di confidenza per μ con σ^2 nota

Per determinare un intervallo di confidenza di grado $1 - \alpha$ per il valore medio μ nel caso in cui la varianza σ^2 è nota viene utilizzato il metodo pivotale e considerata la variabile aleatoria di pivot

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

dove \bar{X}_n risulta essere la media campionaria ed n indica la numerosità del campione utilizzato. La variabile risulta dipendere dal campione casuale e dal parametro non noto μ e quindi può essere interpretata come variabile di pivot. Inoltre tale variabile risulta essere una normale standard, ovvero valore medio nullo e varianza unitaria. Scegliendo nel metodo pivotale $\alpha_1 = -z_{\alpha/2}$ e $\alpha_2 = z_{\alpha/2}$ in modo tale che

$$P(Z_n < -z_{\alpha/2}) = P(Z_n > z_{\alpha/2}) = \frac{\alpha}{2}$$

abbiamo

$$P\left(-z_{\frac{\alpha}{2}} < Z_n < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

ottenendo successivamente

$$P\left(\bar{X}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Ponendo $\underline{C}_n = \bar{X}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ e $\overline{C}_n = \bar{X}_n + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ allora l'intervallo $(\underline{C}_n, \overline{C}_n)$ è un intervallo di confidenza di grado $1 - \alpha$ per μ .

La funzione mostrata sotto è stata utilizzata per graficare gli intervalli di confidenza. Da notare come le popolazioni utilizzate siano quelle generate negli esempi mostrati nel paragrafo precedente e come al crescere del numero di individui l'intervallo di confidenza diventi sempre più piccolo.

```

plot.normalConfidence <- function(population, oneMinusAlpha, sigma){
  n <- length(population)
  alpha <- 1 - oneMinusAlpha
  alphaHalf <- alpha/2
  leftLimit <- mean(population) - qnorm(1-alphaHalf, mean=0, sd=1)*sigma/sqrt(length(population))
  print(c("Left limit: ", leftLimit))
  rightLimit <- mean(population) + qnorm(1-alphaHalf, mean=0, sd=1)*sigma/sqrt(length(population))
  print(c("Right limit: ", rightLimit))
  print(c("Difference: ", abs(rightLimit-leftLimit)))

  quantile <- qnorm(1-alphaHalf, mean=0, sd=1)
  print(c("Quantile 1-alpha: ", quantile))

  #grafic generation code

  curve(dnorm(x, mean=0, sd=1), from=-4, to=4, axes=FALSE, xlab="", ylab="Densità", main="Densità normale intervallo di confidenza 1-alpha per μ con varianza nota")
  axis(1, c(-4, -quantile, 0, quantile, 4), c("", abbreviate(-quantile, 5), 0, abbreviate(quantile, 4), ""))
  vals <- seq(-3, quantile, length = 100)
  x1 <- c(-4, vals, quantile, 4)
  y1 <- c(0, dnorm(vals), 0, 0)
  polygon(x1, y1, density = 20, angle=45)

  vals <- seq(quantile, 4, length = 100)
  x1 <- c(quantile, vals, 4, 1)
  y1 <- c(0, dnorm(vals), 0, 0)
  polygon(x1, y1, density = 20, angle=-45)
  abline(h=0)
  text(-3.2, 0.03, expression(alpha/2))
  text(-3.2, 0.05, "Regione di rifiuto")
  text(3.2, 0.05, "Regione di rifiuto")
  text(3.2, 0.03, expression(1-alpha))
  text(0, 0.03, expression(1-alpha))
  text(0, 0.1, "Regione di accettazione")
}

}

```

Caso 1 – $\alpha = 0.99$ ($z_{0.005} = 2.57$)

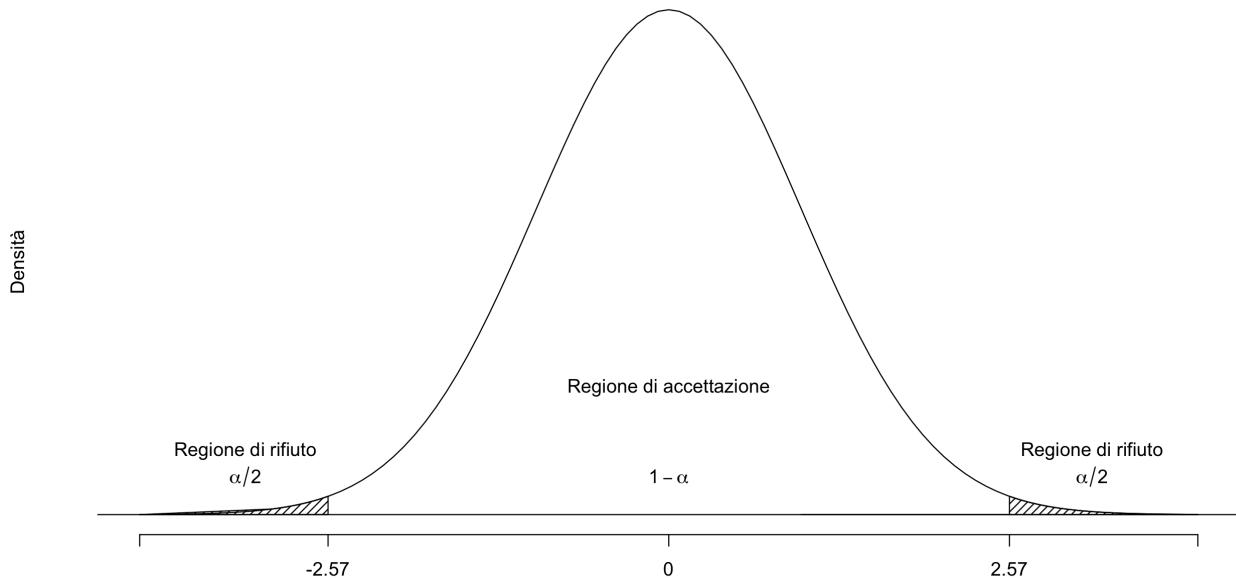
Determiniamo una stima dell'intervallo di confidenza di grado $1-\alpha=0.95$ per la media μ degli elementi. In questo caso $\alpha = 0.01$ e $\alpha/2 = 0.005$.

```

> plot.normalConfidence(popolazione1, 0.99, 25)
[1] "Left limit: "      "295.488184016521"
[1] "Right limit: "     "308.367330534266"
[1] "Difference: "      "12.8791465177445"
[1] "Quantile 1-alpha: " "2.5758293035489"
> plot.normalConfidence(popolazione2, 0.99, 25)
[1] "Left limit: "      "297.609540006766"
[1] "Right limit: "     "300.489404717351"
[1] "Difference: "      "2.87986471058559"
[1] "Quantile 1-alpha: " "2.5758293035489"
> plot.normalConfidence(popolazione3, 0.99, 25)
[1] "Left limit: "      "299.277691273886"
[1] "Right limit: "     "301.314063139641"
[1] "Difference: "      "2.03637186575497"
[1] "Quantile 1-alpha: " "2.5758293035489"
> plot.normalConfidence(popolazione4, 0.99, 25)
[1] "Left limit: "      "299.260075467469"
[1] "Right limit: "     "300.700007822762"
[1] "Difference: "      "1.43993235529285"
[1] "Quantile 1-alpha: " "2.5758293035489"
> plot.normalConfidence(popolazione5, 0.99, 25)
[1] "Left limit: "      "299.142792527203"
[1] "Right limit: "     "300.16097846008"
[1] "Difference: "      "1.01818593287749"
[1] "Quantile 1-alpha: " "2.5758293035489"
> plot.normalConfidence(popolazione6, 0.99, 25)
[1] "Left limit: "      "299.68509973474"
[1] "Right limit: "     "300.405065912386"
[1] "Difference: "      "0.71996617764637"
[1] "Quantile 1-alpha: " "2.5758293035489"

```

Densità normale intervallo di confidenza 1-alpha per μ con varianza nota



Caso 1 – $\alpha = 0.95$ ($z_{0.025} = 1.95$)

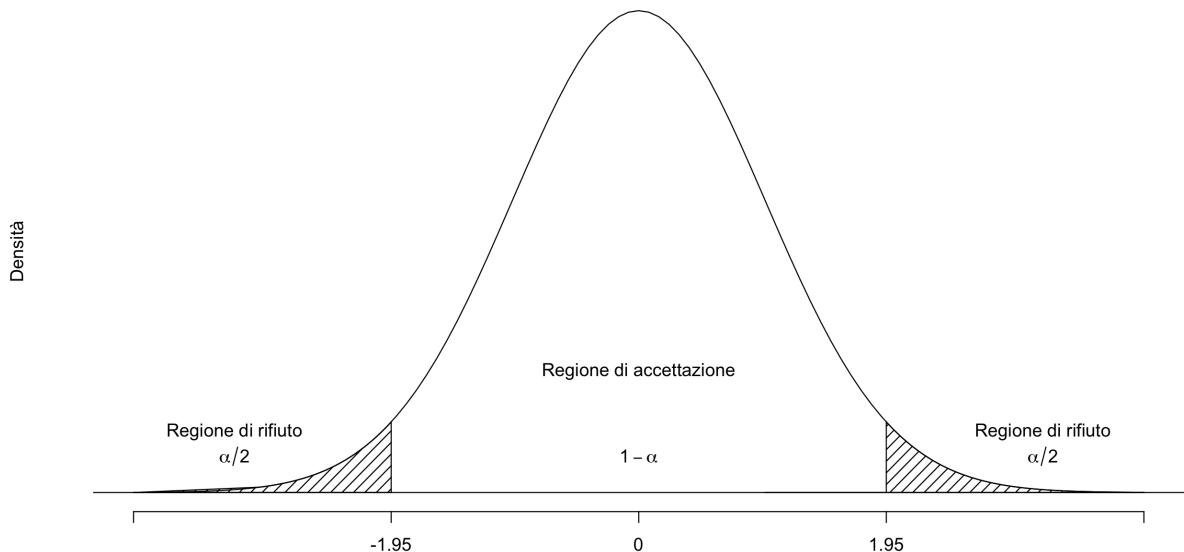
In questo caso determiniamo una stima dell’intervallo di confidenza di grado $1-\alpha=0.95$ per la media μ degli elementi. In questo caso $\alpha = 0.05$ e $\alpha/2 = 0.025$.

```

> plot.normalConfidence(popolazione1, 0.95, 25)
[1] "Left limit: "      "297.027847314043"
[1] "Right limit: "     "306.827667236743"
[1] "Difference: "      "9.79981992270029"
[1] "Quantile 1-alpha: " "1.95996398454005"
> plot.normalConfidence(popolazione2, 0.95, 25)
[1] "Left limit: "      "297.953819186338"
[1] "Right limit: "     "300.145125537779"
[1] "Difference: "      "2.19130635144143"
[1] "Quantile 1-alpha: " "1.95996398454005"
> plot.normalConfidence(popolazione3, 0.95, 25)
[1] "Left limit: "      "299.521133416383"
[1] "Right limit: "     "301.070620997144"
[1] "Difference: "      "1.54948758076137"
[1] "Quantile 1-alpha: " "1.95996398454005"
> plot.normalConfidence(popolazione4, 0.95, 25)
[1] "Left limit: "      "299.432215057255"
[1] "Right limit: "     "300.527868232976"
[1] "Difference: "      "1.09565317572071"
[1] "Quantile 1-alpha: " "1.95996398454005"
> plot.normalConfidence(popolazione5, 0.95, 25)
[1] "Left limit: "      "299.264513598451"
[1] "Right limit: "     "300.039257388832"
[1] "Difference: "      "0.774743790380739"
[1] "Quantile 1-alpha: " "1.95996398454005"
plot.normalConfidence(popolazione5, 0.95, 25)
> plot.normalConfidence(popolazione6, 0.95, 25)
[1] "Left limit: "      "299.771169529633"
[1] "Right limit: "     "300.318996117493"
[1] "Difference: "      "0.547826587860413"
[1] "Quantile 1-alpha: " "1.95996398454005"

```

Densità normale intervallo di confidenza 1-alpha per μ con varianza nota



Comparazione dei risultati

Di seguito viene mostrato il grafico della distribuzione normale della popolazione, confrontando la stima della media per una popolazione di 100 elementi.

Di seguito la funzione e il grafico ottenuto:

```

#0.01          #0.05
plot.confidenceComparison <-function(population,myMean=300,MySd=25,myFrom=200,myTo=400, alphaHalf1=0.005, alphaHalf2=0.025,sigma){

  curve(dnorm(x,mean= myMean,sd= MySd),from=myFrom,to=myTo,ylim=c(0,0.03), ylab="Densità",xlab="x",main="Confronto fra i due intervalli di confidenza")

  arrows(301,0.025,310,0.025,code=1,length = 0.10)
  text(315,0.025,"Media")
  abline(v=myMean)

  #first
  leftLimit <- mean(population)-qnorm(1-alphaHalf1,mean=0,sd=1)*sigma/sqrt(length(population))
  rightLimit <- mean(population)+qnorm(1-alphaHalf1,mean=0,sd=1)*sigma/sqrt(length(population))
  vals<-seq(leftLimit, rightLimit, length =50)
  x1<-c(leftLimit, vals, rightLimit)
  y1<-c(0,dnorm(vals, mean= myMean, sd= MySd),0)
  polygon (x1,y1,density=30,angle=45,col='blue')

  xLegend <-c(350,370,370,350)
  yLegend <-c(0.025,0.025,0.03,0.03,0.025)
  polygon (xLegend, yLegend,density=50,angle=45,col='blue')
  text(382, 0.0275,"Stima a=0.01")

  #second
  leftLimit <- mean(population)-qnorm(1-alphaHalf2,mean=0,sd=1)*sd(population)/sqrt(length(population))
  rightLimit <- mean(population)+qnorm(1-alphaHalf2,mean=0,sd=1)*sd(population)/sqrt(length(population))
  vals<-seq(leftLimit, rightLimit, length =50)
  x1<-c(leftLimit, vals, rightLimit)
  y1<-c(0,dnorm(vals, mean= myMean, sd= MySd),0)
  polygon (x1,y1,density=50,angle=-45,col='red')

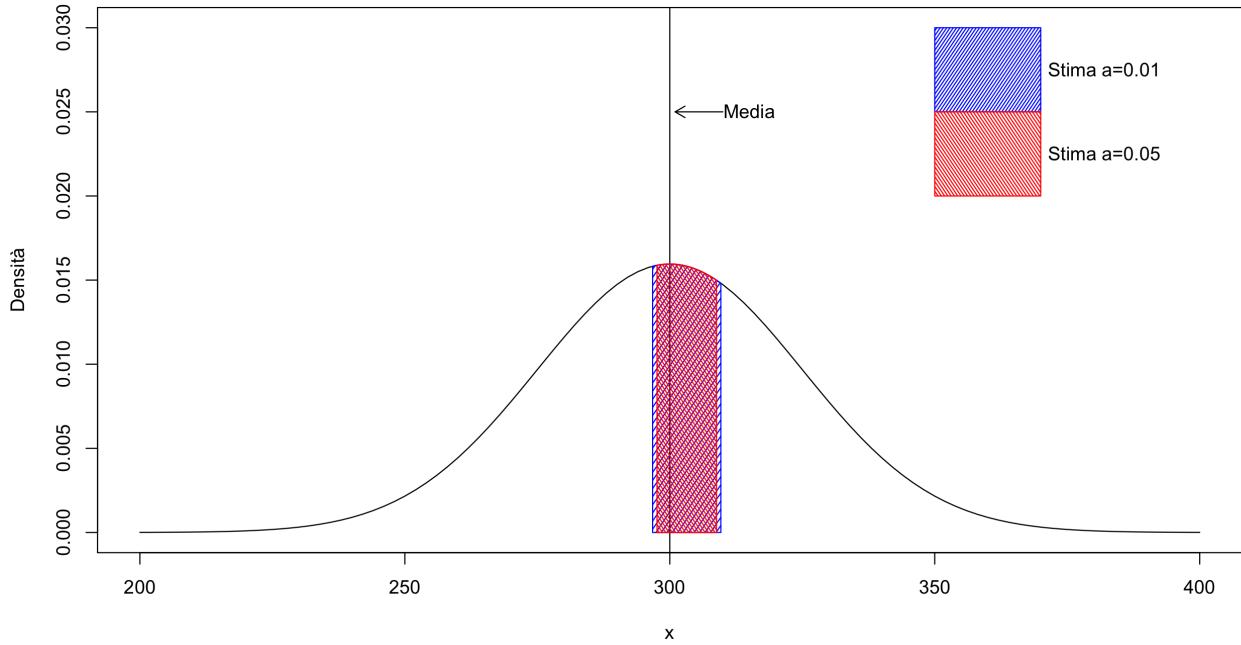
  xLegend <-c(350,370,370,350)
  yLegend <-c(0.02,0.02,0.025,0.025,0.02)
  polygon (xLegend, yLegend,density=50,angle=-45,col='red')
  text(382, 0.0225,"Stima a=0.05")

}

}

```

Confronto fra i due intervalli di confidenza



Come possiamo notare dal grafico, all'aumentare del grado di fiducia aumenta anche la lunghezza dell'intervallo di confidenza.

2.3 Intervallo di confidenza per μ con σ^2 non nota

Per costruire un intervallo di confidenza di grado $1 - \alpha$ per il valore medio μ nel caso in cui la varianza σ^2 della popolazione normale non è nota viene utilizzato il metodo pivotale e considerata la variabile aleatoria di pivot

$$T_n = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}}$$

dove \bar{X}_n ed S_n rappresentano rispettivamente la media e la deviazione standard. T_n dipende dal campione casuale e dal parametro non noto σ^2 e quindi può essere interpretata come variabile di pivot ed è distribuita con legge di Student con $n-1$ gradi di libertà. Scegliendo nel metodo pivotale $\alpha_1 = -t_{\alpha/2, n-1}$ e $\alpha_2 = t_{\alpha/2, n-1}$ tale che

$$P(T_n < -t_{\alpha/2, n-1}) = P(T_n > t_{\alpha/2, n-1}) = \frac{\alpha}{2}$$

con la scrittura

$$P\left(\bar{X}_n - t_{\alpha/2, n-1} \frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + t_{\alpha/2, n-1} \frac{S_n}{\sqrt{n}}\right) = 1 - \alpha$$

indichiamo che nel $100(1 - \alpha)\%$ dei campioni presi in considerazione, la media μ è compresa nell'intervallo di confidenza aventi estremi

$$\left[\bar{X}_n - t_{\alpha/2, n-1} \frac{S_n}{\sqrt{n}}, \bar{X}_n + t_{\alpha/2, n-1} \frac{S_n}{\sqrt{n}}\right]$$

dove $1 - \alpha$ prende il nome di livello di confidenza. Nei seguenti esempi mostriamo la costruzione di intervalli di confidenza al variare del livello di confidenza e il loro confronto.

La funzione mostrata sotto è stata utilizzata per graficare gli intervalli di confidenza:

```

plot.studentConfidence1 <- function(population,oneMinusAlpha=0.95){

  n <- length(population)
  alpha <- 1- oneMinusAlpha

  alphaHalf <- alpha/2

  print(alphaHalf)

  #left
  print(c("Left limit: ",mean(population)-qt(1-alphaHalf,df=n-1)*sd(population)/sqrt(n)))

  #right
  print(c("Left limit: ",mean(population)+qt(1-alphaHalf,df=n-1)*sd(population)/sqrt(n)))

  z <- qt(1-alphaHalf,df=n-1)

  print(c("Quantile: ",z))

  #graphic
  curve(dt(x, df=n-1), from=-4, to=4, axes=FALSE, xlab="", ylab="Densità", main="Densità di Student con n-1 gradi di libertà intervallo di confidenza 1-alpha=0.95 per μ con varianza non nota ")
  axis(1,c(-4,-z,0,z,4), c("",abbreviate(-z,5),0,abbreviate(z,4),""))

  vals<-seq(-3,-z, length =100)
  x1<-c(-4,vals ,-z,-4)

  y1<-c(0,dt(vals,n-1),0,0)

  polygon (x1,y1,density =20,angle=45)

  vals<-seq(z,4, length =100)

  x1<-c(z,vals ,4,1)
  y1<-c(0,dt(vals,n-1),0,0)
  polygon (x1,y1,density =20,angle=45)
  abline(h=0)
  text(0,0.03, expression (1-alpha))
  text(0,0.05, "Regione di accettazione")
  text(2.5,0.06, expression(alpha/2))
  text(2.5,0.09, "Regione di rifiuto")
  text(-2.5,0.06, expression(alpha/2))
  text(-2.5,0.09, "Regione di rifiuto")

}

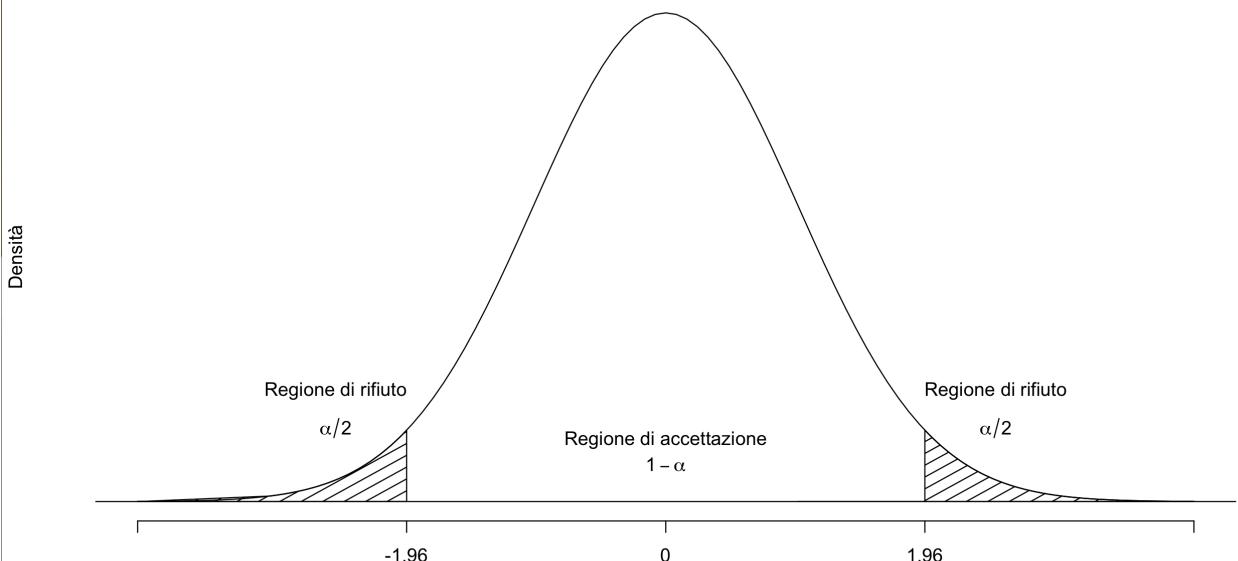
```

Caso 1 – $\alpha = 0.95$ (popolazione 1000 individui)

Supponendo che la popolazione da cui proviene il campione risulta essere normale, calcoliamo un intervallo di confidenza di grado $1 - \alpha = 0.95$ per la media μ degli elementi. In questo caso $\alpha = 0.05$ e $\alpha/2 = 0.025$.

Come possiamo notare $t_{0.025} = 1.96$ e la stima dell'intervallo di confidenza di grado $1-\alpha=0.95$ per la media μ è quindi(298.7836, 301.9327).

Densità di Student con n-1 gradi di libertà intervallo di confidenza 1-alpha=0.95 per μ con varianza non nota

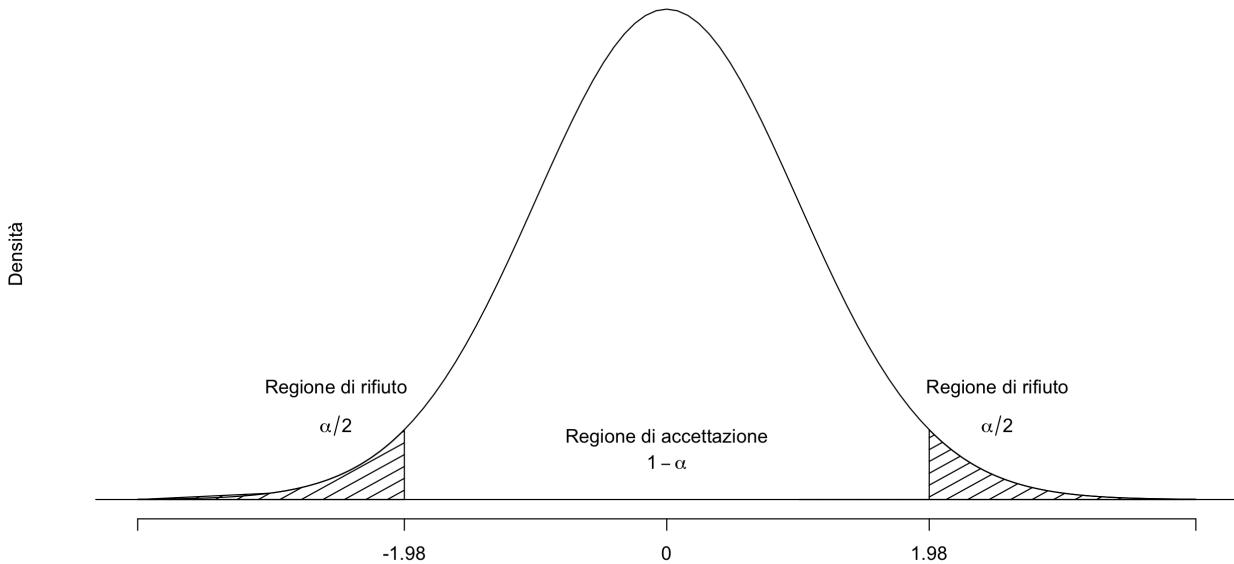


Caso 1 – $\alpha = 0.95$ (popolazione 100 individui)

Supponendo che la popolazione da cui proviene il campione risulta essere normale, calcoliamo un intervallo di confidenza di grado $1 - \alpha = 0.95$ per la media μ degli elementi. In questo caso $\alpha = 0.05$ e $\alpha/2 = 0.025$.

Come possiamo notare $t_{0.025} = 1.98$ e la stima dell'intervallo di confidenza di grado $1-\alpha=0.95$ per la media μ è quindi(296.66, 307.1863).

Densità di Student con n-1 gradi di libertà intervallo di confidenza 1-alpha=0.95 per μ con varianza non nota

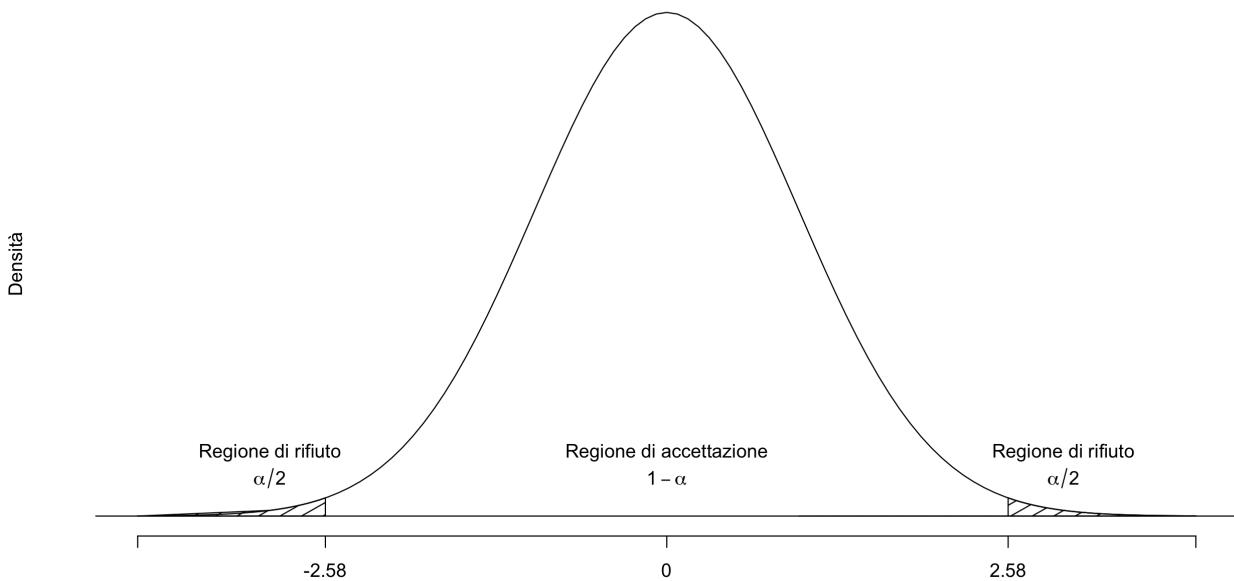


Caso 1 – $\alpha = 0.99$ (popolazione 1000 individui)

In questo caso determiniamo una stima dell'intervallo di confidenza di grado $1-\alpha=0.99$ per la media μ degli elementi e quindi $\alpha = 0.01$ e $\alpha/2 = 0.005$

Come possiamo notare $t_{0.005} = 2.58$ e la stima dell'intervallo di confidenza di grado $1-\alpha=0.99$ per la media μ è quindi(298.28, 302.42).

Densità di Student con n-1 gradi di libertà intervallo di confidenza 1-alpha=0.95 per μ con varianza non nota

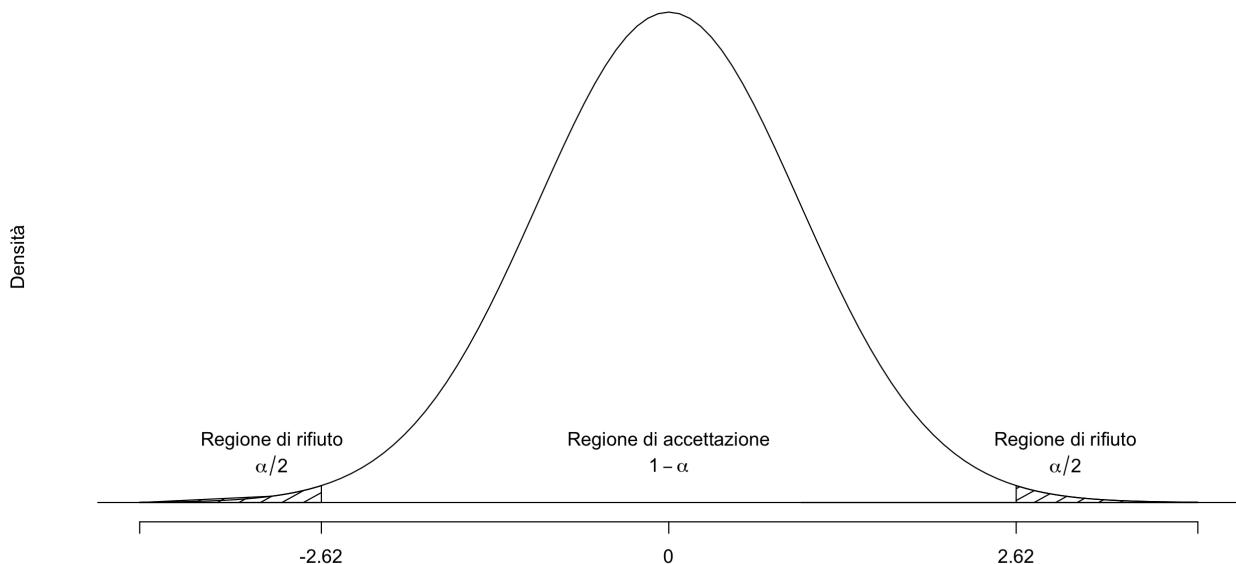


Caso 1 – $\alpha = 0.99$ (popolazione 100 individui)

In questo caso determiniamo una stima dell'intervallo di confidenza di grado $1-\alpha=0.99$ per la media μ degli elementi e quindi $\alpha = 0.01$ e $\alpha/2 = 0.005$

Come possiamo notare $t_{0.005} = 2.62$ e la stima dell'intervallo di confidenza di grado $1-\alpha=0.99$ per la media μ è quindi(294.9672, 308.8883).

Densità di Student con $n-1$ gradi di libertà intervallo di confidenza $1-\alpha=0.95$ per μ con varianza non nota



Comparazione dei risultati

Di seguito viene mostrato il grafico della distribuzione normale della popolazione, confrontando le stime della media nei due precedenti esempi. Da notare come il primo grafico si riferisca alla popolazione di 1000 individui mentre il secondo a quello di 100.

La funzione utilizzata:

```
plot.confidenceComparisonStudent <- function(population,myMean=300,MySd=25,myFrom=200,myTo=400, alphaHalf1=0.005, alphaHalf2=0.025){

  curve(dnorm(x,mean= myMean,sd= MySd),from=myFrom,to=myTo,ylim=c(0,0.03), ylab="Densità",xlab="x",main="Confronto fra i due intervalli di confidenza")

  arrows(301,0.025,310,0.025,code=1,length = 0.10)
  text(315,0.025,"Media")
  abline(v=myMean)

  n <- length(population)

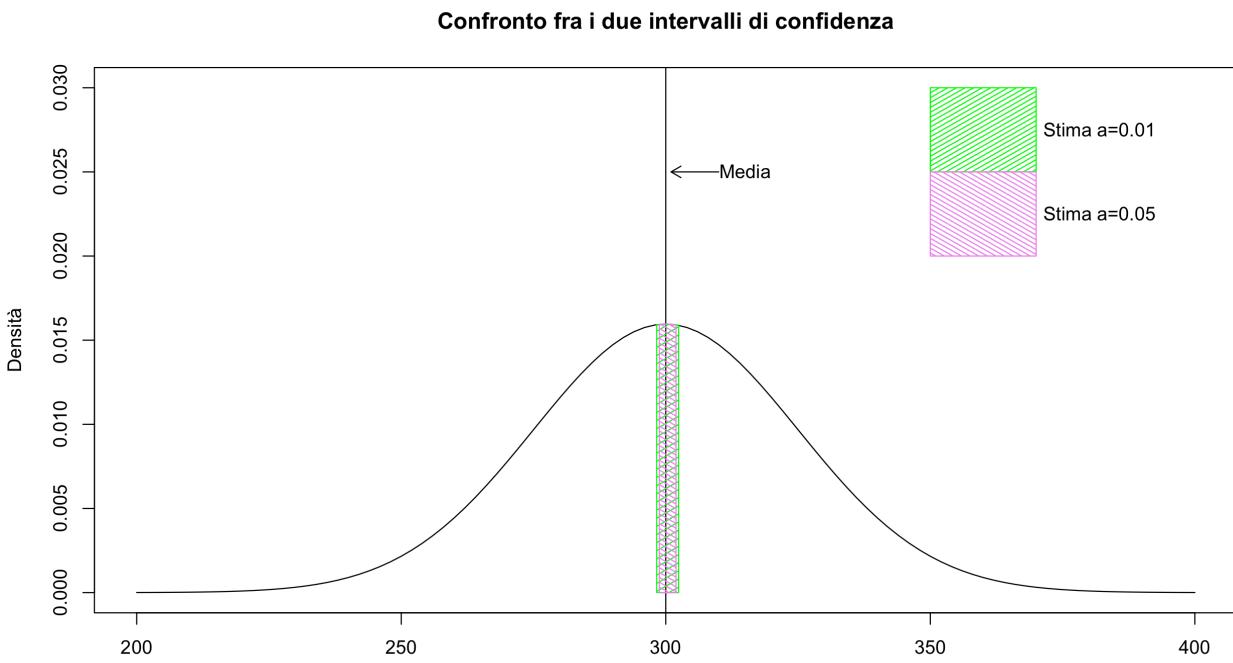
  #first
  leftLimit <- mean(population)-qt(1-alphaHalf1,df=n-1)*sd(population)/sqrt(n)
  rightLimit <- mean(population)+qt(1-alphaHalf1,df=n-1)*sd(population)/sqrt(n)
  vals<-seq(leftLimit, rightLimit, length =50)
  x1<-c(leftLimit, vals, rightLimit)
  y1<-c(0,dnorm(vals, mean= myMean, sd= MySd),0)
  polygon (x1,y1,density=30,angle=45,col='green')

  xLegend <-c(350,370,370,350,350)
  yLegend <-c(0.025,0.025,0.03,0.03,0.025)
  polygon (xLegend, yLegend,density=50,angle=45,col='green')
  text(382, 0.0275,"Stima a=0.01")

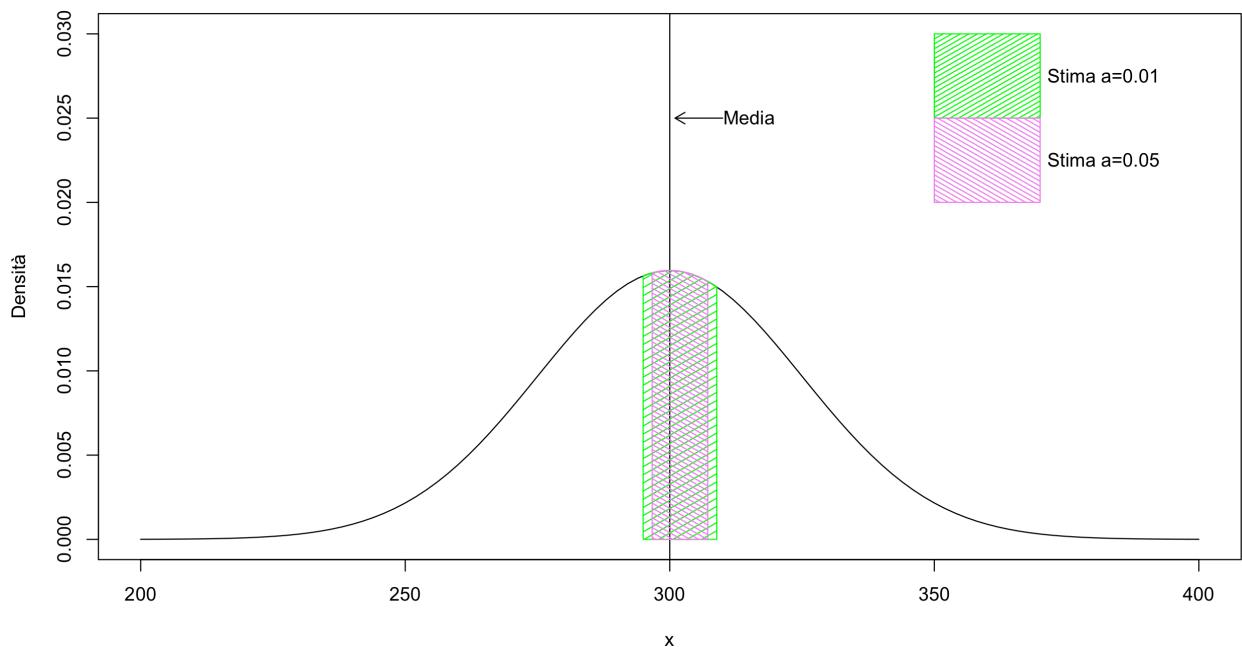
  #second
  leftLimit <- mean(population)-qt(1-alphaHalf2,df=n-1)*sd(population)/sqrt(n)
  rightLimit <- mean(population)+qt(1-alphaHalf2,df=n-1)*sd(population)/sqrt(n)
  vals<-seq(leftLimit, rightLimit, length =50)
  x1<-c(leftLimit, vals, rightLimit)
  y1<-c(0,dnorm(vals, mean= myMean, sd= MySd),0)
  polygon (x1,y1,density=50,angle=-45,col='violet')

  xLegend <-c(350,370,370,350,350)
  yLegend <-c(0.02,0.02,0.025,0.025,0.02)
  polygon (xLegend, yLegend,density=50,angle=-45,col='violet')
  text(382, 0.0225,"Stima a=0.05")

}
```



Confronto fra i due intervalli di confidenza



Come possiamo notare dal grafico, all'aumentare del grado di fiducia aumenta anche la lunghezza dell'intervallo di confidenza.

2.4 Intervallo di confidenza per σ^2 con valore medio μ noto

Per costruire un intervallo di confidenza di grado $1 - \alpha$ per la varianza σ^2 nel caso in cui il valore medio μ della popolazione normale è noto viene utilizzato il metodo pivotale e considerata la variabile aleatoria di pivot:

$$V_n = \sum_{t=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2$$

Tale variabile aleatoria dipende dal campione casuale e dal parametro non noto σ^2 ed è distribuita con legge di chi-quadrato con n gradi di libertà, essendo costituita dalla somma dei quadrati di n variabili aleatorie normali standard. V_n può essere scritta anche nel seguente modo, in termini di media e varianza campionaria:

$$V_n = \frac{(n-1)S_n^2}{\sigma^2} + \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \right)^2$$

Scegliendo nel metodo pivotale $\alpha_1 = X_{1-\alpha/2,n-1}^2$ e $\alpha_2 = X_{\alpha/2,n-1}^2$ in modo tale che

$$P(0 < V_n < X_{1-\alpha/2,n}^2) = P(V_n > X_{\alpha/2,n}^2) = \frac{\alpha}{2}$$

Abbiamo

$$P(X_{1-\alpha/2,n}^2 < V_n < X_{\alpha/2,n}^2) = 1 - \alpha$$

Ottenendo successivamente

$$P\left(\frac{(n-1)S_n^2 + n(\bar{X}_n - \mu)^2}{X_{\alpha/2,n}^2} < \sigma^2 < \frac{(n-1)S_n^2 + n(\bar{X}_n - \mu)^2}{X_{1-\alpha/2,n}^2}\right) = 1 - \alpha$$

Tale scrittura indica che nel $100(1 - \alpha)\%$ dei campioni presi in considerazione, la varianza σ^2 è compresa nell'intervallo di confidenza aventi estremi

$$\left[\frac{(n-1)S_n^2 + n(\bar{X}_n - \mu)^2}{X_{\alpha/2,n}^2}; \frac{(n-1)S_n^2 + n(\bar{X}_n - \mu)^2}{X_{1-\alpha/2,n}^2} \right]$$

e $1 - \alpha$ prende il nome di livello di confidenza. Nei seguenti esempi mostriamo la costruzione di intervalli di confidenza al variare del livello di confidenza e il loro confronto.

Caso 1 – $\alpha = 0.95$

Supponendo che la popolazione da cui proviene il campione preso in considerazione è normale con media uguale a 300 determiniamo una stima dell'intervallo di confidenza di grado $1-\alpha=0.95$ per la varianza σ^2 . In questo caso $\alpha = 0.05$ e quindi $\alpha/2 = 0.025$ e $1 - \alpha/2 = 0.975$.

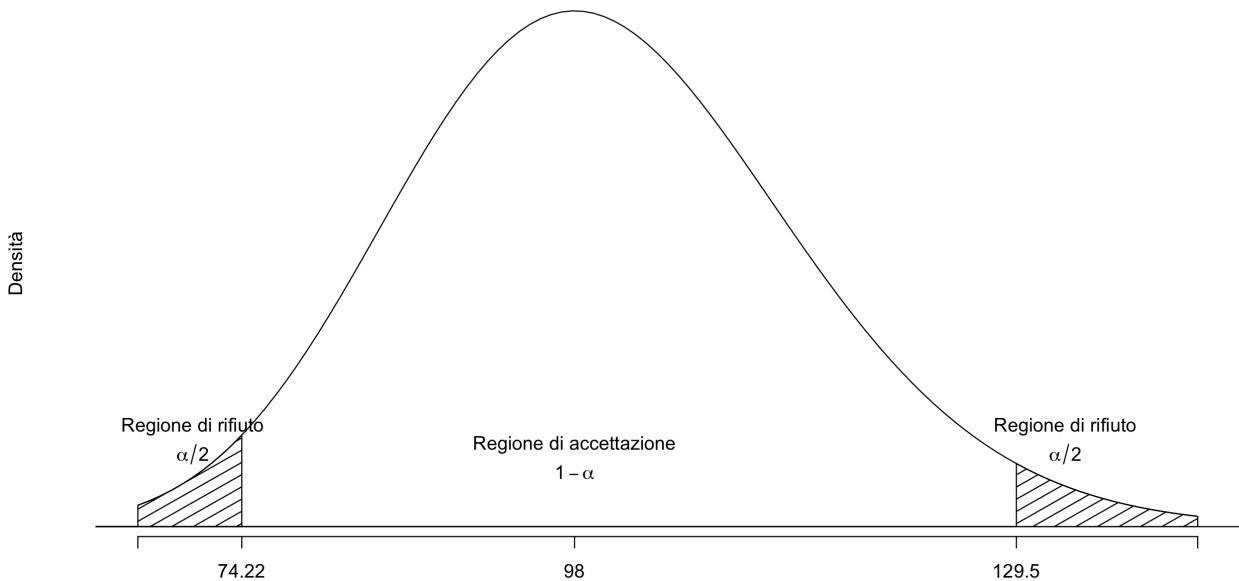
La funzione utilizzata è la seguente:

```
plot.chiQConfidence1 <- function(population,mymean,myalpha){
  n<-length(population)
  x<-population
  alpha<-myalpha
  varianza<-var(x)
  print("Varianza: ",varianza)

  print("Gradi di libertà: ",n)
  z1<-qchisq(1-alpha,df=n)
  print("Z1",z1)
  z2<-qchisq(alpha,df=n)
  print("Z2",z2)
  print("Limit1",alpha,z1)
  print("Limit2",alpha,z2)
  print("((n-1)*varianza)+n*((mean(x)-mymean)**2))/qchisq(1-alpha,df=n))
  print("((n-1)*varianza)+n*((mean(x)-mymean)**2))/qchisq(alpha,df=n))
  curve(qchisq(x, df=n), from=z2-z2/10, to=z1+z1/10, axes=FALSE,xlab="",ylab="Densità chi-quadrato con n-1 gradi di libertà intervallo di confidenza per la varianza con μ non noto,1-alpha=0.95")
  axis(1,c(z2-z2/10,z2-n-2,z1,z1+z1/10),c("",abbreviate(z2,minl=5),n-2,abbreviate(z1,minl=5),""))
  vals<-seq(z2-z2/10,z2, length =100)
  x1<-c(z2-z2/10,vals,z2,0)
  y1<-c(0,dchisq(vals,n),0,0)
  polygon (x1,y1,density =20,angle=45)
  vals<-seq(z1,z1+z1/10, length =100)
  x1<-c(z1,vals,z1+z1/10,1)
  y1<-c(0,dchisq(vals,n),0,0)
  polygon (x1,y1,density =20,angle=45)
  abline(v=0)
  text(z1+3.5,0.004,expression(alpha/2))
  text(z1+3.5,0.0055,"Regione di rifiuto")
  text(z1+3.5,0.003,expression(1-alpha))
  text(z1+3.5,0.0045,"Regione di accettazione")
  text(z2-3.5,0.0055,"Regione di rifiuto")
  text(z2-3.5,0.005,expression(alpha/2))
}

> plot.chiQConfidence1(popolazione1,300,0.025)
[1] "Varianza: "           "702.365096395704"
[1] "Gradi di libertà: "   "100"
[1] "Z1"                  "0.975"                "129.561197185837"
[1] "Z2"                  "0.025"                "74.2219274749237"
[1] "Limit1"              "0.975"                "539.557914505747"
[1] "Limit2"              "0.025"                "941.847938105298"
```

Densità chi-quadrato con $n-1$ gradi di libertà intervallo di confidenza per la varianza con μ non noto, $1-\alpha=0.95$

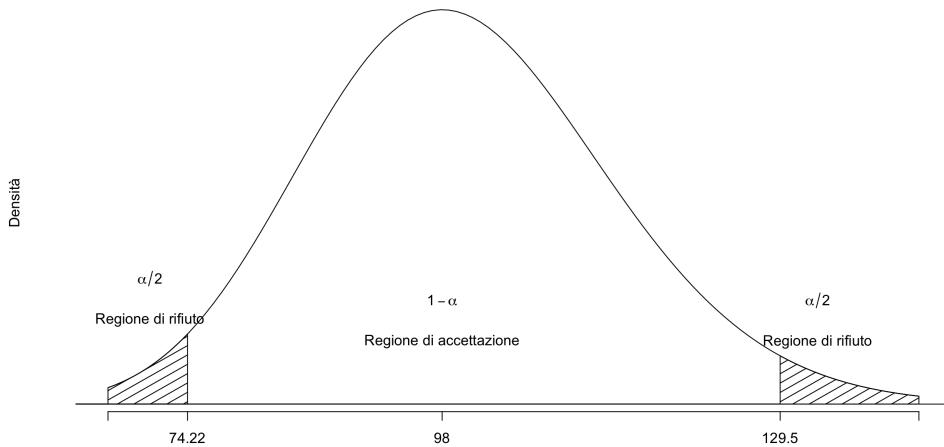


Come possiamo notare $X^2_{0.975,100} = 129.56$ e $X^2_{0.025,100} = 74.22$ e la stima dell'intervallo di confidenza di grado $1-\alpha=0.95$ per la varianza σ^2 è quindi(539.55, 941.84).

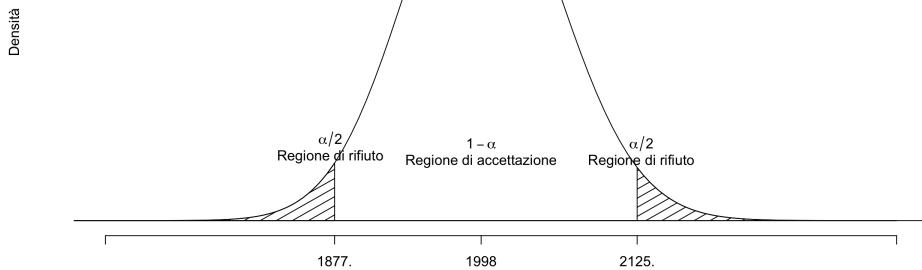
Ulteriori test per $1 - \alpha = 0.95$ (incremento popolazione)

```
> plot.chiQConfidence2(popolazione1, 300, 0.025, 0.001, 0.95, "popolazione1")
[1] "Varianza: "           "702.365096395704"
[1] "Gradi di libertà: "   "100"
[1] "Z1"                  "0.975"          "129.561197185837"
[1] "Z2"                  "0.025"          "74.2219274749237"
[1] "Limit1 "              "0.975"          "539.557914505747"
[1] "Limit2 "              "0.025"          "941.847938105298"
> plot.chiQConfidence2(popolazione2, 300, 0.025, 0.0001, 0.95, "popolazione2")
[1] "Varianza: "           "596.762539602673"
[1] "Gradi di libertà: "   "2000"
[1] "Z1"                  "0.975"          "2125.84230244978"
[1] "Z2"                  "0.025"          "1877.94603681539"
[1] "Limit1 "              "0.975"          "562.005620487435"
[1] "Limit2 "              "0.025"          "636.192573601714"
> plot.chiQConfidence2(popolazione6, 300, 0.025, 0.0001, 0.95, "popolazione6")
[1] "Varianza: "           "634.060780538938"
[1] "Gradi di libertà: "   "32000"
[1] "Z1"                  "0.975"          "32497.7276042898"
[1] "Z2"                  "0.025"          "31506.0609904459"
[1] "Limit1 "              "0.975"          "624.332144150861"
[1] "Limit2 "              "0.025"          "643.983262819479"
```

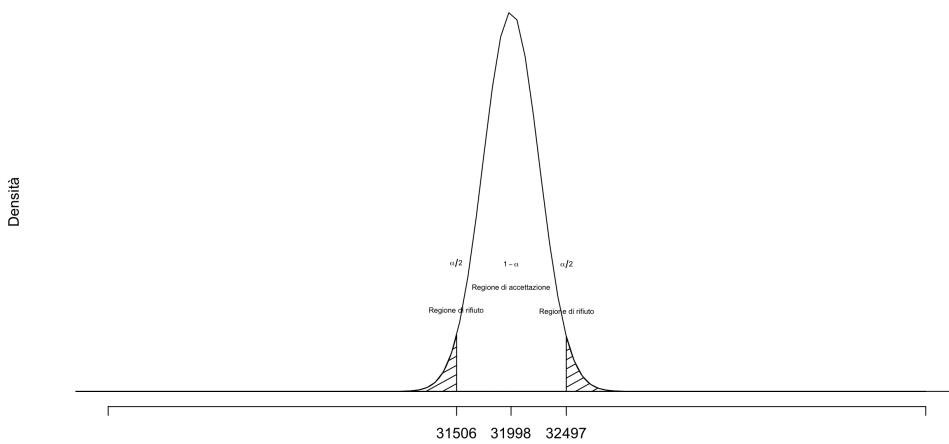
Densità chi-quadrato con $n-1$ gradi di libertà intervallo di confidenza per la varianza con μ non noto, $1-\alpha=0.95$
popolazione1



Densità chi-quadrato con $n-1$ gradi di libertà intervallo di confidenza per la varianza con μ non noto, $1-\alpha=0.95$
popolazione2



Densità chi-quadrato con $n-1$ gradi di libertà intervallo di confidenza per la varianza con μ non noto, $1-\alpha=0.95$
popolazione6

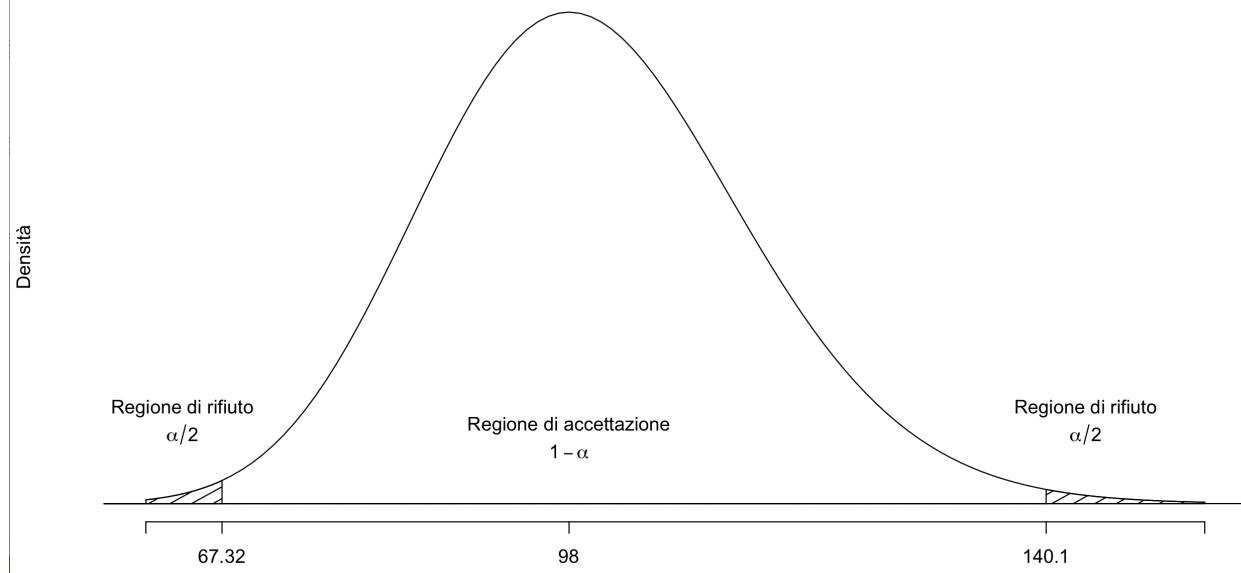


Caso 1 – $\alpha = 0.99$

In questo caso determiniamo una stima dell'intervallo di confidenza di grado $1-\alpha=0.99$ per la varianza σ^2 degli elementi. In questo caso $\alpha = 0.01$ e quindi $\alpha/2 = 0.005$ e $1 - \alpha/2 = 0.995$

```
> plot.chiQConfidence1(popolazione1,300,0.005)
[1] "Varianza: "           "702.365096395704"
[1] "Gradi di libertà: "   "100"
[1] "Z1"                  "0.995"          "140.169489442314"
[1] "Z2"                  "0.005"          "67.3275633054792"
[1] "Limit1 "              "0.995"          "498.723150327428"
[1] "Limit2 "              "0.005"          "1038.29347034707"
```

Densità chi-quadrato con $n-1$ gradi di libertà intervallo di confidenza per la varianza con μ non noto, $1-\alpha=0.99$



Comparazione dei risultati

Come possiamo notare dai grafici ottenuti aumentando il livello di confidenza aumenta anche la dimensione dell'intervallo di confidenza.

2.5 Intervallo di confidenza per σ^2 con valore medio μ non noto

Per costruire un intervallo di confidenza di grado $1 - \alpha$ per la varianza σ^2 nel caso in cui il valore medio μ della popolazione normale non è noto viene utilizzato il metodo pivotale e considerata la variabile aleatoria di pivot:

$$Q_n = \frac{(n-1)S_n^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Questa variabile dipende dal campione casuale e dal parametro non noto σ^2 ed è distribuita con legge chi-quadrato con $n-1$ gradi di libertà. Scegliendo nel metodo pivotale $\alpha_1 = X_{1-\alpha/2,n-1}^2$ e $\alpha_2 = X_{\alpha/2,n-1}^2$ tale che

$$P(0 < Q_n < X_{1-\alpha/2,n-1}^2) = P(Q_n > X_{\alpha/2,n-1}^2) = \frac{\alpha}{2}$$

Abbiamo

$$P(X_{1-\alpha/2,n-1}^2 < Q_n < X_{\alpha/2,n-1}^2) = 1 - \alpha$$

Ottenendo successivamente

$$P\left(\frac{(n-1)S_n^2}{X_{\alpha/2,n-1}^2} < \sigma^2 < \frac{(n-1)S_n^2}{X_{1-\alpha/2,n-1}^2}\right) = 1 - \alpha$$

Tale scrittura indica che nel $100(1 - \alpha)\%$ dei campioni presi in considerazione, la varianza σ^2 è compresa nell'intervallo di confidenza aventi estremi

$$\left[\frac{(n-1)S_n^2}{X_{\alpha/2,n-1}^2}, \frac{(n-1)S_n^2}{X_{1-\alpha/2,n-1}^2} \right]$$

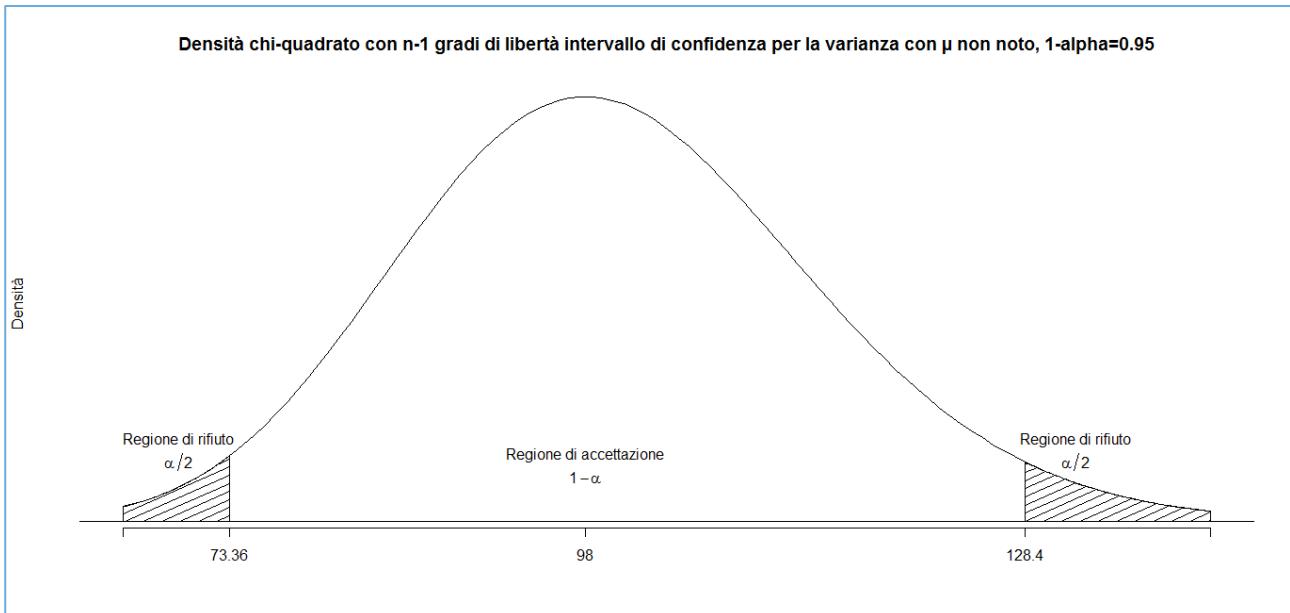
e $1 - \alpha$ prende il nome di livello di confidenza. Nei seguenti esempi mostriamo la costruzione di intervalli di confidenza al variare del livello di confidenza e il loro confronto.

Caso 1 – $\alpha = 0.95$

Supponendo che la popolazione da cui proviene il campione sia normale, determiniamo una stima dell'intervallo di confidenza di grado $1 - \alpha = 0.95$ per la varianza σ^2 degli elementi. In questo caso $\alpha = 0.05$ e quindi $\alpha/2 = 0.025$ e $1 - \alpha/2 = 0.975$.

La funzione utilizzata è la seguente:

```
plot.chiQConfidence3 <- function(population,myalpha){  
  n<-length(population)  
  x<-population  
  alpha<-myalpha  
  varianza<-var(x)  
  z1<-qchisq(1-alpha,df=n-1)  
  print(c("Z1",z1))  
  z2<-qchisq(alpha,df=n-1)  
  print(c("Z2",z2))  
  
  print(c("Left",((n-1)*varianza)/qchisq(1-alpha,df=n-1)))  
  print(c("Right",((n-1)*varianza)/qchisq(alpha,df=n-1)))  
  
  curve(dchisq(x, df=n), from=z2-  
    z2/10,to=z1+z1/10,axes=FALSE,xlab="",ylab="Densità chi quadrato con n-1 gradi di libertà intervallo di confidenza per la varianza con μ non noto, 1-alpha=0.95")  
  axis(1,c(z2-z2/10,z2,n-2,z1,z1+z1/10),c("",abbreviate(z2,minl=5),n-2,abbreviate(z1,minl=5),""))  
  vals<-seq(z2-z2/10,z2, length =100)  
  xl<-c(z2-z2/10,vals ,z2,0)  
  y1<-c(0,dchisq(vals,n),0,0)  
  polygon (xl,y1,density =20,angle=45)  
  vals<-seq(z1,z1+z1/10, length =100)  
  x1<-c(z1,vals ,z1+z1/10,1)  
  y1<-c(0,dchisq(vals,n),0,0)  
  polygon (x1,y1,density =20,angle=45)  
  abline(h=0)  
  text(z1+3.5,0.004,expression(alpha/2))  
  text(z1+3.5,0.0055,"Regione di rifiuto")  
  text(98,0.003,expression(1-alpha))  
  text(98,0.0045,"Regione di accettazione")  
  text(z2-3.5,0.0055,"Regione di rifiuto")  
  text(z2-3.5,0.004,expression(alpha/2))  
}  
  
> plot.chiQConfidence3(popolazione1,0.025)  
[1] "Z1"           "128.42198864384"  
[1] "Z2"           "73.3610801912837"  
[1] "Left"         "541.450457802966"  
[1] "Right"        "947.834251647733"
```

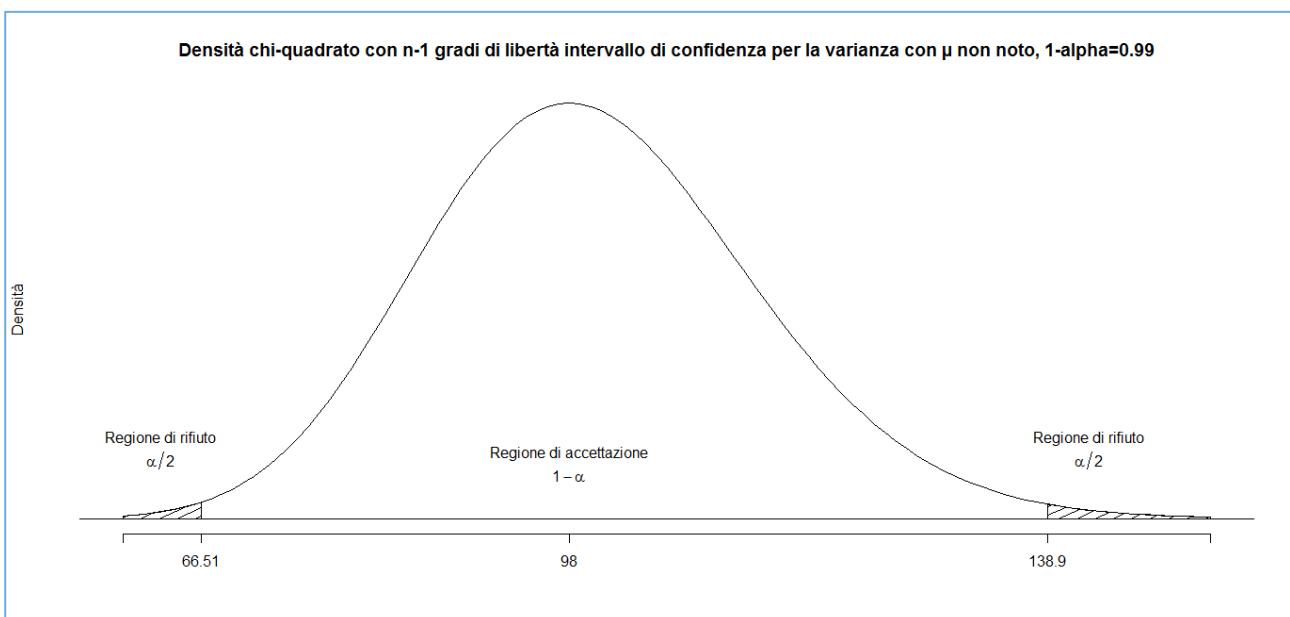


Caso 1 – $\alpha = 0.99$

In questo caso determiniamo una stima dell'intervallo di confidenza di grado $1-\alpha=0.99$ per la varianza σ^2 degli elementi. In questo caso $\alpha = 0.01$ e quindi $\alpha/2 = 0.005$ e $1 - \alpha/2 = 0.995$.

```
> plot.chiQConfidence3(popolazione1,0.005)
[1] "Z1"                  "0.995"                "138.98678345094"
[1] "Z2"                  "0.005"                "66.5101053017374"
[1] "Left"                "500.29321361854"
[1] "Right"               "1045.46736511269"
```

Come possiamo notare $X_{0.995,99}^2 = 66.51011$ e $X_{0.005,99}^2 = 138.9868$ e la stima dell'intervallo di confidenza di grado $1-\alpha=0.99$ per la varianza σ^2 è quindi(500.29, 1054.46).



Comparazione dei risultati

Come possiamo notare dai grafici ottenuti aumentando il livello di confidenza aumenta anche la dimensione dell'intervallo di confidenza.