Laboratorio di ST1
Lezione 6

Claudia Abundo

Dipartimento di Matematica
Università degli Studi Roma Tre

23 Aprile 2010

Stima puntuale (Legge Grandi Numeri)

Vediamo come variano gli stimatori media campionaria e varianza campionaria (corretta) al variare di n.
Generiamo una serie di lanci di una moneta e calcoliamo la media per $n = 1, 2, \ldots 10000$

```
numrep=10000
x=rbinom(numrep,1,0.5)#SI GENERA UN'UNICA SEQUENZA
xm=0
xvar=0
for (i in 1:numrep){
xm[i]=mean(x[1:i])
xvar[i]=var(x[1:i])
}
par(mfrow=c(1,2))
plot(xm,type="l",log="x",main="Media Campionaria",
 xlab="lanci moneta",ylab="Media lancio")
abline(h=0.5,col=2)
plot(xvar,type="l",log="x", main="Varianza Campionaria",
xlab="lanci moneta",ylab="Varianza lancio")
abline(h=0.25,col=3)
```

Si genera un'unica sequenza di estrazioni e si calcola la media sui primi n elementi, $n = 1, \ldots 10000$

Teorema Centrale del Limite

Supponiamo vi aver calcolato (numrep=)10000 volte la media campionaria per $n_1 = 50$ e $n_2 = 500$ lanci di moneta. Analizziamo la distribuzione di $\bar{X}_{n_i}$

```
numrep=10000
n1=20
n2=500
p=0.5
xm1=0
xm2=0
for (i in 1:numrep){xm1[i]=mean(rbinom(n1,1,p))}
for (i in 1:numrep){xm2[i]=mean(rbinom(n2,1,p))}
par(mfrow=c(2,1))

hist(xm1,freq=F,col="gray")
curve(dnorm(x,p,sqrt(p*(1-p)/n1)),min(xm1),
max(xm1),add=T,col=2,lwd=3)

hist(xm2, freq=F,col="gray")
curve(dnorm(x,p,sqrt(p*(1-p)/n2)),
min(xm1),max(xm1),add=T,col=2,lwd=3)
```

modifichiamo p=0.5 con p=0.1, p=0.9,...

Inferenza in Grandi Campioni

Esercizio (Moore)
Il latte materno contiene una certa quantità di calcio, una parte del quale deriva direttamente dal calcio contenuto nelle loro ossa. Alcune donne, quindi, durante l'allattamento possono andare incontro a demineralizzazione ossea. I ricercatori hanno misurato la variazione percentuale di calcio nelle vertebre di 47 mamme nel corso di tre mesi d'allattamento. I dati sono contenuti nel file `calcio.txt`.
Si costruisca un intervallo di confidenza al 99% per la media della variazione percentuale di calcio nella popolazione utilizzando.

E' ragionevole supporre che il campione sia IID. Inoltre la numerosità campionaria $n = 47$ dovrebbe essere sufficientemente elevata per applicare il TCL.

Stima per intervallo

La funzione qnorm serve per ottenere i quantili della distribuzione Normale Standardizzata. Ugualmente la funzione qt serve per ottenere i quantili della distribuzione Student-T e così via per le altre distribuzione disponibili (Cauchy, Lognormale, Gamma...) Per avere un elenco delle distribuzioni disponibili cercate distribution nell'help-on-line.

```
qnorm(0.5)
qnorm(0.95)
```

La varianza della popolazione é ignota, dobbiamo stimarla.

```
XM=mean(cm)
S=sd(cm)
n=length(cm)
Z=qnorm(99.5/100) #QUANTILE
CI=c(XM-Z*S/sqrt(n),XM+Z*S/sqrt(n))
# INTERVALLO DI CONFIDENZA
CI
# STIMA PUNTUALE
XM
```

L'intervallo di confidenza ci indica il grado di fiducia che abbiamo nella stima puntuale. Tanto più é ampio l'intervallo, tanto più la fiducia é bassa. La fiducia aumenta al crescere di n.

Grafico

```
n=length(cm)
curve(dnorm(x,XM,S/sqrt(n)),XM-2,XM+2)
v=seq(CI[1],CI[2],by=0.001)
x=c(CI[1],v,CI[2])
y=c(0,rep(0.05,length(v)),0)
polygon(x,y,col="gray")
text(XM,0.03,"Intervallo di Confidenza")
lines(c(XM,XM),c(0,dnorm(XM,XM,S/sqrt(n))))
curve(dnorm(x,XM-0.5,S/sqrt(n)),XM-2,XM+2,add=T,col=2)
lines(c(XM-0.5,XM-0.5),c(0,dnorm(XM-0.5,XM-0.5,S/sqrt(n))),col=2)
curve(dnorm(x,XM+1,S/sqrt(n)),XM-2,XM+2,add=T,col=3)
lines(c(XM+1,XM+1),c(0,dnorm(XM+1,XM+1,S/sqrt(n))),col=3)
```

Un valore di $\bar{X}$ che cade nell'intervallo di confidenza potrebbe essere generato da una distribuzione la cui media é fuori dall'intervallo (in verde)! Se il TCL fornisce una buona approssimazione, questo accadrà (circa) l'$\alpha\%$ delle volte.

Esercizi

Per costruire un intervallo di confidenza a livello 0.95% per un campione di ampiezza 50 estratto da una distribuzione Normale di media incognita, media campionaria 6.2 e varianza = 4.6 si procede così:

```
Z <- qnorm(0.975)
CI = c(6.2-Z*4.6/sqrt(50),XM+Z*4.6/sqrt(50))
```

In R usando la funzione t.test ed usando l'argomento opzionale conf.level si ottengono intervalli di confidenza:

```
x <- c(1,2,3,4,5,6)
mean(x) +c(-1,1)*qt(0.975,5)*sd(x)/sqrt(length(x))
t.test(x,conf.level=0.95)
```

Per fare l'intervallo di confidenza per la differenza fra le medie di due campioni:

```
y <- c(2,3,4,5,6,7,8)
```

se assumo che le varianze delle 2 popolazioni sono ingognite ma UGUALI allora uso lo stimatore della varianza POOLED:

```
t.test(x,y,var.equal=T)
```

se l'ipotesi di uguaglianza delle 2 varianze non è valida allora R usa il metodo introdotto da WELCH:

```
t.test(x,y)
```

Per costruire intervalli di confidenza o fare prova delle ipotesi su proporzioni si usa il comando prop.test. Per esempio, supponiamo che su un campione di 200 individui il 41% ha detto che voterà a favore di un certo candidato alle prossime elezioni. Per fare un intervallo di confidenza a livello 99% e prova delle ipotesi uso il comando:

```
prop.test(82,200,conf.level=0.99)
```

la stessa risposta si ottiene con

```
prop.test(41,100,conf.level=0.99)
```

R usa la *correzione di continuità* e costruisce inoltre un intervallo non simmetrico