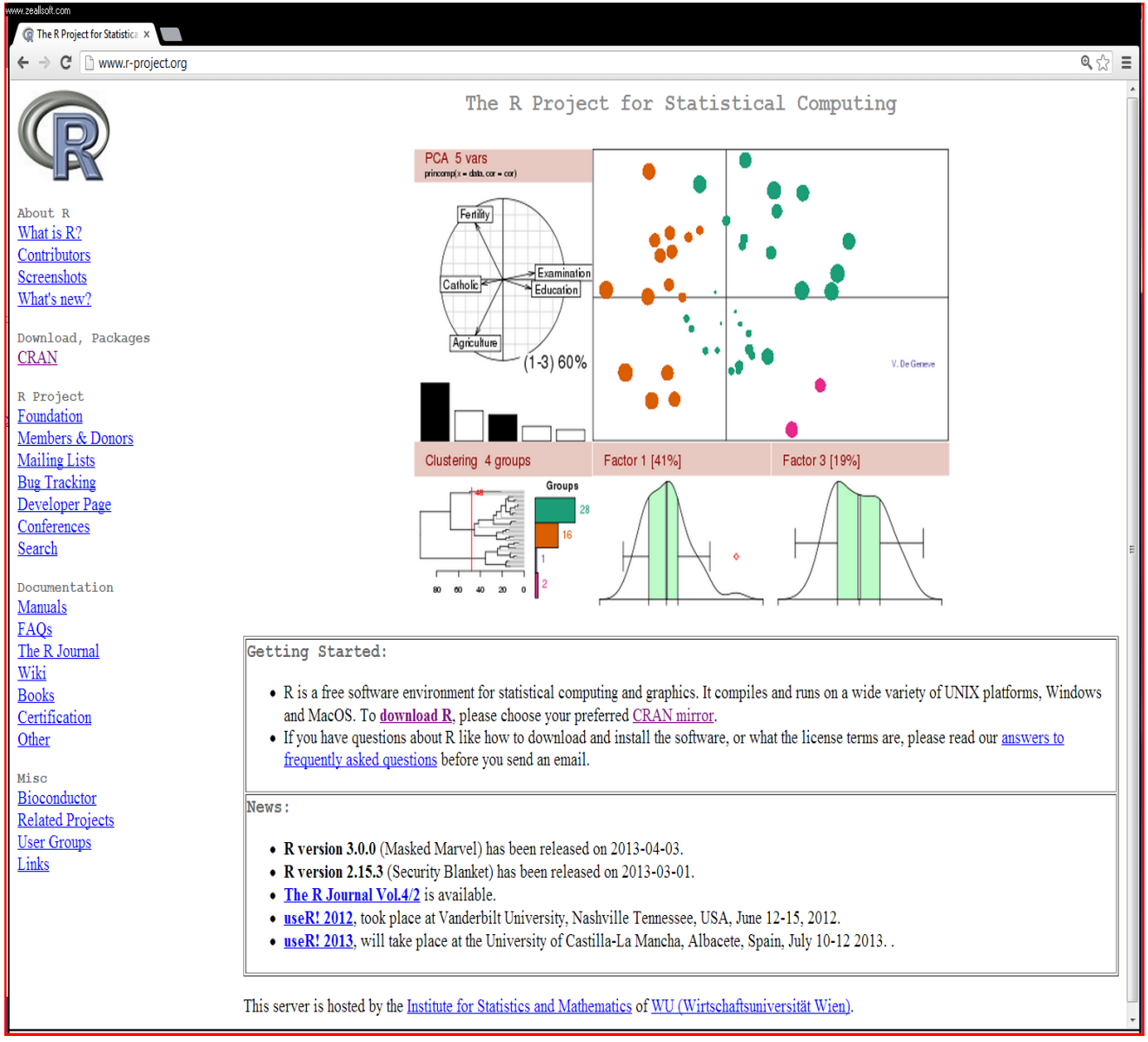
R e il mondo aleatorio

- Parte I -

Variabili casuali discrete e variabili casuali continue

Paola Lecca, CIBIO - UNITN
Corso di Matematica e Stastistica 2

Il software R: http://www.r-project.org/



Testo di riferimento utile

S. M. Iacus, G. Masarotto, Laboratorio di statistica con R, McGraw-Hill.

Variabili casuali discrete

Binomiale, geometrica, Binomiale negativa, ipergeometrica, di Poisson

Variabile casuale binomiale

La variabile casuale X che conta il numero di successi in n prove si chiama Binomiale e si scrive

$$X \sim Bin(n, p)$$

Dove p è tale che 0 <p < 1, ed indica l aprobailità di successo.

X assume tutti i valori da 0 ad n con la seguente distribuzione di probabilità

$$P(X = k) = \binom{n}{k} p^k (1-p)^k, \quad k = 0, 1, 2, \ldots, n$$

La variabile ha le seguenti proprietà

$$E[X] = np \quad Var[X] = np(1-p)$$

La Binomiale in R (1/2)

Se vogliamo calcolare la probabilità

$$P(X \leq 3)$$

dove

$$X \sim Bin(n = 10, p = 0.3)$$

usiamo il comando

```
pbinom(3, 10, 0.3)
```

per calcolare la probabilità

$$P(X > 3)$$

usiamo l'opzione lower.tail=FALSE, come segue:

```
pbinom(3, 10, 0.3, lower.tail=FALSE)
```

La Binomiale in R (2/2)

Se invece vogliamo calcolare la probabilità di X in un punto x = 3, possiamo utilizzare la funzione dbinom, come segue:

```
dbinom(3, 10, 0.3)
```

Il prefisso “d” serve per ricordarci che stiamo calcolando la densità di probabilità della variabile casuale.

Questo prefisso è comune a tutte le altre variabili casuali.

Variabile casuale geometrica (1/2)

In un esperimento Bernouliano ci si può chiedere quanto tempo si deve apsettare pe ravere il primo successo.

Per esempio, se vogliamo sapere con quale probabilità si avrà la prima T (testa) nel lancio di una moneta truccata, tale per cui

$$1 - p = P(C) = 7/8 \quad \text{e} \quad p = P(T) = 1/8.$$

la risposta si ottiene nel seguente modo

$$(1-p)^k \cdot p, \quad k = 0, 1, 2, \ldots$$

In tal caso l'esperiemnto potrebbe avere durata infinita o comunque non prevedibile al contrario, al contrario del modello Binomiale in cui viene fissato a priori il numero n di prove.

La variabile casuale "tempo di attesa per il primo successo" è chiamata *variabile casuale geometrica*.

Variabile casuale geometrica (2/2)

Il seguente codice R disegna una la densità di una geometrica con p = 1/8:

```
k <- 1:10

p <- dgeom(k, 1/8)

plot(k, p, type="h", lwd=10)
```

Variabile Binomiale negativa (1/3)

Una generalizzazione della distribuzione geometrica è data dalla variabile casuale Binomiale negativa per n =1, 2, ... e k = 0, 1, 2, ...

$$P(X = k) = \binom{n+k-1}{k} p^n (1-p)^k$$

e conta il numero di insuccessi k che si devono avere prima di ottenere l'n-esimo successo.

Posto n = 1 si ottiene esattameten la distribuzione geometrica di parametro p.

In R la densità e la distribuzione di probabilità si ottengono rispettivamente con in comandi

dnbinom

e

pnbinom

Variabile Binomiale negativa (2/3)

Esempio: per calcolare

$$P(X \leq 3)$$

e

$$P(X = 3)$$

con

$$X \sim NegBin(n = 5, p = 0.3)$$

scriveremo

pnbinom(3, 5, 0.3)

Variabile Binomiale negativa (3/3)

Provate a calcolare

$$P(X = 3)$$
$$P(Y = 3)$$

sapendo che

$$X \sim NegBin(n = 5, p = 0.3)$$
$$Y \sim Geom(p = 0.3)$$

Suggerimento: usate i comandi dnbinom e dgeom.

Variabile casuale ipergeometrica

Supponiamo di avere una popolazione di N individui di cui K di tipo 1 e gli altri N – k di tipo 2.

Se estraiamo un campione casuale di n individui, ci chiediamo: con quale probabilità k di questi sono ti tipo 1?

Sia X tale numero. Questa probabilità si calcola tramite il rapporto

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

X è detta variabile casuale ipergeometrica di parametri (N, K, n).

Le funzioni di riferimento in R sono

dhyper e phyper.

Variabile casuale di Poisson (1/3)

L'ambito di utilizzo del modello di Poisson è quello di un processo di Bernoulli con eventi rari, cioè con probabilità molto piccola di successo.

Il teorema di Poisson deriva proprio la distribuzione omonina proprio dal processo di Bernoulli.

Se p è prossimo a zero e np = λ rimane costante al crescere di n allora

$$P(X = k) \approx \frac{\lambda^k e^{-k}}{k!}$$

Una variabile casuale X che segue questa legge con $\lambda > 0$ viene detta variabile di Poisson e la indichiamo come

$$X \sim Poi(\lambda)$$

Variabile casuale di Poisson (2/3)

Il codice che segue fornisce una rappresentazione del grafico della densità di probabilità della distribuzione

$$X \sim Poi(5)$$

```
k <- 0:20

p <- dpois(k, lambda=5)

plot(k, p, type="h", lwd=10)
```

Variabile casuale di Poisson (3/3)

Utilizzare R per risolvere il seguente esercizio.

Si supponga che il numero medio di chiamate ad un centralino sia pari a 20 per ora.

Con quale probabilità in 5 minuti non arrivano chiamate oppure che in 10 minuti si abbiano al più 10 chiamate?

Le risposte sono: 0.1888756 e 0.9993085.

Variabili casuali continue

Uniforme, esponenziale, Normale, Gamnma e Beta

Variabile casuale uniforme

$$X \sim Unif(a,b) \qquad f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a,b) \\ 0 & \text{altrimenti} \end{cases}$$

Media e varianza dell'unfrome sono pari a

$$E[X] = \frac{a+b}{2} \quad Var[X] = \frac{(b-a)^2}{12}$$

La densità, la funzione di riaprtizione e i quantili si calcolano atraverso i comandi

dunif
punif
qunif

Variabile casuale esponenziale

$$X \sim Exp(\lambda) \qquad f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0 & \text{altrimenti} \end{cases}$$

Questa varaibile modella i tempi di arrivo di eventi indipendenti.
Media e varianza sono pari a

$$E[X] = \frac{1}{\lambda} \quad Var[X] = \frac{1}{\lambda^2}$$

La densità, la funzione di riaprtizione e i quantili si calcolano atraverso i comandi

dexp
pexp
qexp

Variabile causale Normale

$$X \sim N(\mu, \sigma^2) \qquad f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ è la media e σ^2 è la varianza.

Calcolare probabilità del tipp P(X < x) vuol dire eseguire il calcolo dell'integrale

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

che spesso non è risolvibile in modo analitico ma solo per via numerica.

Variabile casuale normale

Standardizzazione

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

R predispone funzioni per il calcolo della densità, funzione di ripartizione e dei quantili della Normale (sia quella genrica di parametri μ e σ^2, sia quella standard).

Le funzioni sono rispettivamente

dnorm
pnorm
qnorm

Variabile casuale normale

Calcolare P(X > 3) con X ~ N(5, 2).

```
pnorm(3, mean=5, sd=sqrt(2))
```

Oppure passando alla standardizzazione

```
pnorm((3 – 5)/sqrt(2))
```

Il codice seguente invece disegna i grafici della densità della Normale (provate ad eseguirli):

```
curve(dnorm(x, mean=-4), -10, 12, ylab=“”, axes=TURE)
curve(dnorm(x, mean=7), -10, 12, ylab=“”, add=TRUE)
```

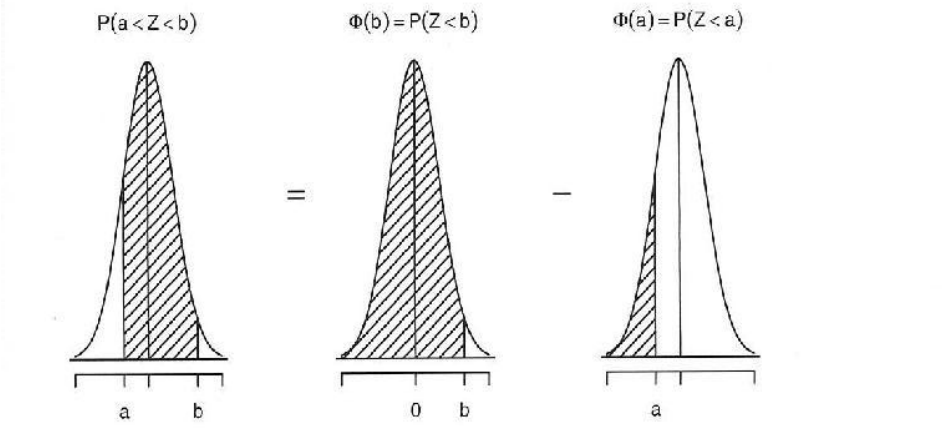
Calcolo grafico delle probabilità

Se volgiamo calcolare le probabilità del tipo

$$P(a < Z < b)$$

Ci possiamo aiutare con i grafici. Infatti rappresentando graficamente P(a < Z < b) notiamo che

$$P(a < Z < b) = P(Z < b) - P(Z < a)$$



Calcolo di $P(a < Z < b) = P(Z < b) - P(Z < a)$. Nell'esempio abbiamo utilizzato la distribuzione Gaussiana, ma la relazione vale per ogni distribuzione continua per il teorema fondamentale del calcolo integrale.

Intervalli notevoli (1/4)

Proviamo ora a calcolare le seguenti probabilità

$$P(\mu - \sigma < X < \mu + \sigma)$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma)$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma)$$

dove

$$X \sim N(\mu.\sigma^2)$$

Intervalli notevoli (2/4)

$$P(\mu - \sigma < X < \mu + \sigma) = P\left(\frac{\mu - \mu - \sigma}{\sigma} < \frac{X - \mu}{\sigma} < \frac{\mu + \sigma - \mu}{\sigma}\right)$$
$$= P(-1 < Z < 1) = \Phi(1) - \Phi(-1)$$
$$= 0.84134 - 0.15866 \approx 0.68$$

dove abbiamo inbtrodotto la seguente notazione

$$\Phi(z) \equiv P(Z < z)$$

Il risultato che abbiamo ottenuto ci dice che tutti i valoripossibili di una Normale si realizzano all'interno dell'intervallo

$$\mu \pm \sigma$$

Intervalli notevoli (3/4)

In R si può eseguir eil calcolo come segue

```
mu <- 5

sigma <- 2

pnorm( mu + sigma, mean=mu, sd=sigma) - pnorm( mu - sigma, mean=mu, sd=sigma)
```

Verificate che si ottiene lo stesso risultato facendo

```
pnorm(1) – pnorm(-1)
```

Usate R per calcolare gli altri intervalli.

Otterrete

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.954$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.997$$

Variabile casuale Gamma

$$f(x) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}, \quad \Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{-x}dx$$

Il valore atteso è $\alpha\beta$ e la varianza è $\alpha\beta^2$.

In R la funzione è Gamma e chiamata nel seguente modo:

gamma(0.5)

Qui calcoliano il valore della Gamma per α = 0.5). La densità, la funzione di ripartizione e i quantili si calcolano attraverso le funzioni

dgamma
pgamma
qgamma

Variabile casuale Beta (1/2)

$$f(x) = \frac{1}{Beta(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1, \quad \alpha > 0, \quad \beta > 0$$

$$Beta(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$$

Per $\alpha = \beta = 1$ si ha la distribuzioen uniforme.

Il valore atteso e la varianza sono

$$E[X] = \frac{\alpha}{\alpha+\beta} \qquad Var[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Variabile casuale Beta (2/2)

Densità, funzione di ripartizione e quantili si calcolano in R attraverso le funzioni

dbeta
pbeta
qbeta

Esempio

```
x <- seq(0, 1, length=21)
dbeta(x, 1, 1)
pbeta(x, 1, 1)
```

Esercizio: graficare sullo stesso plot le curve du desità per una variabile beta avente i seguenti parametri

(0.1, 1)
(1, 0.1)
(0.1, 0.1)
(4, 4)
(2, 6)
(6, 2)
(2, 2)

Generazione di variabili casuali (1/5)

Metodo dell'inversione

A titolo di esempio e come esercizio, supponiamo di voler generare una variabile casuale di Bernoulli di parametro p, cioè

X = 0 con probabilità 1 – p
X = 1 con probabilità p.

Dobbiamo generare una sequenza di 0 e 1.

La funzione di ripartizione è

$$F(x) = \begin{cases} 0, & x < 0 \\ 1-p, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

Generazione di variabili casuali (2/5)

Metodo dell'inversione

Generiamo un numero u da un'uniforme in (0, 1).

Se $u < 1 - p$, allora possiamo definire $F^{-1}(u) = 0$;
se invece $u \geq 1 - p$ definiamo $F^{-1}(u) = 1$.

In sostanza pe rgenerare un variabile casuale di Bernoulli, ci basta generare un numero compreso tra 0 e 1, e se ci viene più piccolo di 1 – p, diciamo che X vale 0, altrimeni che X vale 1.

In R, supponimao di voler generare 5 replicazioni di un aBernoulliana di parametro p =1/3. Questo si ottine tramite il comando

1* runif(5) < 1/3)

Generazione di variabili casuali (3/5)

Metodo dell'inversione

Un altro esempio.

Se vogliamo generare una variabile casuale di Bernoulli di parametri n = 10 e p = 1/3, basterà fare la somma degli 1 nella generazione di 10 replicazioni della variabile di Bernoulli:

```
sum (runif(10) < 1/3)
```

Generazione di variabili casuali (4/5)

Metodo dell'inversione

Supponiamo di voler generare un numero casuale da una variabile casuale X discreta che assume k distinti valori x_i, i = 1, 2, ..., k con distribuzione di probabilità p_1, p_2, ..., p_k.

Supponiamo di aver ordinato i valori xi in ordine crescente, così ch epossimao costruire le frequenze cumulate che rappresentano la funzione di ripartizione di questa variabile.

Se generaiamo un numero casuale u tra 0 e 1, e questo viene più piccolo di p_1 allora diciamo che si è realizzato il valore x_1 di X.

Se il numero u è compreso tra p_1 e p_2, diciamo che è uscito x_2 e così via.

In R:

```
gen.vc <- function(x, p)
        {x[min(which(cumsum(p) > runif(1)))]}
```

Generazione di variabili casuali (5/5)

Metodo dell'inversione

Esercizio: effettuare 1000 simulazioni con R utilizzando la funzione gen.vc e i dati della seguente tabella

x	p
-2	0.2
3	0.1
7	0.4
10	0.2
12	0.1

e visualizzare i risultati.

Suggerimenti:

1. può servirvi un ciclo for: for (i in 1:1000) {.....}
2. può servirvi definire un vettore y, i cui elementi y[i] sono igli output della funzione gen.vc
3. Per visualizzare i risultati potete graficare l'output della funzione table applicata a y.

Intervallo di confidenza per la media

```
# intro_script_3.r
# leggo i valori di x dal file "sample.dat"
x <- read.table("input_script_3.dat", header=FALSE)

# calcolo la varianza di x
s2 <- var(x[,1])

# calcolo la media di x
mx <- mean(x[,1])

n <- length(x[,1])

# calcolo il limite inferiore dell'intervallo
# "qt(0.975, ...)" calcola il quantile di ordine 0.975
# della normale a n-1 gradi di libertà
l.inf <- mx - qt(0.975, df=n-1) * sqrt(s2/n)

# calcolo il limte superiore dell'intervallo
l.sup <- mx + qt(0.975, df=n-1) * sqrt(s2/n)

# stampo in putput l'intervallo di confidenza
c(l.inf, l.sup)
```

File di input: input_script_3.dat

```
0.39
0.68
0.82
1.35
1.38
1.62
1.70
1.71
1.85
2.14
2.89
3.69
```

Esiste anche un metodo diretto per il calcolo dell'intervallo di confidenza, basta scrivere

```
t.test(x[,1])
```

"t.test" è una funzione che segue contemporaneamente un test di ipotesi (test) e il calcolo dell'intervallo di confidenza.

Per l'intervallo al 99%, basta scrivere:

```
t.text(x[,1], conf.lev=0.99)
```

La legge dei grandi numeri (1/2)

Qualunque sia il modello dei dati campionari, purchè si verifichi che

$$E[X_i] = \mu \quad \text{and} \quad Var[X_i] = \sigma^2$$

con X_i un campione di variabili i.i.d., per ogni $\epsilon > 0$ si ha che

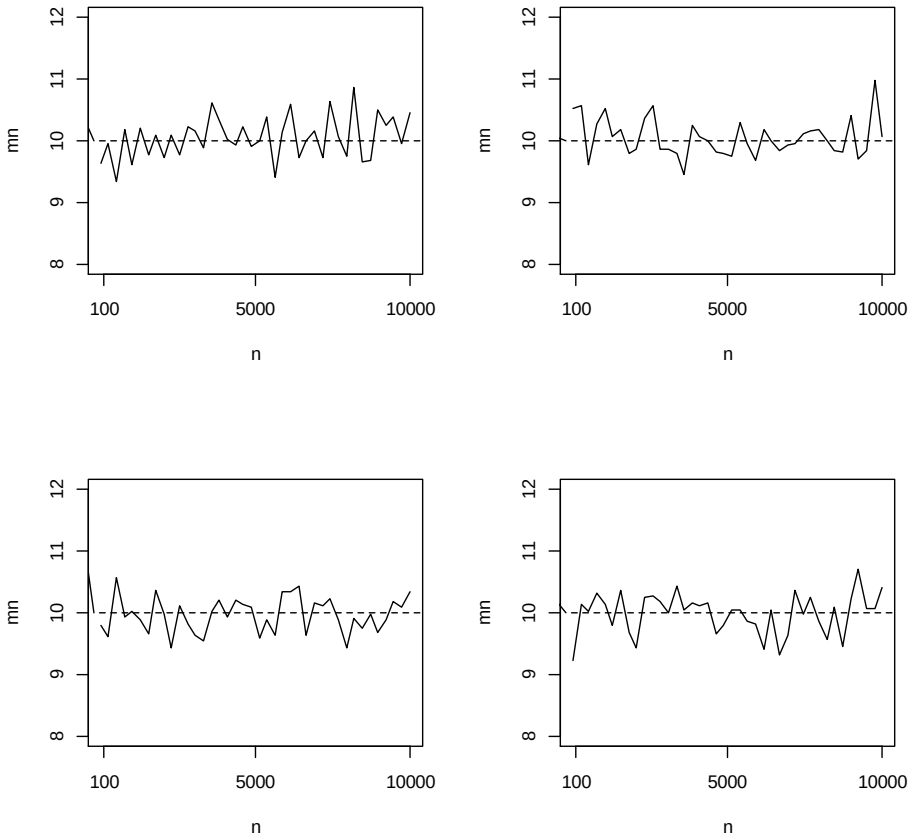
$$P(|\overline{X}_n - \mu| > \epsilon) \xrightarrow{n \to +\infty} 0$$

che significa: quando l'ampiezza campionaria è sufficientemente elevata, allora per quanto piccolo si possa scegleire ϵ, la probabilità che la media campionaria si trovi nell'intervallo $\mu \pm \epsilon$ tende a zero.

La legge dei grandi numeri (2/2)

Si lanci il seguente script e si commenti il risultato.

```
# intro_script_1.r
# divido l'area grafica in 4 sotto-aree (2x2)
par(mfrow=c(2,2))

# definisco il vettore n
n <- seq(10, 10000, length=40)

for (k in 1:4)
   {
    # definisco un vettore di lunghezza 40 e i cui elementi sono tutti uguali a 0
    mn <- numeric(40)

    for (i in 1:40)
       {
        # definisco gli elementi di mn usando la funzione come media
        # di rnorm(...), che genera n numeri random con media "mean"
        # errore standard sd
        mn[i] <- mean(rnorm(n, mean=10, sd=2))
        }
        #
        plot(n, mn, type="l", ylim=c(8, 12), xaxt="n")
        abline(h=10, lty=2)
        axis(1, c(100, 5000, 10000))
       }
```



Il teorema centrale del limite (1/2)

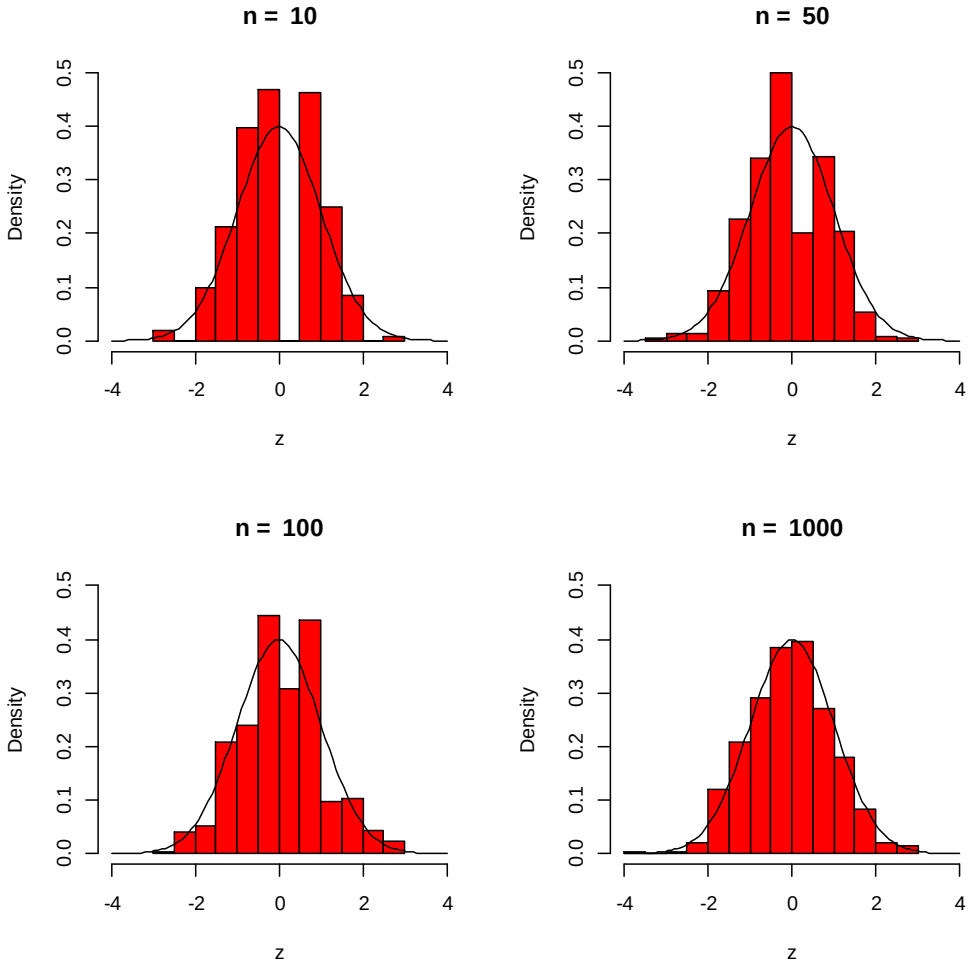
Preso un campione di variabili casuali i.i.d di media μ e varianza σ^2 , sotto condizioni molto generali sul modello probabilistico delle X_i si ha che

$$\frac{\overline{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \xrightarrow{n \to +\infty} \mathcal{N}(0,1)$$

Il teorema centrale limite (2/2)

Si lanci il seguente script e si commenti il risultato.

```
# intro_script_2.r
n <- c(10, 50, 100, 1000)

p <- 0.5

par(mfrow=c(2,2))

for (k in n)
  {
   mn <- numeric(500)
   for (i in 1:500)
      {
       x <- rbinom(k, 1, p)
       mn[i] <- mean(x)
      }

   z <- (mn - p)/sqrt(p*(1-p)/k)

   # disegno l'istogramma
   hist(z, freq=FALSE, ylim=c(0, pnorm(0)), xlim=c(-4, 4),
        col="red",   main=paste("n = ", k))
   curve(dnorm(x), -4, 4, add=TRUE)
  }
```



Intervallo di confidenza per le proporzioni

Se le X_i sono tutte bernoulliane di parametro p incognito, sappiamo che

$$\sum_{i=1}^{n} X_i \sim Bin(n, p)$$

Per la variabile casuale Binomiale vale l'approssimazione alla varibile casuale Guassiana se siamo in presenza di grandi campioni. Per il teorema centrale del limite si ricava che per n elevato:

$$Z = \frac{\hat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0.1)$$

Quindi l'interavllo di confidenza per p ha la seguente forma

$$p \in \left(\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right)$$

Intervallo di confidenza per le proporzioni

Metodo approssimato

Esercizio: alla chiusura del seggio, lo spoglio di n schede (che supponiamo rappresenative del totale delle schede, ha dato il seguente esito: per il SI il 51% e per il NO il 49%. Determinare l'intervallo di confidenza al 95% della percentuale di SI supponendo n=3000.

Soluzione: sappiamo che

$$\hat{p}_n = 51\% = 0.51$$

In R possiamo quindi scrivere

```
pn <- 0.51
n <- 3000
l.inf <- pn – qnorm(0.975) * sqrt(pn*((1-pn)/n)
l.sup<- pn + qnorm(0.975) * sqrt(pn*((1-pn)/n)
int_conf <- c(l.inf, l.sup)
int_conf
```

Intervallo di confidenza per le proporzioni

Metodo approssimato

In R troviamo già predisposto il comando prop.test, che fornisce come risultato un intervallo molto vicino a quello costruito con questo metodo.

```
> prop.test(1530, 3000)

        1-sample proportions test with continuity correction

data:  1530 out of 3000, null probability 0.5
X-squared = 1.1603, df = 1, p-value = 0.2814
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.4919437 0.5280305
sample estimates:
   p
0.51
```

Nota: 1530 è il 51% di 3000.

Intervallo di confidenza per le proporzioni

Metodo esatto

Con R possiamo anche seguire un test esatto utilizando la distribuzione Binomiale anzichè le sue approssimazioni asintotiche. La funzione da utilizzare è binom.test.

```
> binom.test(1530, 3000)

        Exact binomial test

data:  1530 and 3000
number of successes = 1530, number of trials = 3000, p-value = 0.2814
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.4919426 0.5280379
sample estimates:
probability of success
                  0.51
```

Intervallo di confidenza per le proporzioni

Esercizio: Sull abase dei risultati dell'esercizio precedente si calcoli la probabilità che il SI vinca con:

- n=3000
- n=2500
- n=2000
- n=1500
- n=1000
- n=500.

Soluzione: usiamo direttamente al variabile casuale Binomiale $Y = \sum_{i=1}^{n} X_i$.

Il SI vince se si raggiunge almeno la metà più uno dei voti, cioè da n/2 in poi. Ricordiamo poi che, nel metodo approssimato

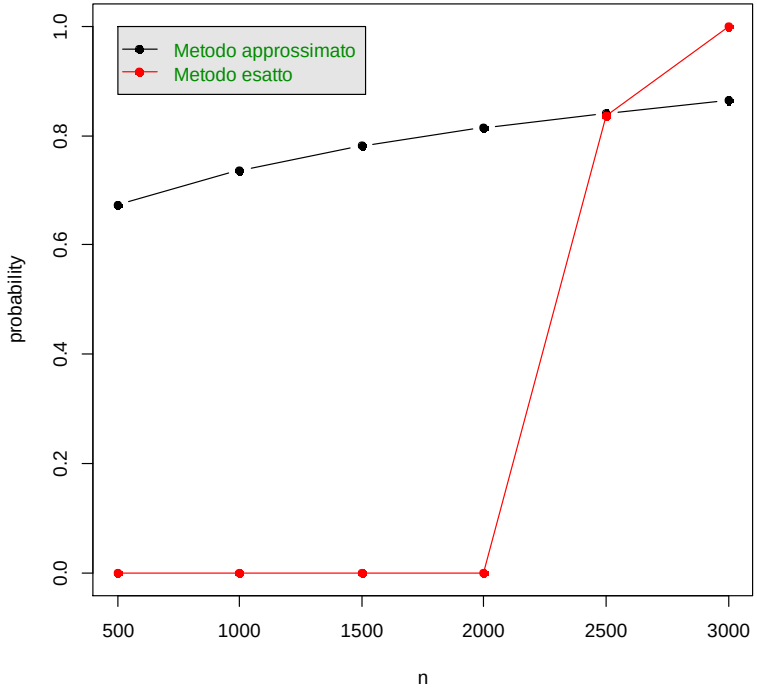
$$Z \sim \frac{Y - n\hat{p}_n}{\sqrt{n\hat{p}_n(1-\hat{p}_n)}}$$

$$P\left(Y > \frac{n}{2}\right) \approx P\left(Z > \frac{\frac{n}{2} - n\hat{p}_n}{\sqrt{n\hat{p}_n(1-\hat{p}_n)}}\right) = 1 - \Phi\left(\sqrt{n}\frac{0.5 - 0.51}{\sqrt{0.51 \times 0.49}}\right) = 1 - \Phi(-0.02\sqrt{n}) = \Phi(0.02\sqrt{n})$$

Intervalli di confidenza sulle proporzioni

Confronto tra metodo approssimato e metodo esatto

```
# intro_script_3.r
k <- 1
i <- 500
p_exact <- numeric(6)
p_approx <- numeric(6)
while (i <=3000)
    {
     # metodo approssimato
     p_approx[k] <- p1 <- pnorm(0.02*sqrt(i))
     # metodo esatto
     p_exact[k] <- pbinom(1250, i, 0.51, lower.tail=FALSE)
     #
     k <- k + 1
     i <- i + 500
    }

# grafico e confronto dei risulatti
n <- seq(500, 3000, 500)
plot(xn p_approx, xlab="n", ylab="probability", type="b", pch=19, ylim=c(0, 1))
points(n, p_exact, col="red", pch=19)
lines(n, p_exact, col="red")

legend(500, 1, c("Metodo approssimato", "Metodo esatto"), col = c(1, 2),
     text.col = "green4", lty = c(1, 1, 1), pch = c(19, 19),
     merge = TRUE, bg = "gray90")
```



Come si nota le probabilità calcolate con il metodo esatto sono inferiori a quelle calcolate con l'approssimazione Normale, in maniera tanto più evidente quanto più piccolo è n

Intervallo di confidenza sulla varianza

$$\sigma^2 \in \left(\frac{(n-1)s_n^2}{\chi^2_\alpha} \right)$$

dove s_n^2 è lo stimatore della varianza.

```
# intro_script_4.r
# Authors: S. M. Iacus, G. Masarotto

ic.var <- function(x, twosides=TRUE, conf.level)
          {
           alpha <- 1 - conf.level
           n <- length(x)
           if(twosides)
            {
             l.inf <- (n - 1) * var(x)/qchisq(1 - alpha/2, df = n -1)
             l.sup <- (n - 1) * var(x)/qchisq(alpha/2, df = n - 1)
            }
           else
            {
             l.inf <- 0
             l.sup <- (n - 1) * var(x)/qchisq(alpha/2, df = n - 1)
            }
           #
           c(l.inf, l.sup)
           }
```

Esempio:

```
# Genero 100 numeri random tra 0 e 10
> x <- rnorm(100, 10)

# applico la funzione ic.var
> ic.var(x, conf.level=0.95)
[1] 0.8237529 1.4420178
```

Stastistica descrittiva minimale

table
hist
plot(table (...))

Esempi

```
## Simple frequency distribution
table(rpois(100,5))

## Histogram
hist(rpois(100,5))

## Barplot
plot(table(rpois(100,5)))
```

Input/output da file e strutture dati

Per leggere da un file di inout input.txt

input.data <- read.table(“/path/input.txt”, ...)

Vettori

X <- array(0, n) # vettore di contenente n zeri
X <- 1:10 # vettore di numeri da 1 a 10

Matrici

M <- matrix(0, n, m) # matrice di 0 avente n righe ed m colonne

M <- matrix(rnomr(100, 10), 10, 10) # matrice 10x10 di 100 numeri random tra 0 e 10