LEZIONE 11 01/06/09 – parte 1ª

<u>Esercizio</u> <u>Studio di un sondaggio elettorale</u> in termini di problema di stima.

È necessario definire alcuni parametri:
-) come si fa un sondaggio?
-) quanto numeroso deve essere il campione?
-) la numerosità del campione deve essere proporzionale alle dimensioni del collegio?

Ipotizziamo di avere 100.000 elettori, x dei quali votano il partito A. Di conseguenza $100.000 - x$ votano diversamente.

In tal modo, se si riesce a prevedere i voti del partito A, si può applicare il procedimento a tutti gli altri casi.

Eseguiamo un'<u>operazione</u> di campionamento che consiste nel selezionare un <u>elettore</u> a caso e nel farsi dire che cosa ha votato ~~(exit poll)~~ o come voterà, l'intervista presuppone che la risposta sia sincera.

Si sta effettuando una prova di Bernoulli (si vota A oppure non si vota A) con probabilità di successo

$$\frac{x}{100.000}$$

Il problema consiste nello stimare la probabilità di successo a partire dai risultati di n prove di Bernoulli.

La soluzione consiste nel determinare la frequenza relativa

$$\hat{h} = \frac{k}{n}, \quad \text{con } k = \text{numero di successi}$$

La frequenza relativa è la media campionaria per la variabile casuale di Bernoulli.

Ciò che interessa davvero è l'intervallo di confidenza, senza il quale i risultati precedenti servirebbero a poco.

Definiamo

$S_n :=$ numero di successi su n prove di Bernoulli (variabile casuale binomiale) con $P(S_n = k) = \binom{n}{k} p^k q^{(n-k)}$ (pag. 31)

$E[S_n] = np$ $\quad Var[S_n] = np(1-p)$ (pag. 68)

ne consegue che, $\hat{p} = \frac{S_n}{n}$

$$E[\hat{p}] = \frac{E[S_n]}{n} = \frac{np}{n} = p$$

$E[\hat{p}]$ è una media campionaria non polarizzata.

$$Var[\hat{p}] = \frac{Var[S_n]}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Si tenga presente che, per il teorema centrale del limite $\hat{p}$ converge in distribuzione ad una v.c. gaussiana.

Si conclude che $\hat{p}$, asintoticamente, è

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Ciò corrisponde a considerare uno stimatore

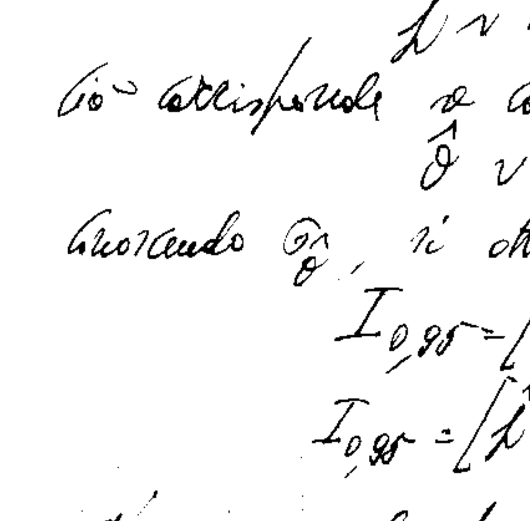
$$\hat{\theta} \sim N(\theta^0, \sigma^2_{\hat{\theta}}) \quad \text{(pag. 164)}$$

Conoscendo $\sigma_{\hat{\theta}}$, si ottiene

$$I_{0,95} = [\hat{\theta} - 1{,}96\,\sigma_{\hat{\theta}},\ \hat{\theta} + 1{,}96\,\sigma_{\hat{\theta}}]$$

$$I_{0,95} = \left[\hat{p} - 1{,}96\sqrt{\frac{p(1-p)}{n}},\ \hat{p} + 1{,}96\sqrt{\frac{p(1-p)}{n}}\right]$$

Non conoscendo p, non si conosce $\sigma_{\hat{\theta}}$



Studiamo la funzione $p(1-p) = p - 2p$

Derivando ed annullando si ottiene che il massimo si ha per

$$p = 0{,}5$$

e vale 0,25.

188 Poiché p è una probabilità $(0 \le p \le 1)$, risulta sempre
$$p(1-p) \le 0{,}25$$
Non potendo calcolare l'esatto intervallo di confidenza, ci mettiamo nelle peggiori condizioni, considerando
$$p = 0{,}5$$
Otteniamo
$$I_{0{,}95} = \left[\hat{p} - 1{,}96\sqrt{\frac{0{,}5(1-0{,}5)}{n}}, \hat{p} + 1{,}96\sqrt{\frac{0{,}5\times 0{,}5}{n}}\right] =$$
$$= \left[\hat{p} - 1{,}96\,\frac{0{,}5}{\sqrt{n}}, \hat{p} + 1{,}96\,\frac{0{,}5}{\sqrt{n}}\right]$$
Con l'approssimazione effettuata, si constata che nel 95% dei casi
$$|p - \hat{p}| \le \frac{1{,}96 \times 0{,}5}{\sqrt{n}} \cong \frac{1}{\sqrt{n}}$$
Ciò permette di determinare quanto deve essere esteso il sondaggio affinché il modulo dell'errore nella percentuale del partito, nel 95% dei casi sia inferiore ad ~~uno~~ un dato valore.

Ad esempio per avere un errore inferiore all'1%, si deve avere
$$\frac{1}{\sqrt{n}} \le 0{,}01$$
$$\sqrt{n} \ge \frac{1}{0{,}01} = 100 \qquad n \ge 10.000$$
Si può costruire la seguente tabellina per $\gamma = 0{,}95$

n	errore (%)
155	± 8%
200	± 7%
625	± 4%
2500	± 2%
10000	± 1%

Osservazioni:

-) per dimezzare l'errore occorre quadruplicare la numerosità del campione

-) la numerosità del collegio è ininfluente sulla precisione poiché quest'ultima non compare in alcuna formula.

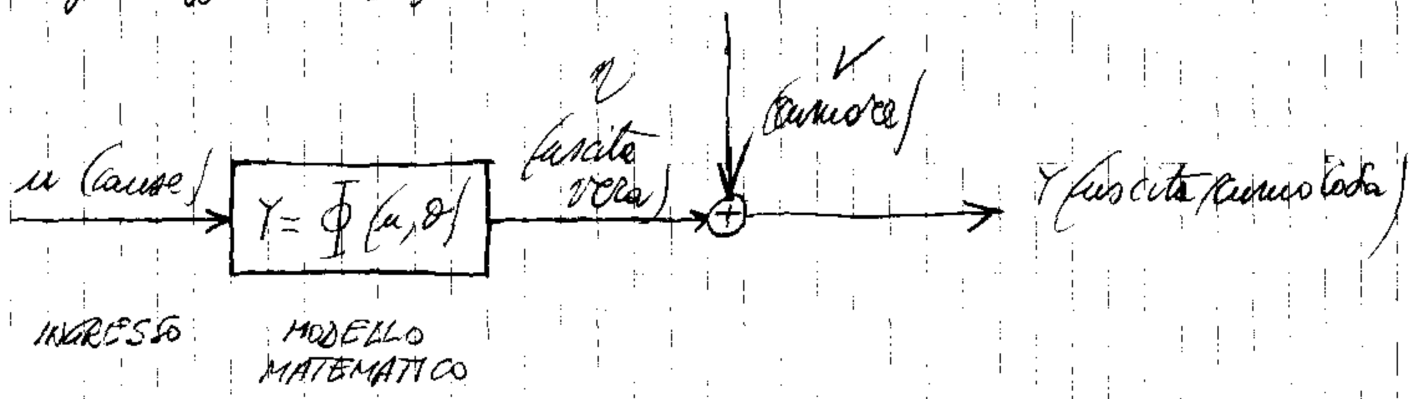
-) cambiando il sistema elettorale le cose si complicano

-) si possono ottenere risultati migliori sfruttando le distoraggiunta dai colleghi mediante strategie di raggruppamento stratificate per zone e suddividendo in sottogruppi.

IDENTIFICAZIONE DEI MODELLI STATICI

Il corso non si occupa dei modelli dinamici, in cui la variabile tempo entra in gioco in maniera rilevante, fornendo equazioni differenziali o equivalenti.

Utilizziamo un modello che lega le cause (u) agli effetti (Y).



Il modello matematico è un sistema di equazioni che permettono di calcolare gli effetti in funzione delle cause.

Il modello contiene ϑ che è un vettore che rappresenta i parametri del modello stesso.

L'uscita è la somma dell'uscita vera e del rumore.

Consideriamo come esempio la legge di Ohm.

$$I = \frac{V}{R}$$

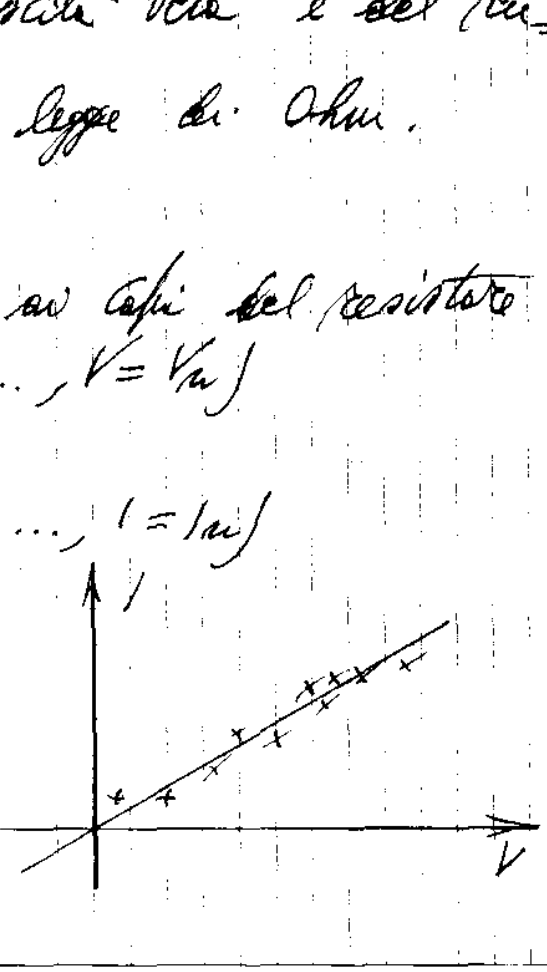
Imponiamo diversi valori di V ai capi del resistore

$$V \quad (V = V_1,\ V = V_2,\ \ldots,\ V = V_n)$$

e misuriamo la corrente

$$I \quad (I = I_1,\ I = I_2,\ \ldots,\ I = I_n)$$

I valori misurati dovrebbero stare su una retta ma ciò è impedito dall'errore di misura.



Definiamo il vettore u delle cause e quello Y delle ~~cau~~ correnti misurate.

$$u = \begin{bmatrix} V_1 \\ V_2 \\ \dots \\ V_n \end{bmatrix} \qquad Y = \begin{bmatrix} I_1 \\ I_2 \\ \dots \\ I_n \end{bmatrix}$$

Il valore I_i misurato conterrà un termine d'errore.

$$I_i = \frac{V_i}{R} + e_i \qquad i = 1, 2, \dots, n$$

Utilizziamo come parametro del modello

$$\vartheta := \frac{1}{R}$$

Ciò permette di scrivere il modello nel modo seguente

$$Y = \vartheta u + e \qquad (191)$$

$$\begin{bmatrix} I_1 \\ I_2 \\ \dots \\ I_n \end{bmatrix} = \frac{1}{R} \begin{bmatrix} V_1 \\ V_2 \\ \dots \\ V_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$$

Si vuol ottenere un modello che lega la corrente in uscita alla causa (tensione V), tramite il modello (legge di Ohm) ed il parametro del modello (R oppure $1/R$).

Abbiamo scritto n equazioni in forma vettoriale.

È opportuno scegliere $1/R$ e non R come parametro del modello per avere un modello lineare.

Infatti,

$$\Phi(u, \vartheta) = \vartheta u$$

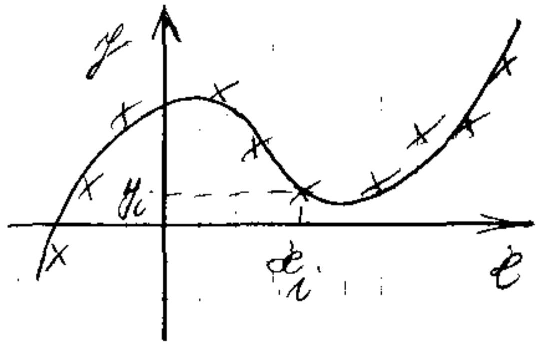
è lineare in ϑ.

Ponendo $\vartheta = R$, si ha

$$\Phi(u, \vartheta) = \frac{1}{\vartheta} u$$

ottenendo un modello non lineare nei parametri.

Avere un modello lineare in ϑ è vantaggioso; gli effetti sono legati alle cause da una formula che dipende linearmente da ϑ.



Esempio

Si raccolgono delle coppie di dati sperimentali

$$(x_i ; y_i)$$

con $i = 1, 2, \dots, n$

L'obiettivo consiste nel determinare una curva che passi per i punti che rappresentano le coppie di dati.

Con la curva e con la formula matematica si possono fare previsioni sui valori di y, dati i valori di x.

Supponiamo che la curva sia ricostruita mediante una cubica con il modello

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + e$$

con e = termine di errore.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \qquad \theta = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Scriviamo le equazioni

$$\begin{cases} y_1 = 1 \cdot a_0 + x_1 a_1 + x_1^2 a_2 + x_1^3 a_3 + e_1 \\ y_2 = 1 \cdot a_0 + x_2 a_1 + x_2^2 a_2 + x_2^3 a_3 + e_2 \\ \dots \\ y_n = 1 \cdot a_0 + x_n a_1 + x_n^2 a_2 + x_n^3 a_3 + e_n \end{cases}$$

Per ogni dato (crocetta nel grafico) viene un'equazione

In forma matriciale si ha

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$$

$$Y = \Phi(n) \times \theta + e$$

con $Y \in \mathbb{R}^{n \times 1}$ $\quad \Phi(u) \in \mathbb{R}^{n \times q}$

$\theta \in \mathbb{R}^{q \times 1}$ $\quad e \in \mathbb{R}^{n \times 1}$

Osservazione

Considerando la relazione (191) di pag. 191 si osserva che

$$\Phi(u,\theta) = \Phi(u) \cdot \theta$$

Tutte le volte che avviene ciò si dice che il modello è lineare nei parametri.

Stiamo approssimando dei dati con una cubica ed il modello non è lineare

I modelli matematici $\Phi(u,\theta)$ si appoggiano su una relazione matematica che dipende dalle cause ma dipende anche dai parametri del modello.

Un modello può essere lineare nel senso che gli effetti dipendono linearmente dalle cause, come nel caso della legge di Ohm.

Questo tipo di linearità interessa poco nei confronti della linearità che riguarda i parametri θ del modello.

Nel nostro caso il modello dipende linearmente dai parametri incogniti a_0, a_1, a_2, a_3 di conseguenza è un modello facile da identificare.

Una volta costruita $\Phi(u)$ si possono trascurare le u.

Definizione

Un modello si dice lineare nei parametri se

$$\Phi(u,\theta) = \Phi(u) \cdot \theta$$

con $\Phi(u)$ = matrice di sensibilità

-) Come identificare un modello lineare nei parametri?

1) metodo dei minimi quadrati

d'ora in poi sarà $q \leq n$, nel senso che il numero dei dati è $\geq$ al numero dei parametri.

L'obiettivo consiste nel cercare un θ tale che

$$Y = \Phi \cdot \theta$$

Si sta trascurando l'errore e si sta cercando di individuare un θ tale che i dati sperimentali siano perfettamente spiegati dal modello.

L'obiettivo non è generalmente raggiungibile nel caso pratico di

$$n > q$$

In pratica ci si accontenta che l'errore

$$\varepsilon := Y - \Phi\theta$$

sia "piccolo".

Di solito si cerca di minimizzare la somma dei quadrati degli scarti:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \|\varepsilon\|^2 = (Y - \Phi\theta)^T (Y - \Phi\theta)$$

Nota: un vettore è piccolo quando la sua norma si avvicina a zero.

-) Teorema (dovuto a Gauss) (1795)

Carl Friedrich Gauss (Braunschweig 1777 - Gottinga, 1855) matematico astronomo e fisico tedesco

Sia $\text{rank}(\Phi) = q$

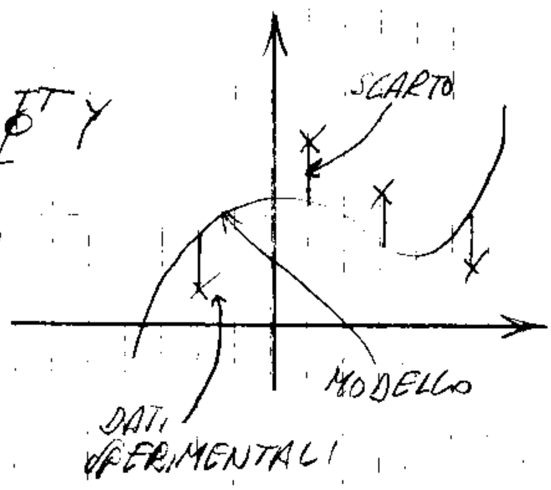
Allora

$$J(\theta) := \varepsilon^T \varepsilon = \|\varepsilon\|^2$$

ha un minimo globale per

$$\theta^{LS} = (\Phi^T \Phi)^{-1} \Phi^T Y$$

Nota: LS = Least Squares



Per "scarto" si intende la differenza tra il dato sperimentale e il modello

Il procedimento è implementabile in Matlab così!

```
ThetaLS = Phi \ Y
```

Identificazione dei Modelli e Analisi dei Dati

Esercitazione 1: esercizi
23 maggio 2007

Esercizio 1: distribuzione binomiale

Si considerino delle prove di Bernoulli con p = 0.5. Calcolare la distribuzione binomiale (ovvero la probabilità di avere k successi su n prove, al variare di k) per n = 2, 10, 100 prove (nota: l'istruzione `conv(a, b)` calcola la convoluzione di due sequenze di numeri `a` e `b`).

Soluzione:

```
f = [0.5 0.5]              % prob. di k = 0, 1 successi in 1 prova

f2 = conv(f,f)             % prob. di k = 0, 1, 2 successi in 2 prove

f10 = f;
for i = 1:9                % prob. di k = 0, 1, ..., 10 successi in 10 prove
   f10 = conv(f10, f);
end

f100 = f;
for i = 1:99               % prob. di k = 0, 1, ..., 100 successi in 100 prove
   f100 = conv(f100, f);
end

figure(1); plot([0 : 2], f2, '-*'); grid;
figure(2); plot([0 : 10], f10, '-*'); grid;
figure(3); plot([0 : 100], f100, '-*'); grid;
```

Quale risultato teorico giustifica l'andamento dei grafici al crescere di n?

Esercizio 2: media campionaria

Si consideri una sequenza X_k, k = 1..N, di variabili casuali indipendenti e identicamente distribuite in modo uniforme in [0, 1], e la loro media campionaria

$$M(N) = \frac{1}{N}\sum_{k=1}^{N} X_k$$

a. Fissato N, estrarre N campioni in modo uniforme (ad esempio, col comando `rand(1, N)`) e calcolarne la media campionaria. Ripetere l'operazione ad esempio 1000 volte, memorizzando ogni volta la media campionaria risultante all'interno di un vettore:

```
N = 200
for u = 1 : 1000
   M(u) = sum(rand(1, N)) / N;
end
```

Queste istruzioni creano automaticamente un vettore M di 1000 elementi. Il valore atteso e la varianza dello stimatore media campionaria M(N) (N fissato) possono quindi essere calcolate con le istruzioni

```
mean(M)
var(M)
```

Usando i dati del vettore M posso costruire un istogramma, il quale approssima la ddp dello stimatore. Ci si potrà aspettare quindi che l'istogramma presenti un valor medio e una varianza simili a quelli appena ricavati:

```
hist(M)
```

b. Ora si provi ad aumentare N di 10 volte, ad esempio, e ripetere le operazioni. Che comportamento ci si può aspettare da media e varianza? Cosa ci si può aspettare che succeda aumentando ancora N? Qual'è il principio teorico alla base di questi risultati?

c. Si confrontino gli istogrammi ottenuti con N = 2 e con un N grande (5 o più). Come varia la forma dell'istogramma? Quale principio teorico giustifica tale risultato?

Esercizio 3: il valore di π

Utilizzando l'istruzione per generare V.C. i.i.d uniformi in [0, 1] (`rand()`), sviluppare un algoritmo per il calcolo approssimato di π.

<u>Suggerimento</u>: si consideri l'esperimento casuale consistente nell'estrazione di un punto scelto in modo equiprobabile nel quadrato $0 \leq x \leq 1$, $0 \leq y \leq 1$. Si consideri il cerchio di diametro 1 all'interno di tale quadrato, e si osservi che ogni punto estratto in modo casuale può cadere dentro o fuori dal cerchio...

Esercizio 1 - soluzione

- f = [0.5 0.5] crea il vettore f delle probabilità
- f2 = conv[f, f] calcola la convoluzione della sequenza f con se stessa.

Il risultato è 0,25 0,5 0,25

Esempio pratico: il vettore f potrebbe contenere i coefficienti di due polinomi da moltiplicare. Il risultato fornisce i coefficienti del polinomio risultante

$$(0.5x + 0.5)(0.5x + 0.5) = 0.25x^2 + 0.5x + 0.25$$

Il vettore f potrebbe rappresentare le probabilità di uscita dei due valori di una moneta.

f2 rappresenta le probabilità di uscita dei valori di due prove di Bernoulli.

0.25 P di uscire due T

0,50 P di uscire una T e una C

0,25 P di uscire due C

```
- f10 = f;
- for i = 1:9
    - f10 = conv(f10, f);
- end
```

Come prima ma vengono eseguite 10 prove di Bernoulli.

Il risultato fornisce la distribuzione binomiale che ricorda un po' la gaussiana.

```
- f100 = f
- for i = 1:99
    - f100 = conv(f100, f);
- end
```

196 come precedentemente ma con 100 prove di Bernoulli.

Risulta più evidente la forma gaussiana. In realtà non si può avere una vera gaussiana poiché la d.d.p. è discreta ed è formata da tante Delta di Dirac.

- figure (1); plot([0:2], f2, '-*'); grid
- figure (2); plot([0:10], f10, '-*'); grid
- figure (3); plot([0:100], f100, '-*'); grid

I disegni vengono rappresentati in tre figure diverse.

Il primo parametro del comando plot rappresenta le etichette dell'asse delle ascisse.

Esercizio 2 - soluzione

- N = 200 200 v.c. uniformi in [0, 1]
- for n = 1 : 1000
  - M(n) = sum(rand(1, N)) / N
- end

Si estraggono 200 valori e si calcola la media campionaria.

L'operazione è ripetuta 1000 volte, memorizzando ogni media entro un vettore.

Lo scopo consiste nel visualizzare la legge dei grandi numeri.

Per N = 200 la d.d.p. sta tra 0,43 e 0,56.

Per N = 1000 sta tutta tra 0,47 e 0,53

Con N ancora maggiore la d.d.p. si stringe ancora di più e tende ad una Delta di Dirac.

Per inserire il valore di N da tastiera:

- N = input('N=? ')
- figure crea ogni volta una nuova figura per poter conservare i grafici precedenti

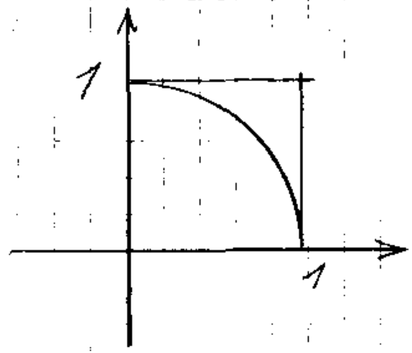
~~- axis([.4 .6~~

Aggiungendo il comando axis dopo hist si blocca la scala degli assi del grafico

- axis([.4 .6 0 250])

L'asse X va da 0,4 a 0,6 e quello delle Y da 0 a 250.

Esercizio 3 - soluzione



Si estrae casualmente un punto nel quadrato e si verifica se esso cade anche all'interno del quarto di cerchio di raggio unitario.

L'area del quadrato vale

$$A_Q = 1$$

L'area del quadrante di cerchio vale

$$A_C = \frac{\pi \cdot 1^2}{4} = \frac{\pi}{4}$$

Si effettua in pratica una prova di Bernoulli con probabilità di successo

$$p = \frac{A_C}{A_Q} = \frac{\pi}{4}$$

Lo stimatore $\hat{p}$ è dato da

$$\hat{p} = \frac{\text{numero di successi}}{\text{numero di tentativi}} = \frac{\pi}{4}$$

Perciò: $\pi \cong 4 \dfrac{\text{num. di successi}}{\text{num. di tentativi}}$

Il codice può essere il seguente

- p=0; inizializza la variabile p per il calcolo della probabilità

- N=input('N=! ') è richiesto il numero di prove

- for i=1:N

198 - x = rand (1,1)

- y = rand (1,1)

Viene generato un punto a caso nel quadrato

- h = h + (((x^2 + y^2)^(.5) ≤ 1);

- end

- h = h / n * 4

Se la condizione $\sqrt{x^2+y^2} \leq 1$ è verificata, il numero dei successi è incrementato di una unità