RATIO MATH. 1.
(1990), 61 - 72

# Interpretazione dei risultati di un campionamento statistico con approccio bayesiano: simulazione su calcolatore. ($)

**Vittorio Colagrande (*)** **Giuseppe Di Biase (**)**
*(*) Istituto Tecnico Commerciale "Galiani", via Ricci 22 - Chieti*
*(**) Dipartimento di Scienze e Storia dell'Arch., viale Pindaro 42 - Pescara*

## 1. INTRODUZIONE AL PROBLEMA

Uno dei problemi dell'inferenza statistica è quello di valutare l'effettiva operatività di ipotesi relative a caratteristiche numeriche incognite di una popolazione a seguito di esperimenti.

L'impostazione bayesiana dell'inferenza permette di utilizzare tutta l'informazione riguardo alle ipotesi in quanto tiene conto di una distribuzione di probabilità iniziale delle ipotesi stesse ed ottiene, a seguito di prove sperimentali, delle distribuzioni che sono sintesi del supporto sperimentale e delle conoscenze preliminari.

Se accade, come nell'esempio qui realizzato, che la distribuzione "a posteriori" ottenuta dopo le prove sperimentali, è dello stesso "tipo" di quella assunta all'inizio, l'effetto dell' "apprendimento dall'esperienza" si riflette sulla distribuzione dei parametri.

($) Lavoro eseguito con il contributo del C.N.R.
(contratto n. 88.00307.01)

Nel seguito si presenta la risoluzione di un problema concreto di *verifica di ipotesi* nel campo della sicurezza delle costruzioni, simulando su calcolatore le prove sperimentali e il conseguente aggiornamento della distribuzione iniziale del parametro considerato.

Uno dei parametri essenziali nella valutazione della sicurezza delle costruzioni è la *resistenza caratteristica* del calcestruzzo.

La definizione del valore di tale grandezza è, però, affetta da notevoli fonti di aleatorietà ed incertezza, riconducibili essenzialmente alle seguenti:

a) *aleatorietà intrinseca,* legata alla variabilità spaziale delle caratteristiche meccaniche del materiale per effetto di alterazioni casuali intervenute durante la fattura originaria ed all'incompletezza delle indagini conoscitive preliminari;

b) *incertezza legata agli errori di misura.*

## 2. LA METODOLOGIA BAYESIANA

Sia X il numero aleatorio rappresentante la resistenza caratteristica del calcestruzzo.

Si attribuisce ad X distribuzione normale con valor medio incognito e varianza assegnata $\sigma^2$.

Sia $\Theta$ il numero aleatorio rappresentante il valor medio (incognito) di X.

Le "ipotesi" di cui ci proponiamo di calcolare la probabilità in seguito ad osservazioni statistiche sperimentali sono eventi che dipendono dalla distribuzione di $\Theta$.

Seguendo l'approccio bayesiano, si assume per $\Theta$ una *distribuzione iniziale* che viene "*aggiornata*" più volte in base ai risultati di prove di schiacciamento su carote di materiale prelevato dalla struttura; le misure ottenute da tali prove sono determinazioni del numero aleatorio X.

Uno dei vantaggi derivanti dalla metodologia inferenziale di tipo bayesiano è che la *distribuzione iniziale* può essere definita in base ad informazioni sia di tipo qualitativo (manufatti simili situati nelle vicinanze) sia di tipo quantitativo (dati di progetto o risultati di provini presi all'epoca del getto).

É possibile, in molti casi, assumere per $\Theta$ una *distribuzione iniziale normale* con valor medio $\mu_0$ e scarto quadratico medio $\sigma_0$ :

$$N_{\mu_o, \sigma_o}(\vartheta)$$

Sia $x = (x_1, x_2, ..., x_n)$ il campione osservato e $\bar{x}$ la media campionaria:

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

Si dimostra (cfr. [8], [9]) che:

- la funzione di verosimiglianza del campione x condizionata a $\Theta$ è una distribuzione normale con varianza $\sigma^2/n$ :

$$\alpha(x_1, x_2, ..., x_n / \vartheta) = \prod_{i=1}^{n} N_{\vartheta,\sigma}(x_i) = N_{\vartheta,\sigma/\sqrt{n}}(\bar{x})$$

– la distribuzione di probabilità "aggiornata" di $\vartheta$ condizionata al campione x (utilizzando il Teorema di Bayes) è ancora normale:

(1) $$\beta(\vartheta / x_1, x_2, ..., x_n) = N_{\mu_n, \sigma_n}(\vartheta).$$

I parametri di tale distribuzione sono dati da:

$$\mu_n = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\bar{x}\,n}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}\ ; \qquad \frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}$$

## 3. OSSERVAZIONI

Con il metodo descritto, la distribuzione iniziale attribuita al valor medio della resistenza caratteristica di progetto viene controllata tramite i risultati dei prelievi effettuati sul manufatto.

Combinando i dati preesistenti, che possono essere anche informazioni qualitative esprimenti il *grado di fiducia* personale sul valore della resistenza, con i dati aggiuntivi di tipo quantitativo e statistico, si riesce ad ottenere una distribuzione di probabilità più attendibile.

Come è naturale, il peso da attribuire al dato $\bar{x}$ dell'esperimento nella valutazione del valor medio $\mu_n$ è tanto maggiore quanto più tale dato è relativo ad un numero aleatorio con distribuzione di probabilità poco dispersa e, analogamente, quanto più precisa è la determinazione iniziale del valor medio $\mu_o$ di $N_{\mu_o,\sigma_o}(\vartheta)$, tanto maggiore è il "peso" da attribuirgli.

*All'aumentare della numerosità* del campione, i dati osservati influenzano i parametri della distribuzione finale di $\Theta$ in modo sempre più preponderante rispetto alla distribuzione iniziale; precisamente *al crescere di n* la distribuzione finale di $\Theta$ tende a concentrarsi intorno al valore $\mu_n$, con la corrispondente varianza $\sigma_{n2}$ tendende a zero.

In conclusione, tenendo conto della forma della curva normale e dell'osservazione precedente, la (1) rende possibile nella pratica interpretare il numero $\mu_n$ come sintesi dell'intera distribuzione di $\Theta$, per cui il numero $\mu_n$ individuato può rappresentare il valore da attribuire con maggior fiducia alla resistenza specifica del calcestruzzo.

## 4. SIMULAZIONE SU CALCOLATORE

Inizialmente si attribuisce a $\Theta$ una distribuzione normale introducendo in input i valori dei parametri caratteristici.

In base alla distribuzione inziale assunta per $\Theta$ siamo praticamente certi che i valori di tale variabile sono compresi fra due numeri reali a e b.. Per la loro determinazione si rimanda a [8].

Nel fare la simulazione abbiamo fissato un intervallo $(c, d) \supset (a, b)$ in modo da aumentare l'incertezza iniziale.

Il modo di procedere è il seguente:
si sceglie una distribuzione normale "vera" di X con scarto quadratico medio $\sigma$ assegnato in input e valor medio:

(2) $$\vartheta_r = (c+d)/2 + r\,(d-c)/6$$

essendo r un numero pseudocasuale con distribuzione normale standardizzata (cfr. [5] ) estratto dal calcolatore.

La (2) si giustifica tenendo presente che $\vartheta_r$ è un valore della variabile $\Theta$ e che quasi certamente quest'ultima è compresa nell'intervallo (c, d) (cfr. [8]).

In tal modo viene simulata una possibile distribuzione di probabilità della resistenza caratteristica del calcestruzzo.

Le prove di schiacciamento di carote vengono simulate mediante estrazione al calcolatore di successioni di numeri pseudocasuali con *distribuzione normale* di parametri $\vartheta_r$ e $\sigma$.

Si possono considerare, ad esempio, (come prescrivono le normative tecniche) tre campioni di numerosità quattro.

## 5. UN TENTATIVO DI DECISIONE

Due maniere significative di utilizzare i risultati ottenuti, allo scopo di arrivare a *decisioni statistiche*, sono le seguenti:

- determinare un intervallo di confidenza $[\vartheta_1, \vartheta_2]$ di $\Theta$, tale, cioè, che:

$$\text{Prob}\,(\vartheta_1 \leq \Theta \leq \vartheta_2) = \gamma$$

essendo $\gamma$ un valore assegnato in $(0, 1)$.

In sostanza, si vogliono ottenere due valori entro cui il valor medio della resistenza caratteristica del calcestruzzo è compresa con un certo livello di probabilità $\gamma$.

- Fissare un *valore di collasso* T e determinare la

$$\text{Prob}(\vartheta \leq T)$$

## 6. APPENDICE

Si allega un esperimento effettuato sul calcolatore ed il programma scritto in linguaggio Turbo Pascal ver. 5.0

```
Introdurre i parametri della distribuzione normale
iniziale del valor medio della resistenza caratteristica
del calcestruzzo:

      valor medio                 mo = 280
      scarto quadratico medio     so = 40


Scarto quadratico medio della resistenza caratteristica
del calcestruzzo: 25
```

estrazione al calcolatore di successioni di numeri pseudocasuali con *distribuzione normale* di parametri $\vartheta_r$ e $\sigma$.

Si possono considerare, ad esempio, (come prescrivono le normative tecniche) tre campioni di numerosità quattro.

## 5. UN TENTATIVO DI DECISIONE

Due maniere significative di utilizzare i risultati ottenuti, allo scopo di arrivare a *decisioni statistiche*, sono le seguenti:

- determinare un intervallo di confidenza $[\vartheta_1, \vartheta_2]$ di $\Theta$, tale, cioè, che:

$$\text{Prob}\,(\vartheta_1 \le \Theta \le \vartheta_2) = \gamma$$

essendo $\gamma$ un valore assegnato in $(0, 1)$.

In sostanza, si vogliono ottenere due valori entro cui il valor medio della resistenza caratteristica del calcestruzzo è compresa con un certo livello di probabilità $\gamma$.

- Fissare un *valore di collasso* T e determinare la

$$\text{Prob}(\vartheta \le T)$$

## 6. APPENDICE

Si allega un esperimento effettuato sul calcolatore ed il programma scritto in linguaggio Turbo Pascal ver. 5.0

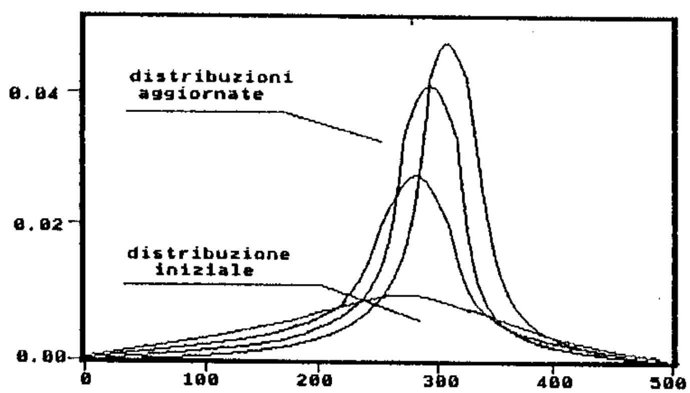
```
Introdurre i parametri della distribuzione normale
iniziale del valor medio della resistenza caratteristica
del calcestruzzo:

      valor medio                  mo = 280
      scarto quadratico medio      so = 40


Scarto quadratico medio della resistenza caratteristica
del calcestruzzo: 25
```

```
Numerosità del oampione  n = 4
319   313   312   323

Numerosità del campione  n = 4
317   320   310   317

Numerosità del campione  n = 4
319   320   319   324

Introduci un livello di confidenza o un valore di collasso ?
Livello di confidenza : 0.95
```

Con i dati di input dell'esperimento, in seguito al campionamento, si ottengono i seguenti risultati:

| | valor medio | scarto quadratico medio |
|---|---|---|
| $1^o$ campionamento | 288 | 12 |
| $2^o$ campionamento | 303 | 9 |
| $3^o$ campionamento | 308 | 7 |

ed i seguenti grafici:




L'intervallo di confidenza ottenuto è il seguente:

$$\text{Prob } (303 \leq \Theta \leq 332) = 0.95$$

**Program Inferenza_bayesiana;**

```
{$N+}
{$IFDEF CPU87E+}
{$E-}
{$ELSE}
{$N+,E+}
{$ENDIF}

uses crt,graph,Integrat,radici;

const
  Num_Aleatorio = 'Resistenza Caratteristica del Calcestruzzo';
  Int = 100;
  rimp = 25;
  Interv = 3;
  xmin = 0;xmax = 20;ymin = -1;ymax = 1.5;
type
  reali = extended;

var
   a1,a2                          :integer;
   a,b,si,gamma,Ta,mr             :reali;
   m0,mN,s0,sN                    :reali;
   ALT,LARG                       :word;
   val_medio,scarto               :reali;
   sx,sy,k,incrx                  :reali;
   drivergraf,modograf,Errcodice  :integer;
   fx,fy,hx,hy                    :integer;
   flag                           :boolean;
   p                              :pointer;
   ch                             :char;


Function F(x,m,s:reali):reali;
begin
   F: = exp(-(x-m)*(x-m)/(2*s*s))/(s*sqrt(2*Pi));
end;


Procedure GRAFICA;
begin
   DriverGraf: = Detect;
   InitGraph(DriverGraf,ModoGraf,'');
   ErrCodice: = GraphResult;
   if Errcodicegrok then
     begin
        writeln('Errore grafica : ',GraphErrorMsg(Errcodice));
        writeln('Il programma abortisce...');Halt(1);
     end;
end;


Procedure PARAMETRI;
begin
   LARG: = GetMaxX;ALT: = GetMaxY;
   sx: = LARG/abs(xmax-xmin);
   sy: = ALT/abs(ymax-ymin);
   fx: = round(abs(xmin*sx));
   fy: = round(abs(ymin*sy));
   k: = 1;
   hx: = round(k*sx);hy: = round(k*sy);
   incrx: = 1/sx;
end;
```

```
Procedure ASSI;
var
 sinx,alty,desx,basy,i :integer;
 nu                    :string;
begin
   line(0,ALT-fy,LARG,ALT-fy);
   line(fx,0,fx,ALT);
   sinx:=fx;desx:=fx;
   alty:=fy;basy:=fy;
   i:=1;
   while (sinx0) or (desxARG) do
      begin
         sinx:=sinx-hx;
         desx:=desx+hx;
         line(sinx,ALT-fy-2,sinx,ALT-fy+2);
         line(desx,ALT-fy-2,desx,ALT-fy+2);
         str(25*i,nu);
         outtextxy(desx-7,ALT-fy+10,nu);
         i:=i+1
      end;
   i:=1;
   while  (alty0) or (basyALT) DO
      begin
         alty:=alty-hy;
         basy:=basy+hy;
         line(fx+3,ALT-alty,fx-3,ALT-alty);
         line(fx+3,ALT-basy,fx-3,ALT-basy);
         str(i/25:2:2,nu);
         outtextxy(fx+10,ALT-basy-3,nu);
         i:=i+1
      end;
end;


Procedure   TRACCIA_FUNZIONE(Val_medio,scarto:reali);
var
 px,py,pxprec,pyprec :integer;
 x,y                 :reali;
begin
   x:=xmin;
   pxprec:=round(x*sx/rimp)+fx;
   y:=rimp*F(x,Val_medio,scarto);
   pyprec:=round(y*sy)+fy;
   repeat
     y:=rimp*F(x,Val_medio,scarto);
     px:=round(x*sx/rimp)+fx;
     py:=round(y*sy)+fy;
     putpixel(px,ALT-py,15);
     line(pxprec,ALT-pyprec,px,ALT-py);
     x:=x+rimp*incrx;
     pxprec:=px;pyprec:=py;
   until xrimp*xmax
end;


Procedure   MEMORIZZA_GRAFICO(x1,y1,x2,y2:word;var punt:pointer);
var
   lunghezza :word;
begin
   lunghezza:=imagesize(x1,y1,x2,y2);
   getmem(punt,lunghezza);
   getimage(x1,y1,x2,y2,punt^);
end;
```

```
Procedure    RICHIAMA_GRAFICO(x1,y1,x2,y2:word;var punt:pointer);
var
  lunghezza :word;
begin
   PutImage(x1,y1,punt^,1);
   lunghezza: = imagesize(x1,y1,x2,y2);
   freemem(punt,lunghezza);
   p: = nil
end;


Procedure    FIN_TESTO;
begin
   window(0,0,20,20);
   RestoreCrtMode;
end;


Function    RAN1: reali;
var
  t:reali;
begin
   t: = (a1*32749 + 3) mod 32749;
   a1: = trunc(t);
   Ran1: = abs(t/32749)
end;


Function    RAN2: reali;
var
  t:reali;
begin
   t: = (a2*10001 + 3) mod 17417;
   a2: = trunc(t);
   Ran2: = abs(t/17417)
end;


Function    SCELTARAN: reali;
var
  f:reali;
begin
   f: = Ran2;
   if f0.5 then SceltaRan: = Ran1
          else  begin Randomize;SceltaRan: = Random end
end;


Procedure    RULLA;
var
  i  :Integer;
  sc:reali;
begin
   a1: = 1;a2: = 203;
   for i: = 1 to 10 do sc: = SCELTARAN
end;


Function    CASNORM:reali;
var
  somma,sc:reali;
  i           :Integer;
begin
   somma: = 0;
   for i: = 1 to 12 do
     begin
        sc: = SCELTARAN;
```

```
        somma: = somma + sc
     end;
   CASNORM: = (somma-6)/12
end;


Procedure    PAR_INIZ;
var
  ch:char;
begin
    clrscr;
    gotoxy(1,3);
    write('Intrudurre i parametri della distribuzione normale iniziale del');
    gotoxy(1,5);write('valor medio della ',Num_Aleatorio,' : ');
    gotoxy(10,8);write('valor medio             m0 =  ');read(m0);
    gotoxy(10,10);write('scarto quadratico medio  s0 =  ');read(s0);
    gotoxy(1,13);
    write('Scarto Quadratico Medio di ',Num_Aleatorio,' : ');
    read(si);
    repeat
      gotoxy(1,24);write('Vuole introdurre un "Livello di confidenza"  o ',
                         'un Valore di collasso ? (L/V) ');
      ch: = readkey;ch: = upcase(ch)
    until ch in ['L','V'];
    gotoxy(1,15);
    if ch = 'L'then begin
                      write('Livello di confidenza  : ');readln(gamma);flag: = true
                    end
              else  begin
                      write('Valore di collasso : ');readln(Ta);flag: = false
                    end;
    a: = m0-interv*s0;b: = m0 + interv*s0;
end;


Procedure    SCELTA;
var
   c,d,sc,cas :reali;
begin
   RULLA;
   sc: = SCELTARAN;c: = int*sc + a-int;
   sc: = SCELTARAN;d: = int*sc + b;
   cas: = CASNORM;
   mr: = (d-c)*cas/6 + (d + c)/2;
end;


Procedure    CAMPIONE;
var
   cas,x,media,va:reali;
   N,I               :integer;
begin
    clrscr;
    gotoxy(3,3);write('Numerosit del campione  N =  ');readln(N);
    writeln;
    media: = 0;
    for I: = 1 to N do
      begin
        cas: = CASNORM;
        x: = si*cas + mr;
        write(x:8:2);
        delay(300);
        media: = media + x/N
      end;
    readln;
    sn: = 1/(s0*s0) + N/(si*si);
```

```
   mn: = (m0/(s0*s0) + N*media/(si*si))/sn;
   sn: = 1/sqrt(sn);
   m0: = mn;s0: = sn;
end;

Procedure   INTERVALLO;
var
   epsi:reall;
begin
   epsi: = sn*soluz4((gamma + 1)/2);gotoxy(10,10);
   write('Intervallo di confidenza  :  (',mn-epsi:3:3,' , ',mn + epsi:3:3,' )')
end;


Procedure   COLLASSO;
var
   Prob:reall;
begin
   Prob: = simpson(-50,(Ta-mn)/sn);
   gotoxy(5,10);writeln('La probabilità che la ',Num_Aleatorio);
   gotoxy(5,12);write('sia insufficiente  : ');write(Prob:3:3)
end;


Procedure   TASTO;
begin
   outtextxy(0,alt-10,' ESC per finire ');
   ch: = readkey;ch: = upcase(ch)
end;


{Programma  Principale}

begin
   clrscr;textcolor(15);
   PAR_INIZ;
   SCELTA;
   GRAFICA;
   PARAMETRI;
   ASSI;
   TRACCIA_FUNZIONE(m0,s0);
   readln;
   repeat
     MEMORIZZA_GRAFICO(0,0,larg,alt,p);
     FIN_TESTO;
     CAMPIONE;
     GRAFICA;
     RICHIAMA_GRAFICO(0,0,larg,alt,p);
     TRACCIA_FUNZIONE(m0,s0);
     TASTO
   until ch = chr(27);
   RestoreCrtMode;
   if flag then INTERVALLO else COLLASSO;
   repeat until keypressed
end.
```

Il programma utilizza due **unit**:

**integrat**, una cui function (chiamata simpson) è utilizzata nella procedura COLLASSO per calcolare con il metodo di Simpson l'integrale tra $-\infty$ e $\frac{T-\mu_n}{\sigma_n}$ della funzione:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

**radici**, una cui function (chiamata soluz4) è utilizzata nella procedura COLLASSO per il calcolo approssimato delle radici dell' equazione che si ricava imponendo che $\text{Prob}(\vartheta_1 \leq \Theta \leq \vartheta_2) = \gamma$.

Di tali **unit** si omette, per brevità, il listato.

## BIBLIOGRAFIA

[1] N. CERA - A. MATURO (1990) Analisi della bontà di alcuni generatori di numeri pseudocasuali per la cifratura di messaggi e la simulazione. Ratio Math. 2, in corso di stampa.

[2] M. CIAMPOLI (1989) Reability Evaluation of Existing Structures: Updating Techniques to account for Experimental Data. Atti della International conference on "Monitoring, Snerveillance and Predective Maintenance of Plants and Structures", Taormina, 16-18 Ottobre 1989, in corso di stampa.

[3] V. COLAGRANDE - G. DI BIASE (1990) Simulazione su calcolatore di alcuni esempi di applicazione del Teorema di Bayes, accettato sul Periodico di Matematiche, in corso di stampa.

[4] A.MATURO (1990) Numeri Pseudocasuali. Libreria dell'Università, Pescara.

[5] L.PICCINATO-N.PINTACUDA (1985) Probabilità e Statistica. Quaderni C.N.R.

[6] H.SCHILD (1988) Turbo Pascal 4.0. Programmazione avanzata. Mc.Graw-Hill.

[7] R.SCOZZAVAVA (1988) Probabilità soggettiva e grandi rischi. Archimede N.40.

[8] R.SCOZZAFAVA (1981) Introduzione alla probabilità e alla statistica. Ed. Veschi, Roma.

[9] R.SCOZZAFAVA (1989) La probabilità soggettiva e le sue applicazioni. Ed. Masson, Milano.

Ringraziamo sentitamente il Prof. Romano Scozzafava per l'incoraggiamento e gli utili suggerimenti fornitici.

Il programma utilizza due unit:

**integrat**, una cui function (chiamata simpson) è utilizzata nella procedura COLLASSO per calcolare con il metodo di Simpson l'integrale tra $-\infty$ e $\frac{T-\mu_n}{\sigma_n}$ della funzione:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

**radici**, una cui function (chiamata soluz4) è utilizzata nella procedura COLLASSO per il calcolo approssimato delle radici dell' equazione che si ricava imponendo che $\text{Prob}(\vartheta_1 \leq \Theta \leq \vartheta_2) = \gamma$.

Di tali **unit** si omette, per brevità, il listato.

## BIBLIOGRAFIA

[1] N. CERA - A. MATURO (1990) Analisi della bontà di alcuni generatori di numeri pseudocasuali per la cifratura di messaggi e la simulazione. Ratio Math. 2, in corso di stampa.

[2] M. CIAMPOLI (1989) Reability Evaluation of Existing Structures: Updating Techniques to account for Experimental Data. Atti della International conference on "Monitoring, Snerveillance and Predective Maintenance of Plants and Structures", Taormina, 16-18 Ottobre 1989, in corso di stampa.

[3] V. COLAGRANDE - G. DI BIASE (1990) Simulazione su calcolatore di alcuni esempi di applicazione del Teorema di Bayes, accettato sul Periodico di Matematiche, in corso di stampa.

[4] A.MATURO (1990) Numeri Pseudocasuali. Libreria dell'Università, Pescara.

[5] L.PICCINATO-N.PINTACUDA (1985) Probabilità e Statistica. Quaderni C.N.R.

[6] H.SCHILD (1988) Turbo Pascal 4.0. Programmazione avanzata. Mc.Graw-Hill.

[7] R.SCOZZAVAVA (1988) Probabilità soggettiva e grandi rischi. Archimede N.40.

[8] R.SCOZZAFAVA (1981) Introduzione alla probabilità e alla statistica. Ed. Veschi, Roma.

[9] R.SCOZZAFAVA (1989) La probabilità soggettiva e le sue applicazioni. Ed. Masson, Milano.

Ringraziamo sentitamente il Prof. Romano Scozzafava per l'incoraggiamento e gli utili suggerimenti fornitici.