ブートストラップ検定の例

標本の大きさ(sample size)が10 である下記のデータがある。

10.9	7.1	8.8	8.4	7.6	9.7	9.0	4.9	5.0	7.6

これは独立かつ同一の正規分布からのランダム標本であると仮定する。

この標本における、標本平均は7.9 であり、標本標準偏差は1.907 である。ゆえに、標本平均を母集団平均の推定量としたとき、標準誤差(standard error)は

$$1.907/\sqrt{10} = 0.603$$

となる。

① ブートストラップ標準誤差の導出

(A) パラメトリック・ブートストラップ【図1, 図3(A)】

平均が7.9、標準偏差が1.907 である正規乱数を10 個発生させ、その標本平均を求める。これをB 回繰り返す。そして計算された標本平均のB 個のデータから、(標本)標準偏差を求める。これが(パラメトリック)ブートストラップ標準誤差となる。

B=59 として計算した結果 $\mathrm{se}(\hat{\mu})_{\mathrm{pb}} = 0.565$ がえられた。

(B) ノンパラメトリック・ブートストラップ【図1, 図3(B)】

上記の10 個のデータからなる標本から、無作為復元抽出で10 個データを取り出し、その標本平均を求める。これをB 回繰り返す。そして計算された標本平均のB 個のデータから、(標本)標準偏差を求める。これが(ノンパラメトリック)ブートストラップ標準誤差となる。

B=59 として計算した結果 $\mathrm{se}(\hat{\mu})_{\mathrm{nb}} = 0.510$ がえられた。

② ブートストラップ検定

$\mathrm{H_0}: \mu = 6.0$ vs $\mathrm{H_1}: \mu > 6.0$ の検定を考える。有意水準を5%とする。

(A) パラメトリック・ブートストラップ【図4】

平均が6.0、標準偏差が1.907 である正規乱数を10 個発生させ、その標本平均を求める。これをB 回繰り返す。そして計算された標本平均のB 個のデータからなる経験分布の上側5%点を求める。B=59 として計算した結果、上側5%点は、B 個のデータを大きい方から並べて3番目の6.717 である。元の標本の大きさ10 のデータから計算した標本平均(=母平均の推定量)は、7.9 であるので、経験分布の上側5%点よりも大きい。ゆえに、5%水準で帰無仮説を棄却する。

(B) ノンパラメトリック・ブートストラップ【図4】

上記の10 個のデータから、それぞれ標本平均7.9 を引き、その10 個の残差データからなる標本を作成する。その標本から無作為復元抽出で10 個データを取り出し、それに帰無仮説の平均6.0 を加え、その標本平均を求める。これをB 回繰り返す。そして計算された標本平均のB 個のデータからなる経験分布の上側5%点を求める。B=59 として計算した結果、上側5%点は、B 個のデータを大きい方から並べて3番目の6.9 である。元の標本の大きさ10 のデータから計算した標本平均(=母平均の推定量)は、7.9 であるので、経験分布の上側5%点よりも大きい。ゆえに、5%水準で帰無仮説を棄却する。

図1. Scilab による計算結果

```
        ___________________________________________
                        scilab-5.2.2
              Scilab コンソーシアム (DIGITEO)
            Copyright (c) 1989-2010 (INRIA)
            Copyright (c) 1989-2007 (ENPC)
        ___________________________________________

立ち上げ開始:
  初期環境の取り込み

x =
10.9
7.1
8.8
8.4
7.6
9.7
9.
4.9
5.
7.6

mu =
7.9

sig =
1.9072959
n =
10.

b =
59.

% 中略 %

mupb =
7.7519354

sigpb =
0.5645161

munb =
7.9413559

signb =
0.5103091
```

図2A. 図1の計算のためのScilab のコード

```
clear
x=[10.9;7.1;8.8;8.4;7.6;9.7;9.0;4.9;5.0;7.6]
mu=mean(x)
sig=stdev(x)
n=10
b=59
//parametric bootstrap
pb=zeros(b,1)
for i=1:b, y=zeros(n,1), y=mu+sig*rand(n,1,'normal'), pb(i,1)=mean(y), end
mupb=mean(pb)
sigpb=stdev(pb)
//nonparamatric bootstrap
nb=zeros(b,1)
for i=1:b, y=zeros(n,1), y=sample(n,x), nb(i,1)=mean(y), end
munb=mean(nb)
signb=stdev(nb)
histplot(10,pb);
clf
histplot(10,nb);
```

図2B. 図１の計算のためのMatlab のコード

```
%%% Matlab code for bootstrap estimation %%%

%% data and descriptive statistics
x=[10.9, 7.1, 8.8, 8.4, 7.6, 9.7, 9.0, 4.9, 5.0, 7.6]';   % population
mu=mean(x);   % to calculate the mean of the values in the population x
sig=std(x);   % to calculate the standard deviation of the values
              % in the population x

%% parameters
n=10;   % sample size
b=59;   % number of repetition

%% parametric bootstrap estimation (PBE)
pb=zeros(b,1);   % box of variables

% to process bootstrap
for i=1:b
    y=mu+sig*randn(n,1);   % randomly drawing n values from N(mu,sig^2)
    pb(i,1)=mean(y);
end

% results
mubp=mean(pb);    % estimated mean by PBE
sigpb=std(pb);    % estimated standard deviation by PBE

%% nonparametric bootstrap estimation (NBE)
nb=zeros(b,1);   % box of variables

% to process bootstrap
for i=1:b
    y=randsample(x,n,true);   % randomly sampling n values from the vlues
    nb(i,1)=mean(y);          % of the population x with replacement
end

% results
munb=mean(nb);   % estimated mean by NBE
signb=std(nb);   % estimated standard deviation by NBE

%% figures

% figure 3A
figure
hist(pb,10);   % to depict the distribution of pb among 10 bins

% figure 3B
figure
hist(nb,10);
```

図3A. パラメトリック・ブートストラップによる標本平均の分布

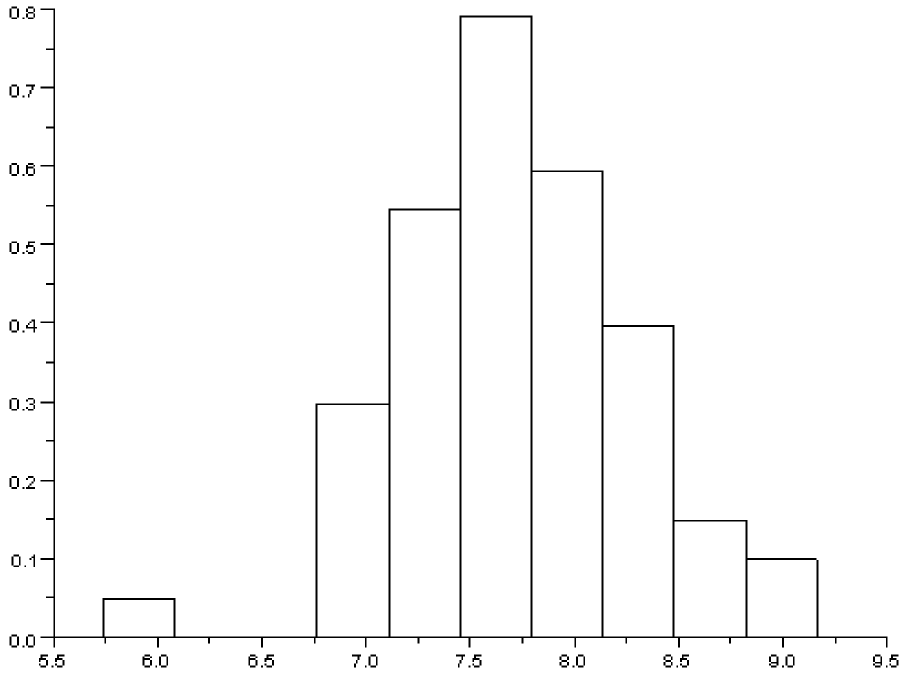

図3B. ノンパラメトリック・ブートストラップによる標本平均の分布

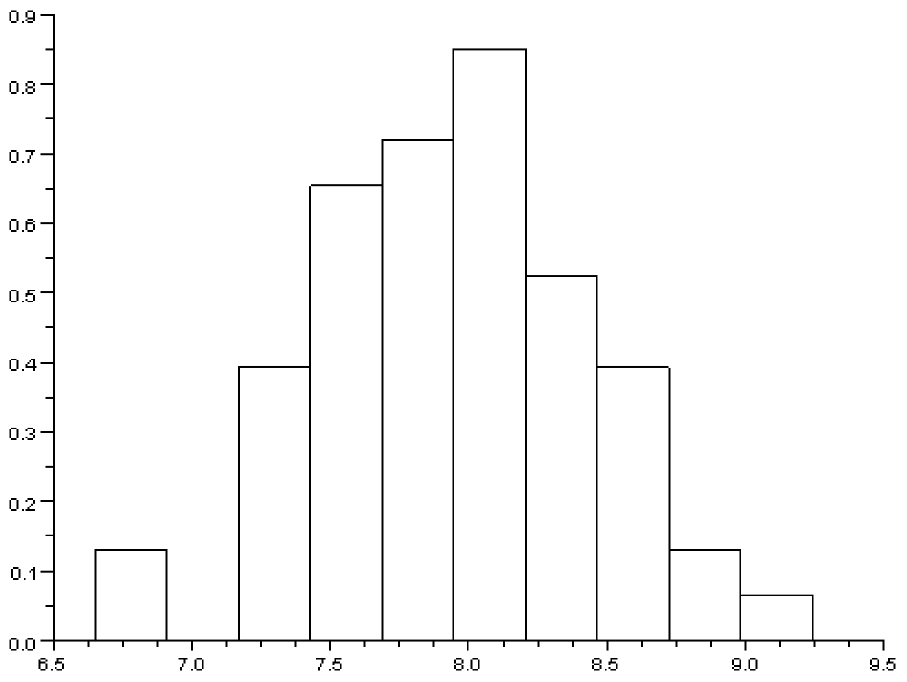

図4. ブートストラップ検定のための経験分布の計算結果（左）とそのScilab コード（右）

```
-->decpb(1:10)
ans =
7.6947878
6.7342402
6.7172983
6.7027829
6.6880651
6.6273286
6.6162382
6.4716517
6.4329194
6.4076237

-->decnb(1:10)
ans =
7.09
6.92
6.9
6.87
6.86
6.85
6.83
6.74
6.74
6.68
```

```
clear
x=[ 10.9;7.1;8.8;8.4;7.6;9.7;9.0;4.9;5.0;7.6 ]
mu=mean(x)
sig=stdev(x)
format(10);mu
format(10);sig
n=10
b=59
mu0=6.0
//parametric bootstrap
pb=zeros(b,1)
for i=1:b, y=zeros(n,1), y=mu0+sig*rand(n,1,'normal'),
pb(i,1)=mean(y), end
mupb=mean(pb)
sigpb=stdev(pb)
decpb=sort(pb)
//nonparamatric bootstrap
nb=zeros(b,1)
for i=1:b, y=zeros(n,1), u=zeros(n,1), u=sample(n,x)-mu,
y=mu0+u, nb(i,1)=mean(y), end
munb=mean(nb)
signb=stdev(nb)
decnb=sort(nb)
```

図5. 図4（左）の計算のためのMatlab コード

```
%%% Matlab code for bootstrap test %%%

%% data and descriptive statistics
x=[10.9, 7.1, 8.8, 8.4, 7.6, 9.7, 9.0, 4.9, 5.0, 7.6]';   % population
mu=mean(x);   % to calculate the mean of the values in the population x
sig=std(x);   % to calculate the standard deviation of the values
                    % in the population x

%% parameters
n=10;   % sample size
b=59;   % number of repetition

%% null hypothesis
mu0=6.0;

%% parametric bootstrap test (PBT)
pb=zeros(b,1);   % box of variables

% to process bootstrap
for i=1:b
      y=mu0*ones(n,1)+sig*randn(n,1);   % randomly drawing n values from
      pb(i,1)=mean(y);                          % N(mu0,sig^2)
end

% results
mubp=mean(pb);   % estimated mean by PBE
sigpb=std(pb);   % estimated standard deviation by PBE
decpb=sort(pb,'descend');   % to sort the elements of pb in descending order
decpb(3)   % reject the null hypothesis if < mu

%% nonparametric bootstrap test (NBT)
nb=zeros(b,1);   % box of variables

% to process bootstrap
for i=1:b
      u=x-mu.*ones(n,1);   % u = vector of deviations of sample mean
      y=randsample(u,n,true)+mu0;
      nb(i,1)=mean(y);
end

% results
munb=mean(nb);   % estimated mean by NBE
signb=std(nb);   % estimated standard deviation by NBE
decnb=sort(nb,'descend');
decnb(3)   % reject the null hypothesis if < mu
```