# ブートストラップ法を用いた統計量の推定に関する研究

熊本大学工学部　学生員　○木時　秀一
〃　　　正　員　三池　亮次
〃　　　正　員　小林　一郎

1．はじめに　アーチダムの竣工後の挙動が安定しており、破壊のおそれがあるかどうかを検証する維持管理上の手法として重回帰分析による方法がある。ダムに埋設された温度計で計測される温度条件と、貯水池の水圧を示す水位とたわみの線形回帰モデルを設定し、これの径年変化の有無と程度を、安全性の判断の指標として使用する。筆者らは先に、ある期間ごとの回帰モデルの差の検定理論に基づいて、安全管理の管理限界を設定する方法を提案した[1]。この方法は回帰偏差が正規分布に従うことを前提としている。

回帰偏差が正規分布に従わないときでも、データサイズnが十分に大きいとき、重回帰における中心極限定理に従って、回帰推定値は正規分布に漸近する。この中心極限定理の存在およびnがいかなる大きさのとき、これが正規分布に漸近するとみなしてよいかを実証するのに、電子計算機によるシミュレーションの一手法であるブートストラップ法（ＢＳ法）の適用を試みる。

すでに単回帰モデルにおいて、データサイズn＝20以上であれば、回帰偏差が正規分布に従わなくても、回帰係数の推定値は正規分布に漸近することを述べた[2]。また、回帰推定値およびその不偏分散の平方根の回帰分析による値と、上記のＢＳ法による値の比は、1.0 に近づくことが分かった。

本研究では、重回帰モデルにおいて、説明変数が一般に３個以上の場合について同様の研究を行なう。

2．重回帰モデルへのブートストラップ法の適用

本法の概要を説明するため説明変数が３個のものを解析モデルとする。回帰式として次式を設定する。

$$y_i=\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+e_i \quad \cdots \quad (1)$$

ただし、$\beta_0=\beta_1=\beta_2=1$とする。

回帰推定値は、

$$\hat{\mu}=\hat{\beta}\mathbb{X} \quad \cdots\cdots\cdots\cdots \quad (2)$$

回帰推定値の不偏分散の平方根は分散を、

$$S_E=\mathbf{y}^{(t)}\mathbf{y}-\hat{\beta}^{(t)}\mathbb{X}^{(t)}\mathbf{y} \quad \cdots\cdots \quad (3)$$

とすると、

$$\hat{\sigma}_{\hat{\mu}_i}=\sqrt{D_{ii}S_E/(n-p)} \quad \cdots\cdots \quad (4)$$

ここに、$D_{ii}$は　$\mathbb{X}\mathbb{S}^{-1}\mathbb{X}^{(t)}$ の対角要素であり、$\mathbb{S}=\mathbb{X}^{(t)}\mathbb{X}$ である。また、$n-p$は自由度である。

$$\bar{\mu}_B=\Sigma\hat{\mu}_{BI}/N_B \quad \cdots\cdots \quad (5)$$

$$\hat{\sigma}_{\hat{\mu}_{BI}}=\sqrt{\Sigma(\mu_{BI}-\bar{\mu}_B)^2/(N_B-1)} \quad \cdots \quad (6)$$

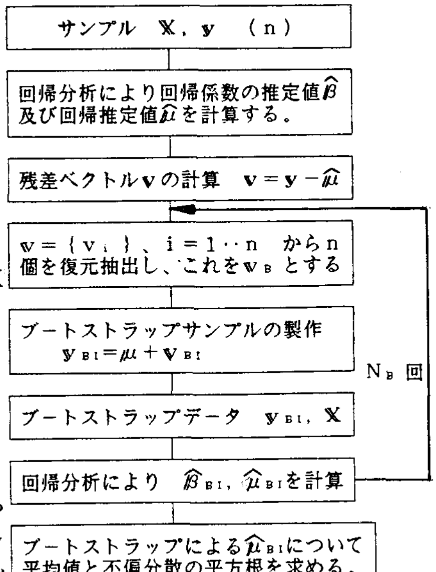
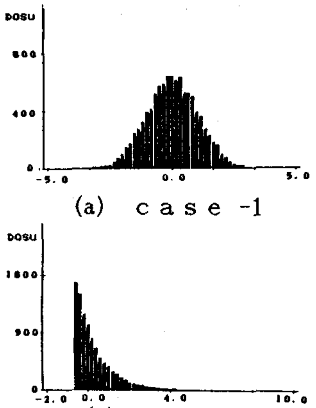



図１　ＢＳ法のフローチャート

ＢＳ法は、図１に示すように、あらかじめ与えられた誤差$e_i$のもとに$\mathbf{X}$，$\mathbf{y}$をサンプルとして用い重回帰分析を行ない、得られた$\hat{\mu}$より、残差ベクトル$\mathbf{v}=\mathbf{y}-\hat{\mu}$を計算する。次に、残差$\mathbf{v}$の乱数を発生させコピーを$N_B$個作り、その中から任意のn個の残差を抽出する。これを$\mathbf{v}_B$とし、（２）式より求めた$\hat{\mu}$を加えて$\mathbf{y}_{BI}$として、ＢＳサンプル〔$\mathbf{y}_{BI}$，$\mathbb{X}$〕を作成する。さらに、回帰分析により$\hat{\beta}_{BI}$、$\hat{\mu}_{BI}$を計算し、$\hat{\mu}_{BI}$（$I=1\cdots N_B$)の分布形を調べ、（５）式によってＢＳ法による$\hat{\mu}_{BI}$の平均値と（６）式に




(a)　case -1

(b)　case-2

図2　誤差の分布型

による$\hat{\mu}_{BI}$の不偏分散の平方根を求め（４）式で求めたものと比較する。

<u>３）数値計算例</u>

サンプル数n＝ 5,10,15,20,40,60,80 の７通り、$N_B$=1000,2000,3000,4000,5000回の５通りについて解析を行なった。

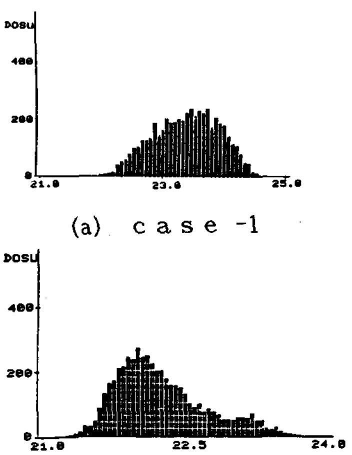
誤差$e_i$ には、つぎに示す２つの異なる分布型を用いた。分布図は図２の通りである。

case-1‥$e_i$ はN〔０，１〕の正規分布に従う。

case-2‥$e_i$ は$e_i = \alpha \log_e(1-u_i) - \alpha \log_e 2$ の指数分布に従う。ただし、$\alpha$=1で$u_i$ は一様乱数である。

$N_B$ =5000 回の時の$\hat{\mu}_B$の分布をサンプル数n=10,20の場合について示したものが図３，４である。n=10 の時ではcase-1、case-2共にばらつきが大きく、n=20 の時では分布形がほぼ正規分布形をしているのが分かる。そこで、残差平方和の平方根の比をサンプル数を変化させて描いたのが図５であり、$\hat{\sigma}_{\mu i}/\hat{\sigma}_{\mu BIi}$＝Pとすると、Pが 1.1以下になるのはn=20 以上の時であることが分かる。これは単帰分析の結果[2]とよく一致している。また、誤差$e_i$ の分布型が異なっていても、サンプル数nが増加すれば、回帰推定値$\hat{\mu}_B$ の度数分布はほぼ同じ形状を示してくる。このことから、中心極限定理が数値実験により確認されたものと考えられる。

説明変数がさらに多くなった場合の解析結果およびBS法の構造部材の信頼性解析への応用についての例題の報告は講演時に行なう予定である。

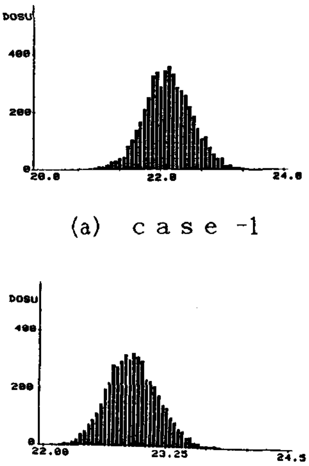



(a)　case -1

(b)　case-2

図３ 回帰推定値$\hat{\mu}_B$の度数分布

（n=10）




(a)　case -1

(b)　case-2

図４ 回帰推定値$\hat{\mu}_B$の度数分布

（n=20）

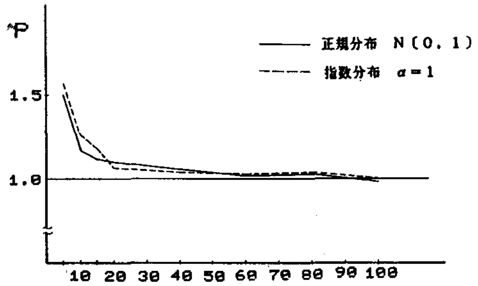



図５　重回帰モデルによる解析結果

参　考　文　献

１）Miike,Kobayashi:Safety Control of Dams by Multi-variate Regression Model, Proc. of ICOSSAR.

２）三池　他：ブートストラップ法の回帰モデル安全管理への応用について、西部支部研究発表会概要集 1986,3.