I-227　データ数を考慮した特性値の推定に関する考察

山梨大学　正員　杉山　俊幸
東京大学　正員　藤野　陽三
東京大学　正員　伊藤　　学

1．はじめに　構造物の安全性照査の規範は，構造物の設計変数を強度Rと荷重Sの二つに分けたとき，一般に　ν（Sd ／Rd ）≦1・・・(1)　のように書ける．ここで，Rd, Sd は各々強度R，荷重Sの設計値，νは構造物の重要度・限界状態に達した場合の社会的影響等を考慮するための全体的安全係数である．構造物の安全性照査が（1）式に依拠している場合には，与えられたデータから設計値として用いる統計的特性値を推定することが要求される．その際，推定する特性値が超過確率あるいは非超過確率のかなり小さい値となる場合も多い．ところが，こうした分布形裾部の値を精度よく推定する場合には，かなり多くのデータが必要である．実際に十分なデータが入手できる設計変数もあるが，少ないデータから特性値を推定せねばならない設計変数も現時点では多い．従って，利用可能なデータが少ない場合に如何に精度よく特性値を推定するかを考えておくことは重要である．また，設計値のわずかな"ずれ"が構造物の信頼性に大きく影響する場合もあることを考えると，より高精度の推定方法を検討しておく必要がある．

そこで本研究では，母集団分布形およびそのパラメータ値の推定には最尤法を用いることを前提として，有限個のデータから特性値を推定する幾つかの方法を定式化し，各々の方法の推定精度の優劣をモンテカルロ法を用いた数値実験により比較検討することにする．

2．推定方法の定式化　超過確率あるいは非超過確率eに対応する特性値を推定する方法として，以下の3通りを考える．

①最適合と判定された理論分布のみを用いて直接eに対応する特性値X_1を算出する［ES－1法］．

②母集団分布形と考えられるm種類の理論分布の各々に関してeに対応する特性値x_i　（i＝1～m）を算出し，尤度比率Li ［＝（i番目の理論分布に関する尤度）／Σ（j番目の理論分布に関する尤度）］で重み付けを行った値を特性値X_2とする．すなわち　$X_2 = \Sigma Li \cdot x_i$　により算出する［ES－2法］

③eよりも値が0．5に近い超過確率あるいは非超過確率e'に対応する特性値X'_3をES－2法で算出し，これにデータのばらつきを考慮した係数ϕを乗じて，eに対応する特性値$X_3 = \phi X'_3$とする．なお，ϕはデータそのものが本来有するばらつきと共に，データが少ないことに起因するばらつきも含む係数で，その評価は

$\phi = 1 - k\sigma$　　（下方裾部特性値の推定）

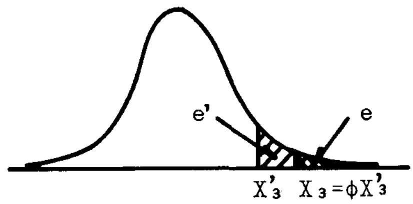
$\phi = 1 + k\sigma$　　（上方裾部特性値の推定）

ただし，σ：標本標準偏差　，　$k = \Sigma Li \cdot ko,i$

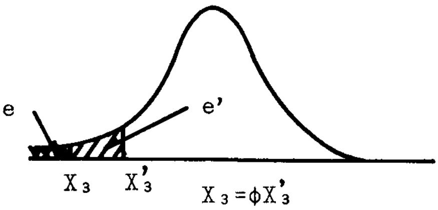
$$ko,i = \frac{\text{超過確率 e に対応する真の特性値}}{\text{超過確率 e' に対応する真の特性値}}$$

（for　i番目の理論分布形）

［ES－3法］　（図1参照）



e'超過確率値よりe超過確率値を推定



e'非超過確率値よりe非超過確率値を推定

図1　係数$\phi = 1 + k\sigma$　あるいは$\phi = 1 - k\sigma$　を用いた特性値の推定

3．推定精度の比較　上記3通りの推定方法の精度を以下のプロセスに従って，数値実験により比較する．

(1) 母集団分布形およびそのパラメータ値がわかっているデータをモンテカルロ法により有限個作成する

(2) 各方法で超過確率eに対応する特性値を推定する

(3) シミュレーションを２００回繰り返すことにより，推定値$X_{i,j}$と真の特性値X_oとの二乗平均誤差R．M．S．（母集団平均値で無次元化した値）を次式で求め，その大小から各手法の優劣を比較する

$$R.M.S.=\sqrt{\sum_{j}(X_{i,j}-X_o)/n/\mu}$$

ただし，$X_{i,j}$：j回目のシミュレーションでのi番目の方法による推定値

X_o：真の特性値　，　n：シミュレーション数　，　μ：母集団平均値

４．数値実験結果および考察　上記２．で定式化した３通りの方法を用いて推定した特性値と真の特性値との二乗平均誤差を２００回のシミュレーションから求めた結果の一部を示したのが図２である．この図は，超過確率e＝０．１％に対応する上方裾部特性値を推定した場合で，横軸にデータ数n，縦軸に二乗平均誤差R．M．S．をとっている．母集団分布形および変動係数は図中に記してある．ただしES－３法においては，あらかじめ算出する特性値の超過確率e'の値をe'＝１０％としている．これらの図より，(i) ３通りの方法の中ではES－３法のR．M．S．が最も小さいこと，(ii) 変動係数の大小によって各方法の優劣の順位が逆転することはほとんどないこと，(iii) １００個程度のデータから推定する場合には，どの方法を用いても推定精度にほとんど差がないこと　がわかる．

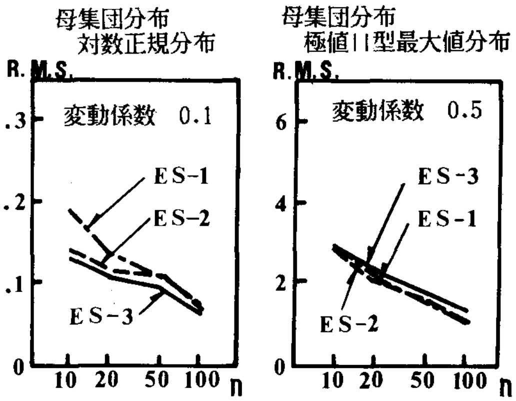


図２　０．１％超過確率値の推定に関する二乗平均誤差R.M.S.とデータ数nの関係

そこで，データが１０～５０個と少ないときの３通りの推定方法に関して，数値実験結果より得られる優劣の比較を母集団分布形別に定性的に示したのが表１である．これらの結果をまとめると以下のことがいえる．

①母集団が極値II型最大値分布の場合を除いて，ES－３法の精度が安定して高い．

②母集団が極値II型最大値分布に従う場合には，ES－１法の精度が高い．これは，他の分布形と比べて少数のデータからでもかなりの高精度で母集団分布を判定できるためと考えられる．

③ES－１法とES－２法の推定精度は，母集団分布形・データ数によって逆転するため，その優劣は判定できない．

方法	母集団分布					
	N	LN	EXT1S	EXT1L	EXT2	EXT3
ES-1	△	×	△	×	○	△
ES-2	△	△	○	△	○	△
ES-3	○	○	○	○	△	○

(a) 上方裾部特性値の推定

方法	母集団分布					
	N	LN	EXT1S	EXT1L	EXT2	EXT3
ES-1	×	×	△	△	○	×
ES-2	△	△	△	△	×	○
ES-3	○	○	△	○	×	○

(b) 下方裾部特性値の推定

表１　特性値推定方法の定性的な比較（○：優　△：中間　×：劣）

５．まとめ　有限個の少ないデータから設計値として用いる特性値を推定するための幾つかの方法を定式化し，各々の方法の推定精度の優劣をモンテカルロ法を用いた数値実験により比較検討した．その結果，超過確率あるいは非超過確率の小さい特性値を推定する場合には，少ないデータからでも比較的精度よく推定できる特性値を推定し，これにデータ不足等による不確実度を含ませた係数を乗じて，所与の超過確率値を算出する方法が相対的に優れていることが明らかとなった．ただし，この係数はデータより得られる尤度比率を含む定式化に基づいて算出すればよい．

なお本研究では，各々の推定方法の妥当性を理論的に裏付けることができないため，モンテカルロ法を用いた数値実験を主として考察を行うにとどまっている．今後，理論的な証明が可能かどうかを検討していく必要があると考えている．