

小標本における標準誤差と信頼区間の推定について

メタデータ	言語: Japanese 出版者: 水産総合研究センター 公開日: 2024-10-02 キーワード (Ja): キーワード (En): bootstrap; confidence interval; standard error; small sample 作成者: 平松, 一彦, 岡村, 寛 メールアドレス: 所属: 遠洋水産研究所, 遠洋水産研究所
URL	https://fra.repo.nii.ac.jp/records/2010787

This work is licensed under a Creative Commons Attribution 4.0 International License.



小標本における標準誤差と信頼区間の推定について

平松一彦*, 岡村 寛*

Estimation of standard errors and confidence intervals in small samples

Kazuhiko HIRAMATSU* and Hiroshi OKAMURA*

Abstract The standard errors and confidence intervals have been used as a measure of the accuracy of estimators. However, it is known that the estimates of standard error using the standard procedures based on the unbiased estimate of variance and the bootstrap method are biased and the bootstrap confidence intervals based on the percentile method are too short when the sample size is small. The amounts and causes of these biases are examined and the corrections of these biases are discussed. The results indicate that the biases in the standard error using the standard procedures would be negligible. However, our Monte Carlo simulations show that the biases in the bootstrap estimate of standard error and confidence intervals would not be negligible when sample size is equal to or less than 10 or the ratio of the number of parameters estimated to the sample size is larger than 0.1. We therefore recommend that researchers carefully consider the effective sample size when applying the bootstrap method. In particular, since the result of hypothesis test can be affected by the bias, bias corrected bootstrap method such as calibration or other method should be used when the effective sample size is small.

Key words: bootstrap, confidence interval, standard error, small sample

水産資源評価をはじめとする水産分野の研究においては、誤差の多い標本（データ）から推定値を得たり仮説を検証したりする必要がある。推定値の信頼性の指標として、推定量の標本分布の標準偏差である標準誤差や信頼区間が広く使われており、得られた点推定値は通常これらの指標を付けて示される。また区間推定と同様の考え方は仮説検定にも利用されている。

しかし、平均値の標準誤差として一般的に使われている不偏分散の平方根から計算される推定量は、不偏ではなく負の偏りを持つことが知られている。特に小標本の場合にはその影響が大きくなる。

また近年ブートストラップ法により標準誤差や信頼区間が計算されることも多いが、ブートストラップ法で推定された平均値の標準誤差や信頼区間は自由度の

修正がされていない。このため、特に標本数が少ない場合には、この影響が無視できず誤った結論に導く危険性がある。

本稿では、小標本における標準誤差や信頼区間の推定の問題点とその原因、修正方法、仮説検定への影響について述べる。特にブートストラップ法を小標本に適用した場合の影響は無視できないことがあるため、ブートストラップ法については詳しく検討する。水産分野においては実験や観測から大標本を得ることが実際上不可能で、小標本から推定値を得たり判断を下したりしなければならないことがある。水産分野において特に注意すべき点についても述べる。

2002年7月22日受理 (Received on July 22, 2002)

水産総合研究センター業績 A 第27号 (Contribution No. A 27 from Fisheries Research Agency)

* 遠洋水産研究所 〒424-8633 静岡県清水市折戸5-7-1 (National Research Institute of Far Seas Fisheries, 5-7-1, Orido, Shimizu, Shizuoka, 424-8633, Japan)

標準誤差の推定

平均値の標準誤差

平均値 μ の標準誤差として, 不偏分散

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2 \quad (1)$$

から計算される s を用いて

$$\hat{SE}[\hat{\mu}] = \frac{s}{\sqrt{n}} \quad (2)$$

の形が一般的に使われている。ここで x_i は標本の値, $\hat{\mu}$ は標本平均, n は標本の大きさである。平均値の場合, 母集団の標準偏差を σ とすると, その標準誤差は

$$SE[\hat{\mu}] = \frac{\sigma}{\sqrt{n}} \quad (3)$$

となるが, (2) 式はこの σ をその推定値 s で置き換えていることに相当する。しかし, s^2 は分散の不偏推定量であるが, s は標準偏差の不偏推定量ではない。このため (2) 式は偏りを持った値となっている。

母集団が正規分布の場合, 標準偏差の不偏推定量 s^* は

$$s^* = \sqrt{\frac{n-1}{2} \frac{\Gamma((n-1)/2)}{\Gamma(n/2)} s} \quad (4)$$

となる (Sokal and Rohlf, 1995)。ここで Γ はガンマ関数である。小標本の場合の s/s^* を Fig. 1 に示す。 $n=5$ で 6%, $n=10$ で 3% 程度過小評価となっている。なお s^* は

$$s^* \cong \sqrt{\frac{1}{n-1.5} \sum_{i=1}^n (x_i - \hat{\mu})^2} \quad (5)$$

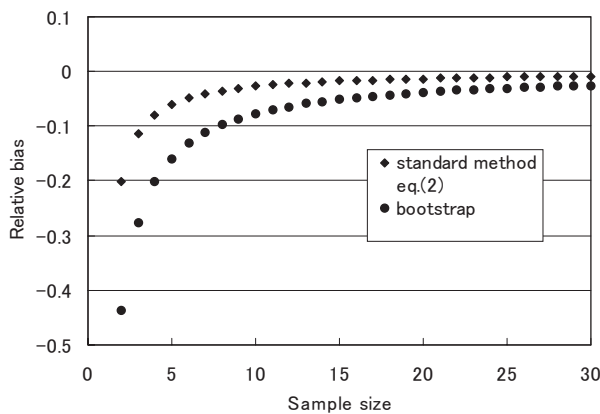


Fig. 1. Relative bias of standard errors of mean estimated by the standard procedure using eq.(2) and the bootstrap method

で (1) 式と類似の形で近似することができる。

母集団の正規性が仮定できない場合には, (4) 式は成立しない。しかし, 一般に

$$E[s^2] \geq (E[s])^2 \quad (6)$$

であり, また,

$$E[s^2] = \sigma^2 \quad (7)$$

であることから

$$\sigma \geq E[s] \quad (8)$$

が成立する。(6) 式で等号が成立するのは s の分散が 0 の時のみであり, 実際には常に負の偏りを持つと考えられる。

一般の統計量の標準誤差

標準誤差は平均値に限らず一般の統計量に対しても計算される。しかし (2) 式などは平均値に対する式であり, それ以外の統計量の場合には個別に計算する必要がある。例えば 2 項分布における p の推定を考えると, p の不偏推定量は

$$\hat{p} = \frac{x}{n} \quad (9)$$

であり, $\hat{p}$ の分散は

$$V[\hat{p}] = \frac{1}{n} p(1-p) \quad (10)$$

である (ただし x をベルヌーイ分布からの観測の和とすると, (9) 式も平均値とみなすこともできる)。ここで $\hat{p}$ を使ってあらわすと,

$$\hat{V}[\hat{p}] = \frac{1}{n-1} \hat{p}(1-\hat{p}) \quad (11)$$

が $\hat{p}$ の分散の不偏推定量である。 $\hat{p}$ の標準偏差, すなわち標準誤差を

$$\hat{SE}[\hat{p}] = \sqrt{\frac{1}{n-1} \hat{p}(1-\hat{p})} \quad (12)$$

で求めるとこれは負の偏りをもった値となる。シミュレーション結果によれば, 例えば, $p=0.5$, $n=3$ の場合には, (12) 式を使うと 14% 程度過小評価となっている。

t分布による信頼区間の推定

平均 μ 、標準偏差 σ の正規母集団から大きさ n の標本が得られた場合、母標準偏差 σ が既知であれば

$$z = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad (13)$$

である。ここで $N(0,1)$ は標準正規分布である。(13)式の σ を推定量 s で置き換えると、

$$t = \frac{\hat{\mu} - \mu}{s/\sqrt{n}} \sim t(n-1) \quad (14)$$

となることが知られている。ここで $t(n-1)$ は自由度 $n-1$ の t 分布である。平均値に関する区間推定や仮説検定はこの t 分布を用いて行われ、例えば $(100-\alpha)\%$ 信頼区間は

$$\hat{\mu} - t(\alpha/2, n-1) \frac{s}{\sqrt{n}} \leq \mu \leq \hat{\mu} + t(\alpha/2, n-1) \frac{s}{\sqrt{n}} \quad (15)$$

となる。ここで、 $t(\alpha/2, n-1)$ は自由度 $n-1$ の t 分布の上側 $\alpha/2\%$ 点である。

(14)式で使われる s は(4)式の σ の不偏推定量ではなく、不偏分散の平方根である点に注意が必要である。(4)式を使って σ の不偏推定量を得ても、 t 分布を用いた信頼区間の推定や仮説検定には使えない。標準誤差の不偏推定量が一般に使われていないのは、計算が複雑なことに加えこのような原因によると考えられる。

ブートストラップ法による標準誤差の推定

標準誤差が解析的に計算できるのは、平均値などの特定の場合のみである。一方、ブートストラップ法を利用すれば、どの推定量に関しても同一の方法で標準誤差や信頼区間を計算できる。このため水産分野においてもブートストラップ法の利用が広がっている(平松, 1994)。

ブートストラップ法では標本を母集団とみなし、これから再度標本抽出(リサンプリング)を行って得られた標本(ブートストラップ標本)から推定量を計算する。そしてこの操作を繰り返し、推定量の標本分布を得る。標本そのものをリサンプリングするノンパラメトリックブートストラップと、母集団の分布形を仮定し標本から得られる推定値を代入した確率分布からリサンプリングを行うパラメトリックブートストラップの2種類がある。いずれの方法でも、リサンプリングにより得られた推定量の標本分布を用いて、推定量の標準誤差や信頼区間を求める(Efron and Tibshirani,

1993)。しかし、通常使われているブートストラップ法では、小標本の場合に問題が生じることが知られている。ここでは、まずブートストラップ法による標準誤差の推定の検討を行う。

平均値の標準誤差

ブートストラップ法では推定量 $\hat{\theta}$ の標準誤差は

$$\hat{SE}[\hat{\theta}] = \sqrt{\frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}_b - \sum_{b=1}^B \frac{\hat{\theta}_b}{B} \right)^2} \quad (16)$$

で計算される。ここで、 $\hat{\theta}_b$ は各ブートストラップ標本に基づく θ の推定値、 B はブートストラップ標本の数(リサンプリングの回数)である。平均値 μ の場合には $B \rightarrow \infty$ の時

$$\hat{SE}[\hat{\mu}] = \sqrt{\frac{1}{n} \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2} \quad (17)$$

となる(Efron and Tibshirani, 1993, p.14)。(1)式、(2)式と比較すると、(17)式で得られる標準誤差は、不偏分散(1)式のかわりに自由度の影響を考慮しない偏りのある分散

$$u^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 \quad (18)$$

を使用して計算していることに相当する。このためブートストラップ法で計算された平均値の標準誤差は、通常の(2)式で計算される標準誤差より

$$\sqrt{\frac{n-1}{n}} \quad (19)$$

だけ小さくなっている。さらに(2)式の標準誤差も負の偏りを持つことから、ブートストラップ法で計算された平均値の標準誤差には二重に負の偏りがかかっており、小標本の場合にはかなり偏りが大きくなる。これらの偏りの影響を標本数の関数としてFig. 1に示す。

2種類の偏りのうち(17)式に起因する偏りは、ブートストラップ法で得られた標準誤差に(19)式の逆数をかけて

$$\hat{SE}[\hat{\theta}] = \sqrt{\frac{n}{n-1} \frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}_b - \sum_{b=1}^B \frac{\hat{\theta}_b}{B} \right)^2} \quad (20)$$

とすることで修正可能である。母集団が正規分布であれば、さらに(4)式を利用して標準偏差の不偏推定量を求めることができる。ただし、正規母集団の平均値の標準誤差の推定であれば、ブートストラップ法を使うまでもなく、解析的な方法で計算すれば十分である。

一般の統計量の標準誤差

一般の統計量の場合には (17) 式のような形には書けないため、ブートストラップ法で推定された標準誤差の性質は自明ではない。しかし、平均値以外の統計量においても負の偏りが生じていることが予想される。

例えば、二項分布の $\hat{p}$ のブートストラップ推定量の分散は (10) 式に相当することが知られている (丹後, 2000)。これはブートストラップで推定された p の標準誤差は、上記の平均値の場合と同様の偏りを持つことを示している。

ブートストラップ法による信頼区間の推定

ブートストラップ法による区間推定の方法として様々なものが考えられているが、現在水産分野で一般的に用いられているのはパーセンタイル法と呼ばれる方法である。これはリサンプリングによって得られた推定量の分布の、上側 $\alpha/2\%$ 点と下側 $\alpha/2\%$ 点の間を、信頼係数 $(100 - \alpha)\%$ の信頼区間とする方法である。

平均値の信頼区間

t 分布を用いた正規母集団の平均値の信頼区間は (15) 式ようになる。自由度の影響は t 分布と s の双方に入っているが、標準誤差の議論からブートストラップ法による信頼区間はこれらの自由度の影響を考慮しておらず、通常の t 分布による信頼区間よりも短くなることが予想される。実際、パラメトリックブートストラップ法でパーセンタイル法を用いて信頼係数 $(100 - \alpha)\%$ の信頼区間を構成すると、(15) 式より短い

$$\hat{\mu} - \frac{u}{\sqrt{n}} z(\alpha/2) \leq \mu \leq \hat{\mu} + \frac{u}{\sqrt{n}} z(\alpha/2) \quad (21)$$

となり、実際の信頼係数は $(100 - \alpha)\%$ より小さくな

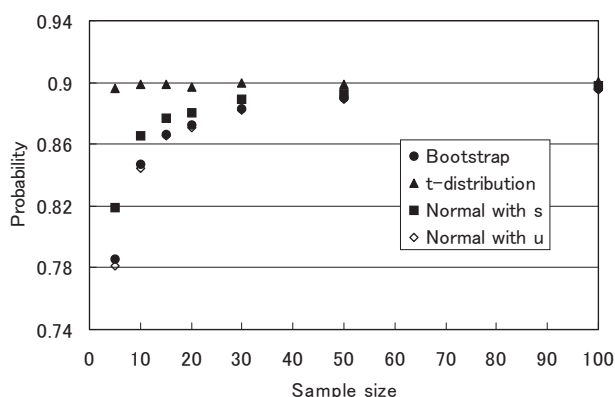


Fig. 2. Probability that the estimated 90% confidence intervals for the mean cover the population mean. The numbers of bootstrap samples and Monte Carlo trials are 100 and 10000, respectively.

ることが示される (付録参照)。ここで $z(\alpha/2)$ は、標準正規分布の上側 $\alpha/2\%$ 点である。

ノンパラメトリックブートストラップ法においても、同様の問題が生じていることが予想される。そこで、

- ①ブートストラップ法 (パーセンタイル法)
- ② (15) 式の通常の t 分布
- ③ (15) 式において t 分布の代りに正規分布を使用
- ④正規分布と u を用いた (21) 式の信頼区間

についてそれぞれの方法で90%信頼区間を計算し、それが実際に母集団の平均値 (真値) を含む割合を求めることにより、各手法で得られる信頼区間の正しさを比較した。シミュレーション結果をFig. 2に示す。これから、ノンパラメトリックブートストラップ法の信頼区間も、④の自由度の影響を全く考えていない推定に対応していることがわかる。ブートストラップ法で計算された「90%信頼区間」は、実際には $n=5$ では79%信頼区間、 $n=10$ では85%信頼区間に対応している。

一般の統計量の信頼区間

一般の統計量の信頼区間に関し、上に示したような偏りの程度を示すことは困難であるが、Efron and Tibshirani (1993) の25章で、2変量正規分布における2つの確率変数の平均値の比の信頼区間の推定においても同様の問題が示されており、一般的にパーセンタイル法を用いるとこのような小標本に起因する問題が存在するものと考えられる。

回帰係数の区間推定では、自由度は $n-2$ となるため平均値の推定より影響が大きくなることが予想される。回帰係数の90%信頼区間が母集団の値を含む割合を、通常の t 分布とブートストラップ法を用いた場合について、 n を変えてシミュレーションにより検討した結果をFig. 3に示す。Fig. 2と比較すれば明らかなように、

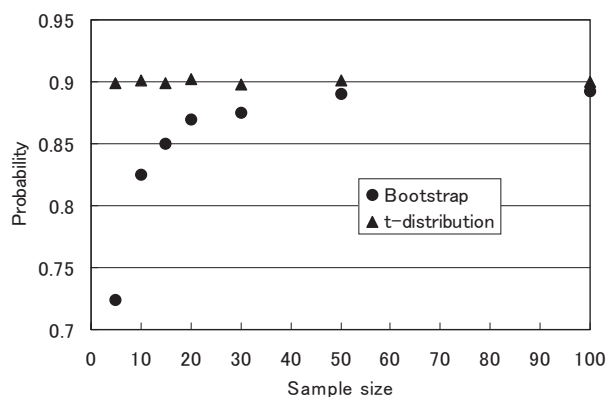


Fig. 3. Probability that the estimated 90% confidence intervals for the regression coefficient cover the true value. The numbers of bootstrap samples and Monte Carlo trials are 100 and 10000, respectively.

Table 1. Comparison of probability that the estimated 90% confidence intervals for the mean cover the true value. The numbers of bootstrap samples and Monte Carlo trials are 100 and 10000, respectively.

Sample size	t-distribution	bootstrap		
		Percentile	BC	BC _a
5	0.8960	0.7853	0.7689	0.7710
10	0.8992	0.8473	0.8305	0.8324
15	0.8984	0.8662	0.8505	0.8513
20	0.8973	0.8721	0.8553	0.8569
30	0.8997	0.8834	0.8614	0.8642
50	0.8984	0.8904	0.8694	0.8702
100	0.9001	0.8961	0.8819	0.8831

平均値に比べより影響が大きくなっている。

信頼区間の修正

パーセンタイル法を改良した区間推定法としてBC法やBC_a法がある (Efron and Tibshirani, 1993)。これらは点推定値の偏りやパラメータの値に依存して分散が変化するという影響を考慮して、より正確に信頼区間を計算する方法である。しかし、シミュレーションの結果によれば小標本に起因する問題はそのまま残る (Table 1)。

ブートストラップ t 法 (Efron and Tibshirani, 1993) は t 分布に対応する統計量

$$Z^*(b) = \frac{\hat{\theta}^*(b) - \hat{\theta}}{\hat{se}^*(b)} \quad (22)$$

をブートストラップ標本毎に計算し、この分布の上側 $\alpha/2\%$ 点 $\hat{t}(\alpha/2)$ と下側 $\alpha/2\%$ 点 $\hat{t}(100 - \alpha/2)$ を用いて

$$\hat{\theta} - \hat{t}(\alpha/2) \cdot \hat{se} \leq \theta \leq \hat{\theta} - \hat{t}(100 - \alpha/2) \cdot \hat{se} \quad (23)$$

によって $(100 - \alpha)\%$ 信頼区間を構成する方法である (t 分布のように左右対称とは限らないため、上側と下側を区別する必要がある)。ここで $\hat{se}^*(b)$ はブートストラップ標本毎の標準誤差で、何らかの方法で (例えばブートストラップ法で) 推定する。ブートストラップ t 法によれば、 $\hat{se}$ として自由度の影響を考慮した (20) 式などによる推定値を用いれば、(15) 式に対応する区間推定、すなわち小標本に起因する問題を修正した区間推定が可能である。ただし、ブートストラップ t 法はパーセンタイル法などに比べるとばらつきの大きい不安定な結果を与えること、スケールに依存するため適切なスケールを選択する必要があるといった問題が知られている (Efron and Tibshirani, 1993)。ま

たブートストラップ標本毎にさらにブートストラップを行って $\hat{se}$ を推定する必要がある、計算は煩雑となる。

信頼区間を補正する方法として Efron and Tibshirani (1993) はキャリブレーションと呼ぶ方法を示している。これは各ブートストラップ標本に対して (さらにブートストラップを行って) 上側および下側 $\alpha/2\%$ 点を推定し、これが実際に真の値 (すなわちもとの標本の点推定値) を含む確率と比較することにより補正を行うという方法である。小標本に起因する問題もこれで補正可能であるが、計算量は膨大になる。

考 察

統計学の理論は漸近的 ($n \rightarrow \infty$) な条件の下で組立てられていることが多い。このため標本数が少ない場合には、偏りを生じることがある。(2) 式が偏りを持つことは統計学の教科書に書かれているが、広く認識されているわけではなく、また不偏な (4) 式もほとんど使われていない。この原因としては

- 偏りは $n=5$ で 5% 程度、 $n=10$ で 2% 程度とそれほど大きくない。
- t 分布を使った区間推定や仮説検定では、不偏な (4) 式ではなく (2) 式が使われる。

といったことが考えられる。小標本の場合には標準誤差自体のばらつきも大きいことを考慮すると、実用上は、不偏ではあるが計算の面倒な (4) 式を使う利点はほとんど無いものと考えられる。

一方、ブートストラップ法の場合にはさらに自由度の影響を無視していることによる影響が存在し、これは小標本の場合には無視できないことがある。標準誤差の場合、不偏推定値に比べ $n=5$ で 11%、 $n=10$ で 5% 程度過小評価となっている。パーセンタイル法による平均値の区間推定では、例えば 90% 信頼区間が、実

際には $n=5$ では79%区間, $n=10$ では85%区間となっている。

これらは標本数を増やせば解消される問題であるが, 水産分野では繰り返し実験したり観測したりすることは実際上不可能なことがある。わずかな標本数で信頼区間を計算したり仮説検定を行ったりする場合には十分な注意が必要である。平均値などの区間推定や仮説検定では, 母集団が正規分布に近いなら通常の t 分布を用いた方がよい。既存の解析的な評価ができないような推定量の場合はブートストラップ法が有力であるが, 単純なパーセンタイル法では n が10以下, あるいはパラメータ数と標本数の比が0.1以上では数%からそれ以上信頼係数が異なっていることに注意が必要である。おおよその推定精度の指標とするために, 標準誤差や信頼区間を推定する場合にはそれほど神経質になる必要はないと考えられる。しかし, 仮説検定を行う場合には, 本来棄却できない仮説を棄却してしまう可能性があり十分な注意が必要である。キャリブレーション法を用いるなど, 小標本の影響を考慮した慎重な検討が必要である。

標本数が少ない場合は比較的明らかであるが, パラメータ数と標本数の比が大きい場合や複数のリサンプリングを行い各リサンプリングにおいては標本数が少ないような場合, 一見標本数が多くとも問題が生じる可能性がある。特に水産分野で注意が必要な場合として以下のような例を挙げることができる。

- ・回帰式, 非線形最小二乗法等で標本数に対するパラメータ数の比率が高い場合。代表的なプロダクションモデルのプログラムであるASPIC (Prager, 1994) では残差のリサンプリングにより区間推定を行っている。推定するパラメータは4個であるので, 例えば20年間のデータを用いたとしても自由度の影響は無視できない可能性がある。またADAPT VPAなどではパラメータのうち漁具能率は解析的に計算可能であるので, 未知パラメータと認識されていないことがある。しかし本来これらも未知パラメータとみなすべきであり, 見かけよりもパラメータ数/標本数の割合が高いことがある。

- ・層化サンプリングで層毎の標本数が少ない時に, 層毎にリサンプリングを行って資源量の区間推定を行う場合。全体の標本数が多くとも, 層内の標本数が少ないと, この影響により問題が生じる。例えば観測点は50点あるが, 5点毎10層にわけてブートストラップ法で計算する場合など, 実質的な標本数は5となり, 大

きな影響がある。なお有限母集団からのサンプリングの場合, 有限修正も考慮する必要がある。これらを考慮したブートストラップ法による層化サンプリングの分散の推定量がSmith and Gavaris (1993) に示されているが, 必ずしもまだ十分ではないようである。

ブートストラップ法の標準的教科書とも言えるEfron and Tibshirani (1993) では, 小標本に起因する問題の指摘もあるが, さほど重要視されているようにはみえない。適用例として挙げられているのは標本数10個程度のものが多いが, 小標本に起因する問題の修正はほとんどなされていない。このためブートストラップ法は小標本でも問題なく適用可能であるとの印象を持ちかねない。水産分野でもブートストラップ法が使われることが多くなったが, その問題点についてはほとんど認識されておらず, 本稿で指摘したような点も含め十分留意して用いるべきであろう。

謝 辞

本論文に対しご助言を頂いた東京水産大学北門利英助手, 遠洋水産研究所庄野 宏研究員に感謝する。

文 献

- Efron B. and Tibshirani R. J., 1993: An Introduction to the Bootstrap, Chapman and Hall, New York, 436pp.
- 平松一彦, 1994: Bootstrap法による信頼区間とバイアスの推定. 遠洋水産研究所ニュース, **94**, 2-6.
- Prager M. H., 1994: A suite of extension to a nonparametric surplus-production model. *Fish. Bull.*, **92**, 374-389.
- Smith S. J. and Gavaris S., 1993: Evaluating the accuracy of projected catch estimates from sequential population analysis and trawl survey abundance estimates, in "Risk evaluation and biological reference points for fisheries management" (ed. by Smith S. J., Hunt J. J., and Rivard D.). *Can. Spec. Publ. Fish. Aquat. Sci.*, **120**, 163-172.
- Sokal R. R. and Rohlf F. J., 1995: Biometry, 3rd ed., Freeman, New York, pp.52-53.
- 丹後俊郎, 2000: 統計モデル入門, 朝倉書店, 東京, pp.32-33.

付 録

正規分布の母集団を仮定し、パーセンタイル法によるパラメトリックブートストラップ信頼区間と通常の t 分布による信頼区間の比較を行い、ブートストラップ法で得られた信頼区間は通常の t 分布から得られる信頼区間に比べて短く、信頼係数は $(100 - \alpha)\%$ より小さくなっていることを示す。

問題の設定

平均 μ 、分散 σ^2 の正規分布 $N(\mu, \sigma^2)$ から大きさ n の標本 $x_1, \dots, x_n$ を得る。この時

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{A1})$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2 \quad (\text{A2})$$

さらに不偏でない分散の推定量として

$$u^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 \quad (\text{A3})$$

とする。

 t 分布による信頼区間

通常の t 分布の理論から

$$\frac{\hat{\mu} - \mu}{s/\sqrt{n}} \sim t(n-1) \quad (\text{A4})$$

である。これから、 $(100 - \alpha)\%$ 信頼区間は t 分布の $\alpha/2\%$ 点を使って

$$\hat{\mu} - \frac{s}{\sqrt{n}} t(\alpha/2, n-1) \leq \mu \leq \hat{\mu} + \frac{s}{\sqrt{n}} t(\alpha/2, n-1) \quad (\text{A5})$$

となる。ここで $\hat{\mu}$ は

$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad (\text{A6})$$

であるので、信頼区間は、中心が μ のまわりに分散 σ^2/n の正規分布でばらつく、長さ

$$2 \frac{s}{\sqrt{n}} t(\alpha/2, n-1) \quad (\text{A7})$$

の区間である。なお、標本によって s も変動するため、信頼区間の長さもばらつく。

パラメトリックブートストラップ法の信頼区間

パラメトリックブートストラップ法では (A1) 式、

(A3) 式でデータから推定された平均と分散を用いて

$$N(\hat{\mu}, u^2) \quad (\text{A8})$$

からさらに大きさ n のデータ x^* を何度も発生させる。そしてこの平均

$$\hat{\mu}^* = \frac{1}{n} \sum_{i=1}^n x_i^* \quad (\text{A9})$$

の分布から信頼区間を推定する。 x^* は (A8) 式の正規分布から発生しているから $\hat{\mu}^*$ の分布は

$$\frac{\hat{\mu}^* - \hat{\mu}}{u/\sqrt{n}} \sim N(0,1) \quad (\text{A10})$$

となる。パーセンタイル法ではこの分布を使って信頼区間が推定される。 $(100 - \alpha)\%$ 信頼区間は、標準正規分布の $\alpha/2\%$ 点 $z(\alpha/2)$ を使って

$$\hat{\mu} - \frac{u}{\sqrt{n}} z(\alpha/2) \leq \mu \leq \hat{\mu} + \frac{u}{\sqrt{n}} z(\alpha/2) \quad (\text{A11})$$

となる。

$\hat{\mu}$ は平均 μ 、分散 σ^2/n の正規分布でばらつくから、結局ブートストラップ法で推定された信頼区間は、中心が μ のまわりに分散 σ^2/n の正規分布でばらつく、長さ

$$2 \frac{u}{\sqrt{n}} z(\alpha/2) \quad (\text{A12})$$

の区間となる ((A7) 式同様 u の変動によりこの長さもばらつく)。

両者の比較

(A7) 式と (A12) 式を比較すると、パラメトリックブートストラップ法の信頼区間は s が u 、 $t(\alpha/2, n-1)$ が $z(\alpha/2)$ となっている点が異なる。パラメトリックブートストラップ法で計算される信頼区間は、自由度の影響を全く考慮しない信頼区間となっており、 t 分布を用いた信頼区間の

$$\sqrt{\frac{n-1}{n}} \frac{z(\alpha/2)}{t(\alpha/2, n-1)} \quad (\text{A13})$$

倍の短い区間となっている。またこの時

$$\begin{aligned} P\left(\left|\frac{\hat{\mu} - \mu}{u/\sqrt{n}}\right| < z(\alpha/2)\right) &= P\left(\left|\frac{\hat{\mu} - \mu}{s/\sqrt{n}}\right| < \sqrt{\frac{n-1}{n}} z(\alpha/2)\right) \\ &< P\left(\left|\frac{\hat{\mu} - \mu}{s/\sqrt{n}}\right| < t(\alpha/2, n-1)\right) = (100 - \alpha)\% \end{aligned} \quad (\text{A14})$$

となることから, パラメトリックブートストラップ信頼区間の信頼係数は $(100 - \alpha)\%$ より小さくなる。ここで, $P(p < a)$ は $p < a$ となる確率を示す。

パラメトリックブートストラップ法において用いら

れる分散の推定量として, (A3) 式のかわりに不偏な (A2) 式を使っても, (A11) 式において u が s となるだけで正規分布の部分は変化がない。このためやはり短い信頼区間となる。