

ブートストラップ法入門

汪 金芳

千葉大学 大学院自然科学研究科

平成 17 年 5 月 13 日

目次

1	ジャックナイフ法による誤差の推定	2
1.1	分散と偏りの伝統的推定法	2
1.2	ジャックナイフ法による偏りの推定	2
1.3	ジャックナイフ法による分散の推定	3
2	ブートストラップ法による誤差の推定	3
2.1	分散のブートストラップ推定	3
2.2	偏りのブートストラップ推定	5
3	ブートストラップ信頼区間	6
3.1	信頼区間の良さの基準	6
3.2	パーセンタイル法	7
3.3	ブートストラップ t 法	7
3.4	BC_a 法	9
3.5	例：平均の比の信頼区間	10
4	平均の差のブートストラップ検定	11

1 ジャックナイフ法による誤差の推定

スカラー母数 θ を n 個の無作為標本に基づいて構成される推定量 $\hat{\theta}$ で推定することを考える。推定量 $\hat{\theta}$ はやはり確率変数であり、 $\hat{\theta}$ の性能を把握するために、 $\hat{\theta}$ の分布を調べなければならない。少なくとも、推定量としての良さを評価するために、 $\hat{\theta}$ の分散や偏りなどの程度を知りたい。この解説では、これらの問いに答えるためのいくつかの方法を紹介する。これらの方法の共通の点は、得られたデータから擬似データを構成し、計算機の繰り返し計算によって、推定量の精度や分布に関する情報を得ることである。

1.1 分散と偏りの伝統的推定法

まず推定量 $\hat{\theta}$ の分散 $\sigma^2 = E(\hat{\theta} - E\hat{\theta})^2$ と偏り $b = E(\hat{\theta} - \theta)$ の伝統的推定法について簡単に復習する。無作為標本 Y_1, \dots, Y_n が密度関数 $f(y|\theta)$ を持つ母集団から得られているとし、また $\hat{\theta}$ を最尤推定量とすると、 $\hat{\sigma}^2 = 1/(nI(\hat{\theta}))$ は n が大きいとき σ^2 の良い近似となることが多い。ここで、 $I(\theta) = E(\partial \log f(Y|\theta)/\partial \theta)^2$ は 1 標本単位あたりのフィッシャー情報量と呼ばれるものである。推定量 $\hat{\sigma}^2$ の計算に期待値の計算が含まれるが、これが面倒な場合には、 $\hat{\sigma}^2 = -1/\sum_{i=1}^n \ddot{\ell}(Y_i|\hat{\theta})$ で近似することもしばしばある。ここで、 $\ddot{\ell}(y|\theta) = \partial^2 \log f(y|\theta)/\partial \theta^2$ である。

$\hat{\theta}$ が最尤推定量でないときに、まったく別の方法を考える必要がある。一般論の展開は他の専門書に譲るが、無作為標本 $(U_1, V_1), \dots, (U_n, V_1)$ から計算される比推定量 $\hat{\theta} = \bar{U}/\bar{V}$ による平均の比 $\theta = EU/EV$ の推定の場合を考える。ここで、 $\bar{U} = \sum_{i=1}^n U_i/n$ 、 $\bar{V} = \sum_{i=1}^n V_i/n$ である。デルタ法と呼ばれる方法を用いて、 $\hat{\theta}$ の分散を

$$\hat{\sigma}^2 = \frac{1}{n} \left(\frac{\bar{U}}{\bar{V}} \right)^2 \left\{ \frac{S_u^2}{(\bar{U})^2} + \frac{S_v^2}{(\bar{V})^2} - \frac{2S_{uv}}{\bar{U}\bar{V}} \right\} \quad (1)$$

で推定される。ただし $S_u^2 = n^{-1} \sum_{j=1}^n (U_j - \bar{U})^2$ 、 $S_v^2 = n^{-1} \sum_{j=1}^n (V_j - \bar{V})^2$ 、 $S_{uv} = n^{-1} \sum_{j=1}^n (U_j - \bar{U})(V_j - \bar{V})$ である。

デルタ法を適用するために、推定量 $\hat{\theta}$ のある意味での滑らかさを持たなければならない。この方法は、刈込み平均や標本中央値などの平均の推定量には適用できない。たとえば、標本中央値 $\hat{\theta}$ の分散の推定量 $(4nf^2(\hat{\theta}))^{-1}$ を得るために、分位点に関する理論を展開しなければならない、また密度関数 $f(\theta)$ に関する知識も必要となる。

偏りの伝統的推定法については、分散の場合と同様に、個別論を展開する必要がある。たとえば、比推定量 $\hat{\theta} = \bar{U}/\bar{V}$ については、デルタ法を適用できるので、偏りは

$$\hat{b} = \frac{1}{n} \frac{\bar{U}}{\bar{V}} \left(\frac{S_u^2}{\bar{U}^2} - \frac{S_{uv}}{\bar{U}\bar{V}} \right) \quad (2)$$

と推定できる。

1.2 ジャックナイフ法による偏りの推定

母数 θ に対する一致推定量 $\hat{\theta} = \hat{\theta}(Y_1, \dots, Y_n)$ を考える。ここで、 Y_1, \dots, Y_n は母集団からの無作為標本で、また $\hat{\theta}$ の値は Y_1, \dots, Y_n の並べ換えに対して不変であると仮定す

る. i 番目の標本 Y_i を除いたときの推定量を, $\hat{\theta}_{(i)} = \hat{\theta}(Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n)$ で表し, $\hat{\theta}_{(\cdot)} = n^{-1} \sum_{i=1}^n \hat{\theta}_{(i)}$ とすると, ジャックナイフ偏り推定量は次のように定義される

$$\hat{b}_J = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta}). \quad (3)$$

この推定量の合理性を示すためには, 次に定義されるジャックナイフ偏り修正済み推定量

$$\tilde{\theta} = \hat{\theta} - \hat{b}_J = n\hat{\theta} - (n-1)\hat{\theta}_{(\cdot)} \quad (4)$$

の偏りが $\hat{\theta}$ のそれに比べると少なくなっていることを示せばよい.

さて, 多くの場合, $\hat{\theta}$ の期待値が次のように展開できる.

$$E\hat{\theta} = \theta + a_1/n + a_2/n^2 + \dots \quad (5)$$

ここで a_1, a_2, \dots は n とは無関係な定数であり, $b = a_1/n + a_2/n^2 + \dots$ が $\hat{\theta}$ の偏りである. また, $n-1$ 個の標本に基づく推定量 $\hat{\theta}_{(i)}$ については, 次が成り立つ.

$$E\hat{\theta}_{(i)} = \theta + a_1/(n-1) + a_2/(n-1)^2 + \dots \quad (6)$$

(5) 式と (6) 式から, $\tilde{\theta}$ の期待値が次のように計算される.

$$E\tilde{\theta} = E[n\hat{\theta} - (n-1)\hat{\theta}_{(\cdot)}] = \theta + a_2/(n(n-1)) + \dots \quad (7)$$

したがって, $\tilde{\theta}$ の偏り $a_2/(n(n-1)) + \dots = O(1/n^2)$ は $\hat{\theta}$ の偏り $a_1/n + a_2/n^2 + \dots = O(1/n)$ と比較して, (漸近的に) 小さくなっていることが分かる.

1.3 ジャックナイフ法による分散の推定

標本平均 $\hat{\theta} = \bar{Y}$ の分散の推定問題を考える. このとき, $\hat{\theta}_{(i)} = (n\hat{\theta} - Y_i)/(n-1)$, $\hat{\theta}_{(\cdot)} = \hat{\theta}$, $\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)} = (\bar{Y} - Y_i)/(n-1)$ とそれぞれ計算できる. また

$$\hat{\sigma}_J^2 = \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2 \quad (8)$$

と定義すると, $\hat{\theta} = \bar{Y}$ の場合, $\hat{\sigma}_J^2 = n^{-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 / (n-1)$ となり, $\hat{\theta}$ の分散 $\sigma^2 = n^{-1} \text{Var}(Y_1)$ の不偏推定量となる.

以上のことから, 一般の $\hat{\theta}$ についても, (8) 式で定義される $\hat{\sigma}_J^2$ を $\hat{\theta}$ のジャックナイフ分散推定量と呼ぶ. ただし前節と同様に, $\hat{\theta}$ の値は Y_1, \dots, Y_n の並べ換えに対して不変であると仮定する.

2 ブートストラップ法による誤差の推定

2.1 分散のブートストラップ推定

分布関数 $F(y)$ を持つ母集団からの無作為標本 Y_1, \dots, Y_n に対して,

$$F_n(y) = n^{-1} \sum_{i=1}^n \delta(Y_i \leq y)$$

を経験分布関数という。ただし、 $\delta(Y_i \leq y)$ は指標関数で、 $Y_i \leq y$ のときに 1、そうでなければ 0 である。図 4.6 では $n = 20$ 人の女子の体重データに基づいた経験分布関数を示してある。詳細は省略するが、Glivenko-Cantelli の定理によれば、 n が大きいときに、 $F_n(y)$ は $F(y)$ の良い近似となる。

さて、 E_F を真の分布 $F = F(y)$ のもとでの期待値とすれば、推定量 $\hat{\theta} = \hat{\theta}(Y_1, \dots, Y_n)$ の分散は $\sigma^2 = E_F(\hat{\theta} - E_F \hat{\theta})^2$ と書ける。もちろん、 F は分からないので、 σ^2 も分からない。ところで、 n がある程度大きいとき、 F を $F_n = F_n(y)$ で近似できるので、

$$\hat{\sigma}^2 = E_{F_n} \left[\hat{\theta}(Y_1^*, \dots, Y_n^*) - E_{F_n} \hat{\theta}(Y_1^*, \dots, Y_n^*) \right]^2 \quad (9)$$

が σ^2 の自然な推定量となる。上の式における Y_1^*, \dots, Y_n^* は、経験分布 F_n からの無作為標本で、ブートストラップ標本とよばれているものである。また $\hat{\theta}^* = \hat{\theta}(Y_1^*, \dots, Y_n^*)$ をブートストラップ推定量とよぶ。(9) 式で定義される推定量 $\hat{\sigma}^2$ を、ブートストラップ分散推定量とよぶ。

例 11.1 標本平均 $\hat{\theta} = \bar{Y} = \sum_{j=1}^n Y_j/n$ の分散に対する推定問題を考える。 $\hat{\theta}$ の分散は $\sigma^2 = n^{-1} \text{var } Y_1$ で、母分散 $\text{var } Y_1$ を標本分散 $S^2 = \sum_{j=1}^n (Y_j - \bar{Y})^2/n$ で差し込めば、 σ^2 は $n^{-1} S^2/n$ で推定できる。さて、 \bar{Y} のブートストラップ分散推定量を計算してみよう。いまブートストラップ推定量は $\bar{Y}^* = (Y_1^* + \dots + Y_n^*)/n$ で、 $E_{F_n} \bar{Y}^* = \bar{Y}$ となる。したがって、ブートストラップ分散推定量は

$$\hat{\sigma}^2 = E_{F_n} (\bar{Y}^* - \bar{Y})^2 = n^{-1} S_n^2$$

となり、この例の場合、従来の差込推定量と一致する。

一般の場合、上の例のようにブートストラップ分散推定量の理論的計算はできないが、次のアルゴリズムによって、どんな複雑な推定量の分散のモンテカルロ近似も算出できる。

アルゴリズム 11.1 (ブートストラップ分散の推定)

- (i) データ y_1, \dots, y_n から無作為復元抽出を n 回行い、大きさ n のブートストラップ標本 $Y_1^{*b}, \dots, Y_n^{*b}$ を構成し、ブートストラップ統計量 $\hat{\theta}^{*b} = \hat{\theta}(Y_1^{*b}, \dots, Y_n^{*b})$ を計算する。
- (ii) ステップ (i) を B 回繰り返すことにより $\hat{\theta}^{*1}, \dots, \hat{\theta}^{*B}$ を計算する。
- (iii) ブートストラップ分散推定量を次式により近似計算する。

$$\hat{\sigma}_b^2 = \sum_{b=1}^B (\hat{\theta}^{*b} - \hat{\theta}^*)^2 / (B - 1). \quad (10)$$

ただし、 $\hat{\theta}^* = B^{-1} \sum_{b=1}^B \hat{\theta}^{*b}$ はブートストラップ統計量のモンテカルロ平均である。

(9) 式の $\hat{\sigma}^2$ 代わりに、そのモンテカルロ近似である $\hat{\sigma}_b^2$ をブートストラップ分散推定量とよぶ場合もあるが、両者の違いに注意しておくことは重要である。ブートストラップ反復回数 B は、標本数 n の大きさと推定量の複雑さに応じて適切に定める必要があり、一般に n が大きければ B も大きくとる必要がある。

ブートストラップ分散推定法を，表 11.1 のダーウィンのとうもろこしデータ (竹内・大橋, 1981) に対して適用してみよう．このデータは，自家受精に対する他家受精の優越性の有無を検証するために取られた「とうもろこしの丈の長さ」を表している．

表 11.1 他家受精・自家受精したとうもろこしの丈に関する観測値 (1/8 インチ単位) ．

他家受精	188	96	168	176	153	172	177	163	平均	標準偏差
	146	173	186	168	177	184	96	161.53		
自家受精	139	163	160	160	147	149	149	122	平均	標準偏差
	132	144	130	144	102	124	144	140.60		

いま，表 11.1 のデータは対になっていると仮定して，比推定量と標本相関係数の場合について考える．比推定量の分散に対するデルタ推定量は，(1) で与えられている．また標本相関係数の分散に対するデルタ推定量も同様に導出できるが，煩雑であるためここでは省略する．表 11.2 には，いくつかの B の値に対するブートストラップ分散推定量の値，デルタ推定量の値およびジャックナイフ分散推定量の値を示している．ブートストラップ推定量とデルタ推定量の値は，比推定量の場合には非常に近く，またジャックナイフ推定量は他の推定量より僅かに小さくなっていることなどが読みとれる．標本相関係数の分散に対するジャックナイフ推定量が他の推定量の値よりかなり大きくなっている．ある種の条件の下で， $\hat{\sigma}_j^2 > \hat{\sigma}_b^2$ となることが証明できる．またこの例からも分かるように，推定量の分散の推定において，通常ブートストラップ反復回数 B をあまり大きくとる必要はない．

表 11.2 とうもろこしデータ (表 11.1) に対する，比推定量，標本相関係数の分散のブートストラップ推定量，ジャックナイフ推定量とデルタ推定量の値の比較 ．

推定量	B	ブートストラップ法	ジャックナイフ法	デルタ法
比推定量	100	0.0044	0.0053	0.0049
	200	0.0049		
	500	0.0051		
	1000	0.0048		
	2000	0.0050		
相関係数	100	0.0383	0.0503	0.0291
	200	0.0385		
	500	0.0418		
	1000	0.0385		
	2000	0.0371		

2.2 偏りのブートストラップ推定

分散のブートストラップ推定量の考え方と同様に，偏り $b = E_F \hat{\theta} - \theta$ のブートストラップ推定量を

$$\hat{b} = E_{F_n} \hat{\theta}(Y_1^*, \dots, Y_n^*) - \hat{\theta} \quad (11)$$

で定義する。また、分散の場合と同様、 \hat{b} は通常次のモンテカルロ近似で計算される。

$$\hat{b}_b = B^{-1} \sum_{b=1}^B \hat{\theta}(Y_1^{b*}, \dots, Y_n^{b*}) - \hat{\theta}. \quad (12)$$

例 11.2 母分散に対する標本分散推定量 $\hat{\theta} = n^{-1} \sum_{j=1}^n (Y_j - \bar{Y})^2$ の偏りのブートストラップ推定量を計算してみる。このとき、 $\hat{\theta}$ の偏りは $b = -\sigma^2/n$ である。経験分布 F_n の分散が $\hat{\theta}$ であることに注意すれば、ブートストラップ偏り推定量は

$$\begin{aligned} \hat{b} &= E_{F_n} \left[n^{-1} \sum_{j=1}^n (Y_j^* - \bar{Y}^*)^2 \right] - \hat{\theta} \\ &= (n-1)n^{-1}\hat{\theta} - \hat{\theta} = -n^{-1}\hat{\theta} \end{aligned}$$

と計算できる。これを $\hat{\theta}$ の真の偏りと比較すれば、ブートストラップ偏り推定量は $\hat{\theta}$ の偏りの差込推定量となっていることが分かる。

偏りの推定量を利用し、ブートストラップ偏り修正済み推定量 $\tilde{\theta} = \hat{\theta} - \hat{b}$ を構成することができる。偏り修正済み推定量の利用は、不偏性が意味をもつことを前提にしている。しかし、推定量の不偏性は、パラメータ変換に関して不変性をもたないことや、不偏推定量が存在しないこと、また尤度原理と矛盾するなどのことから、吟味せずに偏りの少ない推定量を選択することは、必ずしも望ましいこととは限らないことに注意が必要である。

3 ブートストラップ信頼区間

本節では、信頼区間を構成するための3種類のブートストラップ法について述べる。最も簡単な方法は、パーセンタイル法とよばれるものである。第2の方法はブートストラップ t 法とよばれるものであるが、この方法では推定量の分散の推定が必要となる。分散推定値があまり信頼できない場合には、この方法の適用には注意が必要である。第3の方法は BC_a 法とよばれるもので、パーセンタイル・ブートストラップ法を改良した方法である。

3.1 信頼区間の良さの基準

ブートストラップ法により信頼区間を構成する方法やその性質について解説する前に、信頼区間の良さの基準などについて概説しておこう。

n 個の無作為標本に基づいて、パラメータ θ に対する推定量 $\hat{\theta}$ を考える。いま $\hat{\sigma}^2/n$ を $\hat{\theta}$ の分散に対する一致推定量とすれば、多くの場合、中心極限定理により、標本数 n が大きくなれば、 $T = \sqrt{n}(\hat{\theta} - \theta)/\hat{\sigma}$ の分布が標準正規分布 $N(0, 1)$ により近似できる。この近似の程度は、一般的には1次の正確度しかもっていない。すなわち

$$\Pr \left\{ \sqrt{n}(\hat{\theta} - \theta)/\hat{\sigma} \leq t \right\} = \Phi(t) + O(n^{-1/2}) \quad (13)$$

が成り立つ．ここで $\Phi(t)$ は，標準正規分布 $N(0, 1)$ の分布関数を表している．また，記号 $O(n^{-1/2})$ は $n^{-1/2}$ と同程度の小さい量を表している．

さて， z_α を $N(0, 1)$ の 100α 番目のパーセント点，すなわち $\Phi(z_\alpha) = \alpha$ を満たす点とする．(13) より，次式が成り立つことが分かる．

$$\Pr\{\theta \leq \hat{\theta} - n^{-1/2}\hat{\sigma}z_\alpha\} = 1 - \alpha + O(n^{-1/2}) \quad (14)$$

したがって， $I = (-\infty, \hat{\theta} + n^{-1/2}\hat{\sigma}z_{1-\alpha})$ とすると，(14) より，

$$\Pr\{\theta \in I_L\} - (1 - \alpha) = O(n^{-1/2}). \quad (15)$$

このとき信頼度 $1 - \alpha$ の下側信頼区間 I は，1 次の正確度をもつという．ブートストラップ t 法や， BC_α 法などの方法では，上記の信頼区間 I_L の上側端点をブートストラップ標本分布から計算するが，その結果得られるブートストラップ信頼区間 \hat{I} は 2 次の正確度をもつ．すなわち，次が成り立つ．

$$\Pr\{\theta \in \hat{I}\} - (1 - \alpha) = O(n^{-1}) \quad (16)$$

が成り立つ．

3.2 パーセンタイル法

パーセンタイル法による信頼区間は，(14) から導出された正規理論に基づく信頼区間と，同じ 1 次の正確度をもっている．しかしこの方法は，正規理論に基づく方法と比較するとより一般的な状況に適用でき，またアルゴリズムが簡単かつ“変換保存性”をもっているので，標本数 n が小さいか中程度の場合には適用してみる価値があろう．

さて，パラメータ θ に対する信頼度 $1 - \alpha$ の信頼区間を構成するための手順を示しておこう．

アルゴリズム 11.2 (パーセンタイル・ブートストラップ信頼区間)

- (i) 無作為標本 y_1, \dots, y_n から，復元抽出により互いに独立なブートストラップ標本 Y_1^*, \dots, Y_n^* を抽出する．
- (ii) ブートストラップ推定量 $\hat{\theta}^* = \hat{\theta}(Y_1^*, \dots, Y_n^*)$ の値を計算する．これを B 回繰り返して， $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ を求める．
- (iii) $\hat{\theta}_\alpha = \hat{\theta}_{(\alpha B)}^*$ とする．ここで αB は自然数とし， $\hat{\theta}_{(\alpha B)}^*$ は $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ に対する αB 番目の順序統計量とする．このとき $(\hat{\theta}_\alpha, \infty)$ は， θ に対する名目上の被覆確率 $1 - \alpha$ をもつ上側信頼区間となる．下側信頼区間や両側信頼区間も同様にして構成することができる．

3.3 ブートストラップ t 法

前節で述べたパーセンタイル法は，ブートストラップ信頼区間を構成するために用いる推定量 $\hat{\theta}$ の分散推定値を求める必要がない点では，簡便な方法である．しかし，信頼区

間の被覆確率の観点からは満足できるものとは限らず，一般的には本節で述べる“ブートストラップ t 法”の方が精度が良い．ここでは， $\hat{\theta}$ の標準誤差に対する信頼できる推定値 $\hat{\sigma}/\sqrt{n}$ が得られることを仮定する．

(13) より，スチューデント化された量 $T = \sqrt{n}(\hat{\theta} - \theta)/\hat{\sigma}$ の分布に対する正規近似は，1 次の正確度しかもっていない．これに対して，ブートストラップ標本に基づく T^* の分布は，一般には T の分布に対して 2 次の正確度をもっている．ブートストラップ t 法のアイデアは，この事実に基づいたものである．実際比較的ゆるやかな条件の下で，次の 2 つの式が成り立つ．

$$\begin{aligned}\Pr\{T \leq t\} &= \Phi(t) + n^{-1/2}p(t)\phi(t) + O(n^{-1}), \\ \Pr\{T^* \leq t\} &= \Phi(t) + n^{-1/2}\hat{p}(t)\phi(t) + O_p(n^{-1}).\end{aligned}$$

ここで $p(t)$ は 2 次の偶関数多項式であり， $\hat{p}(t)$ は $p(t)$ に含まれる未知の量を標本から計算した推定値で置き換えた関数である．また $\Phi(t)$ ， $\phi(t)$ は，それぞれ標準正規分布の分布関数，密度関数を表している．これらの結果から

$$\Pr\{T \leq t\} - \Pr\{T^* \leq t\} = O_p(n^{-1}) \quad (17)$$

が成立し，したがって $\Pr\{T^* \leq t\}$ は 2 次の正確度をもつことが分かる．同様な議論を行えば， T^* の分布の 100α 番目のパーセント点 w_α は， T の分布の 100α 番目のパーセント点の近似として，2 次の正確度をもつことが分かる．これに対して正規近似に基づくパーセント点 z_α は，1 次の正確度しかもっていない．以上の議論より，

$$\Pr\left\{\theta \leq \hat{\theta} - n^{-1/2}\hat{\sigma}w_\alpha\right\} = 1 - \alpha + O(n^{-1}) \quad (18)$$

が成立することが分かる．したがって，ブートストラップ t 信頼区間 $(-\infty, \hat{\theta} - n^{-1/2}\hat{\sigma}w_\alpha)$ は，2 次の正確度をもつことになる．ブートストラップ t 信頼区間と正規近似による信頼区間の違いは，(14)における正規分布のパーセント点 z_α が，(18)では T^* のブートストラップ分布のパーセント点 w_α によって置き換えられているだけである．

次にブートストラップ t 法の手順をまとめておこう．

アルゴリズム 11.3 (ブートストラップ t 法)

- (i) 無作為標本 y_1, \dots, y_n から復元抽出により，互いに独立なブートストラップ標本 Y_1^*, \dots, Y_n^* を抽出する．
- (ii) ステップ (i) で得られたブートストラップ標本を用いて， $\hat{\theta}$ および $\hat{\sigma}$ に対するブートストラップ推定値 $\hat{\theta}^*$ および $\hat{\sigma}^*$ を計算する．
- (iii) ブートストラップ t 値 $T^* = \sqrt{n}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$ を計算する．これを B 回繰り返し， T_1^*, \dots, T_B^* を求める．
- (iv) $w_\alpha = T_{(\alpha B)}^*$ とする．ここで αB は自然数とし，また $T_{(\alpha B)}^*$ は，アルゴリズム 11.2 と同様な順序統計量である．このとき $(-\infty, \hat{\theta} - n^{-1/2}\hat{\sigma}w_\alpha)$ は名目上の被覆確率 $1 - \alpha$ をもつ下側信頼区間となる．上側信頼区間や両側信頼区間も同様にして構成することができる．

3.4 BC_a 法

パーセンタイル信頼区間は, 推定量 $\hat{\theta}$ の偏りと, $\hat{\theta}$ の分布の歪みの影響を直接的に受け
てしまい, 1 次の正確度しかもたない. 推定量の偏りとその分布の歪みを同時に補正する
BC_a 法による信頼区間は 2 次の正確度を持つ. この方法を適用するためには, 偏り修正
量 z_0 と, 加速定数とよばれる量 a を推定しなければならない.

さて $\hat{\theta}^*$ を, これまでのブートストラップ推定量とする. このとき推定量 $\hat{\theta}$ の値は, $\hat{\theta}^*$
のブートストラップ分布の中央部分に位置する場合もあれば, そうでない場合もあろう.
この食い違いを表す量, $z_0 = \Phi^{-1}(\Pr[\hat{\theta}^* \leq \hat{\theta}])$ は, 推定値 $\hat{\theta}$ の θ に対する偏りの程度を
表す 1 つの尺度と考えられる. ただし Φ は, 標準正規分布の分布関数である. もし $\hat{\theta}^*$ の
ブートストラップ分布の内のちょうど半分が $\hat{\theta}$ の値より小さければ, $\Pr[\hat{\theta}^* \leq \hat{\theta}] = 0.5$ で
あり, したがって $z_0 = 0$ となる. この偏り修正量 z_0 のモンテカルロ推定値は, ブートス
トラップ推定値 $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ を用いて, 次のように計算される.

$$\hat{z}_0 = \Phi^{-1} \left(\#\{\hat{\theta}_b^* < \hat{\theta}\} / B \right) \quad (19)$$

次に加速定数 a について考える. 推定量 $\hat{\theta}$ が正則な場合 (詳細は汪・田栗 (2003) を参
照)), 次のようにして a を推定する. $\hat{\theta}_{(i)}$ を, ジャックナイフ法を適用したときと同様に,
 i 番目のデータ y_i を取り除いたものから計算された $\hat{\theta}$ の値とし, また $\hat{\theta}_{(\cdot)} = n^{-1} \sum_{i=1}^n \hat{\theta}_{(i)}$
とする. 加速定数 a は次式により推定することができる.

$$\hat{a} = \left\{ \sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^2 \right\}^{-3/2} \sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^3 / 6. \quad (20)$$

アルゴリズム 11.4 (BC_a 法)

1. 無作為標本 y_1, \dots, y_n からの独立な復元抽出により, ブートストラップ標本 Y_1^*, \dots, Y_n^*
を求める.
2. ブートストラップ推定値を, $\hat{\theta}^* = \hat{\theta}(Y_1^*, \dots, Y_n^*)$ により計算する. これを多数回 (B
回) 繰り返し, $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ を求める. ここまで, パーセンタイル法と同じ手順である.
3. 偏り修正推定量 \hat{z}_0 および加速定数推定量 \hat{a} を, (19) および (20) により計算する.
4. $(\hat{\theta}_{(\hat{\alpha}B)}^*, \infty)$ を θ に対する信頼度 $1 - \alpha$ の上側信頼区間とする. ここで任意の α ($0 <$
 $\alpha < 1$) に対して, $\hat{\alpha}$ は次により求める.

$$\hat{\alpha} = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z_\alpha}{1 - \hat{a}(\hat{z}_0 + z_\alpha)} \right). \quad (21)$$

下側信頼区間の構成法も同様である. また信頼度 $1 - 2\alpha$ の両側信頼区間は $(\hat{\theta}_{(\hat{\alpha}B)}^*, \hat{\theta}_{(\widehat{1-\alpha}B)}^*)$
で与えられ, ただし $\widehat{1-\alpha}$ は, (21) の z_α を $z_{1-\alpha}$ で置き換えたものである.

BC_a 法による信頼区間の計算量は, 対応するパーセンタイル法の場合と同程度ですむ
が, BC_a 法は 2 次の正確度をもっている. BC_a 法における順序統計量 $\hat{\theta}_{(\hat{\alpha}B)}^*, \hat{\theta}_{(\widehat{1-\alpha}B)}^*$ を
求める際に, (21) により調整を行っている. $\hat{a} = 0$ の場合には, $\hat{\alpha} = \Phi(z_\alpha + 2\hat{z}_0)$ とな
る. また \hat{a} と \hat{z}_0 が両方とも 0 の場合には, $\hat{\alpha} = \alpha$ となる.

3.5 例：平均の比の信頼区間

表 11.3 では， $n = 8$ 人の患者に対して，認可薬，新薬およびプラセボ投与後の血液レベルの測定値を示している．新薬と認可薬の生物学的同等性を検証するため，

$$X_i = \text{認可薬} - \text{プラセボ}, \quad Y_i = \text{新薬} - \text{認可薬}, \quad i = 1, \dots, n$$

とし，平均の比 $r = E(Y)/E(X)$ に対する信頼区間を構築したい．もし r に対する信頼度 $1 - \alpha$ の信頼区間が，あらかじめ定められた区間 $[-a, a]$ ($a > 0$) に含まれている場合には，新薬と認可薬は生物学的同等性をもつと認める．ここで， $(\alpha, a) = (0.1, 0.2)$ とする．

表 11.3 認可薬，新薬およびプラセボ投与後の血液レベル (Efron & Tibshirani (1993, p.373)) .

患者	認可薬	新薬	プラセボ
1	17649	16449	9243
2	12013	14614	9671
3	19979	17274	11792
4	21816	23798	13357
5	13850	12560	9055
6	9806	10157	6290
7	17208	16570	12412
8	29044	26325	18806

表 11.3 のデータから，比推定量の値を計算すると $\hat{r} = \bar{y}/\bar{x} = -0.0713$ となる．この値は，あらかじめ定められた区間 $[-0.2, 0.2]$ の真ん中辺りに入っているため，生物学的同等性があるだろうと想像できよう．また図 11.1 (左) は， $B = 2000$ 回の繰り返しに基づく \hat{r}^* のブートストラップ分布を表しているが，これを見ても生物学的同等性が成り立ちそうである．ところが，図 11.1 (左) による信頼度 90% のパーセンタイル信頼区間は $(-0.208, 0.105)$ となり，標準区間 $[-0.2, 0.2]$ からはみ出してしまい，2 剤に生物学的同等性が否定されることになる．

さて， BC_a 法による信頼区間を求めてみよう．図 11.1 (左) より，偏り修正量の推定値は $\hat{z}_0 = \Phi^{-1} \{ \#(\hat{r}^{*b} < \hat{r}) / B \} = 0.029$ となり， $\hat{z}_0 > 0$ であるから，半数以上のブートストラップ推定値が観測値 \hat{r} の左側にある．またこの図から分かるように，この場合のブートストラップ分布はやや右に裾をひいている．したがって分布の中央部分では左側に確率が多いことになり，分布が対称の場合と比べて信頼区間が左側にずれることになる．加速定数の推定値は，式 (20) より $\hat{a} = 0.024$ となる．

ここで $\alpha = 0.05$ とし，上で求めた \hat{z}_0, \hat{a} の値を用いると， $\hat{\alpha} = 0.0634$ ， $1 - \hat{\alpha} = 0.9619$ となる．この場合， $\hat{\alpha} + 1 - \hat{\alpha} \neq 1$ であることに注意してほしい．以上より， BC_a 法による信頼区間は $(-0.201, 0.131)$ となり，やはり下側の信頼限界が標準区間に入っていない．

比較するため，図 11.1 (右) では， $B = 2000$ のときのブートストラップ t 統計量 $T^* = \sqrt{n}(\hat{r}^* - \hat{r})/\hat{\sigma}^*$ のヒストグラムを示している．ここで $\hat{\sigma}^2$ は (1) によって計算された．いま

の場合、信頼度 90% のブートストラップ t 信頼区間は $(-0.226, 0.319)$ となり、パーセントイル法と BC_a 法による信頼区間よりかなり長くなっていることに注意してほしい。いずれの場合、2 剤間に生物学的同等性があるとの確証は得られない。

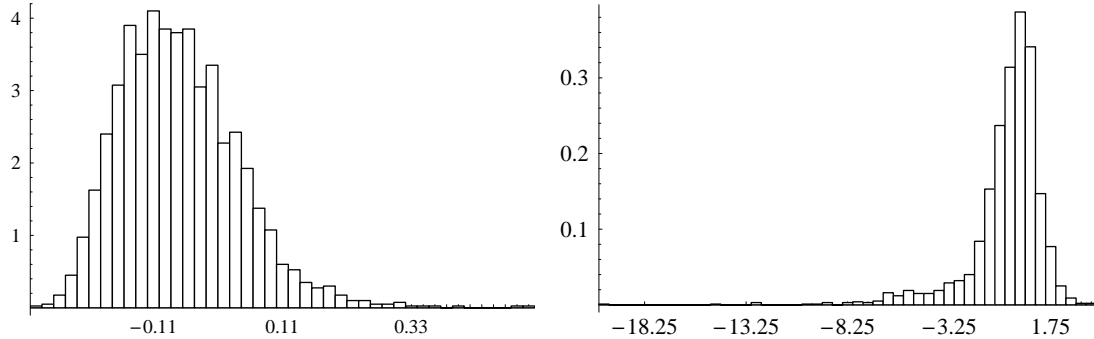


図 11.1 表 11.3 のデータに対する、 $B = 2000$ 回の繰り返しの場合の $\hat{r}^* = \bar{Y}^*/\bar{X}^*$ (左) と $T^* = \sqrt{n}(\hat{r}^* - \hat{r})/\hat{\sigma}^*$ (右) のブートストラップ分布 ($\hat{r} = -0.0713$)。

4 平均の差のブートストラップ検定

最後のこの節では、表 11.2 のダーウィンの実験データを用いて、平均の差のブートストラップ検定について述べる。他家受精と自家受精によるとうもろこしの丈の長さを、それぞれ x, y で表し、それらの平均を μ_x, μ_y とする。いま検定統計量 $t = (\bar{x} - \bar{y})/\sqrt{s_x^2/(m-1) + s_y^2/(n-1)}$ を用いて仮説 $\mu_x = \mu_y$ に対する検定問題を考える。ここで、 $\bar{x} = m^{-1} \sum_{i=1}^m x_i, \bar{y} = n^{-1} \sum_{j=1}^n y_j$ は両群の標本平均であり、 $s_x^2 = m^{-1} \sum_{i=1}^m (x_i - \bar{x})^2, s_y^2 = n^{-1} \sum_{j=1}^n (y_j - \bar{y})^2$ はそれぞれの標本分散である。ダーウィンの例では $t = 2.44$ である。

ブートストラップ検定法では、仮説 $\mu_x = \mu_y$ のもとでの検定統計量 t の分布をブートストラップ標本を用いて近似する。帰無仮説を満たす母集団からの標本抽出を行うため、標本を次のように変換を行う。

$$\begin{aligned} x_i^\dagger &= x_i - \bar{x} + \bar{z}, & i = 1, \dots, m, \\ y_j^\dagger &= y_j - \bar{y} + \bar{z}, & j = 1, \dots, n \end{aligned}$$

ただし、 $\bar{z} = (m+n)^{-1}(m\bar{x} + n\bar{y})$ は全体をプールしたときの標本平均である。変換後のデータに基づく経験分布を、それぞれ

$$F_m(x) = m^{-1} \sum_{i=1}^m \delta(x_i^\dagger \leq x), \quad G_n(y) = n^{-1} \sum_{j=1}^n \delta(y_j^\dagger \leq y)$$

とする。 $F_m(x)$ と $F_n(y)$ は同じ平均 \bar{z} を持ち、帰無仮説の要請を満たす。

次に、ブートストラップ検定統計量を、 F_m からの無作為標本を X_1^*, \dots, X_m^* と G_n からの無作為標本を Y_1^*, \dots, Y_n^* に基づいて、

$$t^* = (\bar{X}^* - \bar{Y}^*)/\sqrt{S_x^{*2}/(m-1) + S_y^{*2}/(n-1)}$$

と計算される。ただし \bar{X}^* , \bar{Y}^* , S_x^{*2} , S_y^{*2} は、それぞれのブートストラップ標本平均と標本分散である。

最後に、 t^* のブートストラップ分布のもとで、 p 値、 $p = \Pr\{t^* \geq t\}$ を計算する。有意水準が α に対して、 $p \geq \alpha$ のとき、帰無仮説を採択し、そうでなければ、帰無仮説を棄却する。ここで、ブートストラップ p 値は、次のモンテカルロ近似により計算される。

$$\hat{p} = \frac{1}{B} \sum_{b=1}^B \delta(t^{*b} \geq t). \quad (22)$$

$B = 10000$ とし、(22) 式を表 11.1 のデータに適用したところ、 $\hat{p} = 0.0406$ となった。したがって有意水準を $\alpha = 0.05$ と設定した場合には、他家受精法が自家受精法より優れていると判断されることになる。

ブートストラップ検定法の詳細については、汪・田栗 (1996, 2003) を参照のこと。

参考文献

- [1] Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*, Cambridge University Press: Cambridge.
- [2] Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall: New York.
- [3] 小西貞則 (1990). ブートストラップ法と信頼区間の構成, 応用統計学, 19, 137–162.
- [4] 小西貞則, 本多正幸 (1992). 判別分析における誤判別率推定とブートストラップ法, 応用統計学, 21, 67–100.
- [5] 久保川 達也, 江口 真透, 竹村 彰通, 小西 貞則 (1993). 統計的推測理論の現状, 日本統計学会誌, 22, 257–312.
- [6] 竹内啓, 大橋靖雄 (1981). 「統計的推測 — 2 標本問題」, 日本評論社, 東京.
- [7] 汪金芳, 大内俊二, 景平, 田栗正章 (1992). ブートストラップ法 — 最近までの発展と今後の展望, 行動計量学, 19, 50–81.
- [8] 汪金芳, 田栗正章 (1996). ブートストラップ法 — 2 標本問題からの考察, 統計数理, 44, 3–18.
- [9] 汪金芳, 田栗正章 (2003). 「統計科学のフロンティア: 新しい計算手法と統計学, 第 11 巻, 計算統計 I — 確率計算の新しい手法; ブートストラップ法の基礎」, 岩波書店.