層化抽出法に基づく母平均の信頼区間の構成について

佐 井 至 道*

Confidence Intervals for Population Mean under Stratified Sampling

Shido SAI

In practical sample surveys, stratified sampling is widely used to estimate population parameters with high precision. Although point estimation for a population mean was investigated from all viewpoints, interval estimation has not been sufficiently studied. This paper is concerned with several kinds of confidence intervals based on stratified random samples.

Firstly we propose a confidence interval procedure for the population mean that uses Student's t -distribution, and also show that t -statistic can be constructed from stratified random sample when the objective variable is normal distributed in each stratum and the sample is drawn with Neyman allocation (the optimum allocation).

If this sample allocation method is not used, the confidence interval is improved using Satterthwaite approximation for the distribution of the statistic unless the strata are terribly skew. In the difficult situation, we obtain the aid of the nonparametric bootstrap method.

The coverage probabilities and some other properties for all methods considered in this paper are investigated through the simulation study for some theoretical distributions and actual data.

1. はじめに

層化無作為抽出は単純無作為抽出と比較して推定の精度が高くなるため、実際の調査においてもしばしば用いられている。(例えば、浅井(1987), NHK放送文化研究所(1996)を参照していただきたい。)層化無作為標本に基づく母平均の点推定については既に様々な角度から議論が行われてきたが、信頼区間の構成法についてはこれまでほとんど議論されていなかった。そのため正規分布または t 分布を用いて単純無作為抽出の理論をそのまま流用しているのが現状であろう。(Levy and Lemeshow (1991)を参照のこと。)

理論が未完成である理由の一つとして、単純無作為抽出の場合に用いられる t 統計量が、層化無作為標本からは容易に構成できない点があげられる。それは分

布の形によるばかりでなく、各層からサンプリングする標本の大きさの決め方(標本配分法)とも関係が深い。Satterthwaite (1946)は、構成された統計量の分布を t 分布で近似する際の自由度の計算法を考案したが、層化抽出法におけるその有効性はAmes and Webster (1991)などによってわずかに議論されているのみである。また、実データに適用するときには発生する種々の問題が十分検討されていないため、実際にはあまり利用されていない。

他方、ブートストラップ法が考案されて間もなくBabu and Singh (1983)などによって層化無作為抽出へも自然な拡張がなされた。その基本となる考え方は、Sitter (1992), Shao and Tu (1995)などに見られるように現在でもほとんど変わらない。

本論文では、区間推定を目的として、層化無作為標本から母平均の信頼区間を構成する方法について考えていく。

まず、単純無作為抽出の場合と同様に t 分布を利用する方法を2節で提案し、 t 統計量の構成が可能となる条件についても考える。3節では、人工的に作られた

* 岡山商科大学商学部 (Faculty of Commerce, Okayama Shoka University)
連絡先氏名 佐井至道
郵便宛先 〒700 岡山市津島京町2-10-1
電話番号 086-252-0642

種々の分布や標本配分法などの組合せに対しモンテカルロシミュレーションを行うことによって、提案した方法の適用可能な状況について検討する。さらに4節では、Satterthwaiteの近似法を提案した方法に応用することによって、信頼区間の改善を行う。

また5節では、改善を行っても t 分布を利用する信頼区間の構成法に頼れない場合に、これに代わる方法であるブートストラップ法による信頼区間の構成法を考える。またこの節でもシミュレーションを行い、提案した方法とブートストラップ法との比較を行う。

最後に6節では、実データに対して提案した方法を含む種々の方法を適用する。統計理論を実際の調査に応用する場合には様々な問題が起きてくる。そのような問題を克服してこそ、初めて理論が生きてくる。複雑なサンプリング法の理論の敬遠される原因の一つがここにあると思われる。この節ではそのような問題を含めて、信頼区間の構成法の善し悪しを考えていく。

2. 母平均の信頼区間の構成

2.1 単純無作為抽出法に基づく信頼区間の構成

一変量の分布から標本が単純無作為抽出される場合、母平均の信頼区間の構成には t 統計量を利用するのが一般的である。(竹内(1975), Thompson(1992)などを参照していただきたい。)提案する方法との比較のため、まず初めにその概略を述べる。

一変量の正規分布 $N(\mu, \sigma^2)$ (μ, σ^2 : 未知)をもつ母集団から、大きさ n の標本 $X_1, X_2, \dots, X_n$ が単純無作為抽出される場合、標本平均と標本分散を $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$, $S^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2$ と定義すると、 $U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$, $V^2 = \frac{nS^2}{\sigma^2} \sim \chi^2(n-1)$ となり、しかも U と V^2 は独立であるから、統計量

$$T = \frac{\sqrt{n-1}U}{V} = \frac{\sqrt{n-1}(\bar{X} - \mu)}{S} \quad (2.1)$$

は自由度 $n-1$ の t 分布に従う。この t 分布の上側確率 $\alpha/2$ のパーセント点を $t_{1-\alpha/2}$ とすると、母平均 μ の信頼度 $1-\alpha$ の信頼区間 I_{srs} は

$$I_{srs} = \left[\bar{X} - t_{1-\alpha/2} \frac{1}{\sqrt{n-1}} S, \bar{X} + t_{1-\alpha/2} \frac{1}{\sqrt{n-1}} S \right] \quad (2.2)$$

と求められる。ただし、“srs”は単純無作為抽出法

(simple random sampling)を意味している。

これから以後、信頼区間としては母平均の不偏推定量を中心とする左右対称な区間のみを考えていく。

2.2 層化無作為抽出法に基づく信頼区間の構成

この節では、層化無作為抽出された標本からの信頼区間の構成法を提案し、それが可能となる条件について考えていく。

母集団 F_0 が L 個の層に分けられていて、第 i 層が一変量の正規分布 $N(\mu_i, \sigma_i^2)$ をもつと仮定する。また第 i 層の重み(母集団全体に対する第 i 層の比率)を w_i と表し、第 i 層から大きさ n_i の標本 $X_{i1}, X_{i2}, \dots, X_{in_i}$ が他の層とは独立に無作為抽出されるものとする。ただし $\sum_{i=1}^L w_i = 1$, $\sum_{i=1}^L n_i = n$ であり、 w_i は既知とし、 μ_i, σ_i^2 は未知とする。実際の調査では、年齢、性別や地理的な特性などによって層化されることが多く、各層の重みは既知または高い精度で推定可能な場合が多い。

このような状況では、母平均 μ の推定には次のような不偏推定量を用いるのが一般的である。

$$\bar{X}_{st} = \sum_{i=1}^L w_i \bar{X}_i \quad \left(\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \right). \quad (2.3)$$

なお、“st”は層化無作為抽出法(stratified random sampling)を意味している。この推定量について

$$\bar{X}_{st} \sim N\left(\mu, \sum_{i=1}^L \frac{w_i^2}{n_i} \sigma_i^2\right) \quad (2.4)$$

となることは容易に求められる。

そこで、単純無作為抽出の場合の統計量 U にならって同様の統計量 U_{st} を次のように定義するのが自然であろう。

$$U_{st} = \frac{\bar{X}_{st} - \mu}{\sqrt{\sum_{i=1}^L \frac{w_i^2}{n_i} \sigma_i^2}}. \quad (2.5)$$

また統計量 V_{st}^2 を

$$V_{st}^2 = \frac{n \cdot \sum_{i=1}^L \frac{w_i^2}{n_i} S_i^2}{\sum_{i=1}^L \frac{w_i^2}{n_i} \sigma_i^2} \quad (2.6)$$

と定義する。ただし

$$S_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \quad (2.7)$$

である。

(2.4) 式より $U_{st} \sim N(0, 1)$ が得られ、明らかに U_{st} と V_{st}^2 は独立なので、もし V_{st}^2 が χ^2 分布に従うならば t 統計量を構成することができる。ここで、(2.6) 式の分子は

$$n \cdot \sum_{i=1}^L \frac{w_i^2}{n_i} S_i^2 = n \cdot \sum_{i=1}^L \frac{w_i^2 \sigma_i^2}{n_i^2} \cdot \frac{n_i S_i^2}{\sigma_i^2} \quad (2.8)$$

と変形され、

$$\frac{n_i S_i^2}{\sigma_i^2} \sim \chi^2(n_i - 1) \quad (2.9)$$

であるから、 i によらず

$$\frac{n \cdot \frac{w_i^2 \sigma_i^2}{n_i^2}}{\sum_{i=1}^L \frac{w_i^2}{n_i} \sigma_i^2} = 1 \quad (2.10)$$

となるのが V_{st}^2 が χ^2 分布に従うための必要十分条件である。ところで、実際の調査において各層の重み w_i を調査者が決定することも場合によっては可能であるが、最もコントロールしやすい量は各層の標本の大きさ n_i である。そこで、上の条件を標本配分法として表すと

$$n_i = \frac{w_i \sigma_i}{\sum_{k=1}^L w_k \sigma_k} n \quad (2.11)$$

となるが、これはネイマン配分法 (最適配分法) である。任意の層別に対して、推定量の分散 $V(\bar{X}_{st})$ はネイマン配分法を用いたときに最小となる。ネイマン配分法については Cochran (1977), Levy and Lemeshow (1991) を参照されたい。

このとき統計量

$$T_{st} = \frac{\sqrt{n-L} U_{st}}{V_{st}} = \sqrt{\frac{n-L}{n}} \cdot \frac{\bar{X}_{st} - \mu}{\sqrt{\sum_{i=1}^L \frac{w_i^2}{n_i} S_i^2}} \quad (2.12)$$

は自由度 $n-L$ の t 分布に従い、この t 分布の上側確率 $\alpha/2$ のパーセント点を $t_{1-\alpha/2}$ とすると、母平均 μ の信頼度 $1-\alpha$ の信頼区間 I_{st} は

$$I_{st} = [\bar{X}_{st} - t_{1-\alpha/2} D, \bar{X}_{st} + t_{1-\alpha/2} D] \quad (2.13)$$

と求められる。ただし

$$D = \sqrt{\frac{n}{n-L}} \sqrt{\sum_{i=1}^L \frac{w_i^2}{n_i} S_i^2} \quad (2.14)$$

である。

(2.6) 式の V_{st}^2 の代わりに統計量

$$V_1^2 = \sum_{i=1}^L \frac{n_i S_i^2}{\sigma_i^2} \quad (2.15)$$

を用いると V_1^2 は常に自由度 $n-L$ の χ^2 分布に従うが、統計量 $\frac{\sqrt{n-L} U_{st}}{V_1}$ の中に母数 σ_i^2 が残ってしまい実際には信頼区間を構成できない。

また Thompson (1992) などに見られるように、過去に層化無作為標本から t 分布に従う統計量を構成しようとしたときには、(2.1) 式を参考に

$$\frac{\bar{X}_{st} - \mu}{\sqrt{\sum_{i=1}^L \frac{w_i^2}{n_i - 1} S_i^2}} \quad (2.16)$$

という形の統計量がいわれてきた。これは (2.6) 式の V_{st}^2 の代わりに

$$V_2^2 = \frac{(n-L) \sum_{i=1}^L \frac{w_i^2}{n_i - 1} S_i^2}{\sum_{i=1}^L \frac{w_i^2}{n_i} \sigma_i^2} \quad (2.17)$$

を用いていることになる。なお、(2.16) 式は本論文の無限母集団の仮定に合わせて書き直したものである。

ネイマン配分法と、一般の調査で用いられる比例配分法 ($n_i = w_i n$)、等配分法 ($n_i = n/L$) の各標本配分法における3つの統計量 V_{st}^2 , V_1^2 , V_2^2 の関係を付録に示す。 V_{st}^2 はネイマン配分法の場合には V_1^2 と一致し、等配分法の場合には V_2^2 と一致する。

各層が正規分布に従う場合には一般に母集団全体は正規分布に従わないため、 I_{srs} を I_{st} と単純には比較できないが、それを無視すると信頼区間の幅は2つの推定量の標準誤差 $SD(\bar{X})$ と $SD(\bar{X}_{st})$ にほぼ比例する。例えば、標準正規分布から単純無作為抽出する場合と、分布を左右対称に2層に分け各層から半分ずつの大きさの標本を無作為抽出する場合を考えると、 $SD(\bar{X}_{st})/SD(\bar{X})$ の値は .6028 となる。層化無作為標本を用いて信頼区間を構成する最も大きな利点はその幅の狭さにあり、その可能となる条件がネイマン配分法であることは、層化無作為抽出によって最も幅の小さい信頼区間が構成できることを保証するものである。

層化抽出法の一般的な理論については Cochran (1977), Thompson (1992) が詳しい。

3. シミュレーションによる信頼区間の検討

3.1 モンテカルロシミュレーションの設計

前節で述べたように、ネイマン配分法を用いると母平均の信頼区間を構成することが可能であるが、各層の標本の大きさ n_i を決定するためには、層内標準偏差 σ_i が既知または推定可能でなければならない。そのため一般の調査では、前節で述べた比例配分法や等配分法などのパラメータの推定を必要としない標本配分法を用いることが多い。しかしその場合 (2.12) 式の T_{st} は t 分布に従わなくなる。また仮にネイマン配分法を用いても、各層が正規分布に従わなければ t 統計量は作れない。このような場合には前節の議論には頼れなくなるが、もし統計量 T_{st} の分布が比較的 t 分布に近ければ、信頼区間 I_{st} を近似的に用いても特に問題がないと思われる。

そこでこの節では、人工的に作った種々の理論分布と層別法に対して、いくつかの標本配分法を用いてサンプリングを行った場合の、統計量 T_{st} の分布と母平均の信頼区間をモンテカルロシミュレーションによって求めていく。

まず、分布と層別法は正規分布と確率密度関数

$$f(x) = \frac{ak^a}{x^{a+1}} \quad (3.1)$$

のパレート分布を用いて次のような場合を考える。なお層の数はすべて2である。

- (A) 正規分布の区間層別
(標準正規分布を $w_1=.5, .6, .7, .8, .9$ となる層別点で2層に区間層別する。)
- (B) 2つの正規分布
(2つの同じ大きさの正規分布を層と考える。 $\sigma_1=1.0$ と固定し、 $\sigma_2=1.0, 2.0, 3.0, 4.0, 5.0$ と変化させる。)
- (C) パレート分布の区間層別
(歪度 $\gamma_1=2.5$ のパレート分布を $w_1=.1, .3, .5, .7, .9$ となる層別点で2層に区間層別する。)
- (D) 2つのパレート分布
(2つの同じ大きさのパレート分布を層と考える。歪度 γ_1 を $2.5, 3.5, 4.5, 5.5, 6.5$ と変化させる。)

なお (C) のパレート分布では (3.1) 式の k の値を10とし、(D) では $k=10, 20$ の2つのパレート分布を用いた。またパレート分布では歪度 γ_1 は母数 a への

み依存し、 $a=14.5677, 6.6635, 5.1363, 4.4844, 4.1355$ のときそれぞれ $\gamma_1=2.5, 3.5, 4.5, 5.5, 6.5$ となる。パレート分布についてはKotz (1970) が詳しい。

次に標本配分法としては、

ネイマン配分法 (Neyman Allocation: NA)

比例配分法 (Proportional Allocation: PA)

等配分法 (Equal Allocation: EA)

の3通りを考える。

また2層からサンプリングする標本の大きさの合計 n としては、20, 100 の2通りを考える。

3.2 被覆誤差

各層が正規分布に従い、かつネイマン配分法を用いる場合を除いては、信頼区間には T_{st} の真の分布の $P\{|T_{st}| \leq v_{1-\alpha/2}\} = 1 - \alpha$ となるパーセント点 $v_{1-\alpha/2}$ を用いなければならない。そのため、 t 分布から求めたパーセント点 $t_{1-\alpha/2}$ を用いて構成した (2.13) 式の信頼区間は、一般に信頼度 $1 - \alpha$ とはならない。

前節のすべての組合せに対してシミュレーションによってサンプリングを行い、 T_{st} を繰り返し計算しその経験分布を求める。また同時に、自由度 $n-L$ の t 分布のパーセント点 $t_{1-\alpha/2}$ を用いて信頼区間 I_{st} も繰り返し求める。

ここで最終的な目標となるものは信頼区間の被覆確率と被覆誤差である。被覆確率は、名目上の信頼度 $1 - \alpha$ の信頼区間 I_{st} を構成した際に、実際にその信頼区間に母平均 μ が含まれる確率 $P\{\mu \in I_{st}\}$ と定義されるが、シミュレーションでは繰り返し求めた信頼区間 I_{st} が母平均 μ を含む比率と考える。また被覆誤差を $P\{\mu \in I_{st}\} - (1 - \alpha)$ とする。

シミュレーションの結果を示す前に、(A), (B), (C), (D) の各分布と層別法に対する2層の重み、標準偏差、歪度 ($\gamma_{1,1}, \gamma_{1,2}$) を表1に、 $n=100$ の場合に、ネイマン配分法と比例配分法によって2層からサンプリングされる標本の大きさを表2に表す。なお層内平均は T_{st} の分布には影響を与えないため表1から除いた。また $n=100$ のとき等配分法では常に $n_1=n_2=50$ となるため、等配分法は表2から省いた。

図表では、標本配分法はNAなどの略称で示す。

(A) では分布が左右対称であるため、 $w_1=.5$ のときにすべての標本配分法は等しくなり、しかもこれが最適層別 ($V(\bar{X}_{st})$ が最小となる層別) となる。 w_1 が大きくなるにつれて各標本配分法の性質が顕著に現れてくる。

表1. 各層の重み, 標準偏差, 歪度

(A) 正規分布の区間層別

w_1	.5		.6		.7		.8		.9	
w_1, w_2	.50	.50	.60	.40	.70	.30	.80	.20	.90	.10
σ_1, σ_2	.60	.60	.65	.56	.70	.51	.76	.47	.84	.41
$\gamma_{1,1}, \gamma_{1,2}$	-1.00	1.00	-.90	1.09	-.79	1.18	-.66	1.27	-.48	1.39

(B) 2つの正規分布

σ_2	1.0		2.0		3.0		4.0		5.0	
w_1, w_2	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50
σ_1, σ_2	1.00	1.00	1.00	2.00	1.00	3.00	1.00	4.00	1.00	5.00
$\gamma_{1,1}, \gamma_{1,2}$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

(C) パレート分布の区間層別

w_1	.1		.3		.5		.7		.9	
w_1, w_2	.10	.90	.30	.70	.50	.50	.70	.30	.90	.10
σ_1, σ_2	.02	.80	.07	.81	.14	.83	.24	.86	.43	.93
$\gamma_{1,1}, \gamma_{1,2}$	.04	2.50	.13	2.50	.26	2.50	.44	2.50	.84	2.50

(D) 2つのパレート分布

γ_1	2.5		3.5		4.5		5.5		6.5	
w_1, w_2	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50
σ_1, σ_2	.79	1.59	2.11	4.22	3.09	6.19	3.86	7.71	4.44	8.88
$\gamma_{1,1}, \gamma_{1,2}$	2.50	2.50	3.50	3.50	4.50	4.50	5.50	5.50	6.50	6.50

(B), (D) では2層の重みが等しいため比例配分法と等配分法が一致し, (B) では更に $\sigma_2=1.0$ のときにネイマン配分法も等しくなる。

(C) では $w_1=.5$ のとき比例配分法と等配分法が等しく, $w_1=.7, .9$ のときそれぞれ等配分法と比例配分法がネイマン配分法に比較的近くなる。なお, $w_1=.1$ のとき $n=20, 100$ とも $n_1=0$ となるが, ここでは $n_1=1$ としてシミュレーションを行った。

図1, 2, 3, 4に, それぞれ (A), (B), (C), (D) の各標本配分法との組合せに対する名目上の信頼度. 95の信頼区間の被覆誤差 $P\{\mu \in I_{st}\}-.95$ を示す。なお, 繰り返しの回数は $n=20, 100$ に対してそれぞれ 500,000 回, 100,000 回である。

横軸は図1, 3では w_1 , 図2では σ_2 , 図4では γ_1 をそれぞれ示す。縦軸は被覆誤差を示している。例えば, 被覆誤差が $-.01$ であれば信頼度は .94 である。

結果を見ると, (A) ではネイマン配分法, 比例配分法の被覆誤差は $n=20$ でも小さく, 層が正規分布に従わないことによる影響はあまり見られない。また $n=100$ 程度ならば等配分法でも誤差は十分小さい。

(B) ではネイマン配分法の被覆誤差は $n=20$ でも十分小さいものの, σ_2 の値が大きい場合, 比例配分法の誤差は比較的大きい。これはネイマン配分法から大きくずれることによるものであり, そのような場合にはある程度大きな標本をサンプリングする必要がある。

(C) ではすべての標本配分法の被覆誤差が大きく, 特に $n=20$ ではネイマン配分法であっても利用することは困難である。これは第2層の歪みの大きさによるものと思われる。表1に示すように第2層の歪度は常に2.5であり, この層の重みが小さくなるにつれて被覆誤差も小さくなることが読み取れる。したがって,

表2. 各層の標本の大きさ n_1, n_2 ($n=100$ の場合)

(A) 正規分布の区間層別

w_1	.5	.6	.7	.8	.9
NA	50 50	64 36	76 24	87 13	95 5
PA	50 50	60 40	70 30	80 20	90 10

(B) 2つの正規分布

σ_2	1.0	2.0	3.0	4.0	5.0
NA	50 50	33 67	25 75	20 80	17 83
PA	50 50	50 50	50 50	50 50	50 50

(C) パレート分布の区間層別

w_1	.1	.3	.5	.7	.9
NA	1 99	4 96	14 86	39 61	81 19
PA	10 90	30 70	50 50	70 30	90 10

(D) 2つのパレート分布

γ_1	2.5	3.5	4.5	5.5	6.5
NA	33 67	33 67	33 67	33 67	33 67
PA	50 50	50 50	50 50	50 50	50 50

層別によって母集団より歪みの小さな層を構成したり、歪みの大きな層の重みを小さくすることができるならば、被覆誤差が改善される可能性があるといえる。

(D) の被覆誤差は更に大きい。ネイマン配分法と比例配分法の結果があまり変わらないことを考え合わせると、各層の歪みが大きい場合には標本配分法を工夫することによっては被覆誤差の改善は望めないであろう。ただし、 $n=500$ として同様のシミュレーションを行ったところ、ネイマン配分法と比例配分法の被覆誤差が $\gamma_1=6.5$ の場合でもそれぞれ $-.0061, -.0099$ であった。したがって歪みの大きな分布に対しては、標本を十分大きくする必要性が指摘される。

3.3 信頼区間の種々の検討

前節で述べたように、信頼区間の良さを評価する場合には被覆確率や被覆誤差が一般的に用いられている。しかし仮に被覆誤差が小さくても、求めた近似信頼区間の信頼限界が真の信頼区間の信頼限界に必ずしも近いとは言えない。統計量 T_{st} の分布は一般には単峰形であるため、被覆誤差が正のときにパーセント点

図1. 信頼区間 I_{st} の被覆誤差
(A) 正規分布の区間層別 図2. 信頼区間 I_{st} の被覆誤差
(B) 2つの正規分布

図3. 信頼区間 I_{st} の被覆誤差
(C) パレート分布の区間層別

図4. 信頼区間 I_{st} の被覆誤差
(D) 2つのパレート分布

は分布の裾の方向にずれ、被覆誤差は小さくても信頼限界のずれは大きいという状況が起こりうる。また区間推定を目的とするならば、被覆誤差よりはむしろ区間の幅を重視すべき場合もある。これらの規準については例えば Hall (1992) を参照していただきたい。

この節では区間推定としての信頼区間の良さを、次の3つの規準によって検討する。

① 信頼区間の正確度

(求めた近似信頼区間の信頼限界と真の信頼区間の信頼限界との差)

② 信頼区間の幅の平均

③ 信頼区間の幅の標準誤差

3.1節の分布、層別法のうち、(A) 正規分布の区間層別に対して、前節と同じシミュレーションによって得られた信頼度.95の信頼区間 I_{st} の①, ②, ③の値をそれぞれ表3, 4, 5に示す。なお、①は上側の信頼限界の差 $(\bar{X}_{st} + t_{.975}D) - (\bar{X}_{st} + v_{.975}D)$ の平均であり、②, ③でいう信頼区間の幅はこの場合 $2 \cdot t_{.975}D$ を意

表3. 信頼区間 I_{st} の正確度

(A) 正規分布の区間層別

n	配分法	w_1				
		.5	.6	.7	.8	.9
20	NA	.0017	.0007	-.0020	-.0030	-.0048
	PA	.0017	.0001	-.0037	-.0045	.0011
	EA	.0017	-.0106	-.0317	-.0481	-.0499
100	NA	-.0001	-.0002	.0003	.0011	.0003
	PA	-.0001	.0004	.0005	-.0001	.0002
	EA	-.0001	-.0002	-.0020	-.0034	-.0045

表4. 信頼区間 I_{st} の幅の平均

(A) 正規分布の区間層別

n	配分法	w_1				
		.5	.6	.7	.8	.9
20	NA	.5559	.5655	.5977	.6559	.7414
	PA	.5559	.5683	.6054	.6686	.7655
	EA	.5559	.5843	.6670	.7991	.9849
100	NA	.2384	.2425	.2555	.2787	.3169
	PA	.2384	.2433	.2580	.2832	.3222
	EA	.2384	.2512	.2879	.3452	.4248

表5. 信頼区間 I_{st} の幅の標準誤差
(A) 正規分布の区間層別

n	配分法	w_i				
		.5	.6	.7	.8	.9
20	NA	.1089	.1100	.1100	.1144	.1200
	PA	.1089	.1104	.1139	.1189	.1260
	EA	.1089	.1233	.1539	.1884	.2260
100	NA	.0203	.0204	.0207	.0214	.0224
	PA	.0203	.0205	.0210	.0218	.0229
	EA	.0203	.0231	.0288	.0350	.0413

味する。

まず表3の信頼区間の正確度を表4の幅の平均と比較しながら見ると、これらの場合では被覆誤差が負のものが多かったため、被覆誤差と比較して信頼限界の差の方が極端に大きいものはない。したがって、信頼限界の差については被覆誤差が正で大きい場合にのみ考え合わせればよいであろう。

標準正規分布から単純無作為抽出した場合の信頼区間 I_{sts} の幅の平均は、 χ^2 分布の性質を用いて正確に求められるが、 $n=20, 100$ に対してそれぞれ .9238, .3958 である。したがって、幅の平均がこの値を超える場合には層化抽出を行う意味はなくなる。表4をみると、ネイマン配分法と比例配分法では常にこの基準の値を下回り、等配分法でも $w_i=.9$ の場合を除き下回る。

幅の平均と同様に、単純無作為標本から得られる信頼区間の幅の標準誤差は、 $n=20, 100$ に対してそれぞれ .1547, .0283 である。ネイマン配分法からずれると、幅の平均と共に標準誤差も大きくなるが、その傾向は標準誤差の方がやや強い。

以上のことを考え合わせると、被覆誤差のみを見たときよりも層化抽出法に基づく信頼区間の良さが分かるが、適当な標本配分法を用いることの重要性も改めて指摘される。

4. Satterthwaite の近似法による信頼区間の改善

4.1 Satterthwaite の近似法の提案した方法への応用

Satterthwaite (1946) は、 χ^2 分布に従う互いに独立な確率変数の一次結合の分布を χ^2 分布で近似する際の自由度を与える式を提案した。それは次のようなも

のである。

$n_i S_i^2 / \sigma_i^2$ ($i=1, 2, \dots, L$) が互いに独立に自由度 n_i-1 の χ^2 分布に従うとき、近似的に

$$H = \frac{\nu \cdot \sum_{i=1}^L \lambda_i \cdot \frac{n_i}{n_i-1} S_i^2}{\sum_{i=1}^L \lambda_i \sigma_i^2} \sim \chi^2(\nu) \quad (4.1)$$

となる。ただし λ_i は任意の係数で、自由度 ν は

$$\nu = \frac{\left(\sum_{i=1}^L \lambda_i \sigma_i^2 \right)^2}{\sum_{i=1}^L \frac{\lambda_i^2}{n_i-1} \sigma_i^4} \quad (4.2)$$

によって決定される。この近似式は、統計量 H の分布の2次のモーメントまで同じモーメントをもつ χ^2 分布で近似するものである。

この近似は (2.17) 式に対しては既に応用されているが、ここでは2.2節で提案した方法に応用していく。(4.1) 式に

$$\lambda_i = \frac{(n_i-1)w_i^2}{n_i^2}$$

を代入し変形すると、

$$H = \frac{\nu \cdot \sum_{i=1}^L \frac{w_i^2}{n_i} \sigma_i^2}{\sum_{i=1}^L \frac{(n_i-1)w_i^2}{n_i^2} \sigma_i^2} \cdot \frac{\sum_{i=1}^L \frac{w_i^2}{n_i} S_i^2}{\sum_{i=1}^L \frac{w_i^2}{n_i} \sigma_i^2} = \nu m \cdot \frac{1}{n} V_{st}^2 \quad (4.3)$$

となる。ただし

$$m = \frac{\sum_{i=1}^L \frac{w_i^2}{n_i} \sigma_i^2}{\sum_{i=1}^L \frac{(n_i-1)w_i^2}{n_i^2} \sigma_i^2} \quad (4.4)$$

である。このとき自由度 ν は

$$\nu = \frac{\left\{ \sum_{i=1}^L \frac{(n_i-1)w_i^2}{n_i^2} \sigma_i^2 \right\}^2}{\sum_{i=1}^L \frac{(n_i-1)w_i^4}{n_i^4} \sigma_i^4} \quad (4.5)$$

によって与えられる。ここで (4.3) 式より統計量 T_{st} は

$$\begin{aligned} T_{st} &= \frac{\sqrt{n-L} U_{st}}{V_{st}} \\ &= \frac{\sqrt{n-L} U_{st}}{\sqrt{\frac{n}{\nu m} H}} \\ &= \sqrt{\frac{m(n-L)}{n}} \cdot \frac{\sqrt{\nu} U_{st}}{\sqrt{H}} \end{aligned} \quad (4.6)$$

と書き直すことができる。 $\sqrt{\nu} U_{st}/\sqrt{H}$ は近似的に自由度 ν の t 分布に従うので、その t 分布の上側確率 $\alpha/2$ のパーセント点を $t_{1-\alpha/2}^+$ とすると、この近似法を用いた母平均 μ の信頼度 $1-\alpha$ の信頼区間 I_{st}^+ は

$$I_{st}^+ = [\bar{X}_{st} - t_{1-\alpha/2}^+ D^+, \bar{X}_{st} + t_{1-\alpha/2}^+ D^+] \quad (4.7)$$

となる。ただし

$$D^+ = \sqrt{m} \sqrt{\sum_{i=1}^L \frac{w_i^2}{n_i} S_i^2} \quad (4.8)$$

である。

4.2 シミュレーションによる近似法の検討

この節では、3節でシミュレーションを行った分布、層別法、標本配分法の組合せに対して前節の近似法を適用し信頼区間の改善を計る。

図5, 6, 7, 8に、それぞれ (A), (B), (C), (D) の各分布に対する Satterthwaite の近似法を利用した名目上の信頼度.95 の信頼区間の被覆誤差を示す。シミュレーションの方法は3.2節と同様であり、ネイマン配分法については3.2節の結果を再び示す。なお、(4.4),

n=20 の場合

n=100 の場合

—■— NA —●— PA —▲— EA

図5. 信頼区間 I_{st}^+ の被覆誤差
(A) 正規分布の区間層別

n=20 の場合

n=100 の場合

—■— NA —●— PA —▲— EA

図6. 信頼区間 I_{st}^+ の被覆誤差
(B) 2つの正規分布

n=20 の場合

n=100 の場合

—■— NA —●— PA —▲— EA

図7. 信頼区間 I_{st}^+ の被覆誤差
(C) パレート分布の区間層別

図8. 信頼区間 I_{st}^* の被覆誤差
(D) 2つのパレート分布

(4.5) 式の m と ν に含まれる各層の母分散 σ_i^2 は、得られた標本の分散 S_i^2 を用いて近似した。

図1, 2, 3, 4と比較すると, (A), (B) では被覆誤差は驚くほど改善されている。標本の大きさが100程度あれば等配分法でも誤差は極めて小さい。

(C), (D) では被覆誤差が大きかった部分は若干改善されネイマン配分法の値に近づいているものの、ネイマン配分法ではほとんど改善は見られない。これは Satterthwaite の近似法が分布の2次のモーメントまでしか近似しないため、統計量 T_{st} の分布の大きな歪みまでは修正できないことによるものである。

したがって、各層の歪みが大きくなければ標本がある程度小さくても Satterthwaite の近似法によって被覆誤差は十分小さく修正できるが、歪みが大きい場合には何かしら他の方法が要求される。

5. ブートストラップ法による母平均の信頼区間の構成

5.1 ブートストラップ信頼区間の構成手順

3節でみたように、通常利用されているような標本

配分法でサンプリングが行われた場合には、各層の歪みが大きくない限り信頼区間 I_{st} を用いても問題はない。また4節でみたように、標本配分法が多少不適切であっても信頼区間 I_{st}^* による改善が見込まれる。しかし各層の歪みが大きい場合には、他の手法を考える必要がある。

その候補であるジャックナイフ法などが, Krewski and Rao (1981), Rao and Wu (1985), (1987), (1988), Kovar, Rao and Wu (1988) で議論されているが、最も有力な候補はブートストラップ法であろう。

単純無作為抽出に対するブートストラップ法が考案されて間もなく、層化無作為抽出への自然な拡張がなされた。これは後で述べるように、サンプリングされた標本と同じ大きさの標本を各層毎にリサンプリングするものであり、現在でも基本となる考え方は変わっていない。

サンプリングが有限母集団からなされる場合、一般にサンプリングが非復元抽出で行われるのに対してリサンプリングは復元抽出で行われ、リサンプリングがサンプリングそのものを再現しない。そのために標本から構成される統計量をブートストラップ標本から構成されるブートストラップ統計量によって推定するときには不偏推定とはならない状況が発生する。層化無作為標本を利用したブートストラップ法では主にこの点に対して焦点が当てられ, Rao and Wu (1988), Sitter (1992) などによって議論されている。その歴史については Shao and Tu (1995), Sitter (1992) が詳しい。また漸近的な理論については Babu and Singh (1983), Bickel and Freedman (1984)などを参照されたい。

単純無作為標本からのブートストラップ法による信頼区間の構成法を、2.2節で提案した方法を基にして、層化無作為抽出に自然に拡張したノンパラメトリック・ブートストラップ法は次のようになる。なお、母集団は無限母集団とする。

1° 母集団 F_0 から標本が層化無作為抽出されるものとするが、第 i 層からサンプリングされた大きさ n_i の標本の観測値を $x_{i1}, x_{i2}, \dots, x_{in_i}$ とする。これらの標本点に同じ確率を持たせた標本全体を経験分布 F_1 とする。

2° 経験分布 F_1 において1°と同じ層別を考え、第 i 層の $x_{i1}, x_{i2}, \dots, x_{in_i}$ から大きさ n_i の標本 $X_{i1}^*, X_{i2}^*, \dots, X_{in_i}^*$ を無作為に復元抽出する。このとき、(2.12) 式と同様にブートストラップ統計量 T_{st}^* を次のように定義する。

$$T_{st}^* = \sqrt{\frac{n-L}{n}} \cdot \frac{\bar{X}_{st}^* - \bar{x}_{st}}{\sqrt{\sum_{i=1}^L \frac{w_i^2}{n_i} S_i^{*2}}} \quad (5.1)$$

ここで

$$\bar{X}_{st}^* = \sum_{i=1}^L w_i \bar{X}_i^* \quad \left(\bar{X}_i^* = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}^* \right), \quad (5.2)$$

$$\bar{x}_{st} = \sum_{i=1}^L w_i \bar{x}_i \quad \left(\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \right), \quad (5.3)$$

$$S_i^{*2} = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij}^* - \bar{X}_i^*)^2 \quad (5.4)$$

である。また

$$s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \quad (5.5)$$

と定義しておく。

3° ステップ2°を B 回繰り返し、計算された T_{st}^* の値を $t_{st,b}^*$ ($b=1, 2, \dots, B$) とし、その値の絶対値の小さい順に並べたものを $|t_{st(1)}^*|, |t_{st(2)}^*|, \dots, |t_{st(B)}^*|$ とする。

4° 整数 k を $k/B=1-\alpha$ を満たす値またはその近似値とすると、 $\hat{v}_{1-\alpha/2} = |t_{st(k)}^*|$ とおく。このパーセント点の推定値を用いて次のように母平均 μ の信頼度 $1-\alpha$ の信頼区間 I_{st}^* を構成する。

$$I_{st}^* = [\bar{x}_{st} - \hat{v}_{1-\alpha/2} D^*, \bar{x}_{st} + \hat{v}_{1-\alpha/2} D^*] \quad (5.6)$$

ただし

$$D^* = \sqrt{\frac{n}{n-L}} \sqrt{\sum_{i=1}^L \frac{w_i^2}{n_i} S_i^{*2}} \quad (5.7)$$

である。

この信頼区間においてブートストラップ法で求めるものは $\hat{v}_{1-\alpha/2}$ のみで、その他の値はすべて標本から計算することになる。

もし母集団 F_0 の各層が正規分布に従う場合には、パラメトリック・ブートストラップ法を用いるべきであろう。その場合には1°で、経験分布 F_1 の第 i 層を正規分布 $N(\bar{x}_i, s_i^2)$ として、2°のリサンプリングを各正規分布からのサンプリングと置き換えるだけであるが、本論文ではノンパラメトリック・ブートストラップ法のみを論じる。

5.2 シミュレーションによるブートストラップ信頼区間の検討

3.1節で考えた分布と層別法のうち、各層が正規分布

に従わずにノンパラメトリック・ブートストラップ法によって改善が期待される(A), (C), (D)について、標本配分法と標本の大きさの組合せに対して、前節のブートストラップ法の手順で信頼区間 I_{st}^* を繰り返し求め被覆誤差を計算する。なお、パーセント点の推定に用いる $t_{st,b}^*$ の個数 B は n の値によらず1,000とし、繰り返しの回数は $n=20, 100$ に対してそれぞれ20,000回、4,000回とする。

図9, 10, 11に、(A), (C), (D)に対する名目上の信頼度.95のブートストラップ信頼区間 I_{st}^* の被覆誤差を示す。

繰り返しの回数が少ないためにやや結果にばらつきがあるが、(A)では図1と比較すると等配分法においてブートストラップ法を用いる効果が認められる。ただし図5のSatterthwaiteの近似法の結果と比較すると、その改善の度合の差は明らかではない。

(C), (D)では図7, 8のSatterthwaiteの近似法の結果と比較しても、各標本配分法において大きな改善が見られる。 $n=100$ 程度あれば十分実用に耐えうる。

前述したようにSatterthwaiteの近似法では各層が大きく歪んでいる場合にはあまり改善が期待できない

図9. ブートストラップ信頼区間 I_{st}^* の被覆誤差 (A) 正規分布の区間層別

図 10. ブートストラップ信頼区間 I_{st}^* の被覆誤差
(C) パレート分布の区間層別

図 11. ブートストラップ信頼区間 I_{st}^* の被覆誤差
(D) 2つのパレート分布

表 6. ブートストラップ信頼区間 I_{st}^* の幅の平均

(A) 正規分布の区間層別

n	配分法	w_1				
		.5	.6	.7	.8	.9
20	NA	.5788	.5882	.6233	.6785	.7390
	PA	.5788	.5939	.6306	.6926	.7726
	EA	.5788	.6298	.7678	.9459	1.1573
100	NA	.2382	.2421	.2561	.2788	.3173
	PA	.2382	.2436	.2584	.2840	.3214
	EA	.2382	.2523	.2928	.3518	.4315

表 7. ブートストラップ信頼区間 I_{st}^* の幅の標準誤差

(A) 正規分布の区間層別

n	配分法	w_1				
		.5	.6	.7	.8	.9
20	NA	.1275	.1287	.1311	.1348	.1304
	PA	.1275	.1319	.1393	.1420	.1413
	EA	.1275	.1725	.2465	.3074	.3368
100	NA	.0216	.0221	.0224	.0234	.0249
	PA	.0216	.0223	.0225	.0240	.0251
	EA	.0216	.0251	.0325	.0387	.0453

が、ブートストラップ法ではそのような場合に大きな効力を発揮するといえる。また、シミュレーションで用いた $B=1,000$ を更に大きくすることによって、誤差が更に改善されることも考えうる。

次に、(A) 正規分布の区間層別におけるブートストラップ信頼区間の幅の平均と標準誤差を、それぞれ表 6, 7 に示す。

信頼区間の幅の平均を見ると、概して表 4 の提案した方法を用いた場合の値より若干大きいだけだが、等配分法で w_1 が大きく、ネイマン配分法からのずれが大きい場合、ブートストラップ信頼区間の幅の平均が大きい。特に標本が小さいときにその傾向は顕著で、被覆確率のみで判断することの危険性を示している。

次に幅の標準誤差を見ると、提案した方法の標準誤差より全般に大きく、 $n=20$ では 1.09~1.63 倍、 $n=100$ でも 1.06~1.13 倍である。すなわち、どのような標本がサンプリングされているかによって、構成されるブートストラップ信頼区間が大きく影響を受けるとい

うことができる。

なお (C), (D) についてもほぼ同様の結果が得られたが、上で述べた傾向は更に強い。

6. 適用例

単純無作為抽出法以外のサンプリング法を用いる際には、一般に何かしら母集団に関する情報を必要とする。しかし、もしその情報が既知であれば、もはや調査する意味のなくなる場合が多い。

例えばネイマン配分法を用いてサンプリングしようとするならば、母集団の層内標準偏差 σ_i が既知でなければならないが、一般にそれは不可能である。したがって過去に大規模な調査があればそれから得られる値を用いるか、または予備調査を行うことによって推定しなければならない。それにも関わらず安定した結果が得られなければ理論を応用することはできないため、そのような状況を想定した検討が必要不可欠である。複雑なサンプリング法が敬遠される理由の一つが、このような議論の不足にあると思われる。

この節では、実際のデータに対して提案した方法を含む種々の方法を用いて信頼区間を構成するシミュレーションを行い、その比較をする。ここで対象にするのは、図12に示すような、ある業種の販売店の年間販売額のデータである。全販売店数が11,291で、1~29人の各従業員規模に対する販売店数と年間販売額の平

均、標準偏差のデータが得られているが、個々の販売店の販売額は明らかにされていない。そこで個々のデータを復元するために、各従業員規模の平均と分散を基に正規乱数を発生させ販売額に相当する値を販売店数分作った。そのため、復元された値の平均、標準偏差は元の値とは異なるが、実用上問題はない。目標は年間販売額の母平均の信頼区間の構成であるが、サンプリングの際には各従業員規模に対する販売店数のみが既知で、個々の年間販売額の値は未知とする。

従業員規模に対する2層への区間層別を行い、大きさ $n=20, 100, 500$ の標本を層化無作為抽出するものとする。

信頼区間の構成方法としては次の5通りを考える。

- (a) ネイマン配分法
- (b) 比例配分法
- (c) 層内標準偏差を推定するネイマン配分法
- (d) Satterthwaiteの近似法を利用した比例配分法
- (e) 比例配分法によるノンパラメトリック・ブートストラップ法

このうち、(a), (b) については3節、(d) については4節、(e) については5節のシミュレーションの方法と同様であるが、各層から最低2以上の大きさの標本をサンプリングするようにする。また、ネイマン配分法では層内標準偏差 σ_i は母数そのものを用いる。したがって (a) は実際には実行不可能であるが、他の方

図12. ある業種の販売店の従業員規模別年間販売額

法の基準として考える。

(c) はネイマン配分法で層内標準偏差 σ_i を予備調査から推定する方法を用いる。その概略は次のようである。

1° 予備調査を層化無作為抽出法で行う。標本の合計を n' として等配分法を用いて各層から大きさ n'/L

の標本をサンプリングする。

2° 予備調査の各層の標本を基に、層内標準偏差の推定値 $\hat{\sigma}_i$ を求める。ここでは標本の層内標準偏差を用いる。その値を用いて近似的なネイマン配分法を次のように決定する。

—●— (a) —●— (b) —●— (c) —×— (d) • (e)

図 13. 各方法による信頼区間の被覆誤差

$$n_i = \frac{w_i \bar{\sigma}_i}{\sum_{k=1}^L w_k \bar{\sigma}_k} n - \frac{n'}{L} \quad (6.1)$$

ただし、 $n_i < 0$ ならば $n_i = 0$ として、その増加分だけ他の層の標本を小さくする。

3° 本来の調査を近似的ネイマン配分法を用いて行う。予備調査と本来の調査で得られた標本に対し提案した方法を用いて、信頼区間 I_{st} の構成を行う。

この方法は予備調査と本来の調査を合わせて大きさ n の標本を近似的にネイマン配分法によってサンプリングするものである。なお、シミュレーションでは $n=20, 100, 500$ に対して予備調査の標本の大きさを $n'=10, 30, 50$ とした。これらの値はやや小さいように思われるが、例えば母平均の推定の際には $n=1,000$ のときでも $n'=50 \sim 100$ 程度で十分であることが経験的に知られている。Sai (1988), (1989) を参照されたい。

元の母集団は有限母集団であるが非常に大きいので $n=20, 100$ の場合には有限修正は行わない。しかし $n=500$ の場合にはその影響が大きい (2.14) 式と

(4.8) 式の $\sum_{i=1}^L \frac{w_i^2}{n_i} S_i^2$ を $\sum_{i=1}^L \frac{w_i^2}{n_i} \cdot \frac{N_i - n_i}{N_i - 1} S_i^2$ で、また

(5.7) 式の $\sum_{i=1}^L \frac{w_i^2}{n_i} s_i^2$ を $\sum_{i=1}^L \frac{w_i^2}{n_i} \cdot \frac{N_i - n_i}{N_i - 1} s_i^2$ でそれぞれ置き換えて有限修正を行う。ただし N_i は母集団の第 i 層の大きさとする。

図13に $n=20, 100, 500$ の、各方法による母平均の名目上の信頼度.95の信頼区間の被覆誤差を示す。ここで被覆誤差は、(a), (b), (c) については $P\{\mu \in I_{st}\} - .95$, (d) については $P\{\mu \in I_{st}^*\} - .95$, (e) については $P\{\mu \in I_{st}^*\} - .95$ である。

図で横軸は第1層の従業員規模の最大値 (層別点 p_1) を表している。なお (e) については、 $p_1=5, 10, 15, 20, 25$ のみを計算した。また繰り返しの回数は、 $n=20, 100, 500$ に対して (a)~(d) ではそれぞれ 500,000 回、100,000 回、20,000 回、(e) ではそれぞれ 20,000 回、8,000 回、4,000 回である。

ここで対象にしているデータの特徴として、右に裾の長いことがあげられる。母集団の歪度は 2.765 であり、3節で考えた (C) パレート分布の区間層別に近い。単純無作為抽出でも、歪度の大きな分布からサンプリングされた標本の T_{srs} は t 分布で近似することの難しいことが知られている。(竹内(1975)を参照のこと。) ちなみに母集団全体から単純無作為抽出した場合の信頼度.95の信頼区間 I_{srs} の被覆誤差を同様のシミュ

レーションで求めると、例えば $n=20$ のときに $-.0393$ で誤差は非常に大きい。

$n=20$ の場合、(e) のブートストラップ法の良さが際だっている。 t 分布を用いる方法では、実行不可能なネイマン配分法以外では (d) の Satterthwaite の近似法の一部を除き被覆誤差が大きく、標本が余りに小さい場合にはどの方法も利用できないことがわかる。ただし、信頼区間 I_{srs} の被覆誤差の $-.0393$ と比較すると概して誤差は小さい。すなわち、元の母集団を区間層別またはそれに近い層別によって、より小さな歪度の層に分割することができれば、被覆誤差が改善される可能性がある。これは層別による副産物と言えよう。なお比例配分法が2つの山をもっているのは、各層から最低2以上の大きさの標本をサンプリングしたためである。比例配分法では実際には $p_1 \geq 15$ のとき第2層からは大きさ1以下の標本しか取られず、もしそれを修正しなければこの部分の被覆誤差は極めて大きくなる。このことは $n=100$ にも当てはまる。

$n=100$ の場合には、比例配分法を除きどの方法もかなり誤差が小さくなる。特に (d) の Satterthwaite の近似法は、層別点の小さいところではネイマン配分法にほぼ匹敵し、層内標準偏差を推定した影響は余り見られない。また (c) の層内標準偏差を推定するネイマン配分法はこの部分ではやや劣るものの、層別点の大きい部分でも安定して小さい誤差を与える。このシミュレーションでは、予備調査の標本の大きさを $n'=30$ とし、残り 70 の大きさの標本を本来の調査でサンプリングしたが、 n' をもう少し大きくすると層内標準偏差を安定して推定できるかも知れない。

$n=500$ の場合にはどの方法を用いても誤差は小さい。したがって、標本が大きければ最も簡単な (b) の

表8. 各方法による信頼区間の幅の平均と標準誤差 ($p_1=10$ の場合)

	n	手 法				
		(a)	(b)	(c)	(d)	(e)
平 均	20	75953	80032	70904	97620	128237
	100	32826	38766	32645	40680	41777
	500	14610	17505	14711	17664	17694
標準誤差	20	19181	29287	19304	56412	123611
	100	3655	7587	3790	8630	11077
	500	714	1552	762	1590	1748

比例配分法を用いれば良いであろう。

ところで、ネイマン配分法と比例配分法の最適層別点(層別点 p_1 のうち $V(\bar{X}_{st})$ を最小とするもの)は、それぞれ 9 と 12 である。したがって一般の調査において母集団を 2 層に分ける場合には、経験からこの付近に層別点をおくことが多い。そこで $n=20, 100, 500$ における $p_1=10$ の場合の各方法による信頼区間の幅の平均と標準誤差を表 8 に示す。

(a) のネイマン配分法を除くと、平均、標準誤差とも (c) の層内標準偏差を推定するネイマン配分法の値が際だって良く、ネイマン配分法そのものと比較しても遜色がない。ブートストラップ法は一般に広い信頼区間を構成してしまい、また理論分布に対するシミュレーションと同じようにサンプリングした標本によってその幅が大きく異なってしまうことが読み取れる。

このデータの難しい性質を考え合わせると、層別法や標本の大きさ、さらには予備調査の可能性の有無などに応じて、(b), (c), (d), (e) の方法を適切に使い分けることにより、提案した方法はかなり幅広い状況に対して適用可能であるといえる。

7. お わ り に

本論文では、層化無作為標本に基づく母平均の信頼区間の数種類の構成方法について議論を行った。提案した方法は、単純無作為標本に基づく信頼区間の構成と同様に t 分布を利用するものである。これについては Satterthwaite による統計量の分布の近似法を用いて、信頼区間の構成法の改善を行った。また、提案した方法にノンパラメトリック・ブートストラップ法を適用した。どの方法についても、人工的な数種類の理論分布、層別法、標本配分法、標本の大きさの各組合せに対してモンテカルロシミュレーションを行うことによって、その有効性を比較した。また実データに各方法を応用することによって、実際の調査での問題点とその克服法、さらには種々の推定誤差を加味した場合の各方法の有効性についても検討した。それを標本配分法に関してまとめると次のようになる。

- ① ネイマン配分法でサンプリングされた標本に対しては、提案した方法による信頼区間 I_{st} の有効性は理論的にもシミュレーションによっても示される。また歪みがそれほど大きくなければ、各層が正規分布に従わないことによる影響は、仮に標本が小さくてもほとんどない。一般の調

査では層内標準偏差を推定する必要があるが、予備調査で大きさ 30 以上程度の標本をサンプリングできれば十分実用に耐えうる。しかし歪みの大きな分布の場合には被覆誤差は大きく、例えば各層の歪度が 2.5 程度であれば、標本の大きさとして 100 は必要である。さらに歪みが大きい場合には、ブートストラップ信頼区間 I_{st}^* によって改善を計ることができる。各層の歪度が 5 程度であっても、標本の大きさとして 100 程度あれば十分である。

- ② 比例配分法では I_{st} の構成は理論的には裏付けられないが、シミュレーションによれば層毎の分散が大きく違ったり層の歪みが大きかったりしなければ特に問題はない。各層の散布度が大きく異なる場合には、Satterthwaite の近似法による信頼区間 I_{st}^* を用いればよい。しかし各層の歪みが大きい場合には I_{st}^* を用いる効果は小さく、ブートストラップ信頼区間 I_{st}^* によって改善を計らねばならない。特に歪みが大きいときには、ネイマン配分法よりも若干大きい標本が必要とされる。
- ③ 等配分法のようにネイマン配分法からずれが大きい場合にも、各層の歪みが小さく、しかも標本が十分大きければ (2 層の場合には 100 以上) I_{st} を用いてもよいが、 I_{st}^* を用いる方が無難であろう。各層の歪みが大きい場合にも、標本が十分大きく 100 程度あればブートストラップ信頼区間 I_{st}^* を用いればよい。

すなわち、各層の歪みが比較的小さい場合には、標本がある程度大きければどのような標本配分法でも問題がないが、やや不適切な標本配分法でかつ標本が小さければ Satterthwaite の近似法によって改善を計ることができる。一方、各層の歪みが大きければ、ネイマン配分法であっても標本を十分大きくとらなければならない。その他の標本配分法では Satterthwaite の近似法による効果は期待できず、ブートストラップ法が唯一の頼りとなる。

以上のように、提案した方法に Satterthwaite の近似法による改善法とブートストラップ法を含めると、適切な標本配分法を用いることによって、かなり広い範囲の状況に対して適用可能であるといえる。

付 録

表 9. 各標本配分法における V_{st}^2 , V_1^2 , V_2^2 の関係

	V_{st}^2	V_1^2	V_2^2
一 般 形	$\frac{n \cdot \sum_{i=1}^L \frac{w_i^2}{n_i} S_i^2}{\sum_{i=1}^L \frac{w_i^2}{n_i} \sigma_i^2}$	$\sum_{i=1}^L \frac{n_i S_i^2}{\sigma_i^2}$	$\frac{(n-L) \sum_{i=1}^L \frac{w_i^2}{n_i-1} S_i^2}{\sum_{i=1}^L \frac{w_i^2}{n_i} \sigma_i^2}$
ネイマン配分法 (NA)	$\frac{n \cdot \sum_{i=1}^L \frac{w_i S_i^2}{\sigma_i}}{\sum_{i=1}^L w_i \sigma_i}$	$\frac{n \cdot \sum_{i=1}^L \frac{w_i S_i^2}{\sigma_i}}{\sum_{i=1}^L w_i \sigma_i}$	$(n-L) \sum_{i=1}^L \frac{1}{1 - \frac{\sum w_h \sigma_h}{w_i \sigma_i n}} \cdot \frac{w_i S_i^2}{\sigma_i}$
比例配分法 (PA)	$\frac{n \cdot \sum_{i=1}^L w_i S_i^2}{\sum_{i=1}^L w_i \sigma_i^2}$	$n \cdot \sum_{i=1}^L \frac{w_i S_i^2}{\sigma_i^2}$	$\frac{(n-L) \sum_{i=1}^L \frac{w_i}{1 - \frac{1}{w_i n}} S_i^2}{\sum_{i=1}^L w_i \sigma_i^2}$
等配分法 (EA)	$\frac{n \cdot \sum_{i=1}^L w_i^2 S_i^2}{\sum_{i=1}^L w_i^2 \sigma_i^2}$	$\frac{n}{L} \sum_{i=1}^L \frac{S_i^2}{\sigma_i^2}$	$\frac{n \cdot \sum_{i=1}^L w_i^2 S_i^2}{\sum_{i=1}^L w_i^2 \sigma_i^2}$

謝 辞

本論文をまとめるにあたり、審査者からは貴重なコメントをいただき、ここに深く感謝致します。

参 考 文 献

- Ames, M.H., & Webster, J.T. (1991). On estimating approximate degrees of freedom. *The American Statistician*, **45**, 45-50.
- 浅井晃 (1987). 調査の技術. 日科技連.
- Babu, G.J., & Singh, K. (1983). Inference on means using the bootstrap. *The Annals of Statistics*, **11**, 999-1003.
- Bickel, P.J., & Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, **9**, 1196-1217.
- Bickel, P.J., & Freedman, D.A. (1984). asymptotic normality and the bootstrap in Stratified Sampling. *The Annals of Statistics*, **12**, 470-482.
- Cochran, W.G. (1977). *Sampling techniques* (3rd ed.). New York: Wiley.
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia: SIAM.
- Hall, P. (1992). *The bootstrap and edgeworth expansion*. New York: Springer-Verlag.

- 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.
- 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.
- Kotz, J. (1970). *Distribution in statistics, continuous univariate distribution*. New York: Wiley.
- Kovar, J.G., Rao, J.N.K., & Wu, C.F.J. (1988). Bootstrap and other methods to measure errors in survey estimates. *The Canadian Journal of Statistics*, **16**, Supplement, 25-45.
- Krewski, D., & Rao, J.N.K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *The Annals of Statistics*, **9**, 1010-1019.
- Lemeshow, S., & Levy, P.S. (1979). Estimating the variances of ratio estimates in complex sample surveys with two primary sampling units per stratum-A comparison of balanced replication and jackknife techniques. *Journal of Statistical Computing and Simulation*, **8**, 191-205.
- Levy, P.S., & Lemeshow, S. (1991). *Sampling of populations: Methods and applications*. New York: Wiley.
- NHK 放送文化研究所編 (1996). 世論調査事典. 大空社.

- Rao, J.N.K., & Wu, C.F.J. (1985). Inference from stratified samples: Second-order analysis of three methods for nonlinear statistics. *Journal of the American Statistical Association*, **80**, 620-630.
- Rao, J.N.K., & Wu, C.F.J. (1987). Methods for standard errors and confidence intervals from sample survey data: Some recent work. in *The Bulletin of the International Statistical Institute*, Proceedings of the 46th Session, Book 3, 5-19.
- Rao, J.N.K., & Wu, C.F.J. (1988). Resampling Inference with complex survey Data. *Journal of the American Statistical Association*, **83**, 231-241.
- Sai, S. (1988). Optimum stratification based on concomitant variables—Comparison with ratio and regression estimators—. *Proceedings of the Fifth Japan and Korea Joint Conference of Statistics*, 120-122.
- Sai, S. (1989). Optimum stratification based on concomitant variables for some typical distributions. *Proceedings of the Sixth Korea and Japan Joint Conference of Statistics*, 95-99.
- Sai, S., & Taguri, M. (1989). Optimum stratification based on a concomitant variable and its application to the current statistics of commerce. *Reports of Statistical Application Research, JUSE*, **36**, 22-31.
- 佐井至道 (1995). 層化抽出法に基づく母平均の信頼区間の構成について. 第 63 回日本統計学会予稿集, 222-223.
- Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, **2**, 110-114.
- Shao, J., & Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer-Verlag.
- 柴田義貞 (1981). 正規分布—特性と応用—. 東京大学出版会.
- Sitter, R.R. (1992). A resampling procedure for complex survey Data. *Journal of the American Statistical Association*, **87**, 755-765.
- Som, R.K. (1996). *Practical sapling techniques*, 2nd edition. New York: Marcel Dekker.
- 竹内啓 (1975). 確率分布と統計解析. 日本規格協会.
- Thompson, S.K. (1992). *Sampling*. New York: Wiley.
- 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**(2), 50-81.

—1996 年 4 月 19 日受付
1996 年 11 月 18 日最終修正—