ブートストラップ法による正規・非正規分布データの相関係数信頼区間の測定

吉　村　治　正*

Measuring Correlation Coefficient Confidence Intervals from Normally and Non-normally Distributed Data.

Harumasa YOSHIMURA

要　旨

ピアソンの積率相関係数は、$\rho \neq 0$ の場合にデータが非正規な分布をしていると区間推定に影響が出ることが、これまでの数理統計学上の研究からわかっている。ただ、その知見は理論的な議論の域を出ず、実際の社会調査データから具体的な影響を測定する方法は長く存在していなかった。この点、1980年代に提唱されたノンパラメトリック・ブートストラップ法は社会調査者にとって革新的であった。だが実際には、ブートストラップ法によって非正規な分布をするデータから相関係数の区間推定を行った事例はほとんどない。そこで本稿では、正規分布・非正規分布をする実際のデータを用い、ノンパラメトリック・ブートストラップ法と通常のパラメトリック推定法で、どちらがより精度の高い推定が可能かを対比してみた。結果は、データが正規分布する場合はパラメトリック法でもブートストラップ法でも概して正確に推定できるが、非正規な分布をするデータの場合、パラメトリック法は信頼区間を過少に算出し厳密性にかける傾向があり、これに対してブートストラップ法はデータが非正規な分布をしている状態でも極めて正確な推定が可能というものであった。

1．問題の所在

本稿では、社会調査データの相関係数信頼区間推定を行う場合、通常のパラメトリック推定法とノンパラメトリック・ブートストラップ法とで、どの程度の差異が生じ、どのような状況でどちらを用いる方がより正確な推定を期待できるのかを、モンテカルロ法を用いて検証してみたい。

社会調査データの分析に携わる者（social researcher：以下、社会調査者と表す）であれば、多かれ少なかれ、数理統計学からの逸脱を自覚する経験を誰もが持っている。だが、これは統計学の知識が不足しているからという場合に限らない。数理統計学の求める条件を厳密に遵守していくとデータの分析が成立しなくなるからという場合も少なくない。例えば、本稿で取り上げるピアソンの積率相関係数が典型例である。ピアソンの積率相関係数は社会調査データの分析ではごく日常的に用いられる統計量であり、初歩的な分析とみなされる。だが、ここにおいて積率相

平成24年７月23日受理　*社会学部准教授

関係数がパラメトリックな統計量であること、つまりデータとして与えられる二変量が少なくとも母集団において正規分布するという条件が与えられていることが考慮されることはむしろ稀である。実際、社会学や心理学で用いられる態度測定項目などでは正規分布するよりも極端な偏り（skewed）を見せる方が多く、二極分化（polarized）することも珍しくない。教育学や政治学のデータは分布が切断されている（truncated）ことが多いし、所得や貯蓄に関するデータは外れ値（outlier）を多分に含んでいる。

これらの特性がパラメトリックな統計量を算出する条件に違反（violate）していることは、ほとんどの場合、データ分析を行っている社会調査者は自覚している。だが、だからといって別のノンパラメトリックな統計量を用いると結果解釈が難しくなる、他の調査結果との対比が困難になる、適用できる分析手法が限定される、適当な代替的ノンパラメトリック統計量が見つからない、などの問題が生じる。それゆえに、前提となる条件に多少違反しても、重大な問題が生じることが明らかでない限り、知名度が高く計算法も比較的簡単なピアソンの積率相関係数を用いることが一般的である。

しかしながら、こうして「片目をつぶる」ことの多い社会調査者も、相関係数の検定となると躊躇する者が少なくない。それは、相関係数が有意かどうかは、一義的に標本数によって決まってしまうからである。標本相関係数の有意性の検定は極めて簡単で、標本数と標本相関係数だけで算出される[1)]。例えば、n =50であれば±0.279、n =100であれば±0.196、n =200であれば±0.138を越えれば95%水準で有意になる。つまり、あるデータから得られた相関係数が有意かどうかを計算する際には、データは必要ない。データの特性に依存しないため、５点ないしは７点のライカート評定法で測定された心理項目から態度測定をする場合も、１グラム・１センチ単位で測定された身長と体重の関係を調べる場合であっても、標本数が同じであれば同じ限界値（critical value）によって一律に判定される。

こうした一律の限界値が与えられるのは、データが二変量とも正規分布にしたがうという条件を満たしていることが前提とされているからである。そのため、逆説的ではあるが、自分の手元にあるデータが正規分布するという確信を持てない、あるいはおそらく正規分布はしないだろうという予測を持っている社会調査者は、ここにおいて二変量正規分布というピアソンの積率相関係数の条件がどれほど大きなものであったかを実感することになる。

したがって、では正規分布しない場合の標本相関係数はどのような分布を描き、どのように検定することができるのか、という疑問が出てくる。これは数理統計学では古典的な問題で、筆者の知る限りでも既に1930年代にはこの問題を扱った論文がみられる（Pearson 1931; Pearson 1932）[2)]。ただし、その内容は社会調査者にはほとんど知られていないし、議論の内容も理論的考察の域を出ず、実際のデータを用いて測定されたことも皆無に近い。理由は簡単であり、通常の統計学の発想では、これは現実的には測定不可能と判断されるからである。

あるデータが与えられた時に、そのデータを正規分布なり二項分布なり既知の確率密度分布に近似させ、その分布曲線の関数式を数理的に展開することでパラメターの値を算出する。これがパラメトリックな統計分析の基本発想である。この時、データの分布がどの確率密度関数に近いと判断するかは、データを扱う者の目視と経験によって決定する。データの分布を示す確率密度

関数式をデータから算術的計算によって直接的に求めることはできない。前述した相関係数のパラメトリックな区間推定法も、二変量正規分布の確率密度関数式

$$df = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\, e^{-\frac{1}{1-\rho^2}\left\{\frac{(x-\mu_x)^2}{2\sigma_x^2}-\frac{2\rho(x-\mu_x)(y-\mu_y)}{2\sigma_x\sigma_y}+\frac{(y-\mu_y)^2}{2\sigma_y^2}\right\}}$$

を展開して得られる（Fisher 1915）ものであり、したがって、もしも非正規な分布をするデータの相関係数区間推定を行いたいのであれば、そのデータの分布を目視し、その形状にマッチする正規分布以外の分布曲線を探し出し、その曲線を関数式で定義する必要がある。ところが実際の社会調査データでよく見られるような、複数のピークを持ったり極端に歪んだような分布にマッチし、なおかつ関数式で表現できるような分布曲線を見つけるとなると、これは極めて困難であり、事実上不可能というしかない。もしこれを行おうとすれば、あらかじめ関数式を定義し、その分布曲線にしたがうデータをコンピュータで派生させる必要がある。これは要するにシミュレーションということである。つまり非正規な分布をするデータの相関係数の区間推計は、理論的・実験的な議論の域を出ず、実際のデータが与えられた状態で算術的に測定することは不可能と考えられてきた。

この点、データを既知の分布に近似させるのではなく、反復計算によってデータから得られた分布曲線を直接に測定しようとするブートストラップ法は、パラメトリックな統計分析に対する、まさに逆転の発想である。それ故に、パラメトリックな手法では極めて困難とされてきた非正規な分布をするデータの相関係数の区間推定も、ブートストラップ法ではいとも容易に行うことができる。これは実際の社会調査データを扱う者にとって、極めて大きな意味をもつ。

ブートストラップ法の基本発想は、モンテカルロ法と極めて近い。つまり両者とも標本の抽出と統計量の算出を反復し、ここから統計量の誤差範囲を推定する。ただし、モンテカルロ法は主としてシミュレーションに用いられるのに対し、ブートストラップ法は実際に調査によって得られた標本から標本を再抽出することで、データの測定に用いられる。この違いは社会調査者にとっては決定的で、特に中央値（median）の信頼区間や非正規な分布をするデータの相関係数の信頼区間など、従来の方法では推定が極めて困難な統計量も容易に算出できる点で期待が大きい。

ブートストラップ法が広く知られるようになったのは、1983年に *Scientific American* 誌上に Diaconis & Efron が掲載した一本の論文による。ここにおいて、Efronらは従来の統計学的な計算では算出が極めて困難な統計量が、計算機を用いた単純な計算の反復で求められることを論証しようとした。この時に事例としてとりあげたのが、相関係数の標本分布であった。それほど相関係数の区間推定は困難とみなされていた。

ところが、ここには今日から見て奇妙な点がある。ブートストラップ法による相関係数の区間推定について、知名度の高いものとしては、この83年の Diaconis & Efron による事例と85年の Lunneborg による研究事例があげられる。標本相関係数の区間推定が極めて困難なのはデータが正規分布しない場合であり、したがってブートストラップ法の有用性は正規分布しないデータを用いた時に発揮されるはずである。ところが、この二つの事例では、両者とも正規分布に極めて

近い分布をすることがあらかじめ分かっているデータを用いて区間推定を行っている。しかもLunneborg によれば、正規分布するデータを選んだのは偶然ではなく、意図的なものであった（Lunneborg 1985）。この理由は、両者がこれらの事例を報告した時はブートストラップ法という発想自体がまだ新規であったため、パラメトリック推定と近似する結果を示すことでブートストラップ法の正確さを印象づける必要があったからである。だが、この点については Rasmussen（1987）が厳しく批判するように、従来の推定法と同じ結果が出せることがブートストラップ法の有用性であるとするならば、最初から従来の推定法を使えばよいのであって、新しい方法をわざわざ持ち出す必要はない。ブートストラップ法が有用であることを証明するには、ブートストラップ法でなければできない推定を行うことが必要であり（Strube 1988）、したがって相関係数の区間推定でも、正規分布するデータではなく非正規な分布をするデータを用いて論証すべきであった。

ところがこの2つの事例研究以降、ブートストラップ法の研究は実際の調査データを用いる実践的なものから、シミュレーションへと関心が移っていったように思える。非正規な分布をする相関係数の区間推定の問題も、シミュレーションでの研究例はいくつも目にする（松田 2006）が、実際の調査データを用いた例はほとんど、というか全くといっていいほど目にすることはない。

したがって本稿では、Efronによって行われた相関係数の信頼区間推定を再現し、さらにEfronの行わなかった非正規に分布する実際のデータの相関係数信頼区間の推定を合わせて行うことで、ブートストラップ法が非正規な分布をするデータの相関係数信頼区間をどこまで正確に推定できているかを論証したいと思う。

2．二つの推定法

二変量正規分布する母集団から抽出された n 組の標本（x_i, y_i）から得られた標本相関係数を r、母集団における相関係数（母相関係数）を ρ（ただし $\rho \neq 0$）とする。そして r および ρ にフィッシャーの z 変換をほどこした変量 r^*および ρ^*を

$$r^* = \frac{1}{2} log \frac{1+r}{1-r}$$

$$\rho^* = \frac{1}{2} log \frac{1+\rho}{1-\rho}$$

とする。この時、r^* の分布は ρ^*を平均、$\frac{1}{\sqrt{n-3}}$ を標準偏差とする正規分布、すなわち

$$\mathrm{N}\left\{ \rho^*, \left(\frac{1}{\sqrt{n-3}}\right)^2 \right\}$$

に近似的にしたがう（$n \geqq 10$の場合）（岸田 1980, 315）[3]。したがって、信頼区間を求める場合は、この標準偏差に ¿水準に相当する z 値をかける、つまり ¿= 0.95であれば、

$$r_1^* = \rho^* - 1.96 \times \left(\frac{1}{\sqrt{n-3}}\right) \quad （下限）$$

$$r_2^* = \rho^* + 1.96 \times \left(\frac{1}{\sqrt{n-3}}\right) \quad （上限）$$

この r_1^* と r_2^* の間が求める信頼区間となる。なお、この値は z 変換した尺度上の値となるので、-1 から $+1$ までを値域とする通常の相関係数の値に直すためには、これに z 変換の逆関数を与える必要がある。すなわち

$$r_1 = \frac{e^{2r_1^*}-1}{e^{2r_1^*}+1} \quad （下限）$$

$$r_2 = \frac{e^{2r_2^*}-1}{e^{2r_2^*}+1} \quad （上限）$$

これが一般的に用いられる、つまりパラメトリックな相関係数の信頼区間の算出法である。

これに対し、ブートストラップ法はデータの反復計算から標本分布を経験的に派生させるところに特徴がある。ブートストラップ法の基本発想はモンテカルロ法と極めて近い。モンテカルロ法は反復抽出、つまり母集団から何度も標本を抽出し、その統計量の分布を調べていくことで解析的な方法に頼らずに統計量の誤差範囲を求める。ただし、母集団からの反復抽出が可能になるためには、母集団そのものが判明していなければならない。現実の社会調査の状況において、調査する以前に母集団の状態がわかっているのは稀であり、さらにそこから標本を反復して抽出することができるとなると、事実上極めて特殊な状況に限られてしまう。それゆえ、モンテカルロ法はシミュレーション、つまり実在の母集団ではなく関数式で定義される仮想の母集団を用いることが多い。

これに対しブートストラップ法は、調査によって得られた標本から再びランダムに標本を抽出（リサンプリング）することで、モンテカルロ法と同様にパラメーターの誤差範囲を測定する[4)]。今、n 個の観察より構成される標本（X, Y）を

$$X = (x_1, x_2, x_3, x_4, x_5, \cdots, x_n),\ Y = (y_1, y_2, y_3, y_4, y_5, \cdots, y_n)$$

とし、ここから得られる標本相関係数を r とする。この n 個のケースの中から、置換（replacement、一つの標本から同じケースを２回以上重複して抽出すること）を認めてランダムに n 個の観察を再抽出（リサンプリング）する。このようにして作られる標本をブートストラップ標本という。例えば $n = 5$ であるとき、ブートストラップ標本（X_b, Y_b）は

$$X_1 = (x_1, x_3, x_3, x_2, x_4),\ Y_1 = (y_1, y_3, y_3, y_2, y_4)$$
$$X_2 = (x_4, x_3, x_5, x_2, x_5),\ Y_2 = (y_4, y_3, y_5, y_2, y_5)$$
$$X_3 = (x_1, x_4, x_1, x_1, x_3),\ Y_3 = (y_1, y_4, y_1, y_1, y_3)$$
$$\vdots$$

といった具合に構成されていく。ブートストラップ標本（X_1, Y_1），（X_2, Y_2），（X_3, Y_3），…（X_b, Y_b）のそれぞれから算出された標本相関係数を $r_1, r_2, r_3, \ldots r_b$ とする。この統計量 $r_1, r_2, r_3, \ldots r_b$ の度数分布から描かれる分布曲線を経験分布曲線と呼び、これをもって標本相関係数の分布とみなす。そして、この $r_1, r_2, r_3, \ldots r_b$ の平均値 $r_{(.)}$ を母集団推定量とする。なお、ブートストラップ推定において標本の再抽出（リサンプリング）の反復回数（B）が十分に大きい場合、$r_{(.)}$ は元の標本統計量 r と理論的に一致する（Efron & Tibshirani 1993）。区間推定については、統計量 $r_1, r_2, r_3, \ldots r_b$ をその値の順にソートし、 ¿水準の上限および下限に相当する個数を数え、その値をもって上限値および下限値とする。これがパーセンタイル法である。

パーセンタイル法による区間推計には複雑な計算をなんら必要としない。ただし、経験分布曲線はデータそのものから描かれるため、偏ったり歪んだりしていることもある。こうした凹凸や歪みのある分布の分位点を数学的に正確に求めようとすると、極めて難しい議論になる。実際、ブートストラップ法の解説書を読み進めると、この辺りで途端にハードルが高くなる。ここから先は数理統計学者や実験統計学者の領域と考えて差支えがないだろう。ブートストラップ法の区間推定法は何種類も考案されているが、その中には高度な数式理解を求めるものもある（DiCiccio & Tibshirani 1987; Efron 1987; DiCiccio & Efron 1992; DiCiccio & Efron 1996）[5)]。ただし、本稿ではこれら実験統計学上の先端研究には立ち入らない。本稿でブートストラップ法による区間推定を行う場合は、もっとも素朴でノンパラメトリック・ブートストラップ法の基本発想に忠実なパーセンタイル法と、標本分布が偏っているような場合に高い評価が安定して与えられているBCa法を用いることにする。

3．データおよび手順

データは正規分布するものと正規分布しないものの二組を用意した。第一は Efron が1983年に *Scientific American* 誌上で相関係数の区間推定を論証した際に用いたデータ、つまり1965年時点での全米82校のロースクールの LSAT スコアと GPA の学校ごとの平均点である。これは統計ソフト「R」の「bootstrap」モジュールに添付されている。第二のデータは、筆者が数年前に担当した科目の履修学生のテストの成績である。この科目は通年の必修科目で、一年生は全員、前期・後期の二回にわたり筆記試験を受けさせられている。Efron のデータも筆者の試験データも共に相関係数は高い。Efron のデータの82校すべてを含めて算出した相関係数は0.760、筆者のデータでは全165人の１年生全員を含めた時の相関係数が0.741であった。（この二つの数字は、以下の分析で母相関係数として扱われる）。ただし、Efron のデータは二変量がそれぞれ極めて正規分布に近いのに対し（図１・図２、なお、図中の破線は平均値と標準偏差から描いた正規分布曲線）、筆者のデータに含まれる二変量は二極に分かれている、つまり明らかに非正規な分布をしている（図３・図４）。なお、こうした違いは偶然ではなく、データの性質と深く関連している。つまりEfronのデータは個々人の点数ではなく学校ごとの平均値なので、当然ながら個人レベルのデータに比べばらつき（variability）が小さい。これに対し筆者のテストは個人レベルのデータであるし、さらに他科目の試験日程の関係で、準備（俗にいう「一夜漬け」）の時間がとれる学生と

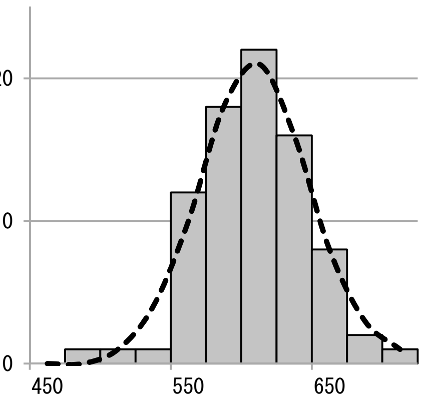

図１　Efron データ（LSAT平均点）の分布

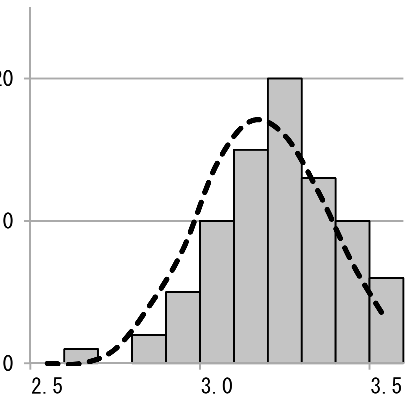

図２　Efron データ（GPA平均点）の分布

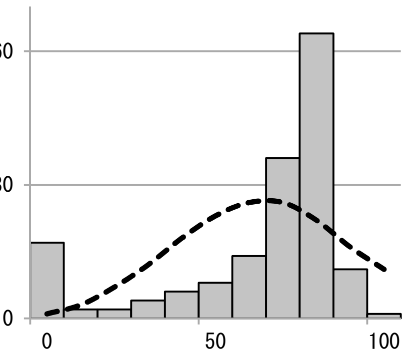

図３　筆者データ（前期試験成績）の分布

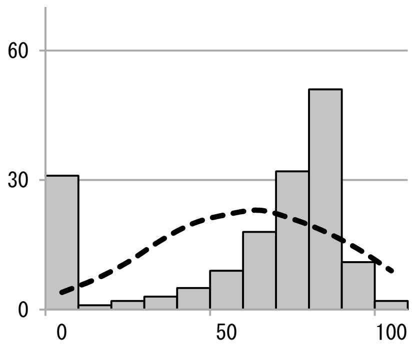

図４　筆者データ（後期試験成績）の分布

とれない学生に分かれていたということも、判明している。

分析の手順は以下のとおりである。

①まず、Efron のデータを構成する全82校をもって母集団と定義する。筆者のデータも165人の一年生全員をもって母集団とする。そしてそれぞれの母集団の示す相関係数（母相関係数＝ ρ ）を算出する。母相関係数は前述のとおり、Efron のデータの場合は ρ ＝0.760、筆者のデータの場合は ρ ＝0.741である。

②次に Efron データの場合は15校、筆者のデータの場合は40人を標本として、置換（replacement）を認めずにランダムに抽出する。これをモンテカルロ標本とし、この標本から標本相関係数を求める。これを10,000回繰り返し、統計量の度数分布を得る。これを真の標本分布とする。なお、10,000個のモンテカルロ標本のうち、最初の10個（＃１～#10）についてはブートストラップ推定で用いるので、その値を保存しておく。

③これに続き、パラメトリックな推定法に基づいて区間推定を行う。パラメトリック推定法による信頼区間幅は標本数から容易に計算され、また z 変換した尺度上では標本相関係数の値に左右されないので、一度の計算で事が足りる。

④次に、ノンパラメトリック・ブートストラップ法を用いた区間推定を行う。つまり、モンテカ

ルロ標本＃1から置換を認めてブートストラップ標本を再抽出し、その都度、標本相関係数を得る。これを10,000回反復し、各ブートストラップ標本から得られる統計量の度数分布からパーセンタイル法およびBCa法で信頼区間を算出する。

⑤上記④を10回繰り返し、モンテカルロ標本#１から#10の信頼区間の平均値を求め、これをパラメトリック推定および真の標本分布から得られた信頼区間と比較する。

4．結　　果

まず Diaconis & Efron（1983）にならって、パラメトリック推定による標本分布、ブートストラップ標本分布[6]、そして真の標本分布を重ねてみた（図５）。なお、図中のＸ軸は Fisher の z 変換を施した尺度となっている。Diaconis & Efronはこれを±１を値域とする通常の尺度に戻して表記しているが、パラメトリック推定法において正規分布に近似するのは z 変換した尺度の上になるので、通常の尺度に戻さない方が信頼区間幅をつかみやすい[7]。

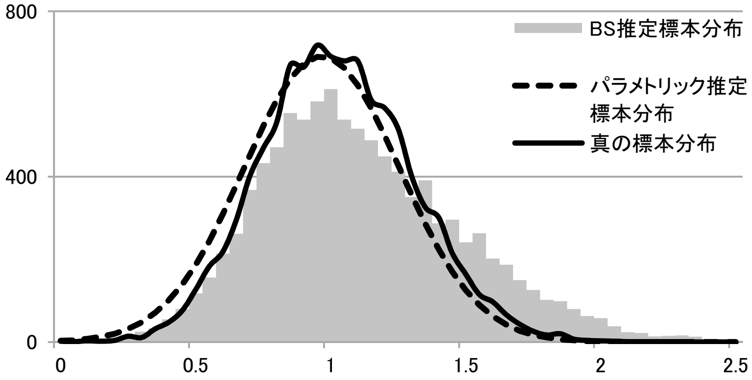

図５　Efron データ（正規分布）の標本相関係数分布

この図を見てまず気づくのが、パラメトリック推定が驚くほど正確に真の標本分布を再現していることである。これに比べるとブートストラップ標本の分布はやや偏り、ばらつきが大きくなっている。なお、パラメトリック推定では信頼区間の幅が標本数だけで決まってしまうのに対し、ブートストラップ推定ではパラメーター推定値もその信頼区間も共に標本から算出されるため、モンテカルロ標本ごとに多少のばらつきがみられる（表１）。これは個々の標本数が15という少なさを考えれば、やむを得ない。しかしながら、10個のモンテカルロ標本の平均値を見ると、特にパーセンタイル法は68%の区間推定でパラメトリック推定よりも真の値に近い。95%の区間推定になるとパラメトリック推定の方が真の値に近くなるが、その差は僅かというか事実上無視しえる範囲とみなすべきであろう。これと比べるとBCa 法による推定は68%でも95%の区間推定でもわずかに真の値よりも広くなっている。とはいえ、全体としては、母集団が正規分布する場合、パラメトリック推定でもブートストラップ推定でも、同じように高い精度で区間推定ができ

ていると言える。この点についてはEfronの指摘する通りである。

表1　Efronデータ（正規分布）の信頼区間幅推定結果

(N=82, n=15)

モンテカルロ標本	標本相関	BS標本相関	BSバイアス	パーセンタイル 68%	パーセンタイル 95%	Bca 68%	Bca 95%	パラメトリック推定 68%	パラメトリック推定 95%	真の標本分布 68%	真の標本分布 95%
#1	0.776	0.770	0.007	0.776	1.585	0.669	1.515				
#2	0.791	0.763	0.028	0.694	1.374	0.694	1.455				
#3	0.787	0.747	0.040	0.709	1.405	0.701	1.480				
#4	0.871	0.832	0.039	0.297	1.053	0.861	1.822				
#5	0.753	0.760	-0.006	0.384	0.831	0.398	0.894	0.577	1.132	0.554	1.098
#6	0.854	0.854	-0.001	0.505	1.040	0.492	0.997				
#7	0.699	0.675	0.024	0.582	1.083	0.588	1.206				
#8	0.584	0.587	-0.002	0.563	1.170	0.542	1.198				
#9	0.880	0.857	0.023	0.623	1.310	0.635	1.323				
#10	0.758	0.780	-0.022	0.415	0.904	0.390	0.815				
（平均）				0.555	1.175	0.597	1.270				

＊ρ=0.760

＊信頼区間幅はすべてZ変換した状態での値である。

では、非正規な分布をするデータの場合はどうか。図6を目視すると、まず真の標本分布の曲線が正規分布からかなり離れていることがわかる。分布の頂点のずれは標本抽出の際に不可避に生じるとしても、分布の形だけをみればパラメトリック推定よりもブートストラップ推定の方が近似しているように見える。パラメトリック推定分布は、真の標本分布よりも明らかにばらつき（variability）が小さい。なお、表2を見ると、標本ごとのブートストラップ推定の信頼区間幅のばらつきは表1よりも目立たなくなっているが、これは標本数が多くなっているためと考えられる。

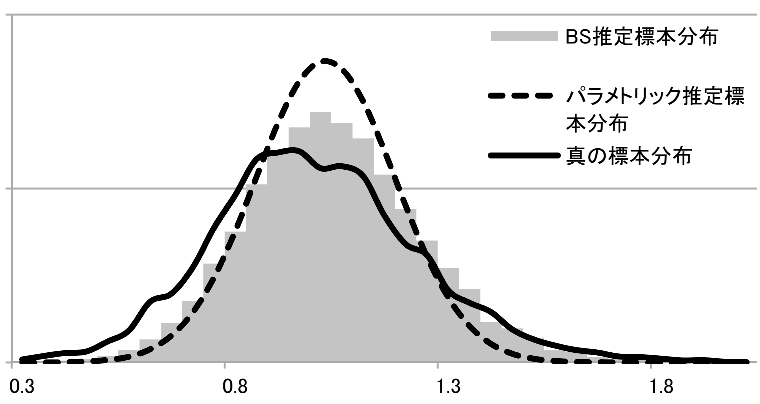

図6　筆者データ（非正規分布）の標本相関係数分布

それよりも目を引くのは、表2において、パラメトリック推定による信頼区間幅が真の標本分布よりも狭くなっている点にある。これに対しブートストラップ推定は、パーセンタイル法、BCa法の両者とも、真の値よりもわずかに過大な区間を推定する傾向が見られる。ただし、10個のモンテカルロ標本の平均値を見れば、パーセンタイル法であってもBCa法であっても、ブート

ストラップ推定の方がパラメトリック推定よりも真の値に近い（68%区間、95%区間ともに）。さらに、10個のモンテカルロ標本の個々の値を見てみると、パラメトリック推定よりブートストラップ推定の方が真の値に近い数値を示しているのが、68%区間でみるとパーセンタイル法で7個、BCa 法で9個となっている。同じく95%区間については、パーセンタイル法、BCa 法ともに10個のモンテカルロ標本全てがパラメトリック推定よりも真の値に近い数値を示している。つまり、標本ごとのばらつきを考慮に入れても、非正規な分布をする場合の区間推定は、パラメトリック推定よりもブートストラップ推定の方が真の値に近いということになる。

表2　筆者データ（非正規分布）の信頼区間幅推定結果

(N=165, n=40)

モンテカルロ標本	標本相関	BS標本相関	BSバイアス	パーセンテイル 68%	パーセンテイル 95%	Bca 68%	Bca 95%	パラメトリック推定 68%	パラメトリック推定 95%	真の標本分布 68%	真の標本分布 95%
#1	0.737	0.731	0.006	0.405	0.852	0.406	0.845				
#2	0.723	0.719	0.004	0.498	0.948	0.490	0.949				
#3	0.703	0.702	0.001	0.381	0.781	0.370	0.773				
#4	0.786	0.784	0.002	0.387	0.790	0.377	0.780				
#5	0.793	0.784	0.009	0.443	0.896	0.449	0.917	0.329	0.644	0.392	0.797
#6	0.730	0.730	0.000	0.442	0.896	0.432	0.872				
#7	0.706	0.702	0.004	0.456	0.924	0.452	0.901				
#8	0.770	0.771	-0.001	0.317	0.659	0.308	0.646				
#9	0.648	0.644	0.004	0.412	0.857	0.418	0.865				
#10	0.766	0.762	0.004	0.406	0.830	0.404	0.826				
（平均）				0.415	0.843	0.411	0.837				

＊ρ=0.741

*信頼区間幅はすべてZ変換した状態での値である。

なお、この差が具体的にどれほどの影響を持つかを考察するために、非正規に分布するデータの10,000個のモンテカルロ標本のうち、実際にどれぐらいが95%区間として推定された領域に入っているかを調べてみた。ブートストラップ推定を行ったモンテカルロ標本#1から#10について個々の標本の ¿(0.025) と ¿(0.975) の平均値を求め、これに点推定の調整を行い（つまり標本相関係数の平均値（$\bar{r}^*$）と ρ^*の差を加え）、これを限界値とした。すると10,000個のモンテカルロ標本相関係数のうち、ブートストラップ法で推定された95%信頼区間におさまるものは9,587個（パーセンテイル法の場合）および9,579個（BCa 法の場合）となった。これはつまり95%信頼区間として推定された領域が真の標本分布の95.8%から95.9%を網羅しているということであり、ブートストラップ法による推定が極めて正確であることを示している。これに対し、パラメトリック法で95%区間として推定された領域におさまったモンテカルロ標本相関係数は、10,000個のうち8,913個、つまり89.1%しかない。これはパラメトリックな前提条件の下で5%とされる棄却域が実際には10%以上の領域を持っていることになる。これは軽視していい問題ではない。というのは、この結果は、これまで社会調査者が行ってきた相関係数の検定が、分布の正規性という問題を考慮しなかったことで有意性を過大に見積もってきた可能性があることを示唆しているからである。

今回の検証から得られる結論は極めて明瞭である。すなわち、正規分布するデータから得られたピアソンの積率相関係数の信頼区間は、パラメトリックな方法を使ってもノンパラメトリック・ブートストラップ法を使っても、極めて正確に推定できる。しかしながら、非正規な分布をするデータから得た標本相関係数の分布は、パラメトリックな推定法が想定するよりもばらつき

が大きくなり、そのためにパラメトリックな推定法では信頼区間を過少に推定してしまう傾向がある。これに対しブートストラップ推定は、全体として信頼区間をわずかに過大に推定する傾向が見られるものの、概して非正規な分布をするデータから得た相関係数の標本分布を精度よく推定することが可能である。

5．展　望

同じ統計データを扱うとしても、社会調査者の発想は数理統計学者とは似て異なるものと筆者は理解している。数理統計学者は推論の理論的根拠・論理的整合性を厳しく求める半面で、データそのものには価値を認めない。そのため、条件に合わなければデータを取り換えることになんら躊躇をしない。これに対し、社会調査者は推論の理論的根拠よりもデータの実測に関心を持つ。目の前に与えられたデータはそれ自体が現実であり、データを取り換えることは現実を差し替えることに等しい。条件に合わないデータは分析する価値がないと言い切れるのは数学者の発想であり、社会調査者は数理的条件に合致しないデータであっても、合致しないことを自覚しながらデータの分析を行わねばならない。とはいえ、あまりにもあからさまに数理統計学的な前提条件や理論的根拠を無視して行う分析は逸脱であり、その結果も信頼には値しない。したがって社会調査者のデータ分析は、数理統計学的な前提条件への違反を自覚しつつ、違反することの影響を考えながら、どこまでその違反に「目をつぶる」ことが許容されるかを自問自答しながら進められる。社会調査者がデータ分析を行う際の難しさは、ここにある。

この点、文頭にも書いたが、ピアソンの積率相関係数は社会調査データを扱う際には極めて日常的に用いられているが、社会調査データの性質として、正規分布しない、値域が極めて狭い（ライカート評定法の場合など)、曲線的な関係が目視できるなど、この統計量の前提条件に合致しない事が多い。この意味で、社会調査者にとってピアソンの積率相関係数はもっとも慎重な取り扱いを求められる統計量の一つであると言ってよい。この過程において、ある条件に違反することでどのような影響がどの程度まで表れているかを測定できるというのは、社会調査者にとって極めて重要である。繰り返すが、Efron によるブートストラップ法の提唱が社会調査データを扱う者にとっては革新的と言えるのは、一般論ではなく具体的な実測の手順を示した点にある。

ノンパラメトリック・ブートストラップ法による測定が実用化されるまで、社会調査者はピアソンの積率相関係数は分布の非正規性に対して頑強（robust）であると指摘するシミュレーションの結果（Pearson 1931 ; Pearson 1932 ; Zeller & Levine 1974）を信じるしかなかった。それ故に、これまではデータの分布形状に関してかなり寛容な態度が保持されてきた。今回の検証の結果として現れた、データが非正規に分布する場合に生じるパラメトリックな方法による過少推定の傾向を深刻な影響とみなすか、それとも無視し得るとみなすかについては、意見が分かれると思われる。しかしながら、この問題は、少なくとも ¿水準という観点から見れば決して軽視できるものではない。つまり、データの分布の影響を考慮しない状態で95%と想定されていた範囲が分布の非正規性を考慮して再測定したところ実は90%にも満たない範囲であったということにな

れば、そもそも95％という有意水準にこだわることの意味が失われてしまう。また、今回の検証に用いたデータは相関係数が比較的高かったために生じなかったが、もっと標本相関係数（r）の値が低いデータであれば、パラメトリックな推定法による信頼区間とノンパラメトリック・ブートストラップ法による信頼区間との相違の中にゼロの地点が含まれる、つまり帰無仮説H_0：$\rho=0$とする有意差検定においてパラメトリック推定法とブートストラップ法とで異なる結果が現れる可能性も否定できない。その点については、今回の検証結果はこれまでの楽観論的な常識に疑問を提起し、社会調査者に対してより慎重な態度を求めることになったといえよう。

註一覧

1）母相関係数（ρ）がゼロであるとする帰無仮説の検定に用いられる統計量（t）は以下の式で与えられる。

$$t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

（ただし、n は標本数、r は標本相関係数）

この統計量 t は自由度 $n-2$ の t 分布に従うとされる（大山・武藤・柳井 1980, 129; 岸根 1980, 314; 芝・渡部・石塚 1984, 148)。この定理には x_i も y_i も出てこない。つまりこの式を見れば、標本相関係数の検定が個々のデータには依存しないことが明らかになる。

2）Pearson はこの二本の論文においてシミュレーション実験を行い、その結果、データが非正規に分布していても $\rho=0$ であれば、標本相関係数の分布は正規分布に極めて近い、つまりデータの非正規性はほとんど影響がないことを明らかにした。これ以降、非正規に分布するデータの標本相関係数の問題は、$\rho\neq0$ の場合に関心が向けられることになった。なお、その後の数理統計学および実験統計学上の展開を概観しておくと、Haldine（1949）は、$\rho\neq0$ の場合、データの非正規性は点推定には影響を与えないが、区間推定に関してはデータの分布形状、特に尖度によって影響を受けると指摘した。同様に、Gayen（1951）は $\rho\neq0$ の場合は標本相関係数の分布が標本の分布形状に大きな影響を受けると結論づけた。さらに Kowalski（1972）は、シミュレーションの際に非正規分布するデータを派生させるために用いた関数によって、影響が異なることを指摘した。つまり複数の正規分布を組み合わせて非正規に分布するデータを派生させた場合、ここから得られる標本相関係数の分布は裾が重たくなり、指数分布や χ^2 分布をもとに非正規分布データを派生させた場合は、逆に分布の裾が軽くなる傾向があると主張した。だが、この指摘に対して Edgell & Noon（1984）は、非正規に分布するデータから得られた標本相関係数の分布は一般的に正規分布の場合と比べてばらつきが大きくなる傾向があると主張し、Kowalski の見出した指数分布や χ^2 分布から派生させた非正規データの標本分布の裾が軽くなる傾向はシミュレーションに用いた標本数が少ないために生じたバイアスであると批判した。総じて、$\rho=0$ の時はデータの非正規性は無視しえるが $\rho\neq0$ の場合は区間推定に影響が出る場合がある、ということまでは合意されているようだが、具体的にどのようなデータを用いた時はどのような影響が出て、それが検定の場面においてどの程度まで深刻な影響を及ぼし得るのかについては、未だ合意には至っていないように思われる。

3）$\rho=0$ のときは、z 変換を必要とせずに区間推定が可能となる。この場合分けの根拠については、岸根（1980, 313）を参照されたい。

4）ブートストラップ法による点推定および区間推定の方法については、統計学者による明瞭で簡潔な紹介が多数なされている（汪・大内・景・田栗 1992；阿久津 2005：2006；金 2007；汪・桜井 2011)。そのため本稿では基礎的な事項の確認にとどめることにした。なお、文系の研究者がおそらく抱くであろう疑問として、標本から標本を抽出することでなぜ母集団の状態が予測できるのかという点だけ触れておき

たい。これについては Moony & Duval（1993）および Lunneborg（1985）に詳しい説明がある。それによると、標本から母集団を推測する場合、常になんらかの手がかりを必要とする。一般的には、母集団がある特定の分布に従うことが経験的にわかっている、あるいは理論的にある特定の分布に従うことが予測される場合が多く、このような場合は、この分布の関数式を手掛かりとして演繹的に推定を行うことができる（パラメトリック推定）。しかしながら、データや統計量の性質によっては、経験的・理論的に母集団の分布を予測することが難しい場合もある。このような場合、母集団の分布を予想する最良の手掛かりは標本そのものにある、と考えるのがブートストラップ法の根底にある発想である。もしも母集団から標本がランダムに（つまり等確率で）抽出されているのであれば、標本の分布は母集団の分布に極めて近い形をしていると考えられる。ならば、この標本からランダムに（等確率で）標本の再抽出を繰り返し行うことで、母集団の分布を再現できるはずである。こうしたブートストラップ法の基本発想の背後には、標本抽出の際に偏りが生じていない事と、標本の数がある程度以上大きく母集団の値域が遺漏なく網羅されている事という二つの前提が存在している。つまり、ブートストラップ法もデータへの厳しい要求が存在する。Lunneborg はこの点を強調し、ブートストラップ法はデータが信頼できる（dependable）場合においてこそ使われるべき手法であるとしている（Lunneborg 1985）。

5）ブートストラップ法は、母集団が正規分布もしくは既知の分布に近似できない場面、あるいは母集団の分布状態が予測できない状態を想定して構築された。つまり本質的にはノンパラメトリックな推定法である。ただし、ノンパラメトリックな適用をしている限り、その推定がどれぐらいの精度を持つかを調べるのは極めて難しい。そのため、分布関数を先に定め、そのパラメトリックな最尤推定値を元にリサンプリングとブートストラップ推定を行うことで、推定の精度を調べる方法が考案された（DiCiccio & Tibshirani 1987）。これが「パラメトリック・ブートストラップ」と呼ばれる方法であり、BCa 法の推定精度などは、パラメトリック・ブートストラップ法によって検証されてきた。

6）図 5 に示したブートストラップ標本の分布は、モンテカルロ標本＃ 1 から導出されたものである。図 6 も同様である。

7）z 変換したままの方が好ましいもう一つの理由が、モンテカルロ標本ごとに標本相関係数の値が異なることである。z 変換は r の値が ± 1 に近づくほど影響が大きくなる。したがって、信頼区間が z 変換した状態で同じ幅であっても、r の値が異なれば、通常の尺度上に戻した時の幅は同じにならない。例えば Efron データのモンテカルロ標本＃ 4 の標本相関係数は0.871、＃ 8 は0.584と差が開いているが、パラメトリック推定において z 変換した状態での95％信頼区間幅は共に1.132である。ところがこれを z 変換を戻して通常の尺度上に位置づけると、＃ 4 の区間幅は0.289、＃ 8 の区間幅は0.742と、大きく異なってしまう。今回の分析のように、複数の標本を比較する場合、標本相関係数の点推定値に差が出てしまうと、通常の尺度上で信頼区間幅を比較することには意味がなくなってしまう。

文献一覧

阿久津洋巳（2005）.「ブートストラップ法入門（1）－標準誤差の推定－」.『日本官能評価学会誌』, 9(2), 122-126.

阿久津洋巳（2006）.「ブートストラップ法入門（2）－信頼区間の推定－」.『日本官能評価学会誌』, 10(1), 43-46.

Chernick, Michael R. (1999). *Bootstrap Methods: A Practitioner's Guide.* Wiley.

Diaconis, Persi & Bradley Efron (1983). Computer-Intensive Methods in Statistics. *Scientific American,* 248, 96-108.

DiCiccio, Thomas & Bradley Efron (1992). More Accurate Confidence Intervals in Exponential Families.

Biometrika, 79(2), 231-245.

DiCiccio, Thomas & Bradley Efron (1996). Bootstrap Confidence Intervals. *Statistical Science,* 11(3), 189-228.

DiCiccio, Thomas & Robert Tibshirani (1987). Bootstrap Confidence Intervals and Bootstrap Approximations. *Journal of American Statistical Association,* 397, 163-170.

Edgell, Stephan E. & Sheila M. Noon (1984). Effect of Violation of Normality on the t-Test of the Correlation Coefficient. *Psychological Bulletin,* 95(3), 576-583.

Efron, Bradrey (1987). Better Bootstrap Confidence Intervals. *Journal of American Statistical Association,* 397, 171-185.

Fisher, R. A. (1915). Frequency Distribution of the Values of the correlation Coefficient in Samples from an Indefinitely Large Population. *Biometrika,* 10(4), 507-521.

Gayen, A.K. (1951). The Frequency Distribution of the Product-Moment Correlation Coefficient in Random Samples of Any Size Drawn from Non-Normal Universes. *Biometrika,* 38, 219-247.

Haldine, J.B.S. (1949). A Note on Non-normal Correlation. *Biometrika,* 36, 467-468.

Hall, Peter (1986). On the Number of Bootstrap Simulations Required to Construct a Confidence Interval. *The Annals of Statistics,* 14(4), 1453-1462.

金明哲（2007）．「Ｒとブートストラップ」．『ESTRELA』，156, 58-63.

岸田卓郎（1980）．『理論応用統計学』．養賢堂

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(3)，137-161.

Kowalski, Charles J. (1972). On the Effects of Non-Normality on the Distribution of the Sample Product-Moment Correlation Coefficient. *Journal of the Royal Statistical Society,* Series C, 21(1), 1-12.

Lunneborg, Clifford E. (1985). Estimating the Correlation Coefficient: The Bootstrap Approach. *Psychological Bulletin,* 98(1), 209-215.

Manly, Bryan F.J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology.* (2nd.ed.) Chapman & Hall.

松田忠之（2006）．「ブートストラップ推定法による相関係数の信頼区間（非正規な場合）」．『経済理論』，334, 87-110，和歌山大学．

Mooney, Christopher Z. & Robert D. Daval (1993). *Bootstrapping: A Nonparametric Approach to Statistical Inference.* SAGE University Paper.

汪金芳・大内俊二・景平・田栗正章（1992）．「ブートストラップ法－最近までの発達と今後の展望－」．『行動計量学』，19(2)，50-81.

汪金芳・桜井裕仁（2011）．『ブートストラップ入門』．共立出版．

大山正・武藤真介・柳井晴夫（1980）．『行動科学のための統計学』．朝倉書房

Pearson, Egon S. (1931). The Test of Significance for the Correlation Coefficient. *The Journal of American Statistical Association,* 26, 128-134.

————. (1932). The Test of Significance for the Correlation Coefficient: Some Further Results. *The Journal of American Statistical Association,* 27, 424-426.

Rasmussen, Jefferey L. (1987). Estimating Correlation Coefficients: Bootstrap and Parametric Approaches. *Psychological Bulletin,* 101(1), 136-139.

Schenker, Nathaniel (1985). Qualms about Bootstrap Confidence Intervals. *Journal of American Statistical Association,* 80(390), 360-361.

Stine, Robert (1990). An Introduction to Bootstrap Methods. *Sociological Methods & Research,* 18(2&3), 243-291.

Strube, Michael (1988). Bootstrap Type I Error Rates for the Correlation Coefficient: An Examination of Alternate Procedures. *Psychological Bulletin,* 104(2), 290-292.
Zeller, Richard A. & Zachary Levine (1974). The Effects of Violating the Normality Assumption underlying *r*. *Sociological Methods & Research,* 2(4), 511-519.

Summary

By means of standard parametric method and nonparametric bootstrap method, Pearson's product-moment correlation coefficient confidence intervals are measured, under Monte Carlo setting, from normally and non-normally distributed real social data. This examination explicates two conspicuous results. First, both parametric and non-parametric methods are capable of estimating confidence interval quite exactly in case data are normally distributed. Secondly, for non-normally distributed data, nonparametric bootstrap method indicates better estimation, while standard parametric method tends to suggest a bit, but conspicuously, too narrow interval.