コンピュータ時代の統計学

高橋　一

一　はじめに

現在、統計学は高校学校では一応数学の中で教えられていると思いますが、幸か不幸か本学を含め多くの大学では入試の範囲から統計学を除く所が多い様に思えます。その結果大学に入り初めて統計学を本格的に学ぶ学生が大部分でしょう。本稿では統計学への招待ということで最近盛んになりつつあるコンピューターを使った新しいタイプの統計手法について説明して行く事にします。

現在およそ科学と名の付いた学問分野で統計学的な処理をせずに学術論文を発表出来る分野はないといえよう。これは科学というものがあくまでも現実を抽象した理論にもとづいているからであることはもちろん、常に現実世界の現象により検証されなければならないからである。多くの場合それらはデータによる理論の検証という形をとることとなり、統計学の助けが必要となるわけである（ここでは、いわゆる量的なデータのみでなく質的なデータも考えている）。しかしすぐに明かになるように単にデータによる理論の検証といってみても各学問分野によりその扱うデータの形も性質も異なり、当然そこで使われる統計手法も異なってくる、故に農業統計、人口統計、生物統計、経済統計、etc. というように（中には統計学批判と呼ばれるものまで）数多くの統計学が存在するわけである。または逆に統計学的な手法を各学問分野へより積極的に取り入れる事により、よりクオンティテイティブ（quantitative）な、新しい学問分野も出来て

きている。計量生物学や計量経済学と呼ばれるものがその例である。このように統計学と一言でいっても実は非常に多くの時には全く異なった意味で使っていることがしばしばあり、すべてを一度に語ることは不可能である。そこでここではそれら多くの統計学のある意味での理論的バックボーンといえる数理統計学に話を限る事としよう。

いわゆる記述統計学が数千年の歴史を持っているのに較べ数理統計学の歴史はやっと百年になるかならないかである。その一つの理由は数理統計学の理論が確率論という比較的新しい数学の理論にもとづいているからであろう。数理統計学に於ける基本的な考えは、与えられたデータは無限に多くのデータの中から（ランダムに）採られた標本と見なすと云うことである。従ってこの考え方を厳密に理論化する為に確率モデルというものが必要となってきた訳である。それでは確率論はいつ、どの様に生まれたのか？　以下それにまつわる話を簡単に紹介しておこう。確率論の起源はフランスの数学者パスカル（Blaise Pascal）とフェルマー（M. Pierre Fermat）との間で一六五四年に交わされた手紙にあるというのが定説である。その手紙の中でパスカルは彼の友人である賭博師ド・メレ（Chevalier de Mere）からの質問について述べている。その中の一つのエピソードを紹介しておこう。

「一個のサイコロを四回ころがして、少なくとも一回六の目が出れば勝ち、しからざるときは負け、という賭の胴元として、ある貴族は思いどおりに儲けていた。彼の確率計算は次の通りであった。六の目が出る確率は六分の一、したがって四回なげればその確率は四倍になって六分の四＝三分の二のはずである。この確率は二分の一より大きいから勝つのは当然である。ただあまり勝ち続けると賭の相手がいなくなる。そこで貴族は、賭の方式を次のように変更した。今度は二個のサイコロを二四回転がして、少なくとも一回六と六のゾロ目が出れば勝ち、さもなくば負けとした。その貴族の思惑によれば、素人の目をごまかすべく若干ルールを複雑にしただけで、勝ちの確率は前の賭と同じはずであった。六と六のゾロ目が出る確率は $1/6\times1/6=1/36$。二四回投げれば $24/36=2/3$ の確率でゾロ目が出るはずである。案の定、以前の賭にはうんざりしていた客も、今回は自分に有利と見

てか、勝負に応じてきた。まさに貴族の思うつぼであった。所が今度は予想に反して、がぜん負けがこみはじめた。サイコロが正しく作られている限り、貴族の論理のどこかに間違いがあるはずである。そこで貴族はパスカルに相談を持ちかけた訳である。パスカルは直ちに貴族の間違いに気付き、正しい確率を次のように与えた。「少なくとも一回何かが起きる確率」を直接計算することは難しいが、この事の逆、つまり「一回も起きない確率」を計算することはたやすい。「少なくとも一回起こる確率」と「一回も起きない確率」を加えれば一となるから、はじめの賭の場合、貴族の勝つ確率は、負ける確率（四回サイコロを投げて一度も六の目が出ない確率）を一から引けばよい。つまり正解は、$1-(5/6)^4 \fallingdotseq 0.518$ である。かくてこの賭は貴族にとって確かに有利であった。しかし、二度目の賭で貴族に勝つ確率は、$1-(35/36)^{24} \fallingdotseq 0.491$ に等しい事が判る。二度目の賭で貴族の負けがこんだのは当り前の話である」。以上『佐和隆光著数量経済分析の基礎より引用』

さて現在我々が使っている公理論的確率論は二〇世紀初めにソ連の数学者コルモゴロフ（Andrei N. Kolmogorov）により確立された。それとは別に確率の解釈については現在に至るまで数多くの説が存在する。ラプラス（Pierre S. Laplace）による先験確率、フォン・ミーゼス（Richard von Mieses）によるコレクティーフの理論にもとづく経験確率、ケインズ（John M. Keynes）やジェフリーズ（Harold Jeffreys）による論理確率、そしてベイジアン（Beysian）と呼ばれる人々に強く支持されている主観確率等がある。それぞれの解釈は一長一短で、すべての場合に当てはまる解釈など存在するはずが無いのは当然であるが、それらの中で一番わかりやすい解釈は経験確率であろう。これは確率を「相対頻度」の極限と考える、即ちコインを投げ表の出る確率が$1/2$であることの意味は続けて何度もコインを投げたとき表の出る回数の相対頻度（表の出た回数÷コインを投げた回数）の極限が$1/2$ということである。これは丁度天気予報の降水確率の意味と同じである。もち論どの解釈をとろうとも、コルモゴロフの公理を満足する様に理論がつくられている事はいうまでも無い。現在多くの統計学者は確率論の解釈には余りこだわらない様である。過去

に於いて（おもに一九五〇年代から一九六〇年代後半にかけ）、主観確率と客観確率との間で大論争が戦わされたが、その不毛な論争も最近では影をひそめている。今でもベイズ統計学者の一部には、統計学はあくまでも主観確率にもとづくべきだと主張し続けている人達もいるが、ここではその様な立場はとらず標準的（？）な解釈（頻度論的解釈）で議論を進めて行く事にする。これは統計学があくまでも客観確率にもとづかなければならないと云う事ではなく、単に説明の便利さのためである。

以下第二節では伝統的な（数理）統計学の基本的な考え方について、第三、四節ではコンピュータを利用した新しい統計手法について述べてゆく。

二　伝統的数理統計学

本節では今世紀初頭ゴセット（W. H. Gosset）やフィッシャー（Ronald A. Fisher）という人々によりイギリスで生まれた近代統計学の基本的な考え方について簡単に説明しておく。前節でも述べた様に数理統計学ではデータはある未知の確率モデルから生み出されると仮定する。我々の目的は逆にデータからその確率モデルを推定することである。具体的には確率モデルは、データと分布関数と呼ばれるデータのとる値をコントロールする関数とにより表される。このフレイムワークの中でデータはそれが採集される前にはどの様な値となるかは我々には判らない。しかしデータがある与えられた値を取る確率は分布関数、F、により決定される。以下その意味で取られる前のデータ列のことを分布Fに従う（独立な）確率変数列、又実際にとられたデータ列を分布Fからの確率変数の実現値（標本）と呼ぶ事がある。従ってここでの目的は標本にもとづき、分布関数Fに関する何等かの推論を行うことと云うことができる。ただここで対象となる分布関数として、任意の分布関数のクラスを考えたとするとこれは範囲が広すぎ一般に意味のある結果を求める事は難しくなる。そこで伝統的には分布関数の形は判っているが、幾つかの未知のパラメータを含んでいるものと仮定する。そして与えられた標本よりそれら未知のパラメータの値を推定したり（推定論）、それがある一定の範囲に含まれるか否かを決定するのである（検定論）。これがパラメトリック統計学といわれるものであり、そしてその中でよく用いられる分布型に「正規分

布」と呼ばれるものがある。これは測定誤差の分布がこの形を持つことが経験的に一八世紀に既にガウス（Carl F. Gauss）等により発見されていたことや、ラプラスにより発見された中心極限定理により、応用の多くの場面で正規分布の使用が正当化されていることによる。さて正規分布は平均と分散という二つのパラメータにより完全に決定される（平均とは分布の中心を、また分散は分布の広がりを測るパラメータである。分散が大きい程分布の広がりは大きい）。それ故伝統的な統計解析に於いてはこの二つのパラメータの推定と検定が中心的な役割を占めてきていたのである。特に z—検定、t—検定と呼ばれる分散が既知の場合及び未知の時の正規分布の平均の検定はその推定問題と並び、長い間統計学の中心的な地位を占めていたといっても言い過ぎではないであろう。さて、ここでのポイントは二つある、一つはこの正規分布にもとづくモデルの広い範囲への応用可能性であり、もう一つは数表等を用いる事により具体的な答えが比較的容易に出せる事である。後者については分析の対象が正規分布の平均という非常に簡単な構造を持ったパラメータであるが故に、その推定値も又データの簡単な関数で表される（実際標本平均を用いればよいのだが、これはデータの線形関数となっている）ということに大きく負っている。しかし実際の応用に於いては話が若干ややこしくなってくる場合もある。その一つの例として次の話がある。

『種々の応用に於いて正規分布のみを仮定しただけではうまく行かないことが数多く見られる事も事実である。実際一九世紀フランスの葡萄栽培家は平均的な収穫量を求めるのに次の様な方法を用いていた。

「例えば過去二〇年間の平均収量を求めるには、二〇年間の単純平均を求めるのではなく、その中で一番収量の多い年と一番少ない年を除いた残りの一八年間の平均を求める」。これは二〇年の中には少なくとも一年ぐらいは例外的に良い年もあれば悪い年もあることを農民は長い間の経験より知っていた。そして彼らが知りたいのは、そのような年を除いた平均的な年の平均収量であるから、この様な方法を使用してきたのである。』

これと同様の話はオリンピックの体操競技等に於ける審判の得点集計にも見られる。この様な決定方式が必要とされる状況を、はたして一つの正規分布のみを用いた

確率モデルで十分表現出来るかといえば、もちろん無理であろう。ここでは平均的な年と異常な年とを別々の分布で説明する方がむしろ合理的であろう。さらにここで使われている推定法も標本の線形関数では無い。問題はこのより複雑なモデルのもとでも、単純な正規性の仮定下の時と同様、少なくとも現実に応用可能な答えが容易に得られるか否かというところである。所でここで議論されているのはロバストネス（頑健性）の問題と言われるもので、次に述べるノン・パラメトリック統計学と共に、ここ二〇年程の間アメリカ、特にカリフォルニア大学のバークレイ分校で盛んに研究されてきた問題である。ノン・パラメトリック法とは、その言葉の最も広い意味では分布型を全く仮定せずに統計的推論を行うものと定義されるが、もち論何等かの意味のあるシャープな結果を得る為には分布関数の形にいくつかの制限を加えねばならない事は当然である。具体的には微分可能であるとか、原点に対し対称である等の仮定を満足する分布のクラス（多くの場合それは正規分布を含む）を考え、そのなかで分布の平均とか分散という特性値についての推論を行う。問題はその様なモデルのもとでも、いかにして意味のある結論を得るのかであろう。さて数学的に高度な（?）手法を用いることによりここで挙げられた幾つかの問題に対する答えを得る事は可能である場合もあるが、それではあまり面白くない。そこで本稿では数学的には非常に簡単であるが、統計学的には非常に奥深い（?）考え方にもとづく手法を次節以降で紹介する事にする。

ここで過去に於ける数理科学の一つの発展方向を見てみると、そこには計算の簡略化という方向がある事は否定出来ないであろう。例えば莫大な量の積分の数値計算を行う代わりに簡単な計算のみで済む適当な近似式を探し求めてきたのである。所が近年に於けるコンピュータ・サイエンスの急速な発展は大量の計算を早く、そして安価に遂行する事を可能とした。これは過去に於いて数理科学では至上命令であった計算の簡略化という方向をそれほど強く追求せずとも良いことを意味している。逆にコンピュータと理論とを結び付けた新しい手法、即ち大量の数値計算に支えられた方法が生まれる素地が出来上がってきている。次節以降ではその様な考えに乗っ取った新しい統計手法について述べて行く。

（これは統計学に於ける分かりにくい用語法の問題であるが、次節以降で確率変数として推定値を考える時には（即ち標本をとる前の事）それを推定量と呼ぶ事にする。一方標本をとった後、即ち事後的に得られるものは推定値と呼びはっきりと区別する事にする。特にその確率論的な性質の分析を議論する際には、推定量を用いる。）

三　ジャックナイフ法

このジャックナイフ（jackknife）という奇妙な名前の付いた手法は一九四〇年代末にクヌーイ（M. Quenouille）という人が時系列分析に於けるパラメータ推定の際のバイアス修正法として考案したものである。（ここでバイアスとはパラメータの推定量の期待値（平均値）と真のパラメータの値との差の事である。）その後一九五〇年代半ばにアメリカの統計学者、トゥーキー（John Tukey）が推定量の分散の推定のための一般的な方法としてジャックナイフ法が使える事を示して以来、統計学の理論・応用の両面に大きなインパクトを与え続けてきた方法である。（ちなみにジャックナイフという名前は、ジャックナイフの様に便利な道具と言う意味をこめてトゥーキーが命名したものである。）ジャックナイフ法の手順及び（最も簡単な場合についてではあるが）その理論的な正当性は容易に証明される。具体的には、ある未知の分布Fにより定まる未知のパラメータをθと書くことにしよう。ここでの問題はn個の観測値（分布Fからの大きさnのランダム・サンプル）すべてを用い得られたθの推定量を$\hat{\theta}_n$と書くとき、$\hat{\theta}_n$の持つバイアス、及び$\hat{\theta}_n$の分散をいかに推定するかにある。これらの問題は推定量$\hat{\theta}_n$の精度を求める上で是非とも必要な量である。ここで推定量の形及び分布の形は共に任意とする。古典的な統計学の世界では前節で述べた様に、Fは未知のパラメータを含む正規分布であったり、又$\hat{\theta}_n$はサンプルの非常に簡単な式で与えられるため、現代の数学や確率論のレベルでもこれらの問題に答えることが可能であった。しかしより複雑な問題にあっては有限標本の理論で完全な解答を与えることは不可能に近い。そこで数理統計の世界では、統計的（汎）関数の理論というものにもとづく漸近理論（推定量の極限分布にもとづく理論の事）等を用い近似的な答えを求めている。

この理論にもとづく方法は推定量が統計的汎関数なるものので書き表せる時、そしてその汎関数の形が明示的に判っている場合には確かに有効な方法である。しかし一般にその形は複雑でたとえ汎関数でかき表せる事が判っていたとしても、それを簡単に数式の形に書き表せない事が多い。況やその微分可能性を調べたり、微係数を求めるなどもっての外であることが多く余り実用的な方法とはいいがたいとの批判が常にあった。そこで、このような高度に（？）数学的な理論を直接使わず、即ち汎関数の形を意識せずに、しかもそれらと同様に強力な結果を与えるのが実はジャックナイフ法なのである。

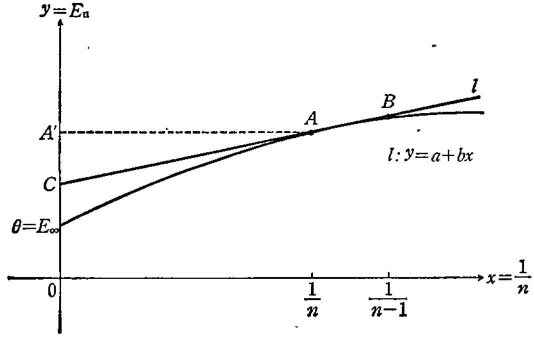
（A） ジャックナイフ・バイアス推定

$\hat{\theta}_n$ を右と同様 n 個のサンプル総てを用い計算された θ の推定量、$\hat{\theta}^{(i)}$ を第 i 番目のサンプル X_i を取り除きつくられた $(n-1)$ 個のサンプルより計算された推定量とする。即ち $\hat{\theta}^{(1)}$ から $\hat{\theta}^{(n)}$ まで、n 個の推定量が出来る訳である。また $\bar{\theta}^{(\cdot)}$ をそれらの算術平均 $(\bar{\theta}^{(\cdot)}=(\hat{\theta}^{(1)}+\cdots\cdots+\hat{\theta}^{(n)})/n)$ としよう。ここでジャックナイフ・バイアス修正済み推定量 $\hat{\theta}^{(J)}$ は

$$\hat{\theta}^{(J)}=n\hat{\theta}_n-(n-1)\bar{\theta}^{(\cdot)}$$

という式で定義される。以下簡単化の為 $\hat{\theta}^{(J)}$ を θ のジャックナイフ推定量と呼ぶことにする。さて一見して何かよく判らないこの式の説明は後で行うが $\hat{\theta}^{(J)}$ が $\hat{\theta}_n$ に較べ（平均して）小さなバイアスしか持たない事は、$\hat{\theta}_n$ の期待値が $(1/n)$ の多項式で表される場合には簡単な代数計算により示される。この時には $\hat{\theta}^{(J)}$ の期待値から $(1/n)$ の項が除去される。この事は図1により直観的に説明される。図1で E_n は $\hat{\theta}_n$ の期待値を表す。横軸は標本数の逆数がとってある、従って原点は無限に多くの標本を取った場合に対応している。図1では暗黙のうちに、いわゆる漸近的不偏性 $(E_\infty=E(\hat{\theta}_\infty)=\theta)$ が仮定されており、図中 A' と θ との間の長さが標本数 n の時のバイアスを表す。さて今点 A と点 B とを結ぶ直線 l の y 切片 C の高さは、$E_{n-1}+n(E_n-E_{n-1})$ で与えられる。さて（少なくともこの図の中では）点 C の方が点 A' より θ に近い、その意味でもしも点 C で θ を推定するならばよりよい推定値が得られるはずである。所で現実には多くの場合 E_n のグラフが図1の様な形をしている事が知られている。そしてジャックナイフ推定量 $\hat{\theta}^{(J)}$ は E_n を $\hat{\theta}_n$ で、E_{n-1} を $\bar{\theta}^{(\cdot)}$ で近似したものに他ならない。そ

図1



れ故にジャックナイフ推定量が元々の推定量より小さなバイアスを持つ事が期待されるのである。例えば、θをある（二次の有限なモーメントを持つ）分布関数の分散とすれば、θの自然な推定量は標本分散 $\hat{\sigma}_n^2$（$\hat{\sigma}_n^2=\sum_{i=1}^n(X_i-\bar{X})^2/n$, $\bar{X}=\sum_{i=1}^n X_i/n$）であろう。所で $\hat{\sigma}_n^2$ はθの不偏推定量でない事は簡単に証明される（実際 $E(\hat{\sigma}_n^2)=\theta-\theta/n$, である）。一方 $\hat{\sigma}_n^2$ のジャックナイフ推定量は、いわゆる分散の不偏推定量 s_n^2（$s_n^2=\sum_{i=1}^n(X_i-\bar{X})^2/(n-1)$）となる。ジャックナイフにより（$1/n$）の項が元の推定量の期待値から消えている事が見られる。但しこの例では $\hat{\sigma}_n^2$ の期待値が数学的に計算されることからも明かな通り、ジャックナイフを用いるには簡単すぎる。応用上ジャックナイフが必要となるのは、例えばθの推定法として四分の三分位値と四分の一分位値の差の二乗を用いるとか、又はかり込み型分散推定量（大きさの順に並べたサンプルの両端の一定割合の標本を捨てた残りの標本の分散）を用いた場合等である。

（B）　ジャックナイフ分散推定

トゥーキーによるジャックナイフ分散推定は次の様に行われる。$\hat{\theta}_n, \bar{\theta}^{(\cdot)}$ 等の記号の意味は右に同じとする。分布Fのもとでの $\hat{\theta}_n$ の分散を $\widehat{Var_F}\{\hat{\theta}_n\}$ であらわす時、そのジャックナイフ推定量は

$$\widehat{Var_J}\{\hat{\theta}_n\}=\{(n-1)/n\}\sum_{i=1}^n(\hat{\theta}^{(i)}-\bar{\theta}^{(\cdot)})^2$$

で与えられる。バイアス修正の時と同様にこの式の意味を考えるために、ジャックナイフ仮性値 $p^{(i)}$（pseudo value）なるものを定義しておく。

$$p^{(i)}=n\hat{\theta}_n-(n-1)\bar{\theta}^{(i)}, \qquad i=1,\ldots\ldots,n\ ,$$

仮性値を用いれば、ジャックナイフ分散推定量は

$$\widehat{Var_J\{\hat{\theta}_n\}}=\sum_{i=1}^{n}(p^{(i)}-p^{(\cdot)})^2/n(n-1),$$

$$p^{(\cdot)}=\sum_{i=1}^{n}p^{(i)}/n,$$

と書かれる。即ち $\widehat{\mathrm{Var}_J\{\hat{\theta}_n\}}$ は $p^{(i)}$ の標本平均 $p^{(\cdot)}$ の分散の推定値となっている。最も簡単な場合：θ が分布の平均値、$\hat{\theta}_n$ が標本平均である時、$p^{(i)}$ は X_i そのものとなり、ジャックナイフ分散推定値は $s_n{}^2/n$ となる。トゥーキーがジャックナイフ分散推定法を提唱した時の直観的な正当化の根拠としたのが今の議論である。しかしながらここでもバイアス修正の時と同様の説明を図2及び図3を用いて出来る。図2に於て、$p^{(i)}$ は点 $\hat{\theta}^{(i)}$ と点Oとを結ぶ直線の y 切片である。それでは次に $p^{(i)}$ の標本分散 $\sum_{i=1}^{n}(p^{(i)}-p^{(\cdot)})^2/(n-1)$ とは一体何なのかを図3を用い考えて見よう。今 $\hat{\theta}_n$ の真の分散が近似的に σ^2/n となっているものと仮定しよう。ここで σ^2 は $\hat{\theta}_1$ の分散である。この関係は $\hat{\theta}_n$ がランダム・サンプルの標本平均である時には正確に成り立つが、もち論一般には、幾つかの条件下で始めて近似的に成立する。話が少々ややこしくなるが $\hat{\theta}_n$ が何らかの意味で、X_i の線形関数の一次結合で近似されるならばうまくいく。さて図3に於て線分 HI の長さは真の分散 σ^2 の大きさに対応しているものとする、又線分 FG は標本数 n の時の分散、σ^2/n に対応している。ここで三つの三角形 $\triangle OFG$, $\triangle ODE$, そして $\triangle OHI$ を考えれば、$\hat{\theta}_{n-1}$ の分散 ($\fallingdotseq \sigma^2/(n-1)$) は線分 DE の長さに対応している。次に三

図2

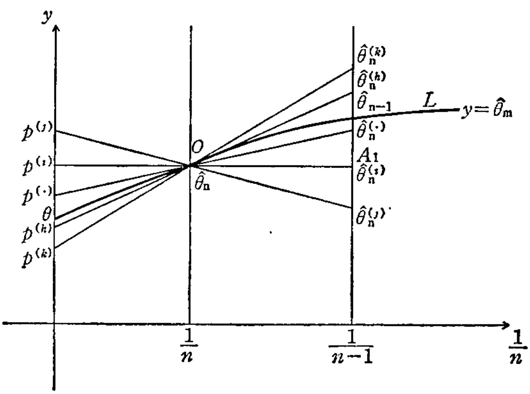
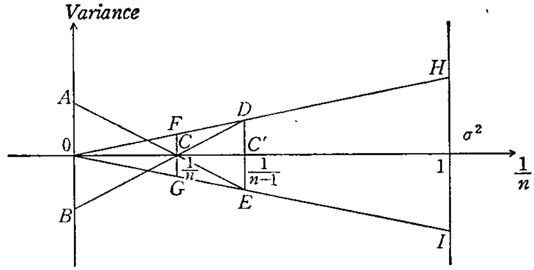


図3



$|HI|\simeq\sigma^2$, $|DE|\simeq\sigma^2/(n-1)$　$|FG|\simeq\sigma^2/n$
$|CC'|=1/n(n-1)$

角形 $\triangle CDE$ と $\triangle CAB$ を比較すれば

$$|AB|=|OC|\cdot|DE|/|CC'|$$

ここで $|AB|$ は線分 AB の長さである。一方

$$|CC'|=1/n(n-1),\ |OC|=1/n,\ |DE|\doteqdot\sigma^2/(n-1)$$

を得る。即ち線分 AB の長さは真の分散 σ^2 に対応している。よって図2、3より

$$\sum_{i=1}^{n}(p^{(i)}-p^{(\cdot)})^2/(n-1)\doteqdot\sigma^2$$

を得る。以上の議論から $\hat{\theta}_n$ の分散のジャックナイフ推定量は、実は σ^2/n の推定量に他ならない事が示されたのである。ここの議論の重要なポイントは $\hat{\theta}_n$ が何らかの意味で線形近似される事にある。これは実関数の世界では、ある関数が微分可能ということに対応している。即ちバイアス修正の場合と同様ジャックナイフ分散推定も推定量にたいし、ある種の滑らかさを仮定した上での漸近理論にもとづいているのである。但し現実にジャックナイフ法を適応する時我々は $\hat{\theta}_n$ の微分可能性、その他の数学的、確率論的な議論とは全く無関係に分析を進めることが出来る所にこの方法の最大の利点の一つがある。一方 $\hat{\theta}_n$ が十分滑かで無い場合にはジャックナイフ分散推定は有効では無い事は直ちに予想される。実際 $\hat{\theta}_n$ が標本中央値 (sample median) である時がその典型的な例である。この様な問題に対しても有効な方法が次節で紹介するブーツトラップ法 (bootstrap method) や一度に複数個のサンプルを取り除くジャックナイフ法 (take-two, take-three jackknife, etc.) である。

四　ブーツトラップ法

ブーツトラップ法（Bootstrap）の目的も基本的にはジャックナイフ法と同じであるが、ジャックナイフ法よりも弱い条件下でも有効である事が報告されている。その事に対する代償はジャックナイフ法を遥かに上回る莫大な量の数値計算である。ブーツトラップ法は一九七〇年代末にこれも又アメリカの統計学者エフロン（Bradley Efron）により提唱された方法で概念的にはジャックナイフ法よりずっと簡単である。エフロンの方法は次の三段階より成る。

第一段階　与えられたn個の標本 $X_1,\ldots\ldots,X_n$ より経験分布関数$\hat{F}$をつくる。さらに $\hat{\theta}_n$ をn個の標本すべてを用い計算されたθの推定値とする。（ここで経験分布関数とは、X_i の各々を $1/n$ の確率でとる分布関数の事である。）

第二段階　分布$\hat{F}$より大きさmの標本 $X_1^*,\ldots\ldots,X_m^*$ をとり、それらより計算されるθの推定値を $\hat{\theta}_m^*$ と書くことにする。具体的には $X_1,\ldots\ldots,X_n$ の中から繰り返しを許してm個のサンプルをランダムに選び、それらにもとづき $\hat{\theta}_m^*$ を計算する。

第三段階　第二段階をB回（例えば $B=100, 500$, etc.）繰り返し $\hat{\theta}_m^{*(1)}, \hat{\theta}_m^{*(2)}, \ldots\ldots, \hat{\theta}_m^{*(B)}$, を得る。そしてそれらを使い $\hat{\theta}_m$ のブーツトラップ・バイアス推定値 $\widehat{B}_{BOOT}$、ブーツトラップ分散推定値 $\widehat{Var}_{BOOT}$、そしてブーツトラップ分布推定値 $\widehat{Dist}_{BOOT}(x)$ を次の様に定義する。

$$
(1)\quad
\begin{aligned}
\widehat{B}_{BOOT} &= \hat{\theta}_m^{*(\cdot)} - \hat{\theta}_n \\
\widehat{Var}_{BOOT} &= \textstyle\sum_{b=1}^{B}(\hat{\theta}_m^{*(b)} - \hat{\theta}_m^{*(\cdot)})^2/(B-1) \\
\widehat{Dist}_{BOOT}(x) &= \#\{b : \hat{\theta}_m^{*(b)} \leq x\}
\end{aligned}
$$

ここで $\hat{\theta}_m^{*(\cdot)} = \sum_{b=1}^{B}\hat{\theta}_m^{*(b)}/B$, 又 $\#\{b : *\}$ は $*$ という性質を満足するbの個数の意味である。即ち $\widehat{Dist}_{BOOT}(x)$ は $\hat{\theta}_m^{*(1)}, \hat{\theta}_m^{*(2)}, \ldots\ldots, \hat{\theta}_m^{*(B)}$ の経験分布関数である。

一般にmは任意であるが応用上 $m=n$ の事が多い故、以下ではその様にしておこう。さて、確率の頻度論的解釈によれば、$\hat{\theta}_n$ の真のバイアス、分散、及び分布は（1）に於て、$\hat{F}$を真の分布関数Fに置き換え、さらにBを無限大へ近づけた時の極限である。一方、適当な正則性の条件のもとで、経験分布関数は標本数を増やしていった時、真の分布関数へ一様に収束することが証明されてい

る（グリベンコ・カンテリの定理）。従って$\hat{\theta}_n$が〃連続〃であれば、ブーツトラップ推定量の一致性（推定量の値が真のパラメータの値に収束する事）は保証される。ここでは推定量の〃微分可能性〃までは必要でなく、連続性のみで十分なのである。もち論直観的には微分可能でないならば、その収束はかなり遅くなると考えられるが大事な事はこの方法がジャックナイフ法よりも広い範囲で応用可能である事である。実際エフロンは、標本中央値のブーツトラップ分散推定量が一致性を持つことを証明した（Efron, B. The Jackknife, the Bootstrap and Other Resampling Plans, 1982, SIAM 参照）。

ブーツトラップ法の確率論的、数学的性質については最近多くの研究が発表されている。しかしBの大きさを如何に決めたらよいのかとか、ブーツトラップ法の有効範囲の明確な設定（例えば、時系列分析や、回帰分析でも果して有効なのか）といったごく基本的な問題に対してすら十分満足のいく答えはまだ与えられていない。少々皮肉な見方ではあろうが、これらの問題に対し数学的な答えが与えられる様な簡単なモデルのもとではブーツトラップ法は不必要であろう。わざわざシミュレイション的な方法に頼らずとも数学的に答えを求めればよいからである。一つの方法として考えられるのはブーツトラップ法をブーツトラップする事であるが、個々の問題毎に分析を行わなければならない等、あまり良い方法とはいいがたい。

とはいうものの、ブーツトラップ法は数学的、確率論的な分析だけで意味のある答えを求めることが非常に困難である様な複雑な、しかし統計学的には十分意味のある手法の分析に現在頻繁に用いられているテクニックである。又通常如何なる問題に於いても、何らかの答えが必ず得られるというブーツトラップ法の長所は実はその最大の欠点でもある。それ故にブーツトラップ法の理論的研究は今急を要する課題といえる。一方ブーツトラップ法とジャックナイフ法との関係については多くの結果が報告されている。ジャックナイフ法が標本からの規則的な再標本法であるのに対し、ブーツトラップ法は完全なランダム標本にもとづいている。この意味でジャックナイフ法はブーツトラップ法の一つの近似である。実際ブーツトラップ法で必要とされる計算量はジャックナイフ法の数倍から多いときには数百倍となっていることよ

り、ジャックナイフで分析が可能であればそれにこした事はない。より厳密には $\hat{\theta}_n$ が〃線形〃であれば

$$\widehat{Var}_J(\hat{\theta}_n)=(n/n-1)\widehat{Var}_{BOOT}(\hat{\theta}_n)$$

が、又 $\hat{\theta}_n$ が〃二次の汎関数〃であれば

$$\hat{B}_J(\hat{\theta}_n)=(n/n-1)\hat{B}_{BOOT}(\hat{\theta}_n)$$

が成立する事がエフロンにより証明されている。

五　終わりに

本稿ではコンピューターを使った新しい統計学についてという事で、ジャックナイッフ法とブーツトラップ法という二つの方法について簡単に紹介してみたが、もち論現在研究が進められているこの手の方法はまだ沢山ある。プロジェクション・パーシューと呼ばれるものは、判別分析と呼ばれる統計学の一部門を新しく作り直すかも知れない。又逐次分析と呼ばれる方法は統計的な決定問題に於て、その決定に至る時間を今までより大幅に短縮するであろう。伝統的な統計学と共にここで述べた様な新しい方法に対しても、興味を持つ学生が増える事を期待しつつ本稿を終わる事とします。

（一橋大学教授）