U-統計量の正規近似の改良とその比較

前園, 伸彦
九州大学経済学部 : 助教授

藤岡, 由美
九州大学数理学研究科 : 院生

https://doi.org/10.15017/4363006

出版情報：經濟學研究. 66 (1), pp.87-100, 1999-06-30. 九州大学経済学会

U- 統計量の正規近似の改良とその比較

前　園　宜　彦
藤　岡　由　美

1. 序

$X_1, X_2, \ldots, X_n$ を母集団分布 $F_\theta(x)$ からの無作為標本とする。即ち $X_1, X_2, \ldots, X_n$ は互いに独立で同じ分布 $F_\theta(x)$ に従う確率変数である。母数 θ に対して,

$$E[h(X_1, X_2, \ldots, X_r)]=\theta$$

となる成分の入れ替えに対して不変な関数 $h(x_1, x_2, \ldots, x_r)$ （r 次のカーネルと呼ばれる）があるとする。このとき U-統計量 U_n は,

$$U_n=\binom{n}{r}^{-1}\sum_{C_{n,r}}h(X_{i_1}, X_{i_2}, \ldots, X_{i_r})$$

で定義される。ここで $\sum_{C_{n,r}}$ は $1\leq i_1<i_2<\cdots<i_r\leq n$ を満たす全ての組み合わせ $(i_1, i_2, \ldots, i_r)$ についての和を表す。U_n は $E(U_n)=\theta$ を満たす θ の最小分散不偏推定量になっている。この U- 統計量のクラスは Hoeffding (1948) によって提案されたクラスで，$h(x)=x$ とおくと U_n は標本平均 $\overline{X}=\sum_{i=1}^n X_i/n$ となり，$h(x, y)=(x-y)^2/2$ とおくと U_n は標本不偏分散 $\sum_{i=1}^n (X_i-\overline{X})^2/(n-1)$ である。このように，実際に使われる統計量の多くがこのクラスに属するか，あるいはこのクラスの統計量で近似される。また Hoeffding (1948) によって標準化した U_n について中心極限定理が成り立つことが示されている。即ち，任意の x に対して，$\sigma_n^2=Var(U_n)$ とおくとき

$$\lim_{n\to\infty}P\left(\frac{U_n-\theta}{\sigma_n}\leq x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{t^2}{2}\right\}dt$$

である。

また分布の近似の精度を上げることを目的とした，エッジワース展開が標準化 U- 統計量について研究され，$r=2$ のとき，Callaert et al. (1980)，一般の r のとき，Maesono (1987) により求められている。それは

$$P\left(\frac{U_n-\theta}{\sigma_n}\leq x\right)=\Phi(x)+\phi(x)\{n^{-1/2}m_1(x)+n^{-1}m_2(x)\}+o(n^{-1})$$

である。ここで $\Phi(x)$，$\phi(x)$ は標準正規分布の分布関数及び密度関数であり，$m_1(x), m_2(x)$ は x の関数で，具体的な式は次節で述べる。また $o(n^{-1})$ は $\lim_{n\to\infty} n|o(n^{-1})|=0$ となる，n^{-1} よりも速く0に収

束する実数列である。さらに標準化するときに分散の代わりにその推定量で置き換えた，スチューデント化 U- 統計量についての n^{-1} の項までのエッジワース展開が，Maesono (1997) により求められている。これは

$$P\left(\frac{U_n-\theta}{\hat{\sigma}_J}\leq x\right)=\Phi(x)+\phi(x)\{n^{-1/2}m_1^*(x)+n^{-1}m_2^*(x)\}+o(n^{-1})$$

である。ここで $m_1^*(x), m_2^*(x)$ は x の関数であり，$\hat{\sigma}_J$ はジャックナイフ分散推定量

$$\hat{\sigma}_J^2=\frac{n-1}{n}\sum_{i=1}^{n}(U_n-U_n^{(i)})^2$$

の正の平方根である。ただし $U_n^{(i)}$ は n 個の標本から X_i を除いた $n-1$ 個の標本に基づく対応する U- 統計量である。

これらのエッジワース展開を利用すれば正規近似の精度を上げることができる。しかし次節で述べるように，各展開の項は未知の母数を含んでおり，実際の応用の場面では，その母数を推定量で置き換える必要がある。従って，単純にエッジワース展開を利用する方法では，理論的には収束のオーダーの改良は得られないことになる。

他方，統計量を変換して正規近似の精度を上げようという正規化変換が提案されている。代表的な例が，相関係数に関するフィシャーの z- 変換，ウィルソン・ヒルファーティ近似，バートレット修正である。Johnson (1978) はスチューデントの t- 統計量について，統計量の修正により近似の精度が上がることを示している。Konishi (1981) は一般の統計量を変換してエッジワース展開の $n^{-1/2}$ の項を消すような正規化変換を議論し，構成法を提案している。Abramovitch & Singh (1985) は，エッジワース展開が与えられた時に $n^{-1/2}, n^{-1}$ の各項を逐次的に消去する方法を提案し，t- 統計量については n^{-1} の項まで消去する変換を求めた。また，Konishi (1991) は推定量が汎関数で表されている時に，von Mises 展開をもとに指数型関数での正規化変換を考察している。この変換は単調なものになっている。さらに Hall (1992) は，多項式の形での単調な変換を提案している。

本論文では，U- 統計量及びスチューデント化 U- 統計量に対しての n^{-1} の項までのエッジワース展開に基づく分布の近似と，$n^{-1/2}$ の項を消去する正規化変換による近似について議論する。近似の良さの理論的な比較とともに，母平均と母分散に対しての信頼区間のシュミレーションによる比較を行う。

2節では，U- 統計量の直交分解を求め，モーメントの評価を得る。そして標準化 U- 統計量及びスチューデント化 U- 統計量のエッジワース展開に基づく分布の近似を述べる。3節では，エッジワース展開の $n^{-1/2}$ の項を消去する正規化変換について紹介するとともに，Hall (1992) の提案した変換の一般化について述べる。最後に4節で正規近似の改良による信頼区間の構成を述べ，母平均及び母分散の信頼区間についてシミュレーションの結果を示す。

2．U- 統計量のエッジワース展開

最初に Lai & Wang (1993) により求められた漸近 U- 統計量の漸近展開の結果を述べる。統計量 T_n が

$$T_n=n^{-1/2}\sum_{i=1}^{n}\alpha(X_i)+n^{-3/2}\sum_{i=1}^{n}\alpha'(X_i)+n^{-3/2}\sum_{1\le i<j\le n}\beta(X_i,X_j)$$
$$+n^{-5/2}\sum_{1\le i<j<k\le n}\gamma(X_i,X_j,X_k)+o_p(n^{-1})$$

と表されていて, $\alpha(X_1)$, $\alpha'(X_1)$, $\beta(X_1,X_2)$, $\gamma(X_1,X_2,X_3)$は

$$E[\alpha(X_1)]=E[\alpha'(X_1)]=0,\ E[\beta(X_1,X_2)\,|\,X_1]=0\ a.s.,$$
$$E[\gamma(X_1,X_2,X_3)\,|\,X_1,X_2]=0\ a.s.,\ E\{|\alpha'(X_1)|^3+|\gamma(X_1,X_2,X_3)|^4\}<\infty$$

の条件を満たすと仮定する。ここで$o_p(n^{-1})$は確率変数列で

$$P\{|o_p(n^{-1})|\ge n^{-1}(\log n)^{-1}\}=o(n^{-1})$$

を満足する。この T_n を Lai & Wang (1993)は漸近 U- 統計量と呼んでいる。多くの統計量が漸近 U-統計量であることが知られており, 標準化 U- 統計量$(U_n-\theta)/\sigma_n$ はもちろん漸近 U- 統計量である。この漸近 U- 統計量に対して Lai & Wang (1993)は n^{-1} の項までのエッジワース展開を求めている。次の条件(C), (D)を準備する。

[条件(C)]: $p>2$に対して$E|\beta(X_1,X_2)|^p<\infty$とする。また$I_A$を事象$A$の定義関数とするとき

$$K(p-2)>4p+(28p-40)I_{\{E|\gamma(X_1,X_2,X_3)|>0\}},$$
$$E[f_v^2(X_1)]<\infty\ (v=1,\ldots,K)$$

となるようなK個のボレル関数$f_v:\mathbf{R}\to\mathbf{R}$が存在する。さらに

$$W_v=(Lf_v)(X_1),\ (Lf)=E[\beta(y,X_2)f(X_2)]$$

とおくとき, $(W_1,\ldots,W_K)$の分散共分散行列は正定値である。

[条件(D)]: 定数c_vとボレル関数$q_v:\mathbf{R}\to\mathbf{R}$が存在し,

$$E[q_v(X_1)]=0,\ E|q_v(X_1)|^p<\infty\ (p\ge 5),\ \beta(X_1,X_2)=\sum_{v=1}^{K}c_v q_v(X_1)q_v(X_2)$$

を満たす。さらに$0<\rho<\min\{1,2(1-11p^{-1}/3)\}$に対して

$$\lim_{|t|\to\infty}\ \sup_{|s_1|+\cdots+|s_K|\le|t|^{-\rho}}\left|E\left[\exp\left(it\left\{\alpha(X_1)+\sum_{v=1}^{K}s_v q_v(X_1)\right\}\right)\right]\right|<1.$$

ここで次の記号を準備する。

$$\sigma^2=E[\alpha^2(X_1)],\ \eta=E[\alpha(X_1)\alpha'(X_1)],$$
$$\kappa_3=E[\alpha^3(X_1)]+3E[\alpha(X_1)\alpha(X_2)\beta(X_1,X_2),]$$
$$\kappa_4=E[\alpha^4(X_1)]-3\sigma^4+4E[\alpha(X_1)\alpha(X_2)\alpha(X_3)\gamma(X_1,X_2,X_3)]$$
$$+12E[\alpha^2(X_1)\alpha(X_2)\beta(X_1,X_2)]+12E[\alpha(X_1)\alpha(X_2)\beta(X_1,X_3)\beta(X_2,X_3)],$$

$$P_1(x)=\frac{\kappa_3(x^2-1)}{6\sigma^3},$$

$$P_2(x)=\left\{\eta+\frac{E[\beta^2(X_1,X_2)]}{4}\right\}\frac{x}{\sigma^2}+\frac{\kappa_4}{24\sigma^4}(x^3-3x)+\frac{\kappa_3^2}{72\sigma^6}(x^5-10x^3+15x).$$

κ_3, κ_4 はそれぞれ T_n の3次及び4次のキュムラントである。これらを使うと漸近 U- 統計量のエッ

ジワース展開が求まる。

[補題１] αがモーメントに対する条件とクラーメル条件

$$E[\alpha^2(X_1)]=\sigma^2>0,\ E[\alpha^4(X_1)]<\infty,\ \limsup_{|t|\to\infty}|E[\exp\{it\alpha(X_1)\}]|<1$$

を満たし，条件(C)または条件(D)のどちらかが成り立つ時，

$$P\left(\frac{T_n}{\sigma}\le x\right)=\boldsymbol{\Phi}(x)-\phi(x)\{n^{-1/2}P_1(x)+n^{-1}P_2(x)\}+o(n^{-1}).$$

(証明) Lai & Wang(1993)を参照。

$n^{-1/2}$の項までのエッジワース展開には条件(C)，(D)は必要ない。

[補題２] αがモーメントに対する条件とクラーメル条件

$$E[\alpha^2(X_1)]=\sigma^2>0,\ E[\alpha^3(X_1)]<\infty,\ \limsup_{|t|\to\infty}|E[\exp\{it\alpha(X_1)\}]|<1$$

を満たすとき

$$P\left(\frac{T_n}{\sigma}\le x\right)=\boldsymbol{\Phi}(x)-n^{-1/2}\phi(x)P_1(x)+o(n^{-1/2}).$$

ここで$o(n^{-1/2})$は$o(n^{-1})$と同様に，$n^{-1/2}$よりも速く0に収束する実数列である。

(証明) Lai & Wang (1993) を参照。

次にU-統計量の漸近的な性質を議論するときに有用な*ANOVA*-分解，またはH-分解と呼ばれる直交分解を準備する。$E|h(X_1,X_2,\ldots,X_r)|<\infty$のとき，次の記号を定義する。

$$a_k(x_1,x_2,\ldots,x_k)=E[h(X_1,X_2,\ldots,X_r)|X_1=x_1,\ldots,X_k=x_k]-\theta\quad(1\le k\le r),$$

$$g_1(x_1)=a_1(x_1),\ g_2(x_1,x_2)=a_2(x_1,x_2)-g_1(x_1)-g_1(x_2),\ldots,$$

$$g_r(x_1,\ldots,x_r)=a_r(x_1,\ldots,x_r)-\sum_{k=1}^{r-1}\sum_{C_{r,k}}g_k(x_{i_1},\ldots,x_{i_k}),$$

$$A_k=\sum_{C_{n,k}}g_k(X_{i_1},\ldots,X_{i_k}).$$

このとき，$U_n-\theta$は次のように書き表される。

$$U_n-\theta=\binom{n}{r}^{-1}\sum_{k=1}^{r}\binom{n-k}{r-k}A_k.$$

定義より

$$E[g_k(X_1,\ldots,X_k)|X_1,\ldots,X_{k-1}]=0\ a.s. \tag{1}$$

となることが分かる。またこの式(1)を使うと

$$\sigma_n^2=\sum_{k=1}^{r}\binom{r}{k}^2\binom{n}{k}^{-1}E[g_k^2(X_1,\ldots,X_k)]$$

$$=\frac{r^2}{n}E[g_1^2(X_1)]+\frac{r^2}{2n(n-1)}E[g_2^2(X_1,X_2)]+\cdots$$

$$+\frac{r!}{n(n-1)\cdots(n-r+1)}E[g_r^2(X_1,\ldots,X_r)]$$

となる。さらに式(1)を使うとA_kがマルチンゲールの性質を持つことが示せ，次のようなモーメント

の評価が求まる。

[補題３] $E|g_k(X_1,X_2,\ldots,X_k)|^p<\infty$とする。このとき次の不等式を満たす，分布と g_k には依存するが，n には依存しない正の定数 c,c' が存在する。

(1) $1\leq p<2$の時，

$$E|A_k|^p\leq cn^k. \tag{2}$$

(2) $2\leq p$ の時，

$$E|A_k|^p\leq c'n^{pk/2}. \tag{3}$$

（証明）Maesono（1997）を参照。

この直交分解とモーメントの評価を利用して，標準化 U- 統計量に対して n^{-1} の項までのエッジワース展開が，Maesono（1987）により求められている。ここで次の記号を準備する。

$$\begin{aligned}
&\xi_1^2=E[g_1^2(X_1)],\ e_1=E[g_1^3(X_1)],\ e_2=(r-1)E[g_1(X_1)g_1(X_2)g_2(X_1,X_2)],\\
&e_3=E[g_1^4(X_1)],\ e_4=(r-1)E[g_1^2(X_1)g_1(X_2)g_2(X_1,X_2)],\\
&e_5=(r-1)^2E[g_1(X_2)g_1(X_3)g_2(X_1,X_2)g_2(X_1,X_3)],\\
&e_6=(r-1)(r-2)E[g_1(X_1)g_1(X_2)g_1(X_3)g_3(X_1,X_2,X_3)],\\
&\kappa_3=\xi_1^{-3}(e_1+3e_2),\\
&\kappa_4=\xi_1^{-4}(e_3-3\xi_1^4+12e_4+12e_5+4e_6).
\end{aligned}$$

さらに

$$P_n(x)=\Phi(x)-\phi(x)\left\{\frac{\kappa_3}{6\sqrt{n}}(x^2-1)+\frac{\kappa_4}{24n}(x^3-3x)+\frac{\kappa_3^2}{72n}(x^5-10x^3+15x)\right\}$$

とおく。ここで，κ_3, κ_4 はそれぞれ標準化 U- 統計量（$(U_n-\theta)/\sigma_n$）の３次及び４次のキュムラントである。このとき次の定理が成り立つ。

[定理１] モーメントに対する条件及びクラーメル条件

$$E|h(X_1,X_2,\ldots,X_r)|^4<\infty,\ E[g_1^2(X_1)]=\sigma^2>0,\ \limsup_{|t|\to\infty}|E[\exp\{itg_1(X_1)\}]|<1$$

を仮定する。さらにβを g_2, γを g_3 でそれぞれ置き換えて，条件(C)または条件(D)が成り立つ時，

$$P\left(\frac{U_n-\theta}{\sigma_n}\leq x\right)=P_n(x)+o(n^{-1}).$$

（証明）Maesono（1987）を参照。

標準化 U- 統計量の σ_n にジャックナイフ推定量

$$\hat{\sigma}_J=\left\{\frac{n-1}{n}\sum_{i=1}^{n}(U_n^{(i)}-U_n)^2\right\}^{1/2}$$

を代入したスチューデント化 U- 統計量$(U_n-\theta)/\hat{\sigma}_J$についてのエッジワース展開を Maesono（1997）が n^{-1} の項まで求めている。次の記号を準備する。

$$\xi_2^2=E[g_2^2(X_1,X_2)],\ \lambda_1=\xi_1^{-3}(2e_1+3e_2),\ \lambda_2=\xi_1^{-3}(e_1+3e_2),$$
$$\nu_1=-\xi_1^{-6}(2e_1+3e_2)^2,$$
$$\nu_2=6\xi_1^{-4}(e_3-6\xi_1^4+12e_4+6e_5+4e_6)-2\xi_1^{-6}(2e_1+3e_2)(2e_1+9e_2),$$
$$\nu_3=3\xi_1^{-6}(4e_1^2+12e_1e_2+3e_2^2)+18\xi_1^{-4}(\xi_1^2\xi_2^2-e_3+2\xi_1^4-4e_4-2e_5),$$

$$Q_n(x)=\Phi(x)+\phi(x)\frac{1}{6\sqrt{n}}(\lambda_1x^2+\lambda_2)+\phi(x)\frac{1}{72n}(\nu_1x^5+\nu_2x^3+\nu_3x).$$

このとき次の定理が成り立つ。

[定理2] モーメントに対する条件及びクラーメル条件

$$E|h(X_1,X_2,\dots,X_r)|^9<\infty,\ E[g_1^2(X_1)]=\sigma^2>0,\ \limsup_{|t|\to\infty}|E[\exp\{itg_1(X_1)\}]|<1$$

を仮定する。さらにスチューデント化 U- 統計量の漸近表現（Maesono（1997）を参照）に現れる β, γ に対して条件(C)または条件(D)が成り立つ時，

$$P\left(\frac{U_n-\theta}{\hat{\sigma}_J}\leq x\right)=Q_n(x)+o(n^{-1}).$$

（証明）Maesono（1997）を参照。

これらのエッジワース展開には，３次及び４次のキュムラントが現れ，それらは未知の母数を含んでいる。漸近展開を利用するためには，未知の母数を推定しなければならず，理論的には収束のオーダーは，標準化では改良されず，スチューデント化では $o(n^{-1/2})$ である。しかし４節のシュミレーション結果に現れているように，スチューデント化 U- 統計量に基づく近似はかなり良いものである。

３．正規化変換

統計量を変換することによって，正規分布への収束のオーダーを改良する正規化変換を考察する。正規化変換は多くの論文で研究されており，エッジワース展開のコーニッシュ・フィッシャーの反転をもとに $n^{-1/2}$ の項を消す正規化変換が Abramovitch & Singh(1985)により提案されている。Konishi (1991) は F の経験分布関数を F_n としたとき，母数 $\theta=\theta(F)$ の推定量を $\hat{\theta}_n=\theta(F_n)$ とし，この汎関数のフォン・ミーゼス展開を使って指数型関数を使った正規化変換を議論している。そしてこれまでに知られている多くの統計量の正規化変換が，これを使って説明できることを示した。また，Hall (1992) は多項式の形の単調な変換で，エッジワース展開の $n^{-1/2}$ の項を0にするような正規化変換を提案している。

ここでは Hall(1992)の正規化変換の修正を行い，それを U- 統計量に適用する。規準化された（スチューデント化）統計量 S が以下のエッジワース展開をもつとする。

$$P(S\leq x)=\Phi(x)+n^{-1/2}\kappa(ax^2+b)\phi(x)+o(n^{-1/2}). \tag{4}$$

ここで，a,b は既知の定数で κ は推定可能な母数を表わす。κ の推定量を $\hat{\kappa}$ とすると，単調な変換（S の関数として T は単調増加）

$$T=S+a\frac{\hat{\kappa}}{\sqrt{n}}S^2+\frac{b\hat{\kappa}}{\sqrt{n}}+\frac{1}{3n}a^2\hat{\kappa}^2S^3$$

は，次を満たす。

$$P(T\leq x)=\mathbf{\Phi}(x)+o(n^{-1/2}).$$

T は $n^{-1/2}$ の項を消す正規化変換であり，S がスチューデントの t- 統計量の時には式(4)の展開を持つ。しかし，一般の統計量では次の修正が必要である。

統計量 S_n が次のように表されたとする。

$$S_n=n^{-1/2}\delta+n^{-1/2}\sum_{i=1}^{n}h_1(X_i)+n^{-3/2}\sum_{1\leq i<j\leq n}h_2(X_i,X_j)+o_p(n^{-1/2}) \tag{5}$$

ただし，$n^{-1/2}\delta$ はバイアスで，$h_1(X_1), h_2(X_1,X_2)$ は，

$$E[h_1(X_1)]=0,\ E[h_1^2(X_1)]=1,\ E[h_2(X_1,X_2)\,|\,X_1]=0\ a.s.,$$
$$E\{|h_1(X_1)|^3+|h_2(X_1,X_2)|^3\}<\infty$$

の条件を満足し，また $o_p(n^{-1/2})$ は確率変数列で

$$P(|o_p(n^{-1/2})|\geq n^{-1/2}(\log n)^{-1})=o(n^{-1/2})$$

を満たす。H-分解を利用すると，多くの統計量が上記の条件を満足することが示せる。このとき次の変換

$$T^*=S_n+\frac{u}{\sqrt{n}}(S_n)^2+\frac{v}{\sqrt{n}}$$

を考える。ただし

$$u=-\frac{1}{6}E[h_1^3(X_1)]-\frac{1}{2}E[h_1(X_1)h_1(X_2)h_2(X_1,X_2)]$$

$$v=\frac{1}{6}E[h_1^3(X_1)]+\frac{1}{2}E[h_1(X_1)h_1(X_2)h_2(X_1,X_2)]-\delta$$

である。このとき補題2により

$$P(T^*\leq x)=\mathbf{\Phi}(x)+o(n^{-1/2})$$

となることが示せる。従って T^* はエッジワース展開の $n^{-1/2}$ の項を0にする正規化変換である。T^* は未知の母数 u, v を含むので，その一致推定量 $\hat{u}, \hat{v}$ を代入した

$$\hat{T}^*=S_n+\frac{\hat{u}}{\sqrt{n}}(S_n)^2+\frac{\hat{v}}{\sqrt{n}} \tag{6}$$

は，

$$P(\hat{T}^*\leq x)=\mathbf{\Phi}(x)+o(n^{-1/2})$$

を満たす正規化変換となるが，$\hat{T}^*$ は単調な変換ではない。そこで式(6)に，$\hat{u}^2(S_n)^3/(3n)$ を付け加えた正規化変換 $\tau(\cdot)$

$$T_n=\tau(S_n)=S_n+\frac{\hat{u}}{\sqrt{n}}(S_n)^2+\frac{\hat{u}^2}{3n}(S_n)^3+\frac{\hat{v}}{\sqrt{n}} \tag{7}$$

を考える。この変換は S_n について単調な変換であり，エッジワース展開の $n^{-1/2}$ の項を消す正規化変換となっている。ここで，u, v の一致推定量 $\hat{u}, \hat{v}$ が次の一致性を持つと仮定する。

$$\frac{\hat{u}}{\sqrt{n}}=\frac{u}{\sqrt{n}}+o_p(n^{-1/2}),\ \frac{\hat{v}}{\sqrt{n}}=\frac{v}{\sqrt{n}}+o_p(n^{-1/2}). \tag{8}$$

この仮定は多くの推定量で満足される。このとき，次の定理が成り立つ。

[定理3] S_n は式(5)を満たし，モーメントに対する条件

$$E\{|h_1(X_1)|^{3+\varepsilon}+|h_2(X_1,X_2)|^3\}<\infty\ \ (\varepsilon>0)$$

と，クラーメル条件

$$\limsup_{|t|\to\infty}|E[\exp\{ith_1(X_1)\}]|<1$$

が成り立つとする。さらに式(8)が満たされるとき，

$$P(T_n\le x)=\Phi(x)+o(n^{-1/2}).$$

(証明) 次の記号を準備する。

$$B_1=\sum_{i=1}^{n}h_1(X_i),\ B_2=\sum_{1\le i<j\le n}h_2(X_i,X_j)$$

このとき

$$\begin{aligned}n^{-1/2}(S_n)^2=&n^{-3/2}\delta^2+n^{-3/2}B_1^2+n^{-7/2}B_2^2+[o_p(n^{-1/2})]^2\\&+2n^{-3/2}\delta B_1+2n^{-5/2}\delta B_2+2n^{-5/2}B_1B_2\\&+2o_p(n^{-1/2})\{n^{-1}\delta+n^{-1}B_1+n^{-2}B_2\}\end{aligned}$$

となる。$o_p(n^{-1/2})$ の定義より，$n^{-3/2}\delta^2=o_p(n^{-1/2})$ である。さらに

$$n^{-3/2}B_1^2=n^{-1/2}E[h_1^2(X_1)]+n^{-1}\sum_{i=1}^{n}\{h_1^2(X_i)-E[h_1^2(X_1)]\}+2n^{-3/2}\sum_{1\le i<j\le n}h_1(X_i)h_1(X_j)$$

と表わせる。モーメントの評価(2)とマルコフの不等式より

$$P\left(|n^{-1}\sum_{i=1}^{n}\{h_1^2(X_i)-E[h_1^2(X_1)]\}|\ge n^{-1/2}(\log n)^{-1}\right)$$

$$\le n^{-(3+\varepsilon)/2}(\log n)^{3+\varepsilon}E|\sum_{i=1}^{n}\{h_1^2(X_i)-E[h_1^2(X_1)]\}|^{(3+\varepsilon)/2}=o(n^{-1/2})$$

だから，$n^{-1}\sum_{i=1}^{n}\{h_1^2(X_i)-E[h_1^2(X_1)]\}=o_p(n^{-1/2})$ となる。同様にモーメントの評価(3)とマルコフの不等式より

$$\begin{aligned}&n^{-7/2}B_2^2=o_p(n^{-1/2}),\ [o_p(n^{-1/2})]^2=o_p(n^{-1/2}),\ 2n^{-3/2}\delta B_1=o_p(n^{-1/2}),\\&2n^{-5/2}\delta B_2=o_p(n^{-1/2}),\ 2n^{-5/2}B_1B_2=o_p(n^{-1/2})\end{aligned}$$

となる。また

$$\begin{aligned}&P(|2o_p(n^{-1/2})n^{-1}B_1|\ge n^{-1/2}(\log n)^{-1})\\&\le P(|2o_p(n^{-1/2})|\ge n^{-1/2}(\log n)^{-1})+P(|n^{-1}B_1|\ge 1)\end{aligned}$$

となり，さらにマルコフの不等式より

$$P(|n^{-1}B_1|\ge 1)\le n^{-2}E|B_1|^2=o(n^{-1/2})$$

となる。他の項も同じように評価できて,

$$n^{-1/2}(S_n)^2=n^{-1/2}E[h_1^2(X_1)]+2n^{-3/2}\sum_{1\leq i<j\leq n}h_1(X_i)h_1(X_j)+o_p(n^{-1/2})$$

となる。同様に $n^{-1}(S_n)^3=o_p(n^{-1/2})$ である。これらの評価式と式(8)の仮定より,同じような評価を行って,

$$\frac{\hat{u}}{\sqrt{n}}(S_n)^2=\frac{u}{\sqrt{n}}(S_n)^2+o_p(n^{-1/2}),\ \frac{\hat{u}^2}{3n}(S_n)^3=o_p(n^{-1/2}),\ \frac{\hat{v}}{\sqrt{n}}=\frac{v}{\sqrt{n}}+o_p(n^{-1/2})$$

が示せる。ゆえに,

$$\begin{aligned}T_n=&n^{-1/2}(\delta+u+v)+n^{-1/2}\sum_{i=1}^{n}h_1(X_i)\\&+n^{-3/2}\sum_{1\leq i<j\leq n}[h_2(X_i,X_j)+2uh_1(X_i)h_1(X_j)]+o_p(n^{-1/2})\end{aligned}$$

となる。このとき,バイアスは

$$\delta+u+v=0.$$

また,T_n の3次のキュムラントは,

$$\begin{aligned}&n^{-1/2}\kappa_3\\&=n^{-1/2}\{E[h_1^3(X_1)]+3E[h_1(X_1)h_1(X_2)\{h_2(X_1,X_2)+2uh_1(X_1)h_1(X_2)\}]\}+o(n^{-1/2})\\&=n^{-1/2}\{E[h_1^3(X_1)]+3E[h_1(X_1)h_1(X_2)h_2(X_1,X_2)]+6u\}+o(n^{-1/2})\\&=o(n^{-1/2}).\end{aligned}$$

従って,補題2より

$$P(T_n\leq x)=\boldsymbol{\Phi}(x)+o(n^{-1/2}).$$

(証明終)

この結果をスチューデント化 U- 統計量に適用してみよう。ここで改めて,

$$f_1(x)=\frac{1}{2}[g_1^2(x)-\xi_1^2]+(r-1)E[g_1(X_2)g_2(x,X_2)],$$

$$h_1(x)=\frac{g_1(x)}{\xi_1},$$

$$h_2(x,y)=\frac{(r-1)g_2(x,y)}{\xi_1}-\frac{1}{\xi_1^3}[f_1(x)g_1(y)+f_1(y)g_1(x)],$$

$$\delta=-\frac{E[f_1(X_1)g_1(X_1)]}{\xi_1^3}$$

とおくと,Maesono (1997) より

$$\frac{U_n-\theta}{\hat{\sigma}_J}=n^{-1/2}\delta+n^{-1/2}\sum_{i=1}^{n}h_1(X_i)+n^{-3/2}\sum_{1\leq i<j\leq n}h_2(X_i,X_j)+o_p(n^{-1/2})$$

の漸近表現が得られる。さらに δ,u,v に対応する値をエッジワース展開に現れた e_1,e_2 を使って表わすと,

$$\delta=-\frac{E[f_1(X_1)g_1(X_1)]}{\xi_1^3}=-\frac{e_1}{2\xi_1^3}-\frac{e_2}{\xi_1^3},$$

$$u=-\frac{1}{6}E[h_1^3(X_1)]-\frac{1}{2}E[h_1(X_1)h_1(X_2)h_2(X_1,X_2)]=\frac{e_1}{3\xi_1^3}+\frac{e_2}{2\xi_1^3},$$

$$v=\frac{1}{6}E[h_1^3(X_1)]+\frac{1}{2}E[h_1(X_1)h_1(X_2)h_2(X_1,X_2)]-\delta=\frac{e_1}{6\xi_1^3}+\frac{e_2}{2\xi_1^3}$$

となる。従って推定しなければならない母数は ξ_1^2, e_1, e_2 となる。これらの母数の，統計量の具体的な形に依存しない，一般的な（ノンパラメトリックな）推定量としては次のジャックナイフ推定量が考えられる。ここで

$$\hat{g}_1(i)=U_n-U_n^{(i)},$$
$$\hat{g}_2(i,j)=-[nU_n-(n-1)(U_n^{(i)}+U_n^{(j)})+(n+2)U_n^{(i,j)}]$$

とおく。ただし $U_n^{(i,j)}$ は n 個の標本から X_i, X_j を除いた $n-2$個の標本に基づく対応する U- 統計量である。このときジャックナイフ推定量は

$$\hat{\xi}_1^2=\frac{n-1}{r^2}\sum_{i=1}^{n}[\hat{g}_1(i)]^2=\frac{n\hat{\sigma}_J^2}{r^2},$$

$$\hat{e}_1=\frac{(n-1)^3}{r^3 n}\sum_{i=1}^{n}[\hat{g}_1(i)]^3,$$

$$\hat{e}_2=\frac{(n-1)^2}{r^3 n}\sum_{i}\sum_{j\neq i}\hat{g}_1(i)\hat{g}_1(j)\hat{g}_2(i,j)$$

で与えられる。これらの推定量が条件(8)を満たすことは Maesono（1998）により示されている。

4．信頼区間への応用

U- 統計量の分布の近似に基づいて母数 θ の近似信頼区間を構成する。先ず正規近似に基づく信頼区間を考える。$z_{\alpha/2}$ を標準正規分布の上側 $\frac{\alpha}{2}$- 点とおくと

$$1-\alpha\approx P\left(-z_{\alpha/2}\leq\frac{U_n-\theta}{\sigma_n}\leq z_{\alpha/2}\right).$$

従って母数 σ_n の推定量 $\hat{\sigma}_J$ を使うと，母数 θ の信頼率1－ α の両側近似信頼区間は

$$U_n-z_{\alpha/2}\hat{\sigma}_J\leq\theta\leq U_n+z_{\alpha/2}\hat{\sigma}_J$$

となる。ここで $\varepsilon>0$に対して

$$\lim_{n\to\infty}P\left(|\sqrt{n}\,\hat{\sigma}_J-r\xi_1|\leq\varepsilon\right)=1$$

である。従って信頼区間の幅は，よく知られているように $n^{-1/2}$ のオーダーで狭くなる。

標準化 U- 統計量のエッジワース展開に基づく信頼区間は，コーニッシュ・フィッシャーの反転公式を使って求めることができる。z_α を標準正規分布の上側 α- 点とし，

$$s_a = z_a + \frac{\hat{\kappa}_3}{6\sqrt{n}}(z_a^2 - 1) + \frac{\hat{\kappa}_4}{24n}(z_a^3 - 3z_a) - \frac{\hat{\kappa}_3^2}{36n}(2z_a^3 - 5z_a)$$

とおく。ここで $\hat{\kappa}_3$, $\hat{\kappa}_4$ はそれぞれ

$$\kappa_3 = \frac{e_1}{\xi_1^3} + \frac{3e_2}{\xi_1^3},$$

$$\kappa_4 = \frac{e_3}{\xi_1^4} + \frac{12e_4}{\xi_1^4} + \frac{12e_5}{\xi_1^4} + \frac{4e_6}{\xi_1^4} - 3$$

の各項 $\xi_1^2, e_1, e_2, \ldots, e_6$ に推定量を代入したものである。$\hat{\xi}_1^2, \hat{e}_1, \hat{e}_2$ は前に提案したジャックナイフ推定量を使い，残りの推定量としては

$$\hat{e}_3 = \frac{(n-1)^4}{r^4 n} \sum_{i=1}^{n} [\hat{g}_1(i)]^4,$$

$$\hat{e}_4 = \frac{(n-1)^3}{r^4 n} \sum_i \sum_{j \neq i} [\hat{g}_1(i)]^2 \hat{g}_1(j) \hat{g}_2(i,j),$$

$$\hat{e}_5 = \frac{(n-1)^3}{r^4 n(n-2)} \sum_i \sum_{j \neq i} \sum_{k \neq i,j} \hat{g}_1(i) \hat{g}_1(j) \hat{g}_2(i,k) \hat{g}_2(j,k),$$

$$\hat{e}_6 = \frac{n(n-1)^2}{r^4(n-2)} \sum_i \sum_{j \neq i} \sum_{k \neq i,j} \hat{g}_1(i) \hat{g}_1(j) \hat{g}_1(k) \hat{g}_3(i,j,k)$$

を使う。ただし

$$\begin{aligned}\hat{g}_3(i,j,k) = {} & nU_n - (n-1)(U_n^{(i)} + U_n^{(j)} + U_n^{(k)}) \\ & + (n-2)(U_n^{(i,j)} + U_n^{(i,k)} + U_n^{(j,k)}) - (n-3)U_n^{(i,j,k)}\end{aligned}$$

で，$U_n^{(i,j,k)}$ は X_i, X_j, X_k を除いた $n-3$ 個の標本に基づく対応する U- 統計量である。このとき，母数 θ の信頼率 $1-\alpha$ の両側近似信頼区間は

$$U_n - s_{\alpha/2}\hat{\sigma}_J \leq \theta \leq U_n - s_{1-\alpha/2}\hat{\sigma}_J$$

で与えられる。

同様にして，スチューデント化 U- 統計量のエッジワース展開に基づく近似信頼区間が構成できる。ここで次の記号を準備する。

$$\hat{\kappa}_{31}^* = \hat{\xi}_1^{-3}(2\hat{e}_1 + 3\hat{e}_2), \quad \hat{\kappa}_{32}^* = \hat{\xi}_1^{-3}(\hat{e}_1 + 3\hat{e}_2), \quad \hat{\kappa}_{41}^* = -\hat{\xi}_1^{-6}(2\hat{e}_1 + 3\hat{e}_2)^2,$$

$$\hat{\kappa}_{42}^* = 6\hat{\xi}_1^{-4}(\hat{e}_3 - 6\hat{\xi}_1^4 + 12\hat{e}_4 + 6\hat{e}_5 + 4\hat{e}_6) - 2\hat{\xi}_1^{-6}(2\hat{e}_1 + 3\hat{e}_2)(2\hat{e}_1 + 9\hat{e}_2),$$

$$\hat{\kappa}_{43}^* = 3\hat{\xi}_1^{-6}(4\hat{e}_1^2 + 12\hat{e}_1\hat{e}_2 + 3\hat{e}_2^2) + 18\hat{\xi}_1^{-4}(\hat{\xi}_1^2\hat{\xi}_2^2 - \hat{e}_3 + 2\hat{\xi}_1^4 - 4\hat{e}_4 - 2\hat{e}_5).$$

ただし

$$\hat{\xi}_2^2 = \frac{n-1}{r^2 n} \sum_i \sum_{j \neq i} [\hat{g}_2(i,j)]^2$$

である。さらに

$$s_{\alpha}^{*}=z_{\alpha}-\frac{\hat{\kappa}_{31}^{*}z_{\alpha}^{2}+\hat{\kappa}_{32}^{*}}{6\sqrt{n}}+\frac{1}{72n}\left[\{4(\hat{\kappa}_{31}^{*})^{2}-2\hat{\kappa}_{31}^{*}\hat{\kappa}_{32}^{*}-\hat{\kappa}_{42}^{*}\}z_{\alpha}^{3}+\{4\hat{\kappa}_{31}^{*}\hat{\kappa}_{32}^{*}-(\hat{\kappa}_{32}^{*})^{2}-\hat{\kappa}_{43}^{*}\}z_{\alpha}\right]$$

とおくと，母数θの信頼率$1-\alpha$の両側近似信頼区間は

$$U_n-s_{\alpha/2}^{*}\hat{\sigma}_J\leq\theta\leq U_n-s_{1-\alpha/2}^{*}\hat{\sigma}_J$$

で与えられる。

最後に正規化変換に基づく信頼区間を構成しよう。Hall (1992)より，式(7)の変換$\tau(\cdot)$の逆変換は

$$\tau^{-1}(t_n)=\frac{\sqrt{n}}{\hat{u}}\left\{1+\frac{3\hat{u}}{\sqrt{n}}\left(t_n-\frac{\hat{v}}{\sqrt{n}}\right)\right\}^{1/3}-\frac{\sqrt{n}}{\hat{u}}$$

となる。即ち

$$\frac{U_n-\theta}{\hat{\sigma}_J}=\tau^{-1}(T_n).$$

この逆変換を使うと，正規化変換に基づく母数θの信頼率$1-\alpha$の両側近似信頼区間は

$$U_n-\tau^{-1}(z_{\alpha/2})\hat{\sigma}_J\leq\theta\leq U_n-\tau^{-1}(-z_{\alpha/2})\hat{\sigma}_J$$

で与えられる。

以上述べたU-統計量の正規近似に基づく近似信頼区間をシュミレーションで比較する。

[例1] 母平均μの信頼区間。

$E(X_1)=\mu$のとき，U-統計量において1次のカーネル$h(x)=x$を考えると

$$U_n=\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

はμの不偏推定量となる。このとき今までの結果を使って，信頼率95％の近似両側信頼区間をシュミレーションで比較した。母集団分布として正規，両側指数，一様，自由度2のχ^2-分布を考えた。繰り返し数はそれぞれ10000回で，表の数値は真の母数を含んだ割合（％）を表わす。従って数値が95に近い程良い手法である。表の各手法は

Nor　：正規近似

Stand：標準化U-統計量のエッジワース展開

Stud　：スチューデント化U-統計量のエッジワース展開

$n^{-1/2}$　：正規化変換

を表わす。

表1　正規分布

手法	$n=10$	$n=15$	$n=20$	$n=30$	$n=50$	$n=100$	$n=200$
Nor	92.22	93.04	93.28	94.51	94.89	94.81	94.79
Stand	92.18	92.92	93.30	94.48	94.86	94.78	94.75
Stud	94.93	94.76	94.63	95.23	95.32	95.13	94.85
$n^{-1/2}$	91.78	92.67	93.08	94.28	94.75	94.82	94.77

表2　両側指数分布

手法	$n=10$	$n=15$	$n=20$	$n=30$	$n=50$	$n=100$	$n=200$
Nor	92.82	93.91	94.22	94.47	94.54	94.95	94.86
Stand	93.40	94.56	94.76	95.08	94.95	95.31	95.00
Stud	94.18	94.28	94.03	93.78	94.02	94.42	94.54
$n^{-1/2}$	87.58	89.60	90.95	91.48	93.05	94.03	94.38

表3　一様分布

手法	$n=10$	$n=15$	$n=20$	$n=30$	$n=50$	$n=100$	$n=200$
Nor	91.62	93.02	92.90	93.67	94.16	94.52	94.65
Stand	90.64	91.99	92.28	93.26	93.93	94.31	94.54
Stud	96.12	96.47	95.89	95.76	95.37	95.28	95.07
$n^{-1/2}$	93.81	94.74	94.58	94.86	94.87	94.91	94.95

表4　自由度2のχ^2-分布

手法	$n=10$	$n=15$	$n=20$	$n=30$	$n=50$	$n=100$	$n=200$
Nor	86.67	89.71	90.43	91.77	93.00	93.83	94.55
Stand	85.55	88.63	89.32	91.00	92.34	93.34	94.33
Stud	91.90	93.52	94.04	94.36	94.96	94.88	95.19
$n^{-1/2}$	88.06	90.56	91.25	92.00	93.35	93.94	94.58

上記のシュミレーションの結果では，スチューデント化 U- 統計量のエッジワース展開に基づく近似信頼区間が良い近似を与え，正規化変換もかなり良いことが分かる。

［例2］ 母分散σ^2の信頼区間。

$Var(X_1)=\sigma^2$の信頼区間を求める。U- 統計量において$h(x,y)=(x-y)^2/2$とおくと

$$U_n=\frac{2}{n(n-1)}\sum_{1\le i<j\le n}\frac{(X_i-X_j)^2}{2}=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

は標本不偏分散である。このとき母平均のときと同じ方法で，4つの母集団分布についてのシミュレーションを行った結果，表5から表8が得られた。

表5　正規分布

手法	$n=10$	$n=15$	$n=20$	$n=30$	$n=50$	$n=100$	$n=200$
Nor	83.11	86.32	88.21	89.96	91.97	93.60	93.90
Stand	82.28	84.73	86.95	88.59	90.86	92.97	93.56
Stud	88.10	90.59	92.51	93.76	94.47	95.40	95.11
$n^{-1/2}$	85.53	88.07	89.79	91.32	92.39	93.85	94.27

表6　両側指数分布

手法	$n=10$	$n=15$	$n=20$	$n=30$	$n=50$	$n=100$	$n=200$
Nor	73.12	78.37	80.25	83.20	86.93	89.90	92.10
Stand	71.86	76.52	78.11	81.22	84.98	88.57	91.28
Stud	80.37	85.92	87.65	90.33	92.76	93.66	94.55
$n^{-1/2}$	76.57	81.22	82.53	84.53	87.22	89.39	91.69

表7　一様分布

手法	$n=10$	$n=15$	$n=20$	$n=30$	$n=50$	$n=100$	$n=200$
Nor	90.15	93.22	93.67	94.43	94.79	94.70	94.84
Stand	89.35	92.32	92.82	93.79	94.41	94.53	94.73
Stud	92.50	95.56	95.57	96.00	95.80	95.27	95.17
$n^{-1/2}$	92.01	95.18	95.49	95.72	95.60	95.23	95.11

表8　自由度2の χ^2-分布

手法	$n=10$	$n=15$	$n=20$	$n=30$	$n=50$	$n=100$	$n=200$
Nor	65.77	70.50	73.40	77.45	81.07	85.88	89.30
Stand	65.07	69.40	72.01	75.67	78.93	84.29	87.73
Stud	72.81	78.43	81.62	85.55	88.49	92.02	93.11
$n^{-1/2}$	68.09	73.54	75.61	78.70	80.77	85.53	88.31

この場合も平均の場合と同じく，スチューデント化 U- 統計量のエッジワース展開に基づく近似信頼区間が良い近似を与え，正規化変換もかなり良いことが分かる。

参 考 文 献

[1] Abramovitch, B.L. and Singh, K.(1985). Edgeworth corrected pivotal statistics and bootstrap. *Ann. Statist.*, **13**, 116-132.

[2] Callaert, H., Janssen, P. and Veraverbeke, N.(1980). An Edgeworth expansion for *U*-statistics, *Ann. Statist.*, **8**, 299-312.

[3] Hall, P.(1992). On the removal of skewness by transformation. *J.R. Statist. Soc.* B, **54**, 221-228.

[4] Hoeffding, W.(1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.*, **19**, 293-325.

[5] Johnson, N.J.(1978). Modified t tests and confidence intervals for asymmetrical populations. *J. Amer. Statist. Assoc.*, **73**, 536-544.

[6] Konishi, S.(1981). Normalizing transformations of some statistics in multivariate analysis. *Biometrika,* **68**, 647-651.

[7] Konishi, S.(1991). Normalizing transformations and bootstrap confidence intervals. *Ann. Statist.*, **19**, 2209-2225.

[8] Lai, T.L. and Wang, J.Q.(1993). Edgeworth expansion for symmetric statistics with applications to bootstrap methods. *Statistica Sinica,* **3**, 517-542.

[9] Maesono, Y. (1987). Edgeworth expansion for one-sample *U*-statistics, *Bulletin of Informatics and Cybernetics,* **22**, 189-197.

[10] Maesono, Y. (1997). Edgeworth expansions of a studentized *U*-statistic and a jackknife estimator of variance. *Journal of Statistical Planning and Inference,* **61**, 61-84.

[11] Maesono, Y.(1998). Asymptotic properties of jackknife skewness estimators and Edgeworth expansions. *Bulletin of Informatics and Cybernetics,* **30**, 51-68.

前園〔九州大学経済学部助教授〕

藤岡〔九州大学数理学研究科院生〕