Kobe University Repository : Kernel

PDF issue: 2026-11-07

回帰モデルにおける有意性検定について : 分布に依存しない検定

谷崎, 久志

(Citation)
国民経済雑誌,182(6):49-61

(Issue Date)
2000-12

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/00045111

(URL)
https://hdl.handle.net/20.500.14094/00045111

回帰モデルにおける有意性検定について：分布に依存しない検定

谷　崎　久　志

1　はじめに

通常，回帰式において，撹乱項は同一の分布に従うと仮定され，特に，回帰係数の有意性検定を行う場合は，撹乱項に正規分布の仮定を置くのが普通である。この仮定のもとで，回帰係数の推定量は自由度$n-k$のt分布に従うことが知られている（ただし，nは標本数，kは回帰係数の数とする）。標本数nが増えると，t分布は標準正規分布$N(0,1)$に近づく。また，中心極限定理から，撹乱項の分布が何であっても，その分散が有限であれば，標本数が大きいとき回帰係数の推定量は正規分布に近づくことが知られている。しかし，撹乱項が非正規分布で小標本の場合，回帰係数の推定量はt分布にならず，通常のt分布による仮説検定を適用することは出来ない。本稿では，撹乱項が非正規分布で小標本の場合でも適用可能な回帰係数の有意性検定を考える。撹乱項の分布関数の仮定に依存しない検定，すなわち，分布に基づかない検定（Nonparametric Test，または，Distribution-Free Test）方法を考察する。

回帰係数の推定量は，基本的には，2標本の相関として捉えることが出来る。よく知られた相関係数のノンパラメトリック検定として，Spearmanの順位相関係数やKendallの順位相関係数等の利用があげられる。これらの検定については，例えば，竹内（1963），Conover（1971），Hollander and Wolfe (1973)，Randles and Wolfe (1979)，Sprent (1989)，Gibbons and Chakraborti (1992)，Hogg and Craig (1995) 等を参照せよ。本稿では，これらの順位相関係数に基づくのではなく，Fisher (1966) の並べ替え検定として知られている検定に基づくものである。それは，あり得るすべての可能な組み合わせを考えて，それぞれの場合について相関係数を計算し，もともとのデータから計算された相関係数がどの位置にあるかを比べるものである。この相関係数の並べ替え検定をそのまま回帰係数に応用することを考える。

本稿の構成としては，次節（2節）では，準備として，まずXとYのデータをn組持っているとき相関がゼロかどうかの並び替えによるノンパラメトリック検定を紹介する。次に，この相関がゼロかどうかの検定は回帰係数がゼロかどうかの検定に，何の修正もなしに直ちに応用出来ることを示す。3節の前半では，もとのデータが非正規分布に従うときに，相関

係数の検出力をノンパラメトリック検定と通常の t 検定で比較する。さらに，3節の後半では，$k=2,3$ について，回帰係数の有意性検定で有意水準を $\alpha=0.10,0.05,0.01$ としたとき，正しくサイズが推定されているかどうかを調べる。

2　回帰係数の検定方法

2.1　相関係数の検定について

大きさ n の標本 (X_1, Y_1)，(X_2, Y_2)，……，(X_n, Y_n) について，X と Y との間に相関があるかどうかの検定（すなわち，相関係数 ρ がゼロかどうかの検定）を行うことを考える。相関係数 ρ は，

$$\rho=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{V}(X)\mathrm{V}(Y)}},$$

と定義される。ただし，$\mathrm{Cov}(X,Y)$，$\mathrm{V}(X)$，$\mathrm{V}(Y)$ はそれぞれ X と Y の共分散，X の分散，Y の分散を表す。このとき，標本相関係数 $\hat{\rho}$ は，

$$\hat{\rho}=\frac{S_{XY}}{S_X S_Y},$$

と書き表される。ただし，S_{XY}，S_X，S_Y はそれぞれ X と Y の標本共分散，X の標本分散，Y の標本分散を表し，

$$S_{XY}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}),\quad S_X^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2,\quad S_Y^2=\frac{1}{n}\sum_{n=1}^{n}(Y_i-\bar{Y})^2,$$

と表される。$\bar{X}$，$\bar{Y}$ は X，Y の標本平均を表すものとする。

X と Y が独立であれば，$\rho=0$ となり，X と Y の結合分布 $f(x,y)$ は X の周辺分布 $f(x)$ と Y の周辺分布 $f(y)$ との積で表される。すなわち，

$$f(x,y)=f(x)f(y),$$

となる。この式の意味するところは，すべての i，j について，X_i と Y_j を適当に組み合わせて，n 組を作ればよいということである。したがって，X_1 は Y_j $(j=1,2,\cdots\cdots,n)$ の n 通りの組み合わせがあり，X_2 は $n-1$ 通りの組み合わせ，……，X_n は1通りとなる。したがって，n 組の X と Y の組み合わせは合計で $n!$ 通りとなる。よって，$n!$ 通りの相関係数を計算することが出来る。もともとの組み合わせで計算された相関係数がどの位置にあるかで，相関があるかないかの検定を行うことが可能となる。このような X と Y の分布に依存しない検定をノンパラメトリック検定（Nonparametric Test，または，Distribution-Free Test）と呼ぶ。この種のノンパラメトリック検定は並べ替え検定（Permutation Test）と呼ばれ，有名な検定としては，Fisher（1966）の2標本検定が有名である。例えば，竹内（1963），竹内・大橋（1981），Stuart and Ord（1991）を参照せよ。

Xの順序はもとの標本のまま固定して，Yを並べ替える。あり得るXとYの組み合わせ（$n!$通りの組み合わせ）に基づいて，XとYとの間に相関があるかないかを検定する。

単純化のために，$n=4$のケースを例として考える。XとYの組み合わせは表1に示されている。この場合，XとYの組み合わせは24（なぜなら，$4!=24$）通りあるので，それぞれについて相関係数を計算することが出来る。24通りの相関係数を，$\hat{\rho}^{(i)}$ $(i=1,2,\cdots\cdots,24)$と表すことにする。また，表1の1番目の組み合わせを，もともとのデータからのXとYの組み合わせとする。このとき，24通りの相関係数から，$\hat{\rho}^{(1)}$より小さい確率，$\hat{\rho}^{(1)}$に等しい確率，$\hat{\rho}^{(1)}$より大きい確率をそれぞれ計算することが出来る。これらの確率で相関係数がゼロかどうかの検定を行うことが出来る。

一般に，n組のデータ(X_i, Y_i), $i=1,2,\cdots\cdots,n$，を考えると，相関係数は$n!$通り計算することが出来る。これらの$n!$個の相関係数を$\hat{\rho}^{(i)}$ $(i=1,2,\cdots\cdots,n!)$とする。$\hat{\rho}^{(1)}$を表1のようにもともとのデータから得られた相関係数とすると，相関係数ρの推定量$\hat{\rho}$の分布は，

$$\mathrm{Prob}(\hat{\rho}<\hat{\rho}^{(1)})=\frac{\hat{\rho}^{(i)}(i=1,2,\cdots,n!)\text{の中で，}\hat{\rho}^{(1)}\text{より小さいものの個数}}{\text{あり得るすべての組み合わせの総数(すなわち，}n!\text{)}},$$

$$\mathrm{Prob}(\hat{\rho}=\hat{\rho}^{(1)})=\frac{\hat{\rho}^{(i)}(i=1,2,\cdots,n!)\text{の中で，}\hat{\rho}^{(1)}\text{に等しいものの個数}}{\text{あり得るすべての組み合わせの総数(すなわち，}n!\text{)}},$$

$$\mathrm{Prob}(\hat{\rho}>\hat{\rho}^{(1)})=\frac{\hat{\rho}^{(i)}(i=1,2,\cdots,n!)\text{の中で，}\hat{\rho}^{(1)}\text{より大きいものの個数}}{\text{あり得るすべての組み合わせの総数(すなわち，}n!\text{)}},$$

として与えられる。

また，S_{XY}は，

表1：XとYの組み合わせ：$n=4$の場合を例にとって

	1	2	3	4	5	6	7	8	9	10	11	12
X_1	Y_1	Y_1	Y_1	Y_1	Y_1	Y_1	Y_2	Y_2	Y_2	Y_2	Y_2	Y_2
X_2	Y_2	Y_2	Y_3	Y_3	Y_4	Y_4	Y_1	Y_1	Y_3	Y_3	Y_4	Y_4
X_3	Y_3	Y_4	Y_2	Y_4	Y_2	Y_3	Y_3	Y_4	Y_1	Y_4	Y_1	Y_3
X_4	Y_4	Y_3	Y_4	Y_2	Y_3	Y_2	Y_4	Y_3	Y_4	Y_1	Y_3	Y_1
相関係数	$\hat{\rho}^{(1)}$	$\hat{\rho}^{(2)}$	$\hat{\rho}^{(3)}$	$\hat{\rho}^{(4)}$	$\hat{\rho}^{(5)}$	$\hat{\rho}^{(6)}$	$\hat{\rho}^{(7)}$	$\hat{\rho}^{(8)}$	$\hat{\rho}^{(9)}$	$\hat{\rho}^{(10)}$	$\hat{\rho}^{(11)}$	$\hat{\rho}^{(12)}$
	13	14	15	16	17	18	19	20	21	22	23	24
X_1	Y_3	Y_3	Y_3	Y_3	Y_3	Y_3	Y_4	Y_4	Y_4	Y_4	Y_4	Y_4
X_2	Y_1	Y_1	Y_2	Y_2	Y_4	Y_4	Y_1	Y_1	Y_2	Y_2	Y_3	Y_3
X_3	Y_2	Y_4	Y_1	Y_4	Y_1	Y_2	Y_2	Y_3	Y_1	Y_3	Y_1	Y_2
X_4	Y_4	Y_2	Y_4	Y_1	Y_2	Y_1	Y_3	Y_2	Y_3	Y_1	Y_2	Y_1
相関係数	$\hat{\rho}^{(13)}$	$\hat{\rho}^{(14)}$	$\hat{\rho}^{(15)}$	$\hat{\rho}^{(16)}$	$\hat{\rho}^{(17)}$	$\hat{\rho}^{(18)}$	$\hat{\rho}^{(19)}$	$\hat{\rho}^{(20)}$	$\hat{\rho}^{(21)}$	$\hat{\rho}^{(22)}$	$\hat{\rho}^{(23)}$	$\hat{\rho}^{(24)}$

$$S_{XY}=\frac{1}{n}\sum_{i=1}^{n}X_iY_i-\overline{XY},$$

と書き直される。標本平均 $\bar{X}$, $\bar{Y}$ は X, Y の順番をどのように入れ替えても同じ値をとる。同様に，S_X, S_Y も X, Y の順番にかかわらず同じ値となる。よって，$\hat{\rho}$ は $\sum_{i=1}^{n}X_iY_i$ の大きさに依存して変化する。すなわち，組み合わせに基づいた分布について，$\hat{\rho}$ の分布は $\sum_{i=1}^{n}X_iY_i$ の分布に一対一に対応することになる。$\hat{\rho}$ の確率は，$\sum_{i=1}^{n}X_iY_i$ の確率として捉えられる。実際の計算では，$\hat{\rho}$ を $n!$ 通り計算する必要はなく，$\sum_{i=1}^{n}X_iY_i$ を Y_i $(i=1,2,\cdots\cdots,n)$ の順番を入れ替えて $n!$ 通り計算すればよい。このようにして計算すると，計算負担をある程度は軽減することが出来る。

特に，(X_i, Y_i), $i=1,2,\cdots\cdots,n$, が正規分布，すなわち，

$$\begin{pmatrix}X_i\\Y_i\end{pmatrix}\sim N\left(\begin{pmatrix}\mu_X\\\mu_Y\end{pmatrix},\begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y\\\rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}\right),$$

のとき，標本相関係数 $\hat{\rho}$ の分布は，帰無仮説 $H_0:\rho=0$ が正しいもとで，

$$\frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}}\sim t(n-2),$$

となることが知られている。ただし，より一般的に，$H_0:\rho=\rho_0$ の検定は t 分布にはならず近似によって検定を行う。例えば，竹内 (1963, 1989), Lehmann (1986), Stuart and Ord (1991, 1994), Hogg and Craig (1995) を参照せよ。しかし，一般には，(X_i, Y_i), $i=1,2,\cdots\cdots,n$, が正規分布であるとは限らない。通常は，得られた標本の分布は未知である。よって，並び替え検定に基づくノンパラメトリック検定が，正規分布を除く他のどの場合にでも適用でき，しかも，より正しい結果を与えることが予想される。このことは，3.1節のモンテ・カルロ実験で確認される。

2.2　回帰係数の検定への応用

前節の相関係数のノンパラメトリック検定と全く同じ様に，回帰係数の有意性検定に関するノンパラメトリック検定を考えることが出来る。

回帰方程式

$$Y_i=X_i\beta+u_i,\quad i=1,2,\cdots,n,$$

について，β の推定量 $\hat{\beta}$ は，

$$\hat{\beta}=(X'X)^{-1}X'Y=\sum_{i=1}^{n}(X'X)^{-1}X_i'Y_i,\qquad(1)$$

と表される。ただし，

$$Y=\begin{pmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{pmatrix},\quad X=\begin{pmatrix}X_1\\X_2\\\vdots\\X_n\end{pmatrix},\quad \beta=\begin{pmatrix}\beta_1\\\beta_2\\\vdots\\\beta_k\end{pmatrix},\quad \hat{\beta}=\begin{pmatrix}\hat{\beta}_1\\\hat{\beta}_2\\\vdots\\\hat{\beta}_k\end{pmatrix},$$

で，Y_i は i 番目の Y でスカラーとし，X_i は i 番目の X で $1\times k$ のベクトルとする。

(1)式の構造から考えると，2.1節の X_i を $(X'X)^{-1}X_i'$ に置き換えることによって，前節で行った議論と全く同じ議論がここでも成り立つことが分かる。異なる点は，2.1節の X_i はスカラーであったのに対して，本節の $(X'X)^{-1}X_i'$ は $k\times 1$ のベクトルであるという点だけである。すなわち，相関係数の場合のノンパラメトリック検定と同様に，Y_i の順番を入れ替えると，$n!$ 通りの回帰係数を得ることが出来る。それらの回帰係数を $\hat{\beta}_j^{(i)}$ $(i=1,2,\cdots\cdots,n!)$ として表すことにする。また，$\hat{\beta}_j^{(i)}$ を $\hat{\beta}^{(i)}$ の j 番目の要素とする。すなわち，$\hat{\beta}^{(i)}=(\hat{\beta}_1^{(i)},\hat{\beta}_2^{(i)},\cdots\cdots,\hat{\beta}_k^{(i)})$ とする。

$\hat{\beta}_j^{(1)}$ をもともとのデータから得られた j 番目の回帰係数の推定量とする。このとき，回帰係数 β の j 番目の要素の推定量 $\hat{\beta}_j$ の分布は，$\beta_j=0$ $(j=1,2,\cdots\cdots,k)$ の仮説が正しいもとで，

$$\mathrm{Prob}(\hat{\beta}_j<\hat{\beta}_j^{(1)})=\frac{\hat{\beta}_j^{(i)}(i=1,2,\cdots,n!)\text{の中で，}\hat{\beta}_j^{(1)}\text{より小さいものの個数}}{\text{あり得るすべての組み合わせの総数(すなわち，}n!)},$$

$$\mathrm{Prob}(\hat{\beta}_j=\hat{\beta}_j^{(1)})=\frac{\hat{\beta}_j^{(i)}(i=1,2,\cdots,n!)\text{の中で，}\hat{\beta}_j^{(1)}\text{に等しいものの個数}}{\text{あり得るすべての組み合わせの総数(すなわち，}n!)},$$

$$\mathrm{Prob}(\hat{\beta}_j<\hat{\beta}_j^{(1)})=\frac{\hat{\beta}_j^{(i)}(i=1,2,\cdots,n!)\text{の中で，}\hat{\beta}_j^{(1)}\text{より大きいものの個数}}{\text{あり得るすべての組み合わせの総数(すなわち，}n!)},$$

によって与えられる。このように，すべての j について，同様の計算を行い，それぞれの確率を計算することが可能となる。それぞれの j について，$\hat{\beta}_j^{(1)}$ が，$n!$ 個の回帰係数の中で，どの位置にあるかを調べることによって，$H_0:\beta_j=0$ の検定を行うことが出来る。Prob $(\hat{\beta}_j<\hat{\beta}_j^{(1)})$，または，Prob $(\hat{\beta}_j>\hat{\beta}_j^{(1)})$ の確率が小さければ，$\hat{\beta}_j^{(1)}$ は分布の十分端に位置すると考えられ，帰無仮説 $H_0:\beta_j=0$ を棄却することが出来る。

より一般的に，帰無仮説 $H_0:\beta=\beta^*$ の検定については，$\hat{\beta}-\beta$ は，

$$\begin{aligned}\hat{\beta}-\beta&=(X'X)^{-1}X'Y-\beta\\&=(X'X)^{-1}X'(Y-X\beta)\\&=\sum_{i=1}^{n}(X'X)^{-1}X_i'(Y_i-X_i\beta),\end{aligned}$$

として変形されるので，帰無仮説 $H_0:\beta=\beta^*$ のもとで，$(X'X)^{-1}X_i'$ と $Y_i-X_i\beta^*$ の相関がゼロかどうかの並び替えによるノンパラメトリック検定を考えればよいことになる。

また，通常のパラメトリックの有意性検定について，撹乱項 u_i $(i=1,2,\cdots\cdots,n)$ は互いに独立に平均ゼロ，分散 σ^2 の正規分布に従うと仮定される。そのため，第 i 要素の最小自乗推定量 $\hat{\beta}_i$ は，帰無仮説 H_0：$\beta_i=\beta_i^*$ が正しいもとで，

$$\frac{\hat{\beta}_i-\beta_i^*}{S\sqrt{a_{ii}}}\sim t(n-k),$$

となる。ただし，a_{ii} は $(X'X)^{-1}$ の第 i 行第 i 列の対角要素を表し，$\beta^*=(\beta_1^*,\beta_2^*,\cdots\cdots,\beta_k^*)'$，$S^2=\frac{1}{n-k}(Y-X\beta)'(Y-X\beta)$ とする。このように，u_i を正規分布と仮定したときにのみ，t 分布を用いることが出来る。しかし，u_i が正規分布でなければ，小標本では，通常の t 検定は誤りとなる。大標本においては，中心極限定理によって，u_i の分散が有限であれば，漸近的に $\sqrt{n}(\hat{\beta}_i-\beta_i)$ は正規分布に従う。このように，大標本では成り立つ検定が，小標本では成り立たない。本稿では，次節で，まず，非正規分布のもとで，相関係数の検出力をモンテ・カルロ実験を通して比較する。次に，回帰分析において，通常用いられる t 検定が，非正規分布のときに，どの程度有効かを調べる。

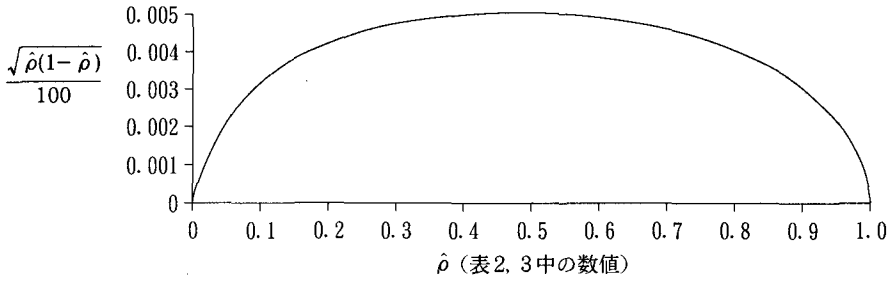
3　モンテ・カルロ実験

3.1　相関係数の検定

表 2 には，10000回のモンテ・カルロ実験の中で，両側確率が α より小さくなるものの回数を比率で表したものである。すなわち，もともとのデータから得られた相関係数が $\alpha/2$ より小さい確率になるか，$1-\alpha/2$ よりも大きい確率になる回数を10000で割ったものを表 2 に示している。さらに，言い換えると，$\hat{\rho}^{(1)}$ をもともとのデータから得られた相関係数として，Prob $(\hat{\rho}<\hat{\rho}^{(1)})$ と Prob $(\hat{\rho}>\hat{\rho}^{(1)})$ の確率を計算し，これを10000回繰り返して，その中で Prob $(\hat{\rho}<\hat{\rho}^{(1)})\leq\alpha/2$ または Prob $(\hat{\rho}>\hat{\rho}^{(1)})\leq\alpha/2$ となる比率を計算したものが表 2 の数値である。$\alpha=0.10,0.05,0.01$ の 3 通りを調べた。N，X，U，L，C は X，Y の分布を意味し，それぞれ，標準正規分布 $N(0,1)$，カイ自乗分布 $\chi^2(1)-1$，一様分布 $U(-1,1)$，ロジスティック分布 $\frac{e^{-x}}{(1+e^{-x})^2}$，コーシー分布 $\frac{1}{\pi(1+x^2)}$ を表す。表中の数値（$\hat{p}$ と表記する）の標準偏差は，$\frac{\sqrt{\hat{p}(1-\hat{p})}}{100}$ として求められる。表中の数値 $\hat{p}$ に対応した標準偏差 $\frac{\sqrt{\hat{p}(1-\hat{p})}}{100}$ は，図 1 に示される。標準偏差は $\hat{p}=0.5$ のとき最大で，そのときの標準誤差の値は0.005である。

X_i，Y_i の求め方は，次のようにした。u_i と v_i は互いに独立で同一の分布に従うものとする。すなわち，それぞれ，標準正規分布 $N(0,1)$，カイ自乗分布 $\chi^2(1)-1$，一様分布 $U(-1,1)$，ロジスティック分布 $\frac{e^{-x}}{(1+e^{-x})^2}$，コーシー分布 $\frac{1}{\pi(1+x^2)}$ から u_i と v_i を生成する。このようにすると，

図1：$\hat{\rho}$ に対応する標準偏差



$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ 0 & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix}$$

のとき，X と Y の相関係数は ρ となる。ただし，コーシー分布の場合は，平均も分散も存在しないので，相関も存在しないことになる。しかし，その場合でも，(X_i, Y_i) は計算することが出来る。

真の相関係数を $\rho=0.0,0.3,0.6,0.9$ として，帰無仮説 H_0：$\rho=0$ の検定を行う。このとき，有意水準は $\alpha=0.10,0.05,0.01$ の両側検定で，10000回の実験の中からの棄却率が表2に示されている。さらに，$n=6,8,10$ について，ノンパラメトリック検定と t 検定の結果を共に示す。

予想される結果は，N については，t 検定が最もよい結果となるはずである。その他の場合は，t 分布には従わず，t 検定はバイアス検定となるので，ノンパラメトリック検定の方が優れているはずである。

$\rho=0$ の場合は，サイズを表し，α に等しい検定がよい検定となる。t 検定について，X と C は，特に $\alpha=0.05,0.01$ の場合，n の値にかかわらず，過剰に棄却されている。例えば，$n=6$ について，$\alpha=0.05$ のとき，X は0.0646，C は0.0748，$\alpha=0.01$ のとき，X は0.0225，C は0.0269 となっている。U についても，$\alpha=0.05,0.01$ の場合，やや過剰棄却される傾向がある。L の場合については，ほぼ α と同じ値となっている。このように，$\rho=0$ のとき，もとの分布に依存して，t 検定では過剰に棄却される場合があるということが言える。これに対して，分布に依存しないノンパラメトリック検定については，$\rho=0$ の場合，もとの分布にかかわりなくすべてほぼ α に等しく正しいサイズが得られている。よって，帰無仮説 H_0：$\rho=0$ のサイズ比較によると，ノンパラメトリック検定の方が，分布に依存しないという意味で，t 検定よりロバストな検定となっている。

次に，$\rho=0.3,0.6,0.9$ の場合を調べる。この場合，それぞれの数値は検出力を表す。X と C については，サイズが過剰推定されているので，検出力を比較することはあまり意味がない。N は t 検定が最もよい検定となるはずであるが，ノンパラメトリック検定も t 検定とほと

表2：帰無仮説 H_0：$\rho=0$ の場合の検出力

n	ρ	α	ノンパラメトリック検定					t 検定				
			N	X	U	L	C	N	X	U	L	C
6	.0	.10	.1008	.0986	.1040	.1035	.1030	.0981	.1019	.1019	.0987	.1180
		.05	.0499	.0509	.0512	.0518	.0513	.0474	.0646	.0507	.0483	.0748
		.01	.0112	.0096	.0113	.0101	.0098	.0099	.0225	.0115	.0093	.0269
	.3	.10	.1491	.1830	.1479	.1621	.2605	.1496	.2284	.1448	.1585	.3055
		.05	.0821	.1108	.0798	.0926	.1639	.0795	.1638	.0794	.0914	.2375
		.01	.0189	.0323	.0191	.0205	.0475	.0203	.0692	.0188	.0225	.1389
	.6	.10	.2876	.3119	.2605	.3014	.3834	.2932	.3841	.2646	.3081	.4384
		.05	.1776	.2145	.1528	.1836	.2696	.1787	.2917	.1568	.1939	.3633
		.01	.0460	.0755	.0386	.0501	.0983	.0481	.1579	.0389	.0564	.2405
	.9	.10	.4674	.4469	.4365	.4682	.4802	.4792	.5144	.4475	.4810	.5325
		.05	.3117	.3206	.2786	.3240	.3636	.3300	.4208	.2887	.3410	.4549
		.01	.0894	.1269	.0745	.1009	.1572	.1060	.2501	.0843	.1240	.3096
8	.0	.10	.1018	.0994	.1040	.1014	.1007	.1020	.0986	.1057	.1025	.1183
		.05	.0523	.0489	.0523	.0549	.0494	.0515	.0609	.0545	.0549	.0791
		.01	.0103	.0111	.0110	.0100	.0093	.0100	.0225	.0126	.0113	.0345
	.3	.10	.1869	.2025	.1748	.1971	.3120	.1866	.2514	.1779	.1978	.3471
		.05	.1066	.1236	.0940	.1096	.2107	.1083	.1821	.0998	.1147	.2780
		.01	.0257	.0385	.0211	.0251	.0749	.0248	.0891	.0250	.0254	.1814
	.6	.10	.3932	.3997	.3605	.4002	.4527	.3996	.4592	.3657	.4049	.4963
		.05	.2602	.2833	.2212	.2611	.3455	.2693	.3695	.2359	.2742	.4205
		.01	.0815	.1148	.0627	.0850	.1615	.0865	.2149	.0714	.0959	.2920
	.9	.10	.6197	.5445	.6256	.6209	.5647	.6289	.6023	.6309	.6313	.5979
		.05	.4653	.4163	.4511	.4850	.4684	.4806	.5125	.4664	.4980	.5315
		.01	.1966	.2034	.1562	.2119	.2646	.2199	.3422	.1795	.2395	.4096
10	.0	.10	.1046	.0987	.0999	.0984	.0992	.1033	.0943	.1006	.0950	.1058
		.05	.0524	.0486	.0520	.0489	.0477	.0510	.0570	.0526	.0479	.0733
		.01	.0106	.0083	.0104	.0102	.0090	.0097	.0220	.0114	.0098	.0348
	.3	.10	.2162	.2243	.2046	.2228	.3468	.2151	.2765	.2065	.2231	.3764
		.05	.1277	.1429	.1179	.1348	.2513	.1284	.2046	.1218	.1348	.3133
		.01	.0338	.0484	.0285	.0351	.1022	.0355	.1037	.0306	.0367	.2149
	.6	.10	.4766	.4501	.4640	.4920	.5089	.4775	.5005	.4962	.4954	.5378
		.05	.3387	.3335	.3118	.3579	.4081	.3416	.4149	.3198	.3599	.4670
		.01	.1325	.1522	.1020	.1424	.2281	.1375	.2604	.1106	.1504	.3610
	.9	.10	.7336	.6128	.7539	.7276	.6075	.7358	.6663	.7558	.7348	.6343
		.05	.6008	.4958	.6038	.6032	.5258	.6116	.5806	.6106	.6129	.5734
		.01	.3116	.2812	.2719	.3286	.3318	.3295	.4128	.2925	.3482	.4553

んど変わらず，検出力は同程度のよさであると言える。

このように，本稿で紹介した並べ替えによるノンパラメトリック検定は，分布の仮定を必要としないにもかかわらず正しい検定結果を得ることが出来るので，有用な検定であると言えるだろう。

3.2　回帰係数の検定

表3では，3.2節で行った分析を，回帰分析に応用する。$X_i=(X_{1,i},X_{2,i},\cdots\cdots,X_{k,i})$ とすると，$X_{1,i}=1$，$j=2,3$ について $X_{j,i}\sim N(0,1)$ として，説明変数 X_i を設定する。ただし，$X_{2,i}$ と $X_{3,i}$ は互いに独立として生成されている。さらに，攪乱項 u_i は，標準正規分布 $N(0,1)$，カイ自乗分布 $\chi^2(1)-1$，一様分布 $U(-1,1)$，ロジスティック分布 $\frac{e^{-x}}{(1+e^{-x})^2}$，コーシー分布 $\frac{1}{\pi(1+x^2)}$ を仮定して，それぞれについてシミュレーションを行う。標本数は $n=6$，8,10，パラメータ数は $k=2,3$ のときを考え，本稿で考察したノンパラメトリック検定と t 検定とをサイズで比較する。表中の数値は，10000回の実験で棄却した回数の比率を表す。そのため，表中の数値は，理論的には，帰無仮説 $H_0:\beta_i=0\ (i=1,2,3)$ が正しいもとで棄却する確率（すなわち，有意水準 α）に等しくなるべきである。

前節と同様に，u_i が独立に平均ゼロの正規分布に従うとき，β_i の推定量 $\hat{\beta}_i$ は自由度 $n-k$ の t 分布に従うことになるので，表3の N の場合は t 検定がノンパラメトリック検定よりよい結果を示すはず（すなわち，α に近い数字になるはず）である。

数値の右肩の * は，有意水準5％の両側検定で，対応する α とは異なることを意味し，** は，有意水準1％の両側検定で，対応する α とは異なることを表す。すなわち，右肩に * や **の付いていない数値が，正しいサイズが得られたと言える。*，**に関する検定については，中心極限定理により，G が大きいとき，$\frac{\hat{p}-p}{p(1-p)/G}\sim N(0,1)$ と近似することが出来る（p は母数，$\hat{p}$ はその推定値とする）。帰無仮説 $H_0:p=\alpha$ を考え，その仮説が正しいもとで，検定統計量 $\frac{\hat{p}-\alpha}{\sqrt{\alpha(1-\alpha)/G}}$ を計算し，有意水準5％と1％で標準正規分布表を用いて，$\hat{p}$ が α と異なるかどうかの検定を行った。その結果が，＊や**で表現されている。ただし，G はシミュレーションの繰り返し回数であり，この場合は $G=10000$ である。そのため，この場合は十分正規近似が可能であると考えてよい。

結果として，X や U の定数項 β_1 の t 検定は，$k=2,3$ の両方とも過大に棄却されている。逆に，C については，定数項 β_1 の t 検定は，$k=2,3$ の両方とも棄却率が過小推定されている。ノンパラメトリック検定については，有意水準の値と統計的に異なるものも少しはあるが，全体的に見れば，ほとんどが α に近い値となっていて，t 検定よりも優れていると言える。

表3：回帰係数の有意性検定：サイズによる比較

	n	α	$k=2$				$k=3$					
			ノンパラメトリック検定		t 検定		ノンパラメトリック検定			t 検定		
			β_1	β_2	β_1	β_2	β_1	β_2	β_3	β_1	β_2	β_3
N	6	.10	.1008	.1008	.1055	.0981	.1054	.1056	.1021	.1030	.1038	.0987
		.05	.0500	.0499	.0534	.0474	.0526	.0559	.0503	.0509	.0517	.0493
		.01	.0113	.0112	.0104	.0099	.0105	.0123	.0089	.0098	.0103	.0101
	8	.10	.0999	.0999	.1021	.1002	.0963	.0964	.0969	.0912**	.0976	.0976
		.05	.0480	.0480	.0524	.0495	.0475	.0489	.0481	.0442**	.0465	.0487
		.01	.0104	.0104	.0120*	.0100	.0099	.0102	.0087	.0075*	.0094	.0101
	10	.10	.0997	.0996	.1033	.0978	.1049	.0980	.1027	.0993	.0964	.1009
		.05	.0492	.0491	.0522	.0492	.0543	.0514	.0514	.0493	.0518	.0509
		.01	.0094	.0094	.0096	.0096	.0116	.0109	.0111	.0092	.0105	.0121*
X	6	.10	.1013	.1013	.2132**	.0991	.1033	.1033	.1004	.1912**	.1032	.1003
		.05	.0517	.0516	.1569**	.0479	.0542	.0536	.0529	.1287**	.0501	.0484
		.01	.0109	.0109	.0806**	.0099	.0115	.0106	.0098	.0542**	.0087	.0096
	8	.10	.1012	.1012	.1925**	.1013	.0963	.1038	.0986	.1866**	.1019	.0957
		.05	.0511	.0511	.1402**	.0477	.0471	.0530	.0463	.1327**	.0485	.0460
		.01	.0103	.0103	.0785**	.0097	.0094	.0101	.0108	.0644**	.0089	.0093
	10	.10	.1029	.1029	.1750**	.1046	.0981	.0999	.0978	.1865**	.0970	.0956
		.05	.0510	.0510	.1263**	.0526	.0526	.0495	.0477	.1346**	.0513	.0475
		.01	.0090	.0090	.0652**	.0086	.0106	.0097	.0114	.0666**	.0106	.0103
U	6	.10	.1082**	.1082**	.1125**	.1042	.1061*	.1056	.1038	.1046	.1014	.1006
		.05	.0565**	.0564**	.0621**	.0541	.0571**	.0553*	.0530	.0556*	.0517	.0508
		.01	.0112	.0112	.0155**	.0106	.0141**	.0121*	.0123*	.0112	.0093	.0099
	8	.10	.0990	.0990	.0998	.1019	.1064*	.1037	.1035	.1092**	.1077*	.1057
		.05	.0493	.0493	.0534	.0514	.0542	.0548*	.0545*	.0579**	.0519	.0522
		.01	.0100	.0100	.0156**	.0102	.0109	.0126**	.0106	.0135**	.0112	.0110
	10	.10	.0980	.0980	.1106**	.0968	.0979	.1035	.1027	.1004	.1029	.1032
		.05	.0495	.0494	.0596**	.0500	.0509	.0511	.0517	.0538	.0527	.0509
		.01	.0117	.0117	.0136**	.0116	.0117	.0098	.0104	.0109	.0116	.0108
L	6	.10	.1080**	.1080**	.0997	.1054	.1053	.1072*	.1036	.0945	.1011	.1001
		.05	.0562**	.0562**	.0491	.0524	.0581**	.0549*	.0535	.0462	.0513	.0525
		.01	.0118	.0118	.0080	.0095	.0143**	.0121*	.0119	.0080*	.0098	.0113
	8	.10	.1003	.1003	.0953	.1019	.1086**	.1067*	.1064*	.1035	.1052	.1049
		.05	.0499	.0499	.0452*	.0531	.0565**	.0511	.0549*	.0514	.0524	.0558**
		.01	.0098	.0098	.0094	.0102	.0101	.0120*	.0104	.0101	.0120*	.0112
	10	.10	.0967	.0967	.1050	.0965	.0981	.1012	.1002	.0996	.1000	.1003
		.05	.0487	.0486	.0527	.0484	.0504	.0519	.0489	.0451*	.0521	.0483
		.01	.0114	.0114	.0089	.0122*	.0108	.0091	.0103	.0076*	.0109	.0111

表3：回帰係数の有意性検定：サイズによる比較－（続き）

	n	α	$k=2$				$k=3$					
			ノンパラメトリック検定		t 検定		ノンパラメトリック検定			t 検定		
			β_1	β_2	β_1	β_2	β_1	β_2	β_3	β_1	β_2	β_3
C	6	.10	.1114**	.1114**	.0709**	.1035	.1045	.1059*	.1008	.0777**	.0987	.1003
		.05	.0566**	.0566**	.0303**	.0505	.0547*	.0561**	.0536	.0362**	.0500	.0523
		.01	.0116	.0116	.0042**	.0104	.0143**	.0127**	.0118	.0073**	.0101	.0104
	8	.10	.0997	.0997	.0697**	.0999	.1054	.1074*	.1064*	.0841**	.1068*	.1067*
		.05	.0490	.0490	.0292**	.0525	.0553*	.0544*	.0547*	.0382**	.0538	.0542
		.01	.0104	.0104	.0038**	.0100	.0107	.0105	.0108	.0066**	.0113	.0117
	10	.10	.0981	.0981	.0700**	.0962	.1010	.1016	.0982	.0737**	.1024	.0992
		.05	.0503	.0503	.0258**	.0484	.0505	.0518	.0469	.0312**	.0528	.0488
		.01	.0103	.0103	.0045**	.0105	.0106	.0102	.0099	.0047**	.0112	.0089

4　おわりに

攪乱項が正規分布に従う場合にのみ，t 分布によって検定が行うことが出来る。元来，攪乱項の分布は未知であるので，攪乱項の正規性が妥当な仮定かどうかを，仮説検定を行う前に，前もって調べておく必要がある。しかし，検定の結果，正規性の仮定が棄却された場合，回帰係数の仮説検定はもはや出来なくなる。この点を改善するために，本稿では，攪乱項がどのような分布に従っていても適用可能な回帰係数の有意性検定を考えた。

モンテ・カルロ実験では，2標本の相関係数がゼロかどうかの検定を行い，その検出力を調べた。標本が正規分布，カイ自乗分布，一様分布，ロジスティック分布，コーシー分布の5種類の分布に従う場合について，それぞれ10000回のモンテ・カルロ実験によって検出力を，ノンパラメトリック検定と t 検定とで比較した。結果として，予想されるように t 検定は標本の分布によってはサイズが過大推定され，検定に歪みが生じる場合がある。しかし，ノンパラメトリック検定は標本がどの分布に対してもサイズは正しく推定され，特に，正規分布のときでさえも，t 検定（標本が正規分布のときにのみ，t 検定が最もよい検定となることに注意）とそれほど劣らないという結果が得られた。

次に，回帰係数の有意性検定のモンテ・カルロ実験（10000回の繰り返し実験）を行った。パラメータ数が $k=2,3$ の場合を調べたが，全般的に，ノンパラメトリック検定の方が t 検定よりも正しいサイズが得られたと言える。

このように，本稿で考えたノンパラメトリック検定は，攪乱項の分布にかかわらず，適用できることがモンテ・カルロ実験を通して確認することが出来た。しかし，問題点は，計算量の多さである。検定には $n!$ の計算量を必要とする（$n=10$ のとき，$n!$ は約363万となる）。

すなわち，回帰係数の有意性検定を行う場合，$n!$ 個の回帰係数を計算する必要がある。$n=13,14$ 程度を越えると，今のコンピュータの計算速度では，あり得るすべての組み合わせを計算するのはほとんど不可能である。この場合，$(X'X)^{-1}X_i'$ と Y_i $(i=1,2,\cdots\cdots,n)$ の次元を $k\times1$，1×1 から，それぞれ $k\times n_i$，$n_i\times1$ $(i=1,2,\cdots\cdots,m)$ とする。そして，n 組の並べ替えでなく，m 組の並べ替えを考えればよい。ただし，$\sum_{i=1}^{m} n_i=n$, かつ，$n_1=n_2=\cdots\cdots=n_m$ となる必要がある。任意の i, j について，$(X'X)^{-1}X_i'Y_j$ が計算できるためには，n_i と n_j が等しくなっていなければならないため，$n_1=n_2=\cdots\cdots=n_m$ という条件が必要となる。すなわち，$(X'X)^{-1}X_i'$ と Y_j $(i,j=1,2,\cdots\cdots,m)$ との次元が合うように，

$$(X'X)^{-1}X'=((X'X)^{-1}X_1',(X'X)^{-1}X_2',\cdots,(X'X)^{-1}X_m')$$

$$Y=\begin{pmatrix}Y_1\\Y_2\\\vdots\\Y_m\end{pmatrix}$$

と分割して，2節の方法をそのまま当てはめる。m を10程度にすれば，計算量は $m!$ となり，計算量を大幅に減らすことが出来る。分割の方法は，$n_1=n_2=\cdots\cdots=n_m$ となれば，どのような分割でも可能である。このようにすれば，標本数 n が13，14を超えた場合にでも計算が出来るようになる。しかし，$n_1=n_2=\cdots\cdots=n_m$ という条件は n に制約を置くことになる。n が素数で13，14より大きい数のときには，適用出来なくなる。例えば，$n=19$ のときは，$n!$ 個の回帰係数（または，相関係数）を計算することは，現時点でのコンピュータの性能では，不可能である。このような場合は，データを1つ減らして，$n=18$ で分析を進める等の操作が必要である（このとき，$n_i=2, m=9$ とすればよい）。

参 考 文 献

W. J. Convor (1980), *Practical Nonparametric Statistics* (Second Edition), John Wiley & Sons.

R. A. Fisher (1966) *The Design of Experiments* (Eighth Edition), New York: Hafner.

J. D. Gibbons and S. Chakraborti (1992), *Nonparametric Statistical Inference* (Third Edition, revised and Expanded), Marcel Dekker.

R. V. Hogg and A. T. Craig (1995), *Introduction to Mathematical Statistics* (Fifth Edition), Prentice Hall.

M. Hollander and D. A. Wolfe (1973), *Nonparametric Statistical Methods*, John Wiley & Sons.

E. L. Lehmann (1986), *Testing Statistical Hypotheses* (Second Edition), John Wiley & Sons.

R. H. Randles and D. A. Wolfe (1979), *Introduction to the Theory of Nonparametric Statistics*, John Wiley & Sons.

P. Sprent (1989), *Applied Nonparametric Statistical Methods*, Chapman and Hall.

A. Stuart and J. K. Ord (1991), *Kendall's Advanced Theory of Statistics, Vol. 2* (Fifth Edition),

Edward Arnold.

A. Stuart and J. K. Ord (1994), *Kendall's Advanced Theory of Statistics, Vol. 1* (Sixth Edition), Edward Arnold.

竹内啓 (1963)『数理統計学』東洋経済新報社。

竹内啓編 (1989)『統計学辞典』東洋経済新報社。

竹内啓・大橋靖雄 (1981)『統計的推測—2標本問題』(入門現代の数学 [11], 数学セミナー増刊) 日本評論社。