ノンパラメトリック検定のモンテ・カルロ実験による小標本特性

谷崎，久志

(Citation)
国民経済雑誌,173(4):95-107

(Issue Date)
1996-04

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/00176059

(URL)
https://hdl.handle.net/20.500.14094/00176059

ノンパラメトリック検定の
モンテ・カルロ実験による小標本特性

谷　崎　久　志

1　は　じ　め　に

多くのノンパラメトリック検定がこれまで考案されてきた。例えば，スコア検定，メディアン検定，符号検定，並べ替え検定等がある。それらの中で2標本検定のみを，本稿では取り上げる。

ノンパラメトリック検定では分布関数に正規分布等の関数型に関する仮定を置かないということが最大の特徴である。対して，t検定に代表されるパラメトリック検定は必ず分布関数の特定化を必要とする。分布関数の関数型に関して制約が少ないことから，パラメトリック検定に比べてノンパラメトリック検定の検出力は小さくなるだろうと当初は予想されていた。しかしながら，Hodges and Lehman (1956) や Chernoff and Savage (1958) によると，「最悪の場合でも，Wilcoxon 順位和検定はt検定と比較してもそれほど検出力は落ちない」，また，「t検定よりも遥かに Wilcoxon の検出力が上回ることもある」ということが報告された。特に，Chernoff and Savage (1958) は，位置パラメータのシフトの検定問題について，正規スコア検定のt検定に対する漸近的相対効率は1以上になるということを証明した。これは正規スコア検定の検出力がt検定よりも大きいということを意味する。

Wilcoxon 検定と同じようなノンパラメトリック検定が Fisher の2標本検定である。Fisher 検定は標本平均の差を検定統計量とするため，t検定の検出力と漸近的に等しいと言われている (Bradley (1968))。スコア検定がスコアの和を検定統計量とし，Fisher 検定が2標本の平均差を検定統計量とすることを除いて，全く同じタイプの検定である。両方の検定統計量は共に離散型分布に従い，可能な全部の組み合わせを調べなければならない。Mehta and Patel (1983, 1986a, 1986b), Mehta, Patel and Tsiatis (1984), Mehta, Patel and Gray (1985), Mehta, Patel and Wei (1988) は Fisher の並べ替え検定 ($r \times c$ 分割表による独立性の検定) のプログラムをネットワーク・アルゴリズムを用いて作成した。

本稿では，2標本に関するノンパラメトリック検定 (スコア検定，Fisher 検定) の小標本特性を通常のパラメトリック検定であるt検定との比較によって考察する。さらに，経済学への応用として，構造変化の検定の例を示す。

2 ノンパラメトリック検定

ノンパラメトリック検定として有名なものには，正規スコア検定，Wilcoxon（1945）の順位和検定，Fisher（1935）の検定等がある。３つはかなり似た検定である。

２つの標本のグループがあるとき，同じ分布から生成された標本かどうかを検定する。グループ１の無作為標本 $x_1, \cdots, x_{n1}$ は互いに独立に $F(x)$ の分布に従い，グループ２の無作為標本 $y_1, \cdots, y_{n2}$ は互いに独立かつxとも独立で $G(y)$ の分布に従う。$F(x)$ と $G(y)$ は連続分布であり，x，yの累積分布関数を表す。この仮定のもとで，「２つの分布に差はない」という帰無仮説 $H_0 : F(x) = G(x)$ を検定する問題を考える。両者の検定は，対立仮説の下では，yの分布がxの分布よりずれる場合にしか適用できないことが知られている（竹内(1963)[1]）。よって，対立仮説は $H_1 : F(x) = G(x-\mu)$，$\mu > 0$，となる。

いま，グループ１の標本数は$n1$個，グループ２は$n2$個とする。このとき，２つのグループを混ぜ合わせて，$n1 + n2$個の標本から$n1$個を無作為に取り出すことを考えると，その取り出し方は ${}_{n1+n2}C_{n1}$ 通りとなる。${}_{n1+n2}C_{n1}$ 個の組み合わせのそれぞれ一つ一つは等確率で起こり得る。両者の検定は，このそれぞれ一つ一つの起こり得る組み合わせと元々の標本を比べるというものである。

2.1 スコア検定

スコア検定では，２つの標本 $\{x_i\}_{i=1}^{n1}$，$\{y_j\}_{j=1}^{n2}$ を混ぜ合わせて，小さい順に番号を付ける。$\{x_i\}_{i=1}^{n1}$，$\{y_j\}_{j=1}^{n2}$ に対応する順位付けられた標本を $\{Rx_i\}_{i=1}^{n1}$，$\{Ry_j\}_{j=1}^{n2}$ とするとき，スコア検定の検定統計量s_0は

$$s_0 = \sum_{i=1}^{n1} a(Rx_i), \tag{1}$$

によって表される。$a(\cdot)$ が特定化されると，$n1 + n2$個の標本から$n1$個を取り出す ${}_{n1+n2}C_{n1}$ 個のすべての組み合わせについて，同様に，スコアの和(1)を求めることができる。このスコアの和(1)を s_m，$m = 1, 2, \cdots, {}_{n1+n2}C_{n1}$，とする[2]。$s_m$はすべての$m$について等確率で起こり得る。$s_0$ と s_mの大きさの比較によって次の確率を計算することができる。

$$\mathrm{Prob}(s < s_0) = \frac{s_m,\ m = 1, 2, \cdots, {}_{n1+n2}C_{n1},\ \text{の中で}\ s_0\ \text{よりも小さいものの個数}}{{}_{n1+n2}C_{n1}},$$

$$\mathrm{Prob}(s = s_0) = \frac{s_m,\ m = 1, 2, \cdots, {}_{n1+n2}C_{n1},\ \text{の中で}\ s_0\ \text{と同じものの個数}}{{}_{n1+n2}C_{n1}},$$

$$\mathrm{Prob}(s > s_0) = 1 - (\mathrm{Prob}(s < s_0) + \mathrm{Prob}(s = s_0)),$$

ただし，sはスコア検定統計量の分布から生成される確率変数とする。

$\mathrm{Prob}(s<s_0)$ が十分小さければ，s_0 は分布の右端にあることを意味し，すべての定数 a について $F(a)<G(a)$ となる。同様に，$\mathrm{Prob}(s>s_0)$ が十分小さければ，s_0 は分布の左端にあることを意味し，すべての定数 a について $F(a)>G(a)$ となる。よって，帰無仮説 H_0：$F(a)=G(a)$，対立仮説 H_1：$F(a)\neq G(a)$ の場合，例えば，$\mathrm{Prob}(s<s_0)\leq 0.05$ または $\mathrm{Prob}(s>s_0)\leq 0.05$ のときに両側検定によって有意水準10%で帰無仮説を棄却することになる。

関数 $a(\cdot)$ の定式化によって，Wilcoxon 順位和検定，正規スコア検定，ロジスティック・スコア検定等に分類される。

Wilcoxon 順位和検定：Wilcoxon の検定統計量 w_0 は，$a(Rx_i)=Rx_i$ のときのスコア検定に対応し，$w_0=\sum_{i=1}^{n1}Rx_i$，によって表される。

コンピュータの発達する以前の時代には，w の検定統計量の正確な分布を求めることは困難なことであった。これまでは w の分布を正規分布で近似されて用いられてきた。この近似は，$n1$，$n2$ が大きいとき，精度が良いことが知られている（竹内（1963)，Mood，Graybill and Boes（1974))。

Wilcoxon 検定の t 検定に対するピットマンの漸近的相対効率[3]がかなり良いことを Hodges and Lehman（1956）は示した。彼らによると，位置パラメータのシフトについて，Wilcoxon 検定のピットマンの漸近的相対効率は0.864以上という結果を得た。これが意味するところは，Wilcoxon 検定は t 検定と比べて，最悪の場合でさえ，0.136（$=1-0.864$）の漸近的相対効率の減少であり，それほど効率の面で悪くはないということである。さらに，このことは漸近的相対効率が無限に大きくなり得ることも意味しているので，t 検定よりも遥かに検出力が大きくなる可能性がある。特に，x の累積分布関数 $F(x)$ の密度関数を $f(x)$ とし，$f(x)$ が以下の密度関数[4]のとき，漸近的相対効率は1.33になることも Hodges and Lehman は示した。

$$f(x)=\frac{x^2\exp(-x)}{\Gamma(3)}, \qquad (2)$$

ただし，$\Gamma(3)$ はパラメータ3のガンマ関数である。一般的に，漸近的な性質によると，幅の広い分布について，Wilcoxon 検定の検出力は高くなると言われている（竹内・大橋（1981))。

正規スコア検定：正規スコア検定統計量 ns_0 は，$ns_0=\sum_{i=1}^{n1}\Phi^{-1}\left(\frac{Rx_i}{n1+n2+1}\right)$，によって表される[5]。ただし，$\Phi(\cdot)$ は累積標準正規分布関数を表す。（1）式の $a(\cdot)$ が $a(x)=\Phi^{-1}\left(\frac{x}{n1+n2+1}\right)$ と仮定されたとき，正規スコア検定と呼ばれる[6]。

正規スコア検定は，位置パラメータのシフトの仮説と正規性の分布の下で，最適な検定である t 検定と漸近的に同等であるといわれている（Chernoff and Savage（1958)，竹内・大橋（1981))。

ロジスティック・スコア検定：検定統計量は，$ls_0=\sum_{i=1}^{n1}F^{-1}\left(\frac{Rx_i}{n1+n2+1}\right)$，によって表される。ただし，$F(\cdot)$ はロジスティック分布関数，すなわち，$F(x)=\frac{1}{1+e^{-x}}$ を表す。

このように，(1)式の s_0 を w_0, ns_0, ls_0 でそれぞれ置き換えることによって，Wilcoxon 検定，正規スコア検定，ロジスティック・スコア検定の統計量が得られる。その他にも様々なスコア検定が考えられるが，本稿では以上に述べた3つのスコア検定と次節の Fisher 検定との比較を考える。

2.2　Fisher の2標本検定

Wilcoxon 検定は標本の順位和を検定統計量とするのに対して，Fisher 検定では，2つの標本 $\{x_i\}_{i=1}^{n1}$, $\{y_j\}_{j=1}^{n2}$ の標本平均（それぞれ，$\bar{x},\bar{y}$ とする）の差を検定統計量とする。すなわち，検定統計量は，$f_0=\bar{x}-\bar{y}$，によって表される。$n1+n2$ 個の標本から $n1$ 個を取り出す ${}_{n1+n2}C_{n1}$ 個のすべての組み合わせについて，同様に，平均の差を求める。この平均の差を f_m，$m=1,2,\cdots,{}_{n1+n2}C_{n1}$，とする。[7] f_mはすべての m について等確率で起こり得る。f_0 と f_mの大きさの比較によって，スコア検定と同様に，$\mathrm{Prob}(f<f_0)$，$\mathrm{Prob}(f=f_0)$，$\mathrm{Prob}(f>f_0)$ を計算することができる。ただし，fは Fisher 統計量の分布から生成される確率変数とする。

このように，Fisher の2標本検定はスコア検定と同じ型の検定（すべての可能な組み合わせを用いるという意味で）であるが，スコア検定より余分な情報を使っている。スコア検定は順位付けられた標本を検定統計量に用いるのに対して，Fisher 検定は元々の標本そのものを用いる。よって，Fisher 検定の方がスコア検定よりも，検出力の面で優れているように予想される。しかしながら，Bradley (1968)，竹内・大橋 (1981) によると，t検定も Fisher 検定も基本的には標本平均の差を検定統計量としているため，両者は全く同等な検定統計量であると述べている。[8] そのため，Fisher 検定は Wilxoson 検定と比較して，ピットマンの漸近的相対効率の面において，良くなるときもあれば悪くなるときもあり得る。

3　検出力の比較（小標本特性）

3.1　モンテ・カルロ実験 I

x_i は標準正規分布 $N(0,1)$ からの疑似乱数，y_i は正規分布 $N(\mu,\sigma^2)$ からの疑似乱数とする。帰無仮説は「2つの標本は同じ分布から生成されている」，すなわち，$H_0: F(x)=G(y)$ である。$n1=n2=5,7$，$\mu=0.0,0.5,1.0,1.5,2.0$，$\sigma=1.0,1.5,2.0,2.5,3.0$ として，$2\times5\times5=50$ 通りのモンテ・カルロ実験を行う。t検定，Wilcoxon 検定，正規スコア検定，ロジスティック・スコア検定，Fisher 検定の結果をそれぞれ t，w，ns，ls，fと表す。ここでは，位置パラメータ μ の変化に対する検出力と共に，尺度パラメータ σ の変化に対する検出力について，正規スコア，ロジスティック・スコア，Wilcoxon，Fisher，t検定の5者の比較を行

表1：検出力の比較：$F(x)=G\left(\frac{x-\mu}{\sigma}\right)$

μ \ σ		$n1=n2=5$					$n1=n2=7$				
		1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	t	.118	.115	.113	.116	.111	.095	.095	.091	.090	.098
	f	.120	.112	.112	.116	.117	.096	.095	.090	.090	.092
0.0	w	.127	.127∘	.135∘∘	.134∘	.140∘∘	.100	.093	.107∘	.108∘	.103
	ns	.117	.118	.117	.108	.106	.095	.090	.092	.094	.088•
	ls	.116	.115	.102•	.088••	.076•••	.099	.087	.089	.091	.085•
	t	.307	.262	.239	.216	.200	.358	.276	.236	.210	.187
	f	.311	.266	.242	.217	.205	.357	.275	.238	.210	.184
0.5	w	.324∘	.269	.243	.229	.219∘	.344	.272	.234	.218	.204∘
	ns	.306	.253	.215•	.194•	.189	.356	.262	.213•	.188•	.165•
	ls	.310	.246•	.197•••	.163•••	.144•••	.357	.256•	.201••	.176••	.158••
	t	.597	.464	.383	.337	.304	.698	.570	.467	.372	.327
	f	.610	.484∘	.392	.344	.307	.700	.570	.464	.371	.319
1.0	w	.622∘	.486∘	.404∘	.355∘	.321∘	.705	.560	.454	.374	.328
	ns	.598	.458	.367•	.316•	.283•	.689	.539•	.418•••	.319•••	.266•••
	ls	.584	.437•	.328•••	.264•••	.216•••	.691	.530••	.406•••	.299•••	.253•••
	t	.851	.710	.574	.478	.421	.927	.811	.675	.577	.493
	f	.853	.716	.581	.489	.428	.929	.810	.670	.576	.481
1.5	w	.851	.707	.572	.489	.435	.928	.795•	.663	.567	.477•
	ns	.843	.682•	.539••	.448•	.389••	.923	.778••	.619•••	.500•••	.403•••
	ls	.830•	.648•••	.476•••	.365•••	.298•••	.920	.777••	.603•••	.487•••	.388•••
	t	.963	.876	.747	.627	.550	.993	.939	.851	.744	.650
	f	.964	.872	.753	.638	.556	.992	.938	.846	.741	.649
2.0	w	.956•	.856•	.744	.626	.550	.987••	.926•	.829•	.712••	.632•
	ns	.953•	.840•••	.710••	.589••	.501•••	.991	.918••	.798•••	.654•••	.552•••
	ls	.949••	.808•••	.648•••	.495•••	.391•••	.990•	.911•••	.790•••	.638•••	.535•••

う。この場合，対立仮説は $H_1 : F(x)=G\left(\frac{x-\mu}{\sigma}\right)$ によって表される。結果は表1である。[9]

理論的には，$\sigma=1$ のケースは2標本についてその分散が同じであるので，t検定の検出力が最も高くなり，また，σが1と異なる値をとるケースは，t検定よりもノンパラメトリック検定の方が高い検出力を示すことが予想される。

$\sigma=1$ の場合，位置パラメータの値にかかわらず，t検定，Wilcoxon 検定，Fisher 検定の3つの検定はそれほど検出力の差は見られないが，位置パラメータが大きくなると，Wilcoxon 検定は Fisher，t検定よりやや劣る傾向がある。尺度パラメータが大きくなれば，Wilcoxon 検定の検出力が大きくなる傾向が見られる。さらに，標本数が少ない場合の方が，tと Wilcoxon 検定との検出力の差が出るようである。

ロジスティック･スコア検定について，$\sigma=1$ の場合は他の検定と同程度の検出力を示して

いるが，μが大きくなるにつれて，また，σが大きくなるにつれてt検定よりもかなり検出力が落ちるということが表れている。標本数が増えるにつれて，t検定の検出力に近づくという結果も見られる。正規スコア検定についても，全く同様の傾向が見られるが，正規スコア検定の検出力の方が全体的にロジスティック・スコア検定よりも上回っている。

表１から判断して，t検定，Wilcoxon 検定，Fisher 検定は正規スコア検定，ロジスティック・スコア検定よりも優れていると結論付けられる。

Fisher の２標本検定は Wilcoxon 検定より余分な情報を用いるため，小標本の場合には，Fisher 検定の方が Wilcoxon 検定よりも高い検出力を示すことが予想されたが，両者の検出力における差はあまり見られなかった。さらに，t検定が利用可能なときでさえ，Wilcoxon 検定は，ややt検定よりも上回ることもあるが，全体的にはt検定と同じ位の検出力が得られた。これらのことから Hoeffding (1952), Hodges and Lehman (1956), Bradley (1968), 竹内・大橋 (1981) 述べられた大標本の特性が小標本にも当てはまると結論づけられる。

3.2　モンテ・カルロ実験II

Wilcoxon 検定がt検定よりも，漸近的相対効率の面で上回ることが知られている分布から生成される確率変数（$\chi^2(6)$ からの確率変数を２で割ったもの，コーシー分布からの確率変数）を取り上げ，前節と同様に，モンテ・カルロ実験を行う。対立仮説については，位置パラメータのシフトのみを対象とする（すなわち，H_1：$F(x)=G(x-\mu)$ として表される）。ここで取り上げる２つの分布はどちらも正規分布ではないので，t検定による検出力は低いことが予想される。結果は表２，３である。

表２：検出力の比較：$F(x)=G(x-\mu)$

	$n1=n2=5$					$n1=n2=7$				
μ	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
t	.101	.206	.355	.533	.684	.098	.236	.423	.644	.817
f	.110	.219°	.359	.535	.689	.097	.236	.423	.641	.819
w	.117°	.235°°	.391°°	.559°	.699°	.101	.237	.452°	.673°	.842°°
ns	.108	.222°	.371°	.541	.684	.097	.243	.453°	.666°	.827
ls	.100	.218	.372°	.532	.664•	.098	.243	.452°	.662°	.820

(2)の分布のとき（表２），小さいμについては５つの検定の間にはあまり差はないが，μが大きくなるにつれて Wilcoxon 検定の検出力は他の検定と比べて大きくなる。Fisher 検定はt検定と同じ位の検出力を持つ。このことから，竹内・大橋 (1981) はt検定はノンパラメトリック検定であると述べている。元の分布に４次までのモーメントが存在すれば，t検定の有意点は標本数が大きいとき分布型に依存しないと言われ（竹内・大橋 (1981)），本稿のモンテ・カルロ実験から小標本でも同じことが結論付けられる。

一方，元の標本がコーシー分布に従う場合（表３），すなわち，$f(x)=1/\pi(1+x^2)$ のとき，Wilcoxon 検定，Fisher 検定共に，t検定の検出力を大きく上回る。全体的に，Wilcoxon 検定，Fisher 検定，t検定の順番に検出力が小さくなっている。分布型の裾が長くなるほど，Wilcoxon 検定の方がt検定よりも高い漸近的相対効率を得るという事実（竹内・大橋（1981））は，表１～３の３種類の分布についてモンテ･カルロ実験を行った結果によって，小標本についても大標本の場合と同じことが妥当することが言えるであろう。

検定統計量に含まれる情報量の差から，Fisher 検定の方が Wilcoxon 検定よりも検出力の面で上回ることが予想されたが，表２の結果から判断すると Wilcoxon 検定の検出力が優れているということになる。この理由として考えられることは，(ｉ) t検定も Fisher 検定も共に標本平均の差を検定統計量とする，(ii) また，Fisher 検定も Wilcoxon 検定も共に組み合わせに基づくノンパラメトリック検定である，(iii) さらに，幅の広い分布に従う標本について，Wilcoxon 検定はt検定より漸近的相対効率が高い，の３つから，Fisher 検定は検出力の面でt検定は Wilcoxon 検定の中間にあると考えることが出来る。

よって，検出力の大きさの順番は，以上の結果から判断すると，全般的に次のようになる。幅の広い分布に従う標本については Wilcoxon$\geq$Fisher$\geq t$ となり，それ以外の分布については Wilcoxon$\leq$Fisher$\leq t$ となる。

表３：検出力の比較：$F(x)=G(x-\mu)$

	$n1=n2=5$					$n1=n2=7$				
μ	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
t	.070	.132	.232	.317	.404	.089	.157	.242	.339	.426
f	.096°°°	.169°°°	.284°°°	.370°°°	.458°°°	.114°°	.189°°	.280°°	.379°°	.461°°
w	.109°°°	.203°°°	.340°°°	.464°°°	.559°°°	.107°	.226°°°	.370°°°	.524°°°	.661°°°
ns	.101°°°	.184°°°	.307°°°	.426°°°	.528°°°	.107°	.205°°°	.332°°°	.453°°°	.584°°°
ls	.102°°°	.180°°°	.290°°°	.371°°°	.440°°	.108°°	.207°°°	.324°°°	.441°°°	.571°°°

Chernoff and Savage（1958）によって証明された「位置パラメータがシフトするという対立仮説のもとでは，正規スコア検定のt検定にたいする漸近的相対効率は１以上になる」という定理は，小標本の場合にも，ある程度は当てはまると考えられる。表１の $\sigma=1$ の場合には，t検定と正規スコア検定は有意な差異は見られなかった。表２においても，やや正規スコア検定がt検定よりも上回るという程度で，あまり有意な差はない。しかし，表３によると，明らかに，正規スコア検定の検出力がt検定を上回っている。総じて，小標本の場合，表２，３からは，Wilcoxon，正規スコア，Fisher の順に検出力は大きいと言える。

4　応用例：構造変化の検定

通常の回帰モデルにおいて，ほとんどの場合，撹乱項は正規分布であると仮定され，仮説検定が行われる。本稿では，正規性の仮定を緩め，撹乱項の分布に関する仮定を置かずに，構造変化の検定を行う。回帰式 $y_t=x_t\beta+u_t$, $t=1,\cdots,T$, を例にとる。ただし，y_t, x_t, β, u_t はそれぞれ t 期の被説明変数, t 期の説明変数ベクトル $(1\times k)$, 推定されるべき未知パラメータ $(k\times 1)$, t 期の撹乱項（平均ゼロ，分散 σ^2）とする。標本数は T とする。このとき，逐次最小自乗法によって基準化された予測誤差 $w_t=\dfrac{(y_t-x_t\beta_{t-1})}{\sqrt{1+x_t(X'_{t-1}X_{t-1})^{-1}x'_t}}$, $t=k+1,\cdots,T$, が推定される。ただし，$X_{t-1}=(x'_1,x'_2,\cdots,x'_{t-1})'$, $Y_{t-1}=(y_1,y_2,\cdots,y_{t-1})'$ とする。β_{t-1} は $t-1$ 期までのデータを用いて最小自乗法で推定されたパラメータ β の値, すなわち, $\beta_{t-1}=(X'_{t-1}X_{t-1})^{-1}\times X'_{t-1}Y_{t-1}$ を表す。この予測誤差は逐次残差と呼ばれ，互いに独立に分布し，しかも，平均ゼロ，分散 σ^2 となるように基準化されたものである。

逐次残差 w_t をもとにして，構造変化の検定を行う[10]。ある時点の前後で, 逐次残差の構造が変化していれば，構造変化が起こったと判断することができる。標本期間を 2 つに分け，$t=k+1,\cdots,n1$, $t=n1+1,\cdots,T$（ただし，$T=n1+n2$）とする。$\{w_t\}_{t=k+1}^{n1}$ と $\{w_t\}_{t=n1+1}^{T}$ が同じ分布から生成された確率変数かどうかの検定を行う。帰無仮説は H_0：$F(w)=G(w)$, 対立仮説は H_1：$F(w)\neq G(w)$ とする。$F(\cdot)$ は最初の $n1$ 個の予測誤差の分布関数，$G(\cdot)$ は最後の $n2$ 個の予測誤差の分布関数を表す。

『国民経済計算年報』から家計最終消費支出（1985年価格，10億円)，家計国民可処分所得（1985年価格，10億円）のデータ（1960年～1991年）が用いられる。C_t を消費，Yd_t を可処分所得とする。以下の 2 つのタイプの消費関数を推定して，逐次残差を求める。

$$C_t=\beta_0+\beta_1 Yd_t, \tag{3}$$

$$C_t=\beta_0+\beta_1 Yd_t+\beta_2 C_{t-1}, \tag{4}$$

ただし，$\beta_0, \beta_1, \beta_2$ は未知パラメータとする。

それぞれの方程式について，1964年～1991年の逐次残差を算出し，この期間を 2 分割して，前半と後半の逐次残差が同じ分布から生成されているかどうかを調べる。分割時点に示される年を含めてその年の以前と以降で標本を 2 つに分割して，分割された 2 つの標本に差があるかどうかの検定を行う。結果は表 4，5 である。表の中の t, Fisher, Wilcoxon, Normal Scores, Logistic Scores はそれぞれ t 検定, Fisher 検定, Wilcoxon 検定, 正規スコア検定, ロジスティック・スコア検定を意味する。また，それぞれの検定に対応する検定統計量を t_0, f_0, w_0, ns_0, ls_0 によって表す。さらに，その検定統計量よりも小さくなる確率は p 値として示されている。

参考のために，通常よく用いられるステップワイズ Chow 検定の結果を 2 つの表に載せ

表４：構造変化の検定：1964年～1991年の逐次残差の利用
（３）式の消費関数

分割	t		Fisher		Wilcoxon		Normal Scores		Logistic Scores		Chow 検定	
時点	t_0	p 値	f_0	p 値	w_0	p 値	ns_0	p 値	ls_0	p 値	F_0	p 値
1964	-0.67	.253	-2914	.357	10	.357	-0.41	.357	-0.66	.357	1.0	.650
1965	-0.96	.171	-2991	.209	21	.264	-0.73	.304	-1.17	.312	1.7	.801
1966	-1.12	.135	-2887	.141	34	.261	-0.86	.301	-1.39	.317	2.8	.926
1967	-1.23	.114	-2783	.113	48	.273	-0.91	.314	-1.46	.330	5.5	.989
1968	-1.65	.054	-3346	.055	53	.131	-1.90	.176	-3.11	.196	7.7	.997
1969	-1.97	.029	-3637	.030	59	.063	-2.76	.103	-4.52	.122	11.2	.999
1970	-2.27	.015	-3892	.016	66	.031	-3.49	.063	-5.72	.081	16.8	1.000
1971	-2.55	.008	-4111	.009	74	.016	-4.11	.041	-6.72	.056	28.6	1.000
1972	-2.85	.004	-4327	.004	83	.009	-4.62	.029	-7.55	.041	56.5	1.000
1973	-3.48	.000	-4880	.001	87	.002	-5.77	.009	-9.50	.015	66.3	1.000
1974	-4.86	.000	-5850	.000	88	.000	-7.87	.000	-13.51	.000	50.5	1.000
1975	-6.44	.000	-6561	.000	90	.000	-9.47	.000	-16.38	.000	47.3	1.000
1976	-9.19	.000	-7261	.000	93	.000	-10.82	.000	-18.70	.000	55.0	1.000
1977	-11.64	.000	-7586	.000	105	.000	-11.05	.000	-19.06	.000	66.5	1.000
1978	-12.78	.000	-7712	.000	120	.000	-11.00	.000	-18.99	.000	84.5	1.000
1979	-8.82	.000	-7245	.000	140	.000	-10.49	.000	-18.16	.000	53.2	1.000
1980	-7.19	.000	-6916	.000	159	.000	-10.07	.000	-17.49	.000	39.0	1.000
1981	-6.77	.000	-6905	.000	176	.000	-9.85	.000	-17.13	.000	36.6	1.000
1982	-5.51	.000	-6509	.000	199	.000	-8.99	.000	-15.72	.000	25.3	1.000
1983	-4.49	.000	-6063	.000	225	.000	-7.65	.000	-13.40	.000	17.4	1.000
1984	-3.76	.000	-5676	.000	250	.000	-6.50	.001	-11.45	.001	12.5	.999
1985	-3.18	.001	-5342	.002	272	.001	-5.76	.002	-10.26	.002	9.1	.999
1986	-2.94	.003	-5407	.003	290	.003	-5.44	.002	-9.75	.002	7.9	.998
1987	-2.29	.015	-4860	.017	317	.003	-3.84	.017	-6.88	.016	5.0	.986
1988	-1.63	.056	-4090	.067	345	.021	-1.74	.145	-2.87	.160	2.8	.922
1989	-1.14	.131	-3525	.169	366	.109	-1.12	.209	-1.86	.216	—	—
1990	-0.40	.343	-1774	.464	390	.464	-0.13	.464	-0.21	.464	—	—

る。Chow 検定の検定統計量を F_0 とし，そのときの p 値が表されている。

表４の(3)式の消費関数によると，t 検定では1971年～1986年, Fisher 検定では1971年～1986年, Wilcoxon 検定では1972年～1986年, 正規スコア検定では1973年～1986年，ロジスティック・スコア検定では1974年～1986年（ステップワイズ Chow 検定では1968年～1986年，このとき F_0 は帰無仮説の下で自由度（2,25）の F 分布に従う）の期間に有意水準１％で構造変化が起こったと判定することが出来る。また，これらの期間では，分割時点前よりも分割時点後の逐次残差[11]が大きいと判断される。

同様に，表５の消費関数(4)式によると，t 検定では1974年～1982年，Fisher 検定では1974年～1982年，Wilcoxon 検定では1974年～1982年，正規スコア検定では1974年～1982年，ロジ

表５：構造変化の検定：1964年～1991年の逐次残差の利用
（４）式の消費関数

分割時点	t		Fisher		Wilcoxon		Normal Scores		Logistic Scores		Chow検定	
	t_0	p値	f_0	p値	w_0	p値	ns_0	p値	ls_0	p値	F_0	p値
1964	-0.26	.396	-705	.464	13	.464	-0.13	.464	-0.21	.464	.1	.052
1965	-0.41	.340	-794	.338	24	.349	-0.45	.378	-0.72	.383	.3	.192
1966	-0.40	.343	-646	.336	38	.362	-0.49	.384	-0.79	.393	.3	.199
1967	-0.31	.378	-438	.372	55	.437	-0.27	.442	-0.43	.448	.4	.307
1968	-0.69	.246	-885	.243	60	.241	-1.26	.269	-2.08	.284	.4	.259
1969	-0.83	.206	-985	.205	69	.168	-1.77	.210	-2.91	.228	1.1	.644
1970	-1.04	.151	-1169	.152	76	.094	-2.51	.139	-4.11	.159	1.7	.823
1971	-1.25	.110	-1328	.111	84	.055	-3.12	.096	-5.12	.116	2.5	.919
1972	-1.35	.094	-1378	.095	94	.038	-3.54	.076	-5.78	.095	7.9	.999
1973	-1.65	.054	-1619	.055	100	.015	-4.40	.040	-7.19	.055	14.7	1.000
1974	-2.79	.004	-2471	.005	101	.002	-6.49	.004	-11.20	.005	8.1	.999
1975	-3.80	.000	-3037	.000	103	.000	-8.10	.000	-14.07	.000	5.8	.996
1976	-5.00	.000	-3526	.000	106	.000	-9.45	.000	-16.39	.000	5.6	.995
1977	-5.26	.000	-3609	.000	118	.000	-9.67	.000	-16.75	.000	6.4	.997
1978	-4.79	.000	-3448	.000	137	.000	-9.26	.000	-16.09	.000	7.6	.999
1979	-3.33	.001	-2780	.001	165	.000	-7.16	.002	-12.08	.003	4.1	.983
1980	-3.38	.001	-2845	.001	181	.000	-7.02	.002	-11.87	.003	4.1	.983
1981	-3.54	.000	-2992	.000	196	.000	-6.98	.001	-11.79	.002	5.3	.994
1982	-2.79	.004	-2581	.004	222	.003	-5.63	.009	-9.47	.013	3.8	.978
1983	-2.41	.011	-2378	.010	245	.010	-4.78	.021	-8.06	.027	3.2	.960
1984	-2.29	.015	-2378	.014	263	.013	-4.46	.024	-7.55	.030	3.0	.953
1985	-2.05	.025	-2286	.023	284	.025	-3.84	.037	-6.55	.044	2.7	.933
1986	-1.92	.032	-2316	.030	304	.040	-3.32	.049	-5.72	.056	3.0	.952
1987	-1.38	.089	-1877	.086	329	.117	-2.17	.122	-3.77	.127	2.0	.872
1988	-0.50	.309	-797	.312	356	.336	-0.57	.367	-0.90	.377	—	—
1989	0.09	.538	185	.545	380	.619	0.41	.619	0.74	.627	—	—
1990	0.85	.799	2231	.892	402	.892	1.15	.892	1.94	.892	—	—

スティック・スコア検定では1974年～1981年（ステップワイズ Chow 検定では1972年～1978年，1981年，このときF_0は帰無仮説の下で自由度（3,24）のF分布に従う）の期間に有意水準１％で構造変化が起こったと判定することが出来る。

5　お　わ　り　に

ノンパラメトリック検定とは，分布に依存しない検定方法であり，従来は，計算時間とプログラム作成の面から，検定統計量が正規近似によって用いられてきた。近年の目ざましいコンピュータの発展によって，正確な検定が行われることが可能になってきている。本稿では，スコア検定（Wilcoxon 順位和検定，正規スコア検定，ロジスティック・スコア検定）と

Fisher の2標本検定の2種類のノンパラメトリック検定を取り上げ，小標本における検出力の比較を行った。

パラメトリック検定である t 検定とノンパラメトリック検定である Wilcoxon 検定と Fisher 検定を比較した場合，小標本での当初の予想は，(ⅰ) 2つの標本の元の分布が等分散で正規分布に従うとき，t 検定の検出力は最も高い，(ⅱ) t 検定が使えない状況では，2つのノンパラメトリック検定が検出力の面で優れているが，Fisher 検定は元々の標本の情報を余分に用いるため，Wilcoxon 検定より高い検出力を示す，であった。

しかしながら，モンテ・カルロ実験によって得られた結果は，2つの標本が等分散の正規分布に従うときでさえ，Wilcoxon 検定は t 検定と同程度の検出力を示した。さらに，元の標本がコーシー分布のとき，明らかに，Wilcoxon 検定は t 検定よりも優れた検定方法であった。全般的に，t 検定が使えない状況においては，Wilcoxon 検定，Fisher 検定，t 検定の順番に高い検出力を示すことが分かった。Chernoff and Savage (1958) によって証明された「位置パラメータがシフトするという対立仮説のもとでは，正規スコア検定の t 検定にたいする漸近的相対効率は1以上になる」という定理は，小標本の場合にも，ある程度は当てはまると考えられる。総じて，小標本の場合，表2，3からは，Wilcoxon，正規スコア，Fisher の順に検出力は大きいと言えるだろう。

さらに，ノンパラメトリック検定の応用例として，構造変化の検定を取り上げた。このように，ノンパラメトリック検定を用いることによって，撹乱項の分布関数の形状にかかわらず，仮説検定を行うことが可能になる。

注釈

1 分布関数の形状が異なるかどうかのより一般的なノンパラメトリック検定としては，例えば，連による検定が考えられる (Kendall and Stuart (1979))。

2 s_m, $m=1,2,\cdots,{}_{n1+n2}C_{n1}$, の中で少なくとも一つは s_0 に等しい。

3 $N=n1+n2$ のときの t 検定と同じだけの検出力を得るために必要な標本の大きさを $N_0=n1_0+n2_0$ としたとき，N/N_0 の極限値が，ピットマンの漸近的相対効率と定義される (Kendall and Stuart (1979) 参照)。ただし，$n1/N=n1_0/N_0$, $n2/N=n2_0/N_0$ とする。このとき，$\mu=\mu'/\sqrt{N}$ というように N に応じて小さく μ をとると，ピットマン漸近的相対効率は μ' に依存しない形になる。

4 自由度6の χ^2 分布に従う確率変数を z とする。さらに，$x=\frac{z}{2}$ を定義する。このとき，x は密度関数 $f(x)$ からの確率変数となる。

5 以下のモンテ・カルロ実験で用いられるプログラムにおいて，$\frac{Rx_i}{n1+n2+1}$ でなく，$\frac{Rx_i-0.5}{n1+n2}$ を用いる。前者は $n1+n2$ 個のデータそれぞれは $\frac{1}{n1+n2+1}$ で起こりうるものとするが，後者は $1n1+n2$ の確率で生起するものと考える (すなわち，累積分布関数が $\frac{i}{n1+n2}$ と $\frac{i+1}{n1+n2}$, $i=0,1,\cdots,n1+n2-1$, とのちょうど中間の値に対応する実現値が得られるとするものである)。

6　$a(\cdot)$ を一様分布の累積分布関数の逆関数とするとき，すなわち，$a(x)=\dfrac{x}{n1+n2+1}$ と定義されたとき，スコア検定は Wilcoxon 検定と同値である。

7　f_m, $m=1,2,\cdots,{}_{n1+n2}C_{n1}$，の中で少なくとも一つは f_0 に等しい。

8　正規性の仮定のもとでさえ，Fisher 検定は t 検定と漸近的に同等な検出力を持つということを Hoeffding (1952) は示した。

9　t 検定の２つの標本平均の差の検定について，その検定統計量は $t_0=\dfrac{\bar{x}-\bar{y}}{s\sqrt{1/n1+1/n2}}$ として表される（自由度は $n1+n2-2$）。ただし，$\bar{x}=\sum_i x_i$, $\bar{y}=\sum_j y_j$, $s^2=\dfrac{\sum_i(x_i-\bar{x})^2+\sum_j(y_j-\bar{y})^2}{n1+n2-2}$ とする。

10　表１～表３に関する注意点は以下の通りである。(i) $m=1000$ 回のシミュレーションのうちで，$\mathrm{Prob}(t<-t_0)\leq 0.1$, $\mathrm{Prob}(k<k_0)\leq 0.1$, $k=w, ns, ls, f$, となるものの個数を比率で表したものが表に示されている。それぞれの検定に対応する比率を $\hat{p}_k$, $k=t, w, ns. ls, f$ とし，それらは帰無仮説を棄却する確率，すなわち，検出力を表す。さらに，k_0, $k=t, w, ns, ls, f$, を，各シミュレーションで得られた元々の標本からの統計量とする。また，$n1=n2=5$ のとき $t_0=1.3968$ で，$n1=n2=7$ のとき $t_0=1.3562$ となる。(ii) 表におけるそれぞれの値の分散は，$\mathrm{Var}(\hat{p}_k)=p_k(1-p_k)/m$, $k=t, w, ns, ls, f$ となる。(iii) 数字の右の $^\circ$, $^{\circ\circ}$, $^{\circ\circ\circ}$, $^\bullet$, $^{\bullet\bullet}$, $^{\bullet\bullet\bullet}$ は t の値との比較を表す。$\dfrac{\widehat{p_k}-\widehat{p_t}}{\sqrt{\mathrm{Var}(\widehat{p_t})}}$, $k=w, ns, ls, f$, が１より大きいとき $^\circ$，２より大きいとき $^{\circ\circ}$，３より大きいとき $^{\circ\circ\circ}$，-1 より小さいとき $^\bullet$ とし，-2 より小さいとき $^{\bullet\bullet}$，-3 より小さいとき $^{\bullet\bullet\bullet}$ を対応する数字の右肩に付ける。

11　逐次残差でなく，通常の残差を用いることも考えられるが，通常の残差は互いに独立ではない。よって，本稿では，互いに独立である逐次残差に基づいて検定を行う。

参 考 文 献

Bradley, J. V. (1968) *Distribution-Free Statistical Tests*, Englewood Cliffs, New Jersey: Prentice-Hall.

Chernoff, H. and I. R. Savage (1958) "Asymptotic Normality and Efficiency of Certain Nonparametric test Statistics," *Annals of Mathematical Statistics*, Vol.29, pp.972-994.

Fisher, R. A. (1935) *The Design of Experiments* (eighth edition, 1966), New York: Hafner.

Hodges, J. L. and E.L. Lehman (1956) "The Efficiency of Some Nonparametric Competitors of the t Test," *Annals of Mathematical Statistics*, Vol.27, pp.324-335.

Hoeffding, W. (1952) "The Large Sample Power of Tests Based on Permutations of Observations," *Annals of Mathematical Statistics*, Vol.23, pp.169-192.

Kendall, M. and A. Stuart (1979) *The Advanced Theory of Statistics, Vol.2, Inference and Relationship* (fourth edition), Charles Griffin & Company Limited.

Mehta, C. R. and N. R. Patel (1983) "A Network Algorithm for Performing Fisher's Exact Test in $r\times c$ Contingency Tables," *Journal of the American Statistical Association*, Vol.78, No.382, pp.427-434.

Mehta, C. R., N. R. Patel and A. A. Tsiatis (1984) "Exact Significance Testing to Establish Treatment Equivalence for Ordered Categorical Data," *Biometrica*, Vol.40, pp.819-825.

Mehta, C. R., N. R. Patel and R. Gray (1985) "On Computing an Exact Confidence Interval for the Common Odds Ratio in Several 2×2 Contingency Tables," *Journal of the American Statistical Association*, Vol.80, No.392, pp.969-973.

Mehta, C. R. and N. R. Patel (1986a) "A Hybrid Algorithm for Fisher's Exact Test in Unordered $r\times c$ Contingency Tables," *Communications in Statistics*, Vol.15, No.2, pp. 387-403.

Mehta, C. R. and N. R. Patel (1986b) "FEXACT: A Fortran Subroutine for Fisher's Exact Test on Unordered $r\times c$ Contingency Tables," *ACM Transactions on Mathematical Software*, Vol.12, No.2, pp.154-161.

Mehta, C. R., N. R. Patel and L. J. Wei (1988) "Computing Exact Permutational Distributions with Restricted Randomization Designs," *Biometrica*, Vol.75, No. 2, pp. 295-302.

Mood, A. M., F. A. Graybill and D. C. Boes (1974) *Introduction to the Theory of Statistics* (third edition), McGraw-Hill.

Wilcoxon, F. (1945) "Individual Comparisons by Ranking Methods," *Biometrics*, Vol.1, pp.80-83.

加納悟・浅子和美（1992）『入門　経済学のための統計学』 日本評論社。

竹内啓（1963）『数理統計学』 東洋経済。

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。