The Japan Statistical Society

日本統計学会誌
第38巻，第2号，2009年3月
151頁～164頁

パレート分布を用いた経済格差の検定

西埜晴久*，各務和彦*，大鋸　崇*

A simple test for economic inequalities using Pareto distribution

Haruhisa Nishino*, Kazuhiko Kakamu* and Takashi Oga*

これまで，経済格差の計測などに用いられるジニ係数は，記述統計的に計算され，その数値に基づいて分析が行われてきた．本稿の目的は，パラメトリックに所得分布を推定し，その推定値の変化から統計的に格差の拡大（縮小）を検定する簡便な方法を提案することにある．本稿では，発展途上国のようにデータの整備が不十分な地域でも入手可能な分位データを用いて，（部分）順序統計量に基づくパレート分布のパラメータの推定値が陽表的に得られることを示した．また，この推定量が最尤推定と比較して効率性がほぼ同等であることを確認した．さらに，順序統計量に基づく推定量の漸近正規性を用いて，経済格差が異なるかを検定する検定統計量を提案した．あわせて最尤推定から計算できる尤度比検定により格差の変化を検定する方法をも示した．最後に，これらの手法を用いて総務省統計局の家計調査のデータで実証分析を行った．分析の結果，統計的に有意な所得分配の不平等度の拡大や縮小を確認することができた．

Gini coefficients, calculated as descriptive statistics, are widely used to measure economic inequalities. This paper aims at estimating parameters of income distribution using Pareto distribution, and proposing a simple test for change of inequality. The paper shows that the estimator of the parameter is calculated explicitly based on selected order statistics from grouped data, which are simple and available also in developing countries. We also make sure that the estimator is as efficient as maximum likelihood estimators. Moreover, the paper proposes a test statistic which we use its asymptotic normality to derive and a likelihood ratio test based on maximum likelihood estimation, in order to test a change of economic inequality. Finally, we apply these estimations and tests to the Family Income and Expenditure Survey data given by the Ministry of Internal Affairs and Communications, finding statistically significant increases and decreases in income inequality.
Key Words and Phrases: ジニ係数，順序統計量，パレート分布，平均の差の検定

1. はじめに

経済格差・所得格差は古くから研究されてきたテーマである．特に，格差拡大は大きな問題として取り上げられている．例えば，大竹（1994）では1980年代の資産・所得分配について分析し，人口高齢化，年齢内・年齢間賃金格差の拡大，地価高騰などによる資産格差の拡大，税制改革による税金の累進度が弱まったことの4つが格差拡大の要因であると結論付けている．また，大竹（2000）では1990年代の所得格差の要因を分析し，格差拡大の主な要因は人口高齢化と世帯構造の変化であることを示している．そして，橘木（1998）をきっかけとして，所得格差の拡大が深刻な問題であるとして取り上げられるようになり，近年の研究は，大竹（2005）にまとめられている．

格差の測定にあたっては，測定の尺度が必要となる．例えば，変動係数や，Theil尺度，

* 千葉大学法経学部：〒263-8522 千葉市稲毛区弥生町1-33

NII-Electronic Library Service

The Japan Statistical Society

Atkinson (1970) 尺度などが用いられている．中でも広く使われている方法はジニ係数である．ジニ係数とはローレンツ曲線をもとに計測する方法である．係数の範囲は0から1で，係数の値が0に近いほど格差が少ない状態で，1に近いほど格差が大きい状態であることを意味する指標である．先行研究では，分位データや個票データから記述統計的に計算された値を用いて，格差拡大を議論している．実際に，上述した日本の研究でも，分位データからジニ係数によって，所得分配の不平等度が計測され，それらの値の変化によって議論がなされている．しかしながら，記述統計を用いた分位データによるジニ係数の計算は近似計算であるため，過小推計されることが指摘されている．

そこで，パラメトリック・ノンパラメトリックにローレンツ曲線を推定し，ジニ係数を求める方法も提案されている（例えば，Ryu and Slottje (1999)，Hasegawa and Kozumi (2003)，Chotikapanich and Griffiths (2005)，などを参照）．これらの分析では，分布の形状，特に，分布の裾の形状に注目して分析が行われている．標本からは，母集団の最小値や最大値は推測せざるを得ず，ジニ係数の値がそれらの値の影響を受けることが指摘されている．そこで，対数正規分布や，パレート分布などをあてはめるなどして，より正確にローレンツ曲線を推定する試みがなされている．ただし，これらの方法は個票データに基づく分析である．一方，区間データを使ったパラメトリックな分布の推定としては，McDonald and Ransom (1979) およびMcDonald (1984) にあるようにgrouped dataとして多項分布を当てはめて尤度を構成して推定する方法がよく知られており，この方法は分位データに対しても同様に利用できる．さらに，分位データを用いた分布の推定は，発展途上国のようにデータの整備が不十分な地域においても分位データが利用可能であることが多いため，一定の有用性を持つと考えられる．

以上のように，推定に関してはさまざまな方法が提案されているものの，推定されたローレンツ曲線から計算されたジニ係数が変化したか否かを検定する方法についてはこれまであまり研究がなされてない[1)]．日本においても，大竹・齊藤 (1996) はDeaton and Paxson (1994) のモデルを応用して，高齢化にともなって消費のバラツキがどのように変化したかを分析し，統計的に検定しているが，不平等度の変化を直接的に検定したものではない．しかしながら，研究の目的が，格差が拡大したか否かであることを考慮すると，格差が拡大したか否かを検定する方法が必要となるであろう．

本稿の目的は，分位データからパラメトリックに所得分布を推定し，その推定値の変化から統計的に格差の拡大（縮小）を検定する簡便な方法を提案することにある．そこで，本稿では，所得分布にパラメトリックな確率分布の一つであるパレート分布を仮定して，パレート分布のパラメータを推定する方法を提案し，推定されたパラメータを用いて所得分配の不平等度が大きく（小さく）なったか否かを検定する方法を提案する．検定はいわゆる平均の差の検定によって可能であることが示された．さらに，上記の検定を補完する方法として，最尤推定から得られる尤度比検定の検定統計量を計算することによって格差の変化を検定する方法を提案する．そして，提案された方法と家計調査の全世帯と勤労者世帯の分位データを用いて，1985，1990，1995，2000，2005年の日本の所得分配の不平等度を推定し，所得分配の不平等度の拡大と縮小を統計的に検定した．分析の結果，大竹 (2005) と同様に，データの種類で傾向は異なるものの，2000年まで所得分配の不平等度が拡大した後，2005年にかけて縮小しているという傾向を確認することができた．また，統計的に有意な所得分配の不平等度の拡大や縮小を確認することができた．

本稿の構成は以下のとおりである．次節では，推測に用いるパレート分布について説明し，所得分配の不平等度を計測するジニ係数とパレート分布の関係について示す．3節では，分位データに利用できる順序統計量を用いた推定量について説明する．さらに，この推定量の性質

NII-Electronic Library Service

The Japan Statistical Society

についてシミュレーションを用いて確認し，合わせて，従来使われてきた最尤法との比較も行い，これらの推定量がほぼ同等に効率的であることを確認する．さらに4節では，2つのジニ係数が得られたとき，ジニ係数が大きく（小さく）なったかを統計的に検定する方法を提案する．また，最尤推定から得られる尤度比検定により格差の変化を検定する方法も説明する．5節では，3節で得られた結果をもとに，1985，1990，1995，2000，2005年の日本の所得分配の不平等度を推定する．さらに，4節の結果を用いて年によって格差が変化しているかを検定する．6節では，本稿で得られた結論と，今後の課題について述べる．

2. パレート分布

所得分布を推定するための確率分布としては，さまざまな分布が仮定されているが，本稿ではパレート分布を用いる．パレート分布は所得分布の推定にしばしば用いられる分布で（例えばSingh and Maddala (1976) など），所得分布の下限は正であるということを考慮しているなどの特徴がある．パレート分布にもいくつか種類があるが，本稿ではJohnson, Kotz and Balakrishnan (1994) でいう第一種のパレート分布 $P(I)(\theta, \alpha)$ を用いる．この第一種のパレート分布は2つのパラメータを持ち，以下の分布関数と密度関数を持つ．

パレート分布の分布関数：

$$F(x)=1-\left(\frac{x}{\theta}\right)^{-\alpha}, \quad x\geq\theta>0, \alpha>0. \tag{2.1}$$

確率密度関数：

$$f(x)=\alpha\theta^{\alpha}x^{-\alpha-1}, \quad x\geq\theta>0, \alpha>0. \tag{2.2}$$

第一種のパレート分布はパラメータが2つある．一つは分布の下限を表すパラメータ θ であり，もう一つは分布の形状を表すパラメータ α である．

パレート分布を前提とし，$\alpha>1$ の条件の下で，ジニ係数 G はDorfman (1979, (1)) の式を利用して計算できる．期待値が $E(X)=\alpha\theta/(\alpha-1)$ であるので，パレート分布の確率分布関数を $F(x)$ とすると，

$$\begin{aligned} G&=1-\frac{1}{E(X)}\int_0^{\infty}(1-F(x))^2dx \\ &=1-\frac{1}{\alpha\theta/(\alpha-1)}\left[\int_0^{\theta}1dx+\int_{\theta}^{\infty}\left(\frac{x}{\theta}\right)^{-2\alpha}dx\right]=\frac{1}{2\alpha-1}, \end{aligned} \tag{2.3}$$

となる．（その他，Kleiber and Kotz (2003, (3.53)) を参照）．(2.3) よりジニ係数は形状パラメータ α だけによって決まることが分かる．つまり，不平等度についての情報は形状パラメータ α に集約される．その結果，所得分配の不平等度の計測にあたっては，形状パラメータ α の推定が重要になる．

3. 順序統計量を用いたパラメータ推定

パレート分布におけるパラメータの推定法は，Johnson, Kotz and Balakrishnan (1994) 20章に紹介されているように最尤法などいくつか知られている．本稿では，selected order statisticsによる推定法を用いる．なぜなら，経済格差を計測するために利用する所得分布の統計が分位点が分かっている分位データであることが多いため，順序統計量を使った推定法が使いやすいからである．なお，この推定法は，Koutrouvelis (1981) によって提案されたもので，パレー

NII-Electronic Library Service

The Japan Statistical Society

ト分布を対数変換して指数分布とした場合には，Saleh and Ali (1966) などによりABLUE (Asymptotically BLUE) が示されているものである．さらにこの推定量は近年ではBrazauskas and Serfling (2000) で他の推定法との比較が行われている．

3.1 Selected order statisitcs based estimation

まず，母集団より全体の標本がサイズ n で得られ，さらにその部分標本がサイズ k で得られて，サイズ k の部分標本が小さい順に並び替えられているとする．つまり，$1\leq n_1<n_2<\ldots<n_k\leq n$ に対し，$\{X(n_1)\leq X(n_2)\leq,\cdots,\leq X(n_k)\}$ の順序統計量が得られるとする．ただし，i 番目の順序統計量 $X(n_i)$ は，ある一定の比率 $p_i=n_i/n$ に対応して得られているものとする．また，$n\to\infty$ の大標本では，k を固定して，$\lim_{n\to\infty}n_i/n=p_i$ を仮定する．さらに，母集団の分布がパレート分布である仮定の下で，計算の便宜上 $u_i=-\log(1-p_i)$ の表記を導入する．

この時の推定のアイデアを説明すると以下のようになる．確率変数 X が密度関数 (2.2) のパレート分布 $P(I)(\theta,\alpha)$ に従うとき，対数変換した確率変数 $Y=\log X$ は，確率密度関数

$$f(y)=\alpha\theta^{\alpha}e^{-\alpha y}=\frac{1}{\sigma}\exp\left\{-\frac{(y-\mu)}{\sigma}\right\},\quad y>\mu,\ \sigma>0, \tag{3.1}$$

の指数分布に従う．なお，$(\mu,\sigma)=(\log(\theta),1/\alpha)$ とパラメータを変換した．そこで，Mosteller (1946, Theorem 2) より，$(\log X(n_1),\log X(n_2),\ldots,\log X(n_k))'$ の同時確率分布は $n\to\infty$ の漸近的に平均 $(\mu+\sigma u_1,\mu+\sigma u_2,\ldots,\mu+\sigma u_k)'$，分散共分散行列 $(\sigma^2/n)W$ の多変量正規分布にしたがう．ただし分散共分散行列 W の (i,j) 成分は

$$W_{ii}=W_{ij}=W_{ji}=\frac{p_i}{1-p_i}=e^{u_i}-1,\quad (i<j),\quad i,j=1,2,\ldots,k, \tag{3.2}$$

であり，W の逆行列 W^{-1} の (i,j) 成分は，

$$W'_{i,i}=\frac{(e^{-u_{i-1}}-e^{-u_{i+1}})e^{-2u_i}}{(e^{-u_i}-e^{-u_{i+1}})(e^{-u_{i-1}}-e^{-u_i})},$$

$$W'_{i,i-1}=W'_{i-1,i}=-\frac{1}{(e^{u_i}-e^{u_{i-1}})},$$

$$W'_{i,j}=0,\quad for\ |i-j|>1,$$

である．なお，指数分布における順序統計量による推測では，Saleh and Ali (1966) が (3.2) の分散共分散行列およびその逆行列を具体的に示している．

$(\log X(n_1),\log X(n_2),\ldots,\log X(n_k))'$ の同時確率分布が漸近的に多変量正規分布になるので，上記のモデルは以下のように書き換えられる．

$$\log X(n_i)=\mu+\sigma u_i+\varepsilon_i,\quad (i=1,\ldots,k) \tag{3.3}$$

ここで，$(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_k)'$ は，大標本で平均0で分散共分散行列 $(\sigma^2/n)W$ の多変量正規分布で近似できる．そこで，(3.3) を線形回帰式と見立ててやれば誤差分散のウェイトが既知の不均一分散・相関がある回帰モデルの推定になるので一般化最小2乗法で推定すれば良い．この時の推定量は

$$\hat{\mu}=\log X(n_1)-\hat{\sigma}u_1, \tag{3.4}$$

$$\hat{\sigma}=\sum_{i=1}^{k}b_i\log X(n_i) \tag{3.5}$$

NII-Electronic Library Service

The Japan Statistical Society

となる．なお，

$$b_1=-\frac{1}{L}\frac{u_2-u_1}{e^{u_2}-e^{u_1}},$$

$$b_i=-\frac{1}{L}\left[\frac{u_i-u_{i-1}}{e^{u_i}-e^{u_{i-1}}}-\frac{u_{i+1}-u_i}{e^{u_{i+1}}-e^{u_i}}\right],\quad (i=2,\ldots,k-1)$$

$$b_k=-\frac{1}{L}\frac{u_k-u_{k-1}}{e^{u_k}-e^{u_{k-1}}},$$

$$L=\sum_{i=2}^{k}\frac{(u_i-u_{i-1})^2}{e^{u_i}-e^{u_{i-1}}},$$

である．したがって，(3.4) と (3.5) から，$(\hat{\mu},\hat{\sigma})$ は陽表的に計算できることが分かる．つまり，簡便に推定量の計算が可能となるので実証分析を行う研究者にとっては有用である．

3.2 推定量の漸近分散

(3.4)，(3.5) で示された推定量は Koutrouvelis (1981) において漸近正規性を持つことが示されている．合わせて漸近分散も求められており，

$$Var(\hat{\sigma})=\frac{1}{L}\frac{\sigma^2}{n},\tag{3.6}$$

$$Var(\hat{\mu})=\left[\frac{u_1^2}{L}+\{e^{u_1}-1\}\right]\frac{\sigma^2}{n},\tag{3.7}$$

$$Cov(\hat{\sigma},\hat{\mu})=\frac{u_1}{L}\frac{\sigma^2}{n},\tag{3.8}$$

となっている．

したがって，$\hat{\alpha}=1/\hat{\sigma}$, $\hat{\theta}=\exp(\hat{\mu})$ と $\hat{G}=1/(2\hat{\alpha}-1)$ を使って，パラメータ α,θ とジニ係数 G を推定したときに，2つの推定量は共に一致性をもち，漸近分散は，

$$Var(\hat{\alpha})=\frac{\sigma^2}{nL},\tag{3.9}$$

$$Var(\hat{\theta})=\frac{\theta^2[u_1^2/L+\{e^{u_1}-1\}]}{n\alpha^2},\tag{3.10}$$

$$Var(\hat{G})=\frac{4\sigma^2}{(2\alpha-1)^4nL},\tag{3.11}$$

と計算される．なお，$\hat{G}$ の漸近分散の導出にあたってはデルタ法を用いた．

3.3 最尤推定

従来，所得分布のパラメータの推定にあたっては，区間データを grouped data として扱い，多項分布を当てはめて最尤法で行う方法が一般的に良く用いられている．この推定法はいわゆる多項最尤法（MML）であり，後に説明する順序統計量の同時確率密度から得られる最尤法に対しては，擬似最尤法となっている．多項最尤法（MML）は，例えば，McDonald and Ransam (1979) および McDonald (1984) において言及され，日本のデータでの実証分析としては，Atoda, Suruga, and Tachibanaki (1988) および Tachibanaki, Suruga, and Atoda (1997) がある．これらの研究では次のように推定を行っていた．つまり，$1\leq n_1<n_2<\ldots<n_k\leq n$ に対し，$\{X(n_i);\ i=1,2,\ldots,k\}$ の順序統計量の観測された値が $(x_1,\ldots,x_k)$ と得られているとする．

このときこれらの k 個の順序統計量で区切られたセルを考えると，セルの数は $g=k+1$ と

NII-Electronic Library Service

The Japan Statistical Society

なり，また，$m_i=n_1$, $m_i=n_i-n_{i-1}$, $m_g=n-n_k$ とおく．（ただし，$\sum_{i=1}^{g}m_i=n$）．すると多項分布の尤度は，

$$L=n!\prod_{i=1}^{g}\frac{\pi_i^{m_i}}{m_i!} \tag{3.12}$$

となる．ただし，各セルの確率は

$$\begin{aligned}&\pi_1=F(x_1),\\&\pi_i=F(x_i)-F(x_{i-1}),\quad (i=2,\ldots,g-1),\\&\pi_g=1-F(x_{g-1}),\end{aligned}$$

と確率分布関数 $F(x)$ から求めることができる．また，McDonald and Ransom (1979) では，(3.12) を最大化する推定法と同等の推定法としてピアソンの最小カイ2乗推定法を提案していた．

実際に (3.12) を最大化するには，パレート分布の分布関数 (2.1) を代入して対数尤度を計算すれば，

$$\begin{aligned}\log L=&C+n\alpha\log\theta+n_1\log(\theta^{-\alpha}-x_1^{-\alpha})\\&+\sum_{i=2}^{k}(n_i-n_{i-1})[\log(x_{i-1}^{-\alpha}-x_i^{-\alpha})]+(n-n_k)\log(x_k^{-\alpha})\end{aligned} \tag{3.13}$$

となるので，(3.13) を最大化する (θ,α) を求めれば良い．

尤度関数を正確に求めるならば，順序統計量 $(X(n_i);\ i=1,2,\ldots,k)$ の同時確率密度を求める必要がある．そこで，David and Nagaraja (2003, (2.2.3)) より，順序統計量の同時確率密度を求めると

$$n!\left[\prod_{j=1}^{k}f(x_j)\right]\prod_{j=0}^{k-1}\left\{\frac{[F(x_{j+1})-F(x_j)]^{n_{j+1}-n_j-1}}{(n_{j+1}-n_j-1)!}\right\}\times\frac{[1-F(x_k)]^{n-n_k}}{(n-n_k)!} \tag{3.14}$$

となる．（ただし，$F(x_0)=0$, $n_0=0$）．したがって，対数尤度は

$$\begin{aligned}\log L=&C'+\sum_{j=1}^{k}\log f(x_j)+(n_1-1)\log(F(x_1))\\&+\sum_{j=2}^{k}(n_j-n_{j-1}-1)\log(F(x_j)-F(x_{j-1}))+(n-n_k)\log(1-F(x_k))\end{aligned} \tag{3.15}$$

と書くことができる．したがって，対数尤度 (3.15) を最大化することで最尤推定値が得られる．そして，パレート分布の場合には対数尤度は，

$$\begin{aligned}\log L=&C'+\left[k(\log\alpha+\alpha\log\theta)-(\alpha+1)\sum_{j=1}^{k}\log x_j\right]\\&+(n-k)\alpha\log\theta+(n_1-1)\log(\theta^{-\alpha}-x_1^{-\alpha})\\&+\sum_{i=2}^{k}(n_i-n_{i-1}-1)[\log(x_{i-1}^{-\alpha}-x_i^{-\alpha})]+(n-n_k)\log(x_k^{-\alpha})\end{aligned} \tag{3.16}$$

となる．(3.16) を最大化する (θ,α) が推定値となる．なお，この対数尤度は n が十分大きければ (3.16) 式右辺第二項が無視しうるほど小さくなるので，(3.13) 式で近似することができる．さらに，前述の多項最尤法（擬似最尤法）は最尤法と漸近的に同等である．

NII-Electronic Library Service

The Japan Statistical Society

3.4 モンテカルロ実験

3.1節で紹介したμ, σ, α, Gの推定量に関しては, Koutrouvelis (1981) などの先行研究で漸近正規性が示されている. そこでここでは, まずはじめに, 実証分析で用いる総務省統計局家計調査のデータと同じ標本サイズ$n=10000$を使って漸近正規性を確認する. 漸近正規性が確認できれば, 正規分布への近似を使うことでさまざまな検定を行うことが可能になる. さらに前節で説明した最尤推定法についてもモンテカルロ実験を行い, 推定結果を比較する. なお計算に当たってはOx Version 4.1 (Doornik, 2006) を用いてプログラムを作成し計算した.

モンテカルロ実験のデザインは, 以下の通りである.

(i) パラメータを$\alpha=1.35$, $\theta=300$として, パレート分布から$n=10000$の標本をサンプリングする.

(ii) 得られた標本から, 20%, 40%, 60%, 80%分位点を求める.

(iii) 得られた分位点をもとに, $\hat{\mu}, \hat{\sigma}, \hat{\alpha}, \hat{G}$を推定する.

(iv) 上記の1.3をM回繰り返し, $\hat{\mu}, \hat{\sigma}, \hat{\alpha}, \hat{G}$についての経験分布を作製する.

得られた標本について, 本稿では標本平均と標本分散を求めた. なお, シミュレーションの繰り返し数Mは, 1000と5000とした. シミュレーションにおけるパラメータは$\alpha=1.35$, $\theta=300$で行ったが, これは, 表3 (総務省統計局家計調査：全世帯) の分析において推定されたパラメータに準拠し, 標準的な値であると考えられる.

表1には, モンテカルロ実験における母パラメータと推定量の平均及び分散[2]をまとめてある. 表1の母パラメータの平均とシミュレーションの結果を比較すると, 本稿で提案している推定法に問題なく, パラメータは正確に推定されていることが分かる. また, $n=10000$の標本サイズで, 十分に漸近分布に収束していることが示されている.

図1には, 確認のため, 繰り返し数5000回の時の$\hat{\sigma}, \hat{\mu}, \hat{\alpha}, \hat{\theta}, \hat{G}$の分布をヒストグラムとして描いた. また, 実線で描かれているのは, 正規分布の確率密度関数である. この図より, すべての推定量の分布が正規分布に収束しているのが見てとれる. 以上, モンテカルロ実験の結果から, $n=10000$の標本において, 推定量の漸近正規性は確認できた.

最後に, 先に漸近正規性を確認した3.1節のSelected order statistics based estimationと3.3節で紹介した (3.13) を最大化する多項最尤法および (3.16) を用いた最尤推定法のパフォーマンスをそれぞれ, SoE, MML, MLとしてモンテカルロ実験によって比較した. これまでと同様に, $\alpha=1.35$, $\theta=300$, とし, $g=k+1=5$を固定し, $n=50, 100, 1000, 10000$の場合について, 繰り返し数$M=5000$でそれぞれの推定値の標本平均, 標本分散, MSEを計算した. 結果が表2にまとめてある. まず, $n=10000$の時をみると, SoE, MML, MLの3つの推定法で推定量のMSEは3つの推定量で同じ程度に十分小さい, αで0.0003程度であり, θで1.68程度である.

さらに, nが小さいときには, バイアスが生じ, またMSEも大きくなっていることが分か

表1 推定量のシミュレーション結果

		σ	μ	α	θ	G
母パラメータ	平均	0.74074	5.7038	1.35	300	0.58824
理論	値漸近分散	9.3302e-05	1.8363e-05	3.0990e-04	1.6527	1.4842e-04
1000回	平均	0.74113	5.7038	1.3495	300.00	0.58882
	標本分散	9.4726e-05	1.9161e-05	3.1431e-04	1.7250	1.5097e-04
5000回	平均	0.74064	5.7039	1.3504	300.03	0.58820
	標本分散	9.3766e-05	1.8826e-05	3.1205e-04	1.6955	1.4918e-04

NII-Electronic Library Service

The Japan Statistical Society

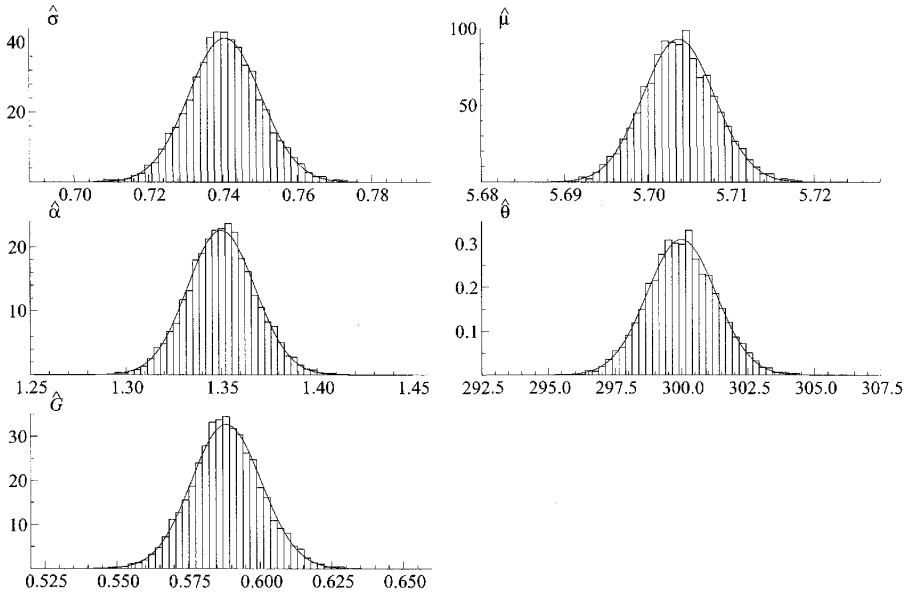

図1 シミュレーションによる推定量の分布（$n=10000$, $M=5000$）

る．しかしながら，$g=5$ の5分位データでは，$n=10000$ なら大標本になっていて，どの推定法を用いても MSE にほとんど差がない．細かく見ると SoE の MSE が MML および ML に比べやや小さく見えるが，SoE では一般化最小2乗法により陽表的に推定量が得られているのに対し，MML および ML では数値的最適化を用いて推定値を求めているために精度が落ちているためと考えられる．しかしながら，SoE, MML, ML の3つの推定法のパフォーマンスはほぼ同じと言えるだろう．つまり，3つの推定量はどれも同程度に効率的であると判断できる．

4. 経済格差の検定

モンテカルロ実験の結果から，$n=10000$ のケースにおいて，Selected order statistics based estimation では十分に大標本とみなすことができることが示された．この結果により，大竹（2005）などの従来の研究のように，記述統計として大きさの変化を追うだけでなく，ジニ係数の違いが統計的に有意であるか検定を行うことが可能となる．以下ではまず，その検定方法

表2 推定法の比較

	SoE		MML		ML	
	α	θ	α	θ	α	θ
$n=50$	1.4370	301.17	1.4389	301.24	1.4505	306.96
	0.076836	324.93	0.077045	324.90	0.078331	318.83
	0.084401	326.31	0.084952	326.44	0.088432	367.32
$n=100$	1.3915	300.42	1.3924	300.45	1.3982	303.38
	0.035308	169.75	0.035357	169.74	0.035647	167.90
	0.037027	169.93	0.037157	169.94	0.037966	179.31
$n=1000$	1.3544	300.08	1.3545	300.08	1.3551	300.38
	0.0031390	16.316	0.0031396	16.325	0.0031422	16.306
	0.0031586	16.323	0.0031600	16.332	0.0031680	16.451
$n=10000$	1.3505	300.03	1.3505	300.03	1.3506	300.06
	0.00031787	1.6817	0.00031818	1.6853	0.00031820	1.6852
	0.00031811	1.6828	0.00031843	1.6865	0.00031852	1.6893

上段：標本平均，中段：標本分散，下段：MSE，繰り返し数 $M=5000$

NII-Electronic Library Service

The Japan Statistical Society

を示す．さらに，3.3節の最尤推定法を応用することで，対数尤度比を用いた尤度比検定によって，格差の変化を検定する方法を合わせて提案する．

4.1 σの変化の検定

ジニ係数の推定量$\hat{G}$は，$\hat{\sigma}>0$において単調変換であるので，ジニ係数の変化を検定することと，σの変化を検定することは同義である．また，パレート分布のパラメータ$\hat{\alpha}$およびジニ係数$\hat{G}$は$\hat{\sigma}$から変換して得られたものなので，元々推定したパラメータσについて検定を行ったほうが良い．

2つのパラメータσ_tとσ_sについて，帰無仮説，対立仮説をそれぞれ，

$$\text{帰無仮説}\ (H_0)\colon \sigma_t=\sigma_s \quad ; \quad \text{対立仮説}\ (H_1)\colon \sigma_t\neq\sigma_s$$

と設定する．つまり，帰無仮説では格差の程度は変わらないであり，対立仮説では格差の程度は異なっているである．前節の議論より$\hat{\sigma}$は標本の大きさnが十分に大きいと正規分布で近似でき，漸近分散も（3.6）を使うことで分かる．そこで，σ_t, σ_sを推定した際の標本サイズをn_t, n_sとし，推定に用いたLをそれぞれL_t, L_sとすれば，

$$\hat{\sigma}_t \sim \mathcal{N}\left(\sigma_t, \frac{\sigma_t^2}{L_t n_t}\right), \quad \hat{\sigma}_s \sim \mathcal{N}\left(\sigma_s, \frac{\sigma_s^2}{L_s n_s}\right), \tag{4.1}$$

となる．正規分布の平均の差の検定を考えると，$\hat{\sigma}_t, \hat{\sigma}_s$が独立に分布するとの仮定の下で，$\hat{\sigma}_t-\hat{\sigma}_s$の分布は，

$$\hat{\sigma}_t-\hat{\sigma}_s \sim \mathcal{N}\left(\sigma_t-\sigma_s, \frac{\sigma_t^2}{L_t n_t}+\frac{\sigma_s^2}{L_s n_s}\right), \tag{4.2}$$

と近似できる．したがって，帰無仮説：$\sigma_t=\sigma_s$の下では検定統計量Zは，以下のように，

$$Z=\frac{\hat{\sigma}_t-\hat{\sigma}_s}{\sqrt{\dfrac{\hat{\sigma}_t^2}{L_t n_t}+\dfrac{\hat{\sigma}_s^2}{L_s n_s}}} \sim \mathcal{N}(0,1) \tag{4.3}$$

となり，大標本で標準正規分布で近似できることが分かる．よって，両側検定の場合には有意水準5%の棄却域を$|Z|>1.96$とし，有意水準1%の棄却域が$|Z|>2.576$と設定することで検定が構成できる．すなわち，検定統計量Zが棄却域に入れば帰無仮説：$\sigma_t=\sigma_s$は棄却されることになり，有意に格差が異なっていると判断できる．

さらに，同様に片側検定を考えることもできる．右片側対立仮説を考えると，帰無仮説：$\sigma_t=\sigma_s$に対し，対立仮説：$\sigma_t>\sigma_s$である．そこで，有意水準5%の棄却域を$Z>1.645$とし，有意水準1%の棄却域を$Z>2.326$と設定することで片側検定が構成できる．

以上のように，標本サイズnが十分に大きい場合，ジニ係数の違いの検定は，平均の差の検定を応用することによって可能となり，検定統計量の分布は，近似的に標準正規分布に従うことが示されたので，標準正規分布を用いて検定を行うことができる．

4.2 尤度比検定による格差の変化の検定

3.3節で説明した最尤推定で計算される対数尤度を用いれば，尤度比検定を使うことで格差の変化を検定することができる．パレート分布を仮定しているので，格差の変化は形状パラメータαがα_tとα_sで変化したかを見ればよい．つまり，

$$\text{帰無仮説}\ (H_0)\colon \alpha_t=\alpha_s \quad ; \quad \text{対立仮説}\ (H_1)\colon \alpha_t\neq\alpha_s$$

NII-Electronic Library Service

The Japan Statistical Society

の検定問題を考えれば良い．帰無仮説では格差の変化はないとなる．

ここでα_tとα_sを別々に推定したときの対数尤度（$\log L(H_1)$）と$\alpha_t=\alpha_s=\alpha$と格差が変化しないことを仮定して推定したときの対数尤度（$\log L(H_0)$）を計算して，その差から対数尤度比を求めて検定統計量とする．具体的には，(3.13) もしくは (3.16) で定義される対数尤度関数を使って，検定統計量λ

$$\lambda=2\log\left\{\frac{L(H_1)}{L(H_0)}\right\}=2[\{\log L(\theta_t,\alpha_t)+\log L(\theta_s,\alpha_s)\}-\{\log L(\theta_t,\alpha)+\log L(\theta_s,\alpha)\}], \tag{4.4}$$

を得る．このときの制約式は1本であるから自由度1のカイ2乗分布から棄却域を定めればよい．実際の臨界値は有意水準5%では，3.841であり，1%では6.635である．たとえば，$\lambda>6.635$なら有意水準1%で格差の変化がないという帰無仮説が棄却されることになる．

ただし，この尤度比検定においてはZ検定とは異なり，片側検定を考えることはできない．また，検定統計量を求めるにあたっては，最尤推定の時に最適化の計算を必要とするため，平均の差の検定を行う場合に比べ，計算がやや複雑となる．そこで，尤度比検定は前節のσの変化の検定と補い合って，格差の変化を検定する方法として利用することができるだろう．

5. 実証分析

本節では，実際に総務省統計局による家計調査に基づき，実証分析を行う．所得分布の推定に利用される統計としては，他に良く用いられているものとして，より大きな標本をもち，10分位データの全国消費実態調査がある．本稿では，少ない分位データしか利用可能でない発展途上国への応用も念頭におき，5分位データの家計調査を用いる．本節では，1985，1990，1995，2000，2005年の5年分についての所得分布について，パラメータ及びジニ係数の推定を行った．また，データは，全世帯と勤労者世帯の両方を用いた．これらのデータは，モンテカルロ実験の時と同様，10000世帯（$n=10000$）について調査を行ったものである．これらのデータは，2000人毎の5階級に区切って，20%，40%，60%，80%分位点を公表している．

5.1 家計調査における推定

全世帯および勤労者世帯についての推定結果をまとめたものが表3である．なお，α, θ, Gの標準誤差は (3.9)，(3.10)，(3.11) の平方根により計算した．そして3.3節の最尤推定の結果は表4にまとめてある．ちなみに，(3.13) に対応するのがMML，(3.16) に対応するのが，MLである．表4のαとθの推定値および標準誤差は，表3の推定結果とほぼ同じであるので，表3の数字に即して説明する．

まず，全世帯の推定結果をみると，ジニ係数Gは，2000年まで上昇の一途をたどり，その後，2000年から2005年の間に0.05353小さくなっていることがわかる．このことから，日本では，バブル前から所得格差が広がっており，その勢いはバブル崩壊後も衰えることなく2000年ごろまで続いたと考えられる．ただし，1990年から1995年のバブル崩壊期におけるジニ係数の伸びが0.05217と大きく，1985年から1990年が0.0092，1995年から2000年が0.01661と比較的小さな値となっている．

勤労者世帯の推定結果をみると，全世帯での結果と異なり，ジニ係数Gは，1995年まではほぼ横ばいであり，その後，2000年にかけて0.02941上昇した．その後，0.00138下がってはいるが，2000年からはほぼ横ばいといって差し支えないであろう．

以上まとめると，全世帯についての調査では，所得分配の不平等度は2000年をピークに緩和されつつあると結論付けることができる．また，勤労者世帯についても同様であるが，不平

NII-Electronic Library Service

The Japan Statistical Society

表3　家計調査（全世帯および勤労者世帯）における推定結果

年	全世帯					勤労者世帯				
	σ	μ	α	θ	G	σ	μ	α	θ	G
1985	0.7113	5.568	1.406	261.9	0.552	0.6031	5.726	1.658	306.8	0.432
	9.276e-03	4.115e-03	0.0183	1.078	0.0157	7.865e-03	3.489e-03	0.0216	1.071	0.0134
1990	0.7189	5.739	1.391	310.9	0.561	0.6073	5.902	1.647	365.9	0.436
	9.375e-03	4.159e-03	0.0181	1.293	0.0159	7.920e-03	3.513e-03	0.0215	1.286	0.0134
1995	0.7604	5.832	1.315	341.0	0.613	0.6050	6.045	1.653	422.0	0.434
	9.915e-03	4.399e-03	0.0171	1.500	0.0170	7.889e-03	3.500e-03	0.0216	1.477	0.0134
2000	0.7730	5.773	1.294	321.5	0.630	0.6330	6.009	1.580	407.2	0.463
	1.008e-02	4.472e-03	0.0169	1.438	0.0173	8.254e-03	3.662e-03	0.0206	1.491	0.0140
2005	0.7313	5.715	1.367	303.2	0.577	0.6317	5.950	1.583	383.9	0.462
	9.537e-03	4.231e-03	0.0178	1.283	0.0162	8.237e-03	3.654e-03	0.0206	1.403	0.0139

上段：推定値，下段：標準誤差

表4　家計調査（全世帯および勤労者世帯）におけるMMLとMLの推定結果

年	全世帯				勤労者世帯			
	MML		ML		MML		ML	
	α	θ	α	θ	α	θ	α	θ
1985	1.410	262.066	1.410	262.092	1.663	306.916	1.663	306.942
	0.0183	1.075	0.0183	1.073	0.0216	1.067	0.0216	1.067
1990	1.395	311.047	1.395	311.078	1.651	366.031	1.651	366.061
	0.0181	1.289	0.0181	1.288	0.0215	1.283	0.0215	1.281
1995	1.320	341.147	1.320	341.182	1.658	422.168	1.658	422.203
	0.0171	1.493	0.0172	1.494	0.0215	1.471	0.0216	1.472
2000	1.298	321.649	1.298	321.683	1.585	407.393	1.585	407.429
	0.0169	1.434	0.0169	1.433	0.0206	1.486	0.0206	1.485
2005	1.371	303.372	1.371	303.401	1.588	384.043	1.588	384.076
	0.0178	1.279	0.0178	1.279	0.0206	1.398	0.0206	1.399

上段：推定値，下段：標準誤差

等度の緩和の程度は，全世帯と比べて小さいように感じられる．

参考のため，大竹（2005）の推定結果と比較してみよう．表5は，大竹（2005）の第1章表1-1を抜粋転載したものである．データは，本稿で用いたデータと同様，総務省統計局家計調査5分位階級データ，全世帯および勤労者世帯である．また，ジニ係数は，再配分を考慮したデータを用いて，記述統計によって算出している．

大竹（2005）において示されているジニ係数は，全世帯においては0.27-0.28，勤労者世帯においては0.22-0.23となっている．本稿で推定された値に比較し全体的に小さな値となっている．これには，大きく2つの原因が考えられる．ひとつは，大竹（2005）では，再配分を考慮してジニ係数を計算しているので，再分配を考慮する前よりジニ係数の値が小さく計測されることによるものである．もう一つは，パレート分布をあてはめたことによって，真の分布とは異なった分布で推定されたことによるバイアスによるものであると考えられる．また，後者の影響の方が大きいということも予想されるが，所得分配の不平等度の傾向について注目すると，本稿での推定結果と同様，2000年をピークに，所得分配の不平等度は縮小を始めていることがわかる．以上のことから総合的に判断すると，本稿で提案している推定方法は，旧来のジニ係数と比較すると若干バイアスを含んでいる可能性があるが，推定値の推移については，旧来のジニ係数の変動特性を周到していると考えられる．つまり，不平等度の傾向を見るとい

NII-Electronic Library Service

The Japan Statistical Society

うことに注目すれば，有用な尺度となりうる．

5.2 家計調査における格差の変化の検定

従来の分析では，以上のようにジニ係数の変化を時間とともに追うことに終始している．本稿では，経済格差が変化したかについての検定方法を提案した．以下では，異時点間のジニ係数の変化について有意性検定を行う．これにより，どの時点における変化が重要であったかを検討することが可能となる．検定の結果および統計量は，表6に示してある．前節で，検定は$\hat{\sigma}_t-\hat{\sigma}_{t-1}$が漸近的に正規分布に従うことを利用することを説明した．

全世帯についての検定統計量は表6の前半に示している．ジニ係数は，2000年までは増大傾向にあり，ジニ係数の変化の値は正値となり，2000-2005年においては負値となっている．検定統計量によると，2000年までの中でも，1990-1995年において，3.036という値をとり，これは，1%有意水準で有意に変化している．すなわち，1985年から2000年にかけての期間の格差拡大期において，バブル崩壊後の1990-1995年の期間のみが有意な変化であるといえる．また，2000-2005年の期間においては，ジニ係数の変化は負であり，検定統計量は−3.002なので，1%有意水準で帰無仮説を棄却する．この結果は，大竹（2005）によって導かれている結論を，統計的に支持する結果である．

勤労者世帯についての検定統計量は，表6の後半に示している．勤労者世帯に注目すると，前述したように1985-2005年の期間において上下動している．検定統計量を見ると，その中でも，1995-2000年の期間において，2.456となり，5%有意水準で有意に変化したといえる．その他の期間は統計的に有意な変化はないという結論を得た．この有意な期間は，所謂，『ネットバブル期』であり，直感的にも富裕層が台頭してきた時期と重なる．一方，全世帯の推定結果とは異なり，2000-2005年の期間ではジニ係数自体は下がっているものの，統計的に有意な格差の縮小は見られず，少なくとも格差が拡大しているという主張は，本稿の推定結果からは導かれなかった．以上まとめると，バブル後の期間において統計的に有意な格差の拡大が見られるものの，近年特に格差が広がっているというマスコミでの主張を支える証拠は得られなかった．

最後に，尤度比検定ににより格差が変化したかをみた．なお，対数尤度関数には（3.13）と（3.16）のどちらを使っても尤度比の数値は同じものが得られたので表は1つだけである．

表7にその結果が掲げられている．格差が変化したと判断されるのは，***(有意水準1%)および**(有意水準5%）の印が検定統計量の値に付されている時である．すなわち，全世帯

表5 大竹（2005）の第1章表1－1より抜粋したジニ係数

年	1985	1990	1995	2000	2003
全世帯	0.272	0.276	0.281	0.284	0.271
勤労者世帯	0.227	0.227	0.229	0.234	0.231

表6 格差の変化の検定（全世帯および勤労者世帯）

	全世帯				勤労者世帯			
年	ΔG	$\Delta\hat{\sigma}$	標準誤差	検定統計量	ΔG	$\Delta\hat{\sigma}$	標準誤差	検定統計量
1985-1990	0.0092	0.0076	0.013188	0.576	0.0043	0.00418	0.011162	0.375
1990-1995	0.0522	0.0414	0.013645	3.036***	−0.0025	−0.00237	0.011178	−0.212
1995-2000	0.0166	0.0126	0.014139	0.893	0.0294	0.02804	0.011418	2.456**
2000-2005	−0.0535	−0.0417	0.013876	−3.002***	−0.0014	−0.00129	0.011662	−0.111

ΔG ：ジニ係数の変化，$\Delta\hat{\sigma}=\hat{\sigma}_t-\hat{\sigma}_{t-1}$，***(1%有意)，**(5%有意)．

NII-Electronic Library Service

The Japan Statistical Society

表7　尤度比検定（全世帯および勤労者世帯）

年	全世帯	勤労者世帯
1985-1990	0.342	0.144
1990-1995	9.164***	0.048
1995-2000	0.832	6.014**
2000-2005	8.908***	0.012

***(1%有意)，**(5%有意)．

では1990-1995年および2000-2005年の2つの期間で格差が変化し，勤労者世帯では，1995-2000年の1つの期間で格差が変化したと判断できる．これらの結果は表6における検定の結果と整合的である．

6. 結論と今後の課題

本稿では2パラメータの第一種のパレート分布を用いることで所得分布を推定した．分位データを用いた推定法は最適化を用いないので，推定は比較的容易である．そのため実証分析家にとっては利用しやすいものと考えられる．また，発展途上国においては，所得分布に関するマイクロデータが利用可能でなく，分位データだけしか利用できない状況がありうる．そうした場合にも有用であるだろう．さらに，推定量の漸近正規性を用いることで，格差が異なっているかの検定を正規分布の平均の差の検定を使って検定できることを示した．さらに最尤推定を利用した尤度比検定も示した．これまでは記述統計で計算されたジニ係数では変化が有意か判断できなかったが，これらの検定で統計的に格差の変化を検定できるようになった．実際に，総務省統計局の家計調査を使った実証分析では記述統計のジニ係数で観察される格差の傾向をジニ係数の推定および格差の変化の検定の両面より裏付けることができた．

しかしながら，課題も残っている．本稿では，2パラメータのパレート分布によって所得分布の推定を行ったので，精密に所得分布を推定することには限界がある．このことは，ジニ係数の値が記述統計で計算されたジニ係数に比べ大きめになった理由の一つとして指摘した．パレート分布のパラメータを増やすなどして，分布を複雑にすることで，ジニ係数の推定をより高い精度で行えるかもしれない．他にも，共分散行列に系列相関を入れて時系列の構造を考えることもできうると考えられる．これらのことは今後の検討課題であるが，格差の拡大（縮小）を簡便な方法で統計的に検定できるようになったことは，この分野における第一歩となるであろう．

謝　辞

本論文の元となる論文は2008年3月に金沢で開催された「複雑現象のモデル化と統計理論的発展」コンファレンスにおいて報告され，さらに，2008年5月の日本経済学会春季大会において報告された．それらの報告においては，福重元嗣大阪大学教授，星野伸明金沢大学准教授，矢島美寛東京大学教授，村澤康友大阪府立大学教授，清野一治早稲田大学教授，吉田建夫岡山大学教授などの参加者から有益なコメントをいただいた．さらに，統計学会誌に投稿するにあたり2名の査読者からも有益なコメントをいただいた．記して感謝を申し述べたい．なお，本研究は科研費（20730144）および（19730157）の助成を受けたものである．

参　考　文　献

Atoda, N., T. Suruga and T. Tachibanaki (1988). Statistical Inference of Functional Form for Income Distribution, *Economic Studies Quarterly*, **39**, 14-40.

NII-Electronic Library Service

The Japan Statistical Society

Atkinson, A. B. (1970). On the Measurement of Inequality, *Journal of Economic Theory*, **2**, 244–263.
Brazauskas, V. and R. Serfling (2000). Robust Estimation of Tail Parameters for Two-Parameter Pareto and Exponential Models via Generalized Quantile Statisitcs, *Extremes*, **3**, 231–249.
Chotikapanich, D. andW.E. Griffiths (2005). Averaging Lorenz curves *Journal of Income Inequality*, **3**, 1–19.
David, H. A. and H.N. Nagaraja (2003). *Order Statistics* (3rd. ed.), Wiley, New York.
Deaton, A. and C. Paxson (1994). Intertemporal Choice and Inequality, *Journal of Political Economy*, **102**, 437–467.
Doornik, J. A. (2006). *Ox: Object Oriented Matrix Programming Language*, Timberlake Consultants Press, London.
Dorfman, R. (1979). A Formula for the Gini Coefficient, *The Review of Economics and Statistics*, **61**, 146–49.
Hasegawa, H. and H. Kozumi (2003). Estimation of Lorenz curves: A Bayesian nonparametric approach, *Journal of Econometrics*, **115**, 277–291.
Johnson, N. L., S. Kotz and N. Balakrishnan (1994). *Continuous Univariate Distributions* (2nd ed.), **1**, Wiley, New York.
Kleiber, C and S. Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley, New York.
Koutrouvelis, I. A. (1981). Large-Sample Quantile Estimation in Pareto Laws, *Communications in Statistics-Theory and Methods*, A, **10**, 189–201.
McDonald, J. B. (1984). Some Generalized Functions for the Size Distribution of Income, *Econometrica*, **52**, 647–663.
McDonald, J. B. and M.R. Ransom (1979). Functional Forms, Estimation Techniques and the Distribution of Income, *Econometrica*, **47**, 1513–1525.
Mosteller, F. (1946). On Some Useful "Inefficient" Statistics, *Annals of Mathematical Statisitcs*, **17**, 377–408.
Murasawa, Y. and K. Morimune (2004). Distribution-Free Statistical Inference for Generalized Lorenz Dominance Based on Grouped Data, *Mathematics and Computers in Simulation*, **64**, 133–142.
Ryu and Slottje (1999). Parametric approximations of the Lorenz curve, *Handbook of Income Inequality Measurement*, (eds Silber, J.), Kluwer Academic Publisher, Boston, 291–312.
Saleh, A. K. Md. E. and Ali, M. M. (1966). Asymptotic Optimum Quantiles for the Estimation of the Parameters of the Negative Exponential Distribution, *Annals of Mathematical Statisitcs*, **37**, 143–151.
Singh, S. K. and G. S. Maddala (1976). A Function for Size Distribution of Incomes, *Econometrica*, **44**, 963–970.
Tachibanaki, T. Suruga and N. Atoda (1997). Estimation of Income Distribution Parameters for Individual Observations by Maximum Likelihood Method, *Jornal of the Japan Statistical Society*, **27**, 191–203.
大竹文雄 (1994).「1980 年代の所得・資産分配」『季刊理論経済学』, **45**, 385–402.
大竹文雄 (2000).「90 年代の所得格差」『日本労働研究雑誌』, **480**, 2–11.
大竹文雄 (2005).『日本の不平等』日本経済新聞社.
大竹文雄・齊藤誠 (1996).「人口の高齢化と消費の不平等度」『日本経済研究』, **33**, 11–35.
橘木俊昭 (1998).『日本の経済格差』岩波書店.

註

1）所得水準と不平等度の両方を考慮して，所得分布が改善したか否かを検討するローレンツ・ドミナンス（優越）を検定する方法は提案されているものの（例えば，Murasawa and Morimune (2004) など），日本の格差拡大に関する議論は，ジニ係数が拡大したか縮小したかである．

2）漸近分散の解析値については，(3.6)～(3.11) 式において，$n=10000$，$L=0.58809$ を代入することで，表1中の値が得られる．

NII-Electronic Library Service