論文の内容の要旨

グループデータからの所得分布の推定

西埜晴久

1　はじめに

経済格差・不平等度は日本経済においても大きな論点であり，不平等度の変化はこれまでは記述統計的に計算されたジニ係数の変化によって議論されていた．また，個票データや階級別データ（グループデータ）を使ってローレンツ曲線を推定する研究はいくつかある．本研究でははじめからパラメトリックな分布を仮定して所得分布を推定することを考える．こうすることで不平等度の変化の検定や時系列モデルへの応用が行いやすくなるという利点がある．また，利用しやすいグループデータを用いることを前提とする．まず，1 章では，2 章以降の準備として日本の所得データおよび所得分布に使われる確率分布などを説明する．

2　パレート分布を用いた経済格差の検定

2 章のパレート分布の推定の話は，Koutrouvelis (1981) の議論そのままであり，部分順序統計量を使った SOE(selected order statistics based estimator) は陽表的に得ることができる．そして，この SOE を用いて経済格差が変化したかを検定する方法を提案する．なお，Koutrouvelis (1981) はパレート分布の確率変数を対数変換すると指数分布になることを利用して，指数分布の部分順序統計量の漸近正規近似を使った推定をパレート分布へと応用していた．指数分布の部分順序統計量にもとづく推定の研究は 1960 年代に盛んに行われていたが，Saleh and Ali (1966) では，この推定量を ABLUE(Asymptotically BLUE) としている．第一種のパレート分布 $P(I)(\theta, \alpha)$ はパラメータが $(\theta,\, \alpha)$ と 2 つある．パレート分布のジニ係数 G は $G = 1/(2\alpha - 1)$ となるので，ジニ係数は形状パラメータ α だけによって決まる．

そして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，$1 \le n_1 < n_2 < \ldots < n_k \le n$ に対し，$\{X(n_1) \le X(n_2) \le, \cdots, \le X(n_k)\}$ 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，計算の便宜上 $u_i = -\log(1-p_i)$ の表記を導入する．確率変数 X がパレート分布 $P(I)(\theta, \alpha)$ 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 $f(y) = (1/\sigma)\exp\{-(y-\mu)/\sigma\}$，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，不平等度についての情報は形状パラメータ σ に集約される．

(μ, σ) の推定量は Koutrouvelis (1981) より，

$$\hat{\mu} = \log X(n_1) - \hat{\sigma} u_1, \tag{1}$$

$$\hat{\sigma} = \sum_{i=1}^{k} b_i \log X(n_i) \tag{2}$$

となる．なお,

$$
\begin{aligned}
b_1 &= -\frac{1}{L}\frac{u_2-u_1}{e^{u_2}-e^{u_1}}, \\
b_i &= \frac{1}{L}\left[\frac{u_i-u_{i-1}}{e^{u_i}-e^{u_{i-1}}}-\frac{u_{i+1}-u_i}{e^{u_{i+1}}-e^{u_i}}\right], \quad (i=2,\dots,k-1) \\
b_k &= \frac{1}{L}\frac{u_k-u_{k-1}}{e^{u_k}-e^{u_{k-1}}}, \\
L &= \sum_{i=2}^{k}\frac{(u_i-u_{i-1})^2}{e^{u_i}-e^{u_{i-1}}},
\end{aligned}
$$

である．したがって，(1) と (2) から，$(\hat{\mu},\hat{\sigma})$ は陽表的に計算できることが分かる．また，2 章ではモンテカルロ実験によってこの推定量が最尤推定量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．$\hat{\sigma}$ は標本の大きさ n が十分に大きいと漸近正規性が成り立つ．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,

$$
\hat{\sigma}_t \sim \mathcal{N}\left(\sigma_t, \frac{\sigma_t^2}{L_t n_t}\right), \quad \hat{\sigma}_s \sim \mathcal{N}\left(\sigma_s, \frac{\sigma_s^2}{L_s n_s}\right), \tag{3}
$$

と近似できる．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できて, 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,

$$
Z = \frac{\hat{\sigma}_t-\hat{\sigma}_s}{\sqrt{\frac{\hat{\sigma}_t^2}{L_t n_t}+\frac{\hat{\sigma}_s^2}{L_s n_s}}} \sim \mathcal{N}(0,1), \tag{4}
$$

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．そこで，検定統計量 Z に対し標準正規分布により棄却域を設定して検定を構成できる．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い，その結果，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．

3　対数正規分布を用いたジニ係数のグループデータ推定と検定

2 章のパレート分布を対数正規分布に変えるというアイデアが 3 章の議論である．指数分布は分位点が陽表的に得られるが，現在では，正規分布の分位関数も容易に求めることができる．また現実の所得データでは対数正規分布の当てはまりが良いことを考えれば，対数正規分布を考えることは有用であると考えている．対数正規分布 $LN(\mu,\sigma)$ のジニ係数 G は $\Phi(\cdot)$ を標準正規分布の分布関数として，$G=2\Phi\left(\sigma/\sqrt{2}\right)-1$, と計算される．ジニ係数 G はパラメータ σ だけに依存し，$G(\sigma)$ は σ の単調増加関数である．そこで不平等度の変化に興味があれば σ だけを推定すればよいことになる．

以下でグループデータからの部分順序統計量に基づいて得られる推定量を selected order statistics にもとづく推定量 (SOE) と呼ぶ．順序統計量の漸近理論によって $\lim_{n\to\infty} n_i/n = p_i$ を満たすとき，部分順序統計量 $(\log X(n_1), \log X(n_2), \dots, \log X(n_k))'$ の同時分布が $n\to\infty$ の時，平均 $(\mu+\sigma u_1, \mu+\sigma u_2, \dots, \mu+\sigma u_k)'$，分散共分散行列 $(\sigma^2/n)\mathbf{W}$ の多変量正規分布

になる．分散共分散行列 $\mathbf{W}$ の (i,j) 成分は

$$\begin{aligned} W_{ii} &= \frac{p_i(1-p_i)}{\varphi(\Phi^{-1}(p_i))^2} \\ W_{ij} &= W_{ji} = \frac{p_i(1-p_j)}{\varphi(\Phi^{-1}(p_i))\varphi(\Phi^{-1}(p_j))} \quad (i<j), \quad i,j=1,2,\ldots,k, \end{aligned} \tag{5}$$

であり，$\varphi(x)$ は標準正規分布の確率密度関数を表わす．$\mathbf{y}=(\log X(n_1), \log X(n_2), \ldots, \log X(n_k))'$ の同時分布が漸近的に多変量正規分布で近似できることより，

$$\log X(n_i) = \mu + \sigma u_i + \varepsilon_i, \quad (i=1,\ldots,k), \tag{6}$$

と書ける．ただし，$(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)'$ は近似的に平均 $(0,0,\ldots,0)'$，分散共分散行列 $(\sigma^2/n)\mathbf{W}$ の多変量正規確率ベクトルである．したがって，(6) は $(\sigma^2/n)\mathbf{W}$ の分散の誤差項を持つ正規線形回帰モデルとみなすことができる．$\mathbf{W}$ が既知であるので，一般化最小 2 乗法 (GLS) によって，パラメータ (μ,σ) を推定することができる．被説明変数ベクトルを $\mathbf{y}$ として，説明変数 $\mathbf{Z}$ を

$$\mathbf{Z} = \begin{pmatrix} 1 & u_1 \\ 1 & u_2 \\ \vdots & \vdots \\ 1 & u_k \end{pmatrix}.$$

とし，$\gamma=(\mu,\sigma)'$ とすると，線形回帰モデルは

$$\mathbf{y} = \mathbf{Z}\gamma + \boldsymbol{\epsilon}, \tag{7}$$

と書ける．ただし，$\boldsymbol{\epsilon}=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_k)'$ とする．すると $\hat{\gamma}$ の GLS 推定量は，

$$\hat{\gamma} = \left(\mathbf{Z}'\mathbf{W}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{W}^{-1}\mathbf{y}, \tag{8}$$

となる．$\hat{\gamma}=(\hat{\mu},\hat{\sigma})'$ は明示的に得られ，これを SOE とする．

さらに，経済格差の変化を検定するために，(8) の GLS 推定量である SOE の漸近正規性を用いた検定を提案する．最後にこれらの推定および検定を日本の家計調査データに対して応用する．

4 対数正規 SV モデルによる持続的な所得不平等度のベイズ推定

そして，3 章で導かれた部分順序統計量を漸近正規近似することで得られた線形モデルに確率的ボラティリティ(SV) モデルをいれることで時系列構造を入れたモデルが 4 章の話になる．この対数正規分布にもとづく SV モデルはジニ係数を含む不平等度を含むモデルであり，不平等度の持続性を考慮したモデルになる．

まず，先に述べた (6) は線形回帰モデルとみなすことができ，誤差分散行列 $(\sigma^2/n)W$ は既知である．そこで，ウェイト行列 W が既知なので一般化最小 2 乗法を用いて推定できる．所得不平等度の持続性を把握するために，部分順序統計量にもとづく線形回帰モデル (6) を動学モデルへと拡張する．これは以下の SV モデルで表される．

$$\begin{aligned} y_{i,t} &= \mu_t + \sigma_t u_i + \sigma_t \eta_{i,t}, & (9) \\ \sigma_t &= \exp(h_t/2) & (10) \\ h_t &= \omega + \phi(h_{t-1}-\omega) + \xi_t, & (11) \end{aligned}$$

ただし，

$$
\begin{aligned}
&y_{i,t} = \ln(X(n_i)), \\
&\boldsymbol{\eta}_t = (\eta_{1,t}, \eta_{2,t}, \ldots, \eta_{k,t})' \sim \mathcal{MVN}(\mathbf{0}, W/n), \\
&\xi_t \sim \mathcal{N}(\mathbf{0}, \sigma_\xi^2).
\end{aligned}
$$

である．$X(n_i)$ が対数正規分布に従う場合に，$u_i = \Phi^{-1}(p_i)$ とし，W_{ij} は (5) で定義され，$\Phi(\cdot)$ は標準正規分布の累積分布関数とする．(9) の μ_t は SV モデルでは各 t で別々に推定され，不平等度に影響する σ_t は (10) により時変構造をもつ．σ_t からジニ係数を得ることができる．(9), (10) および (11) による SV モデルをマルコフ連鎖モンテカルロ (MCMC) 法によって推定する．

日本のデータを用いて，SV モデルをマルコフ連鎖モンテカルロ (MCMC) 法によって推定し，周辺尤度を用いたモデル比較を行った．その結果，SV モデルといった時系列の構造が入ったモデルが，そうでないモデルよりも選択された．つまり，このことは不平等度の持続性を示している．

5 一般化ベータ分布について

本章では McDonald and Xu (1995) で提案された 5 つのパラメータをもつ一般化ベータ分布について，そのジニ係数を数値計算を用いて計算することと MCMC 法によるベイズ推定を行い，日本の家計調査データを用いて実際に推定を行った．

6 今後の課題

本研究の今後の課題として，以下のようなことを考えている．第一に，グループデータの同時密度を尤度として MCMC によるベイズ推定を行う．第二に，3 つ以上のパラメータを持つより複雑な分布の推定である．第三に，局所モーメントの情報を用いた推定も今後の研究課題として考えている．第四に，さらには，経済の不平等および所得格差はどのように生じるのかといったことも今後に追究すべき問題であると考えている．

References

[1] Koutrouvelis, I. A. (1981). Large-Sample Quantile Estimation in Pareto Laws, *Communications in Statistics-Theory and Methods, A*, **10**, 189–201.

[2] McDonald, J. B. and Y. J. Xu (1995). A generalization of the beta distribution with applications, *Journal of Econometrics*, **66**, 133–152.

[3] Saleh, A. K. Md. E. and M. M. Ali (1966). Asymptotic Optimum Quantiles for the Estimation of the Parameters of the Negative Exponential Distribution, *Annals of Mathematical Statisitcs*, **37**, 143–151.