

順序統計量からのジニ係数の推定

経済学研究科博士後期課程

竹内 俊子

1 はじめに

最近、日本の所得分配の構造変化に関する議論が再びなされるようになってきた。(たとえば、橋木 [10]。) そのとき、所得分布の不平等をどのように測るかは大きな問題になろう。不平等を測る指標の種類はいろいろあり、ローレンツ曲線、ジニ係数、平方変動係数、平均対数偏差、アトキンソン尺度等がある。こうした不平等尺度の定義、解釈、性質については高山 [9]、豊田 [11] [12]、青木 [8] 等に詳細な説明がある。その中でも、直感的でわかりやすいこと、いろいろなよい性質を持つことによって、ジニ係数が多く用いられている。

ところで、ジニ係数を求めるためには所得データが必要である。信頼性の高いジニ係数を求めるためには多くの標本が必要となる。しかし、所得分布を表す個票データは、個人ではなかなか入手できない。一般的に、得ることができるのは集約された統計量、たとえば、5分位・10分位等の分位データである。このような集約された統計量により求められた不平等尺度の精度はどの程度信頼性があるだろうか。不平等尺度の精度の問題がここで生じる。

ジニ係数はその背後に分布型を考えない記述的な尺度として一般的に捉えられているが、分布型を仮定し、その性質を調べることもできる。ところで、所得分布は通常、歪んだ分布であることが多い。したがって、等分位に基づくデータよりも順序統計量をいくつか用いることにより精度の高い推定量が得られる可能性がある。そこで、本論文では順序統計量からジニ係数を推定し、その精度について考察する。

最初に、ジニ係数や順序統計量に関するいくつかの定義、命題等を述べ、次に、背後にある分布としてパレート分布を仮定し、順序統計量を用いてジニ係数の最尤推定量を求めることを考える。その際、(1)最小値と最大値、(2)任意の順序統計量、(3)分位データの3つの場合について考察する。また、(1)、(2)に関しては求めた最尤推定量の精度をブートストラップ法を用いて評価する。

2 ジニ係数

ジニ係数はその背後に分布型を仮定することなく計算できる記述的な尺度であるが、分布型を仮定し、その性質を明らかにすることもできる。まず、ジニ係数に関する定義を述べよう。

X を確率変数とする。

定義1 ジニ係数 (Gini index)

$$\gamma(X) = \frac{\frac{1}{2} E[|X_1 - X_2|]}{E[X]} \quad (1)$$

X_1, X_2 は X と同一の確率法則に従う互いに独立な確率変数である。

上の定義において X は確率変数として扱われ、何らかの分布型が仮定されているが、 X の背後に分布型を考えず、ジニ係数を単に記述統計量として扱う場合について定義する。

同一分布に従う確率変数の n 個の実現値を x_1, x_2, \dots, x_n としたとき、これに対応する順序統計量を $x_{1:n}, x_{2:n}, \dots, x_{n:n}$ と表すことにする。

定義2 階級データ 標本の数を n 、階級の数を q 、各階級内の標本の数を n_1, n_2, \dots, n_q とすると、 i 番目の階級データを

$$X_{(i)} = \frac{\sum_{j=N_{i-1}+1}^{N_i} x_{j:n}}{n_i} \quad \text{for } i = 1, 2, \dots, q \quad (2)$$

と定義する。ここで、 $N_i = \sum_{j=1}^i n_j$ であり、便宜上 $N_0 = 0$ とする。

定義3 階級データに基づくジニ係数 階級データに基づくジニ係数は、定義1に従い次のように定義される。

$$G(X) = \frac{\frac{1}{2} \frac{1}{q^2} \sum_{i=1}^q \sum_{j=1}^q |X_{(i)} - X_{(j)}|}{\frac{1}{q} \sum_{i=1}^q X_{(i)}} \quad (3)$$

また、階級内の標本数がすべて等しいとき、 $n_1 = \dots = n_q$ の場合には分位データと呼ぶ。

3 順序統計量に関するいくつかの命題

順序統計量についての命題をいくつか述べよう。

同一分布に従う n 個の確率変数を X_1, X_2, \dots, X_n とする。これらを小さい順に並びかえたものを順序統計量といい、 $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ と表す。すなわち、最小値が $X_{1:n}$ 、最大値が $X_{n:n}$ である。

命題 1 順序統計量の分布関数・密度関数 順序統計量 $X_{i:n}$ の分布関数は、

$$\begin{aligned} F_{i:n}(x) &= P(X_{i:n} \leq x) \\ &= \sum_{r=i}^n \binom{n}{r} \{F(x)\}^r \{1-F(x)\}^{n-r} \end{aligned} \quad (4)$$

で与えられる。

密度関数は、

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!} \{F(x)\}^{i-1} \{1-F(x)\}^{n-i} f(x) \quad (5)$$

で与えられる。

命題 1 では、単一の順序統計量の分布関数・密度関数について述べたが、2つの順序統計量の同時分布関数・密度関数について次の命題 2 で述べる。

命題 2 2つの順序統計量の同時分布関数・密度関数 順序統計量 $X_{i:n}$ と $X_{j:n}$ ($1 \leq i \leq j \leq n$) の同時分布関数は、

$$F_{i,j:n}(x_i, x_j) = \sum_{s=j}^n \sum_{r=i}^s \frac{n!}{r!(s-r)!(n-s)!} \{F(x_i)\}^r \{F(x_j) - F(x_i)\}^{s-r} \{1-F(x_j)\}^{n-s} \quad (6)$$

で与えられる。

同時密度関数は、

$$\begin{aligned} f_{i,j:n}(x_i, x_j) &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \\ &\quad \times \{F(x_i)\}^{i-1} \{F(x_j) - F(x_i)\}^{j-i-1} \{1-F(x_j)\}^{n-j} \\ &\quad \times f(x_i) f(x_j) \end{aligned} \quad (7)$$

で与えられる。

特に最小値 ($X_{1:n}$) と最大値 ($X_{n:n}$) の同時密度関数は、

$$f_{1,n:n}(x_i, x_j) = n(n-1)f(x_i)f(x_j)\{F(x_j)-F(x_i)\}^{n-2} \quad (8)$$

となる。

2つの順序統計量の同時分布関数・密度関数の考え方を拡張して、 m 個の順序統計量の同時分布関数・密度関数について命題3で述べる。

命題3 m 個の順序統計量の同時分布関数・密度関数 順序統計量 $X_{i_1:n}, \dots, X_{i_m:n}$ ($1 \leq i_1 < \dots < i_m \leq n$) の同時分布関数は、

$$F_{i_1, \dots, i_m:n}(x_{i_1}, \dots, x_{i_m}) = \sum_{u=i_m}^n \dots \sum_{s=i_2}^i \sum_{r=i_1}^s \frac{n!}{r!(s-r)! \dots (n-s)!} \\ \times \{F(x_{i_1})\}^r \{F(x_{i_2})-F(x_{i_1})\}^{s-r} \dots \\ \times \{1-F(x_{i_m})\}^{n-u} \quad (9)$$

で与えられる。

同時密度関数は、

$$f_{i_1, \dots, i_m:n}(x_{i_1}, \dots, x_{i_m}) = \frac{n!}{(i_1-1)!(i_2-i_1-1)! \dots (n-i_m)!} \\ \times \{F(x_{i_1})\}^{i_1-1} \{F(x_{i_2})-F(x_{i_1})\}^{i_2-i_1-1} \dots \\ \times \{1-F(x_{i_m})\}^{n-i_m} \\ \times f(x_{i_1})f(x_{i_2}) \dots f(x_{i_m}) \quad (10)$$

で与えられる。

特に $m=n$ の場合は、

$$f_{1,2,\dots,n:n}(x_1, x_2, \dots, x_n) = n! \prod_{i=1}^n f(x_i) \quad (11)$$

となる。

次に、順序統計量の関数の密度関数がどのように表されるかを考えてみよう。

命題4 順序統計量の関数の密度関数 順序統計量 $X_{i_1:n}, \dots, X_{i_m:n}$ ($1 \leq i_1 < \dots < i_m \leq n$) の関数 $Z = \sum_{r=i_1}^{i_m} a_r X_{r:n}$ の密度関数は、

$$f(z) = \int_{x_{i_{m-1}}}^{\infty} \dots \int_1^{\infty} f(z_{i_1}, \dots, z_{i_m}) dz_{i_2} \dots dz_{i_m} \quad (12)$$

順序統計量からのジニ係数の推定 (竹内)

となる。ここで、

$$\begin{aligned}
 f(z_{i_1}, \dots, z_{i_m}) &= \frac{n!}{(i_1-1)!(i_2-i_1-1)! \cdots (n-i_m)!} \\
 &\times \left\{ F\left(\frac{z_{i_1} - \sum_{r=i_2}^{i_m} a_r z_r}{a_{i_1}}\right) \right\}^{i_1-1} \{F(z_{i_2}) - F(z_{i_1})\}^{i_2-i_1-1} \cdots \\
 &\times \{1 - F(z_{i_m})\}^{n-i_m} \\
 &\times f\left(\frac{z_{i_1} - \sum_{r=i_2}^{i_m} a_r z_r}{a_{i_1}}\right) f(z_{i_2}) \cdots f(z_{i_m}) \cdot \frac{1}{a_{i_1}}
 \end{aligned} \tag{13}$$

である。

(証明)

$$\begin{aligned}
 z_{i_1} &= \sum_{r=i_1}^{i_m} a_r x_{r:n} \\
 z_{i_2} &= x_{i_2:n} \\
 &\vdots \\
 z_{i_m} &= x_{i_m:n}
 \end{aligned}$$

とすると、

$$\begin{aligned}
 x_{i_1} &= \frac{z_{i_1} - \sum_{r=i_2}^{i_m} a_r z_r}{a_{i_1}} \\
 x_{i_2} &= z_{i_2} \\
 &\vdots \\
 x_{i_m} &= z_{i_m}
 \end{aligned}$$

となる。ヤコビアンは、

$$\begin{aligned}
 |J| &= \begin{vmatrix} \frac{\partial x_{i_1:n}}{\partial z_{i_1}} & \cdots & \frac{\partial x_{i_1:n}}{\partial z_{i_m}} \\ \vdots & & \vdots \\ \frac{\partial x_{i_m:n}}{\partial z_{i_1}} & \cdots & \frac{\partial x_{i_m:n}}{\partial z_{i_m}} \end{vmatrix} \\
 &= \begin{vmatrix} \frac{1}{a_{i_1}} & -\frac{a_{i_2}}{a_{i_1}} & \cdots & -\frac{a_{i_m}}{a_{i_1}} \\ & 1 & & 0 \\ & & \ddots & \\ 0 & & & 1 \end{vmatrix} \\
 &= \frac{1}{a_{i_1}}
 \end{aligned}$$

となる。式(10)より、 Z_{i_1}, \dots, Z_{i_m} の同時密度関数は、

$$\begin{aligned} f(z_{i_1}, \dots, z_{i_m}) &= \frac{n!}{(i_1-1)!(i_2-i_1-1)! \dots (n-i_m)!} \\ &\times \left\{ F\left(\frac{z_{i_1} - \sum_{r=i_2}^{i_m} a_r z_r}{a_{i_1}}\right) \right\}^{i_1-1} \{F(z_{i_2}) - F(z_{i_1})\}^{i_2-i_1-1} \dots \\ &\times \{1 - F(z_{i_m})\}^{n-i_m} \\ &\times f\left(\frac{z_{i_1} - \sum_{r=i_2}^{i_m} a_r z_r}{a_{i_1}}\right) f(z_{i_2}) \dots f(z_{i_m}) \cdot \frac{1}{a_{i_1}} \end{aligned}$$

となり、求めたい関数 Z の密度関数は、

$$\begin{aligned} f(z) &= f(z_{i_1}) \\ &= \int_{x_{i_m-1}}^{\infty} \dots \int_1^{\infty} f(z_{i_1}, \dots, z_{i_m}) dz_{i_1} \dots dz_{i_m} \end{aligned}$$

となる。

(証明終)

4 順序統計量に基づく推定

所得分布として考えられる分布型はいくつもあるが、ここでは特にパレート分布^{注1}をとりあげる。

4.1 パレート分布

所得分布におけるパレート法則は、1897年に発表されたPareto[7]までさかのぼり、その法則は次のように表せる。

$$N = Ax^{-a}$$

ここで、 N は所得が x 以上的人数であり、 A 、 a はパラメータである。この法則に従う所得分布がパレート分布である。

パレート分布にはいくつかの類型^{注2}があるが、ここでは、その中でももっとも一般的に用いられている $P(I)(a, 1)$ について考えることにする。

パレート分布 $P(I)(a, 1)$ の密度関数は、

$$f(x) = ax^{-(a+1)} \quad (14)$$

分布関数は、

$$F(x) = 1 - x^{-a} \quad (15)$$

^{注1} 所得分布がパレート分布に従うことについては、Champernown[1]、Hayakawa[4]等を参照。

^{注2} Johnson, Kotz and balakrishnan[6]を参照。

順序統計量からのジニ係数の推定 (竹内)

で与えられる。

4.2 ジニ係数の最尤推定

所得分布にパレート分布を仮定した場合のジニ係数は、定義1より、

$$\gamma(X) = \frac{1}{2a-1} \quad (16)$$

となる。

そこで、ジニ係数の最尤推定量を考えてみよう。

補題1 β を θ の関数 ($\beta = f(\theta)$) とする。 $\hat{\theta}$ を θ の最尤推定量とすると、関数 β が単調関数ならば、 $\hat{\beta} = f(\hat{\theta})$ も最尤推定量である。

(証明)

$\ell(x|\beta)$ と $L(x|\theta)$ を β と θ のそれぞれの尤度関数とする。

$$\begin{aligned} \ell(x|\beta) &= L(x|\theta) \\ &= L(x|f^{-1}(\beta)) \quad \text{where } \beta = f(\theta) \\ \frac{\partial \ell}{\partial \beta} &= \frac{\partial}{\partial \beta} L(x|f^{-1}(\beta)) \\ &= \frac{\partial}{\partial \theta} L(x|\theta) \times \frac{\partial}{\partial \beta} f^{-1}(\beta) \\ &= 0 \quad \left(\frac{\partial L}{\partial \theta} = 0 \text{ より} \right) \end{aligned}$$

f^{-1} の存在のためには単調性が必要である。したがって、最尤推定量を代入した関数が単調ならばその関数も最尤推定量である。

(証明終)

\hat{a} を a の最尤推定量とする。式(16)において、ジニ係数 $\gamma(X)$ はパラメータ a に関して単調関数なので、補題1により、次の定理が得られる。

定理1 パレート分布のジニ係数の最尤推定量は、

$$\hat{G} = \frac{1}{2\hat{a}-1} \quad (17)$$

である。

次に、ジニ係数の分散を求めよう。

一般に、確率変数 X の関数 $g(X)$ の分散は、

$$\text{var}\{g(X)\} \doteq \left(\frac{dg}{dX}\right)_0^2 \text{var}X \quad (18)$$

と一次近似できる。

この結果を用いると、ジニ係数 (\hat{G}) の分散は次の通りである。

$$\text{var}(\hat{G}) \doteq \frac{4n^2a^2}{(n-1)^2(n-2)(2a-1)^4} \quad (19)$$

標本数が10000 ($n=10000$) の場合の標準偏差を求めたのが表1である。

a	ジニ係数	標準偏差
1.75	0.40	0.0056
1.93	0.35	0.0047
2.17	0.30	0.0039
2.50	0.25	0.0031
3.00	0.20	0.0024
3.83	0.15	0.0017

表1 ジニ係数の標準偏差

ジニ係数が大きくなるほど、あるいはパラメータ a が小さくなるほど標準偏差が大きくなっていることがわかる。

それでは、ジニ係数の最尤推定量を求めるために、まず、順序統計量を用いてパレート分布 $P(I)(a, 1)$ のパラメータ a の最尤推定量を求めよう。

4.3 最小値と最大値に基づく最尤推定

パレート分布 $P(I)(a, 1)$ に従う確率変数の実現値の最小値 $x_{1:n}$ と最大値 $x_{n:n}$ から最尤法で a を推定しよう。

式(8)より、尤度関数は、

$$g(X \leq x_{1:n}, Y \leq x_{n:n}) = n(n-1)a^2x^{-(a+1)}y^{-(a+1)}(x^{-a}-y^{-a})^{n-2}$$

となる。

a について偏微分し、0 とおくと、

$$\begin{aligned} & \frac{\partial}{\partial a} g(X \leq x_{1:n}, Y \leq x_{n:n}) \\ &= n(n-1)a^2 x^{-(a+1)} y^{-(a+1)} (x^{-a} - y^{-a})^{n-2} \\ & \quad \times \{2a^{-1} - \log x - \log y - (n-2)(x^{-a} - y^{-a})^{-1}(x^{-a} \log x - y^{-a} \log y)\} \\ &= 0 \end{aligned}$$

となり、 $x \neq y$ とすると、

$$2a^{-1} - \log x - \log y - (n-2)(x^{-a} - y^{-a})^{-1}(x^{-a} \log x - y^{-a} \log y) = 0$$

となる。これを a について解けばよい。

最小値と最大値のデータから求めた最尤推定値 \hat{a} の分散を求めることは困難である。そこで、最尤推定値 \hat{a} の精度 (標準誤差) をブートストラップ法を用いて評価してみよう。

ジニ係数の理論値 $G = (0.40, 0.35, 0.30, 0.25, 0.20, 0.15)$ の6つの値に対応するパレート分布 $P(I)(a, 1)$ のパラメータ $a = (1.75, 1.93, 2.17, 2.50, 3.00, 3.83)$ に対し、パレート分布 $P(I)(a, 1)$ に従う乱数を各10000個ずつ発生させた。以下の分析では、その固定させたデータにおいてブートストラップ法を適用する。

表2と表3の a は、パレート分布のパラメータ、 G はそれに対応するジニ係数の理論値であり、各 $a (G)$ の値に応じて、ブートストラップ法を適用している。ブートストラップ標本の数 (B) を25~2000とし、 a 、およびジニ係数の推定量の標準誤差のブートストラップ推定を行なった結果が表の中程である。また、 $B=2000$ 、 $a=0.05$ の場合の信頼区間 (percentile法)、発生させ固定した乱数10000個から推定した a の推定値 (\hat{a})、およびジニ係数の推定値 (\hat{G}) を表の右にまとめた。

なお、ブートストラップ標本を $\hat{\theta}^*(b), b=1 \sim B$ としたとき、表2の \hat{a} 、および表3のジニ係数は $\hat{\theta}^*(\cdot) = \frac{\sum_{b=1}^B \hat{\theta}^*(b)}{B}$ 、表2と表3の \hat{s}_{eB} は $\hat{s}_{eB} = \left[\frac{\sum_{b=1}^B |\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)|^2}{B-1} \right]^{\frac{1}{2}}$ により求めたものである。

a/G	B	25	50	100	250	500	1000	2000	信頼区間	\bar{a}
1.75	\hat{a}	1.8434	1.8284	1.8591	1.8459	1.8396	1.8362	1.8330	1.6878	1.7378
0.40	$\hat{s}e_B$	0.1748	0.1681	0.1745	0.1601	0.1545	0.1509	0.1510	2.0460	
1.93	\hat{a}	2.0315	2.0150	2.0488	2.0342	2.0273	2.0236	2.0200	1.8600	1.9151
0.35	$\hat{s}e_B$	0.1927	0.1852	0.1923	0.1764	0.1703	0.1663	0.1665	2.2547	
2.17	\hat{a}	2.1557	2.1794	2.1699	2.1714	2.1674	2.1595	2.1600	1.9753	2.0531
0.30	$\hat{s}e_B$	0.1874	0.1840	0.1751	0.1703	0.1721	0.1691	0.1710	2.4456	
2.50	\hat{a}	2.6335	2.6120	2.6558	2.6369	2.6280	2.6232	2.6186	2.4112	2.4825
0.25	$\hat{s}e_B$	0.2497	0.2401	0.2492	0.2287	0.2207	0.2156	0.2158	2.9228	
3.00	\hat{a}	3.1602	3.1344	3.1870	3.1643	3.1535	3.1478	3.1423	2.8934	2.9790
0.20	$\hat{s}e_B$	0.2997	0.2881	0.2991	0.2745	0.2649	0.2587	0.2589	3.5074	
3.83	\hat{a}	4.0380	4.0050	4.0722	4.0433	4.0295	4.0222	4.0151	3.6971	3.8065
0.15	$\hat{s}e_B$	0.3829	0.3682	0.3822	0.3507	0.3384	0.3305	0.3309	4.4816	

表2 最小値と最大値のデータから求めた a の最尤推定

表3は、ブートストラップ法を適用する際、 \hat{a} からジニ係数への変換式 $\hat{G} = \frac{1}{2\hat{a}-1}$ を用いて、ジニ係数、および標準誤差を求めたものである。

a/G	B	25	50	100	250	500	1000	2000	信頼区間	\bar{G}
1.75	ジニ係数	0.3779	0.3819	0.3736	0.3765	0.3779	0.3786	0.3796	0.3234	0.4039
0.40	$\hat{s}e_B$	0.0456	0.0444	0.0450	0.0419	0.0404	0.0395	0.0396	0.4209	
1.93	ジニ係数	0.3311	0.3345	0.3275	0.3300	0.3312	0.3318	0.3326	0.2849	0.3533
0.35	$\hat{s}e_B$	0.0387	0.0376	0.0382	0.0355	0.0343	0.0336	0.0336	0.3676	
2.17	ジニ係数	0.3055	0.3011	0.3026	0.3022	0.3029	0.3043	0.3042	0.2570	0.3219
0.30	$\hat{s}e_B$	0.0323	0.0316	0.0304	0.0297	0.0298	0.0293	0.0295	0.3389	
2.50	ジニ係数	0.2373	0.2395	0.2349	0.2365	0.2373	0.2378	0.2383	0.2064	0.2522
0.25	$\hat{s}e_B$	0.0259	0.0252	0.0256	0.0238	0.0230	0.0225	0.0225	0.2616	
3.00	ジニ係数	0.1901	0.1919	0.1883	0.1895	0.1902	0.1905	0.1909	0.1663	0.2017
0.20	$\hat{s}e_B$	0.0200	0.0195	0.0198	0.0184	0.0177	0.0174	0.0174	0.2089	
3.83	ジニ係数	0.1428	0.1441	0.1415	0.1424	0.1429	0.1431	0.1434	0.1256	0.1512
0.15	$\hat{s}e_B$	0.0145	0.0141	0.0143	0.0133	0.0128	0.0126	0.0126	0.1564	

表3 最小値と最大値のデータから求めたジニ係数の最尤推定

最小値と最大値の2つのデータからのジニ係数の推定は、精度が悪く、標準誤差 ($\hat{s}e_B$) は約0.01~0.04の大きな値である。また、標準誤差の理論値である表1よりもかなり精度が落ちている。すなわち、最小値と最大値の2つのデータからジニ係数を推定した場合には、ジニ係数の標準誤差以上の差がなければ不平等度に変化があったとは結論づけられないことになる。たとえば、ジニ係数が0.25の場合では0.0225以上の差がなければ不平等度の変化については何も言えないということである。

ブートストラップ法は、経験分布を元に行っているため、データが正しく分布型を表していないと、すなわち、外れ値などがない状態でないと、正しい結果が得られない。特に、最小値と最大値の2つのデータからジニ係数を推定するこの場合は、外れ値の影響が出やすいので、ブートストラップ法を適用する前にデータを調べる必要がある。

4.4 任意の順序統計量に基づく最尤推定

次に、パレート分布 $P(I)(a, 1)$ に従う確率変数の実現値 $x_{n_1:n}, x_{n_2:n}, \dots, x_{n_m:n}$ から最尤法で a を推定する。

式(10)より、尤度関数は、

$$\begin{aligned} g(X_1 \leq x_{n_1:n}, X_2 \leq x_{n_2:n}, \dots, X_m \leq x_{n_m:n}) \\ &= \frac{n!}{(n_1-1)!(n_2-n_1-1)! \cdots (n-n_m)!} \\ &\quad \times (1-x_1^{-a})^{n_1-1} (x_1^{-a}-x_2^{-a})^{n_2-n_1-1} \times \cdots \times (x_m^{-a})^{n-n_m} \cdot a^m \prod_{i=1}^m x_i^{-(a+1)} \end{aligned}$$

となる。

$x_0=1, x_{m+1}=\infty, n_0=0, n_{m+1}=n+1$ とすると、

$$\begin{aligned} g(X_1 \leq x_{n_1:n}, X_2 \leq x_{n_2:n}, \dots, X_m \leq x_{n_m:n}) \\ &= \frac{n!}{\prod_{i=0}^m (n_{i+1}-n_i-1)!} \prod_{i=0}^m \{(x_i^{-a}-x_{i+1}^{-a})^{n_{i+1}-n_i-1}\} \cdot a^m \prod_{i=1}^m x_i^{-(a+1)} \end{aligned}$$

となる。 a について偏微分し、0 とおくと、

$$\begin{aligned} \frac{\partial}{\partial a} g(X_1 \leq x_{n_1:n}, X_2 \leq x_{n_2:n}, \dots, X_m \leq x_{n_m:n}) \\ &= \frac{n!}{\prod_{i=0}^m (n_{i+1}-n_i-1)!} \\ &\quad \times \left[\sum_{j=0}^m \{(n_{j+1}-n_j-1) (x_j^{-a}-x_{j+1}^{-a})^{n_{j+1}-n_j-2} (x_{j+1}^{-a} \log x_{j+1} - x_j^{-a} \log x_j) \right. \\ &\quad \left. \prod_{i=0, i \neq j}^m (x_i^{-a}-x_{i+1}^{-a})^{n_{i+1}-n_i-1}\} \cdot a^m \prod_{i=1}^m x_i^{-(a+1)} \right. \\ &\quad \left. + \prod_{i=0}^m (x_i^{-a}-x_{i+1}^{-a})^{n_{i+1}-n_i-1} \{m a^{m-1} \prod_{i=1}^m x_i^{-(a+1)} - a^m \sum_{j=1}^m (x_j^{-(a+1)} \log x_j \prod_{i=1, i \neq j}^m x_i^{-(a+1)})\} \right] \\ &= \frac{n!}{\prod_{i=0}^m (n_{i+1}-n_i-1)!} \left[\prod_{i=0}^m (x_i^{-a}-x_{i+1}^{-a})^{n_{i+1}-n_i-1} \right] \left[a^m \prod_{i=1}^m x_i^{-(a+1)} \right] \\ &\quad \times \left[\sum_{j=0}^m \{(n_{j+1}-n_j-1) (x_j^{-a}-x_{j+1}^{-a})^{-1} (x_{j+1}^{-a} \log x_{j+1} - x_j^{-a} \log x_j)\} + \frac{m}{a} - \sum_{j=1}^m \log x_j \right] \\ &= 0 \end{aligned}$$

となり、 $x_j \neq x_{j+1} (j=0, \dots, m)$ とすれば、

$$\sum_{j=0}^m \{(n_{j+1}-n_j-1) (x_j^{-a}-x_{j+1}^{-a})^{-1} (x_{j+1}^{-a} \log x_{j+1} - x_j^{-a} \log x_j)\} + \frac{m}{a} - \sum_{j=1}^m \log x_j = 0$$

となる。これを a について解けばよい。

前と同様の方法で、データが10000個の場合における2000,4000,6000,8000,10000番目のデータから求めた最尤推定値 \hat{a} とジニ係数をブートストラップ法を用いて評価して

みよう。

表の見方は前と同じである。

a/G	B	25	50	100	250	500	1000	2000	信頼区間	\bar{a}
1.75	\hat{a}	1.7900	1.7910	1.7897	1.7897	1.7879	1.7886	1.7891	1.7539	1.7936
0.40	$\hat{s}e_B$	0.0211	0.0204	0.0214	0.0207	0.0214	0.0212	0.0215	1.8244	
1.93	\hat{a}	1.9726	1.9737	1.9713	1.9723	1.9703	1.9711	1.9716	1.9329	1.9766
0.35	$\hat{s}e_B$	0.0233	0.0224	0.0236	0.0229	0.0236	0.0234	0.0237	2.0106	
2.17	\hat{a}	2.2393	2.2330	2.2330	2.2344	2.2345	2.2345	2.2343	2.1872	2.2373
0.30	$\hat{s}e_B$	0.0310	0.0328	0.0298	0.0319	0.0309	0.0296	0.0291	2.2831	
2.50	\hat{a}	2.5571	2.5585	2.5553	2.5567	2.5541	2.5551	2.5558	2.5056	2.5622
0.25	$\hat{s}e_B$	0.0302	0.0291	0.0306	0.0296	0.0305	0.0303	0.0307	2.6063	
3.00	\hat{a}	3.0685	3.0703	3.0664	3.0681	3.0649	3.0661	3.0670	3.0067	3.0747
0.20	$\hat{s}e_B$	0.0362	0.0349	0.0367	0.0356	0.0367	0.0364	0.0368	3.1275	
3.83	\hat{a}	3.9209	3.9231	3.9182	3.9203	3.9163	3.9178	3.9189	3.8419	3.9287
0.15	$\hat{s}e_B$	0.0462	0.0446	0.0469	0.0454	0.0468	0.0465	0.0471	3.9963	

表4 2000,4000,6000,8000,10000番目のデータから求めた a の最尤推定

表5は、ブートストラップ法を適用する際、 \hat{a} からジニ係数への変換式 $\hat{G} = \frac{1}{2\hat{a}-1}$ を用いて、ジニ係数、および標準誤差を求めたものである。

a/G	B	25	50	100	250	500	1000	2000	信頼区間	\hat{G}
1.75	ジニ係数	0.3877	0.3874	0.3881	0.3878	0.3883	0.3881	0.3880	0.3775	0.3865
0.40	$\hat{s}e_B$	0.0063	0.0061	0.0065	0.0063	0.0065	0.0064	0.0065	0.3987	
1.93	ジニ係数	0.3396	0.3394	0.3399	0.3397	0.3401	0.3400	0.3398	0.3310	0.3386
0.35	$\hat{s}e_B$	0.0054	0.0052	0.0055	0.0053	0.0055	0.0054	0.0055	0.3489	
2.17	ジニ係数	0.2876	0.2886	0.2886	0.2884	0.2884	0.2884	0.2884	0.2804	0.2878
0.30	$\hat{s}e_B$	0.0051	0.0055	0.0050	0.0053	0.0051	0.0049	0.0048	0.2963	
2.50	ジニ係数	0.2431	0.2429	0.2433	0.2432	0.2435	0.2434	0.2433	0.2374	0.2425
0.25	$\hat{s}e_B$	0.0036	0.0034	0.0036	0.0035	0.0036	0.0036	0.0036	0.2493	
3.00	ジニ係数	0.1947	0.1946	0.1949	0.1947	0.1950	0.1949	0.1948	0.1903	0.1942
0.20	$\hat{s}e_B$	0.0027	0.0026	0.0028	0.0027	0.0028	0.0028	0.0028	0.1994	
3.83	ジニ係数	0.1462	0.1461	0.1463	0.1462	0.1464	0.1463	0.1463	0.1430	0.1458
0.15	$\hat{s}e_B$	0.0020	0.0019	0.0020	0.0019	0.0020	0.0020	0.0020	0.1496	

表5 2000,4000,6000,8000,10000番目のデータから求めたジニ係数の最尤推定

5つの順序統計量(2000,4000,6000,8000,10000番目)から求めたジニ係数の最尤推定の表5をみると、標準誤差は約0.002~0.007という値である。やはり、標準誤差は5つの順序統計量(2000,4000,6000,8000,10000番目)から求めた値の方が最小値と最大値から求めた値よりも小さい。さらに、理論値である表1と比べると理論値よりも精度は悪いがその差は0.0005~0.0009であり、かなり理論値に近い値である。この値ならば小数点以下2桁目に変化があれば、不平等度が変化したと結論づけられるだろう。

また、5つの順序統計量(2000,4000,6000,8000,10000番目)から求めたジニ係数の最尤推定の表5をみると、ジニ係数が増えると標準誤差も増えている。

比較のため、同じデータを用いて、定義2より5分位データを求め、さらに定義3よりジニ係数を求め、ブートストラップ法を用いて評価してみよう。結果が表6である。

a/G	B	25	50	100	250	500	1000	2000	信頼区間	\bar{G}
1.75	ジニ係数	0.3524	0.3525	0.3515	0.3525	0.3531	0.3534	0.3533	0.3410	0.3536
0.40	$\hat{s}e_B$	0.0054	0.0072	0.0072	0.0069	0.0076	0.0075	0.0076	0.3659	
1.93	ジニ係数	0.3121	0.3121	0.3114	0.3122	0.3127	0.3130	0.3129	0.3028	0.3131
0.35	$\hat{s}e_B$	0.0044	0.0060	0.0059	0.0056	0.0062	0.0061	0.0062	0.3234	
2.17	ジニ係数	0.2657	0.2658	0.2661	0.2669	0.2669	0.2671	0.2670	0.2586	0.2671
0.30	$\hat{s}e_B$	0.0049	0.0045	0.0050	0.0051	0.0051	0.0053	0.0052	0.2754	
2.50	ジニ係数	0.2273	0.2273	0.2269	0.2274	0.2278	0.2279	0.2278	0.2217	0.2279
0.25	$\hat{s}e_B$	0.0028	0.0038	0.0038	0.0035	0.0038	0.0038	0.0038	0.2345	
3.00	ジニ係数	0.1832	0.1832	0.1830	0.1833	0.1836	0.1837	0.1836	0.1791	0.1837
0.20	$\hat{s}e_B$	0.0021	0.0029	0.0028	0.0026	0.0028	0.0028	0.0028	0.1885	
3.83	ジニ係数	0.1384	0.1383	0.1382	0.1384	0.1386	0.1387	0.1386	0.1355	0.1387
0.15	$\hat{s}e_B$	0.0015	0.0021	0.0020	0.0018	0.0020	0.0019	0.0020	0.1421	

表6 5分位データから求めたジニ係数

5つの順序統計量(2000,4000,6000,8000,10000番目)からジニ係数を求めた表5と5分位データからジニ係数を求めた表6を比較すると、ほとんどの場合、5つの順序統計量からジニ係数を求めた場合の精度の方が若干ではあるがよい。

4.5 分位データに基づく最尤推定

分位データとは、順序統計量の関数である。ここでは、 q 分位データを用いて、最尤法で分布のパラメータ(パレート分布 $P(I)(a, 1)$ を仮定した場合は a である。)を推定する方法を考えよう。

q 分位データ $X_{(1)}, \dots, X_{(q)}$ をそれぞれ

$$\begin{aligned}
 X_{(1)} &= \frac{q}{n} \sum_{i=1}^{\frac{n}{q}} X_{i:n} \\
 X_{(2)} &= \frac{q}{n} \sum_{i=\frac{n}{q}+1}^{\frac{2n}{q}} X_{i:n} \\
 &\vdots \\
 X_{(q)} &= \frac{q}{n} \sum_{i=\frac{(q-1)n}{q}+1}^n X_{i:n}
 \end{aligned}$$

とする。

命題 4 より、それぞれの密度関数は、

$$\begin{aligned}
 f(z_1) &= \int_{z_{\frac{n}{q}-1}}^{\infty} \cdots \int_1^{\infty} f(z_1, \dots, z_{\frac{n}{q}}) dz_2 \cdots dz_{\frac{n}{q}} \\
 f(z_2) &= \int_{z_{\frac{2}{q}n-1}}^{\infty} \cdots \int_{z_{\frac{n}{q}+1}}^{\infty} f(z_{\frac{n}{q}+1}, \dots, z_{\frac{2}{q}n}) dz_{\frac{n}{q}+2} \cdots dz_{\frac{2}{q}n} \\
 &\vdots \\
 f(z_q) &= \int_{z_{n-1}}^{\infty} \cdots \int_{z_{\frac{q-1}{q}n+1}}^{\infty} f(z_{\frac{q-1}{q}n+1}, \dots, z_n) dz_{\frac{q-1}{q}n+2} \cdots dz_n
 \end{aligned}$$

となる。ここで、

$$\begin{aligned}
 f(z_{i_1}, \dots, z_{i_m}) &= \frac{n!}{(i_1-1)!(i_2-i_1-1)! \cdots (n-i_m)!} \\
 &\times \left\{ F\left(\frac{z_{i_1} - \sum_{r=i_2}^{i_m} a_r z_r}{a_{i_1}}\right) \right\}^{i_1-1} \left\{ F(z_{i_2}) - F(z_{i_1}) \right\}^{i_2-i_1-1} \cdots \left\{ 1 - F(z_{i_m}) \right\}^{n-i_m} \\
 &\times f\left(\frac{z_{i_1} - \sum_{r=i_2}^{i_m} a_r z_r}{a_{i_1}}\right) f(z_{i_2}) \cdots f(z_{i_m}) \cdot \frac{1}{a_{i_1}}
 \end{aligned}$$

である。

このとき、背後の分布としてパレート分布 $P(I)(a, 1)$ を仮定した場合は、 $f(x) = ax^{-(a+1)}$ 、 $F(x) = 1 - x^{-a}$ をそれぞれ代入すればよい。

尤度関数は、

$$g(Z_1 \leq X_{(1)}, Z_2 \leq X_{(2)}, \dots, Z_q \leq X_{(q)}) = f(z_1) \cdot f(z_2) \times \cdots \times f(z_q)$$

となり、この式を分布のパラメータについて偏微分し、0 とおき、そのパラメータについて解けばよい。

5 結論

ジニ係数は、背後に分布型を仮定した場合、順序統計量を用いることによって、分布のパラメータの最尤推定を通じ、その最尤推定量を求めることが可能となる。

本論文では、背後に仮定する分布型としてパレート分布をとりあげ、データの(1)最小値と最大値、(2)任意の順序統計量からパラメータ a を最尤推定する方法を述べ、ブートストラップ法によりその精度を評価した。表 7 は、ジニ係数の標準偏差(理論値)、ブートストラップ法により求めた(1)最小値と最大値と(2)5つの順序統計量(2000,4000,6000,8000,10000番目)のジニ係数の標準誤差、同じく、5分位データからブートストラップ法により求めたジニ係数の標準誤差をジニ係数が0.25の場合において比較したものである。4つの値を見てみよう。

理論値	(表 1)	0.0031
最小値と最大値	(表 3)	0.0225
5つの順序統計量	(表 5)	0.0036
5分位データ	(表 6)	0.0038

表 7 ジニ係数0.25における標準誤差

最小値と最大値から求めたジニ係数の精度(標準誤差)はかなり悪い。しかし、5つの順序統計量(2000,4000,6000,8000,10000番目)から求めたジニ係数の精度は、すべてのデータから求めた最尤推定量の理論上の精度にかなり近い値であり、5分位データから求めたジニ係数の精度よりも若干ではあるがよい。このことから所得データの公表に際し、分位データのみを表象するのではなく、順序統計量をも基本統計量として開示することは十分意味があるといえるであろう。

また、背後にある分布型として様々なものを仮定して同様に分析することもできる。その場合、順序統計量によりジニ係数の最尤推定量を理論的に導くことは一般に難しくなるが、数値計算で解決することができる。

参考文献

- [1]Champernown,D.G.(1952), "The graduation of income distribution", *Econometrica*, 20, 591-615
- [2]Efron,B. and Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman & Hall
- [3]Efron,B. (1979), "Bootstrap methods:another look at the jackknife", *Annals of Statistics*, 7, 1-26
- [4]Miyoji Hayakawa (1951), "The Application of Pareto's Law of Income to Japanese Data", *Econometrica*, 19, 2, 174-183
- [5]Kendall,M.G. and Stuart,A (1991), *The Advance Theory of Statistics*, vol.2, fifth edition, Edward Arnold
- [6]Johnson,N.L., Kotz,S. and Balakrishnan,N.(1994), *Continuous Univariate Distribution*, vol.1, second edition, John Wiley & Sons, Inc.
- [7]Pareto,V.(1896-1897), *Cours d'économie politique*, 2 vols., Lausanne: Rouge
- [8]青木昌彦(1979),『分配理論』,筑摩書房
- [9]高山憲之(1974),「所得不平等の尺度：再検討」,『国民経済』, 131, 41-69
- [10]橋木俊詔(1998),『日本の経済格差』,岩波書店

[11] 豊田敬(1975), 「所得分布の不平等度－不平等度の比較と尺度－」, 『国民経済』, 135, 15-41

[12] 豊田敬(1996), 「情報量と不平等尺度」, 『経営志林』, 32, 4, 197-201