所得の不平等度尺度のリデザイン
～カルバックライブラー情報量の試用～

馬 場　晶 悟(福岡大学大学院)　　山 崎　好 裕(福岡大学経済学部教授)　　藤 本　浩 明(福岡大学経済学部教授)

●10月25日(土)1-A分科会 I:「経済システム分析」

報告要旨： 経済学では、所得格差を測るものさし、すなわち、所得の不平等度尺度として、ジニ係数が重宝されている。ここで、その係数とは、一辺の長さが 1 のエッジワースボックスに描かれた、均等所得線とみなされている 45° 線と座標 $(0,0)$ ならびに座標 $(1,1)$ を通る凸のローレンツ曲線とで囲まれる面積の値のことを言う。しかし、ジニ係数は、完備性が付与されているものの、その限界が指摘されている。それは、例えば、2 つの異なる所得分布間でジニ係数の値が一致する場合、どのようにして両者間の不平等度判定を行えば良いのだろうかということである。

そこで、我々は、任意の所得分布 $\boldsymbol{Y}$ から導出されるローレンツ曲線を、数学の軌跡問題としてデザインし直すことにより、x-L 平面上の曲線、すなわち、軌跡の座標 (x,L) を形成する 2 本の式「$x=$ 累積世帯数割合」および「$L=$ 累積所得額割合」を表す累積分布関数、特に、これらの基となる確率密度関数に着目する。そして、ジニ係数の代わりに、カルバックライブラー情報量の概念を用い、当該確率密度関数間の距離を測定して、この係数の上記限界を不問とする、所得の不平等度尺度の提示を試みる。

1　はじめに

所得の平等不平等の先行研究としてヴィルフレド・パレート、コッラド・ジニ、マックス・ローレンツが挙げられる。彼らはそれぞれパレート分布、ジニ係数、ローレンツ曲線と現代の所得の平等不平等の基礎概念の土台となる理論を構築している。特に、ジニ係数とローレンツ曲線の組合せは一見に値し、ローレンツ曲線と均等分配線の面積の2倍がジニ係数と同値である。すなわち、不平等度を数値で表したジニ係数を可視化したものがローレンツ曲線である。したがって、現代の所得研究基盤においてこれらの理論を無視することはできないだろう。

しかし、ジニ係数やローレンツ曲線を過信し過ぎるのもよくないのである。なぜならば、図1の所得分布Aと所得分布C、もしくは、所得分布Bと所得分布Cのローレンツ曲線は平等不平等度が可視化され、所得分布Cの方が不平等であると直感的に訴えかけられる。それに対して、所得分布Aと所得分布Bのようにローレンツ曲線が交差している場合には判断に躊躇するであろう。そこで、対角線とローレンツ曲線が囲む面積が、対角線の下側になる二等辺三角形の面積に占める割合として定義されたジニ係数を用いるとしても、ジニ係数が一致している場合は、これら所得分布に対してどのように平等不平等度を判断すればよいのであろうか。そこで我々は不平等度尺度をリデサインすべくカルバックライブラー情報量を提案する。

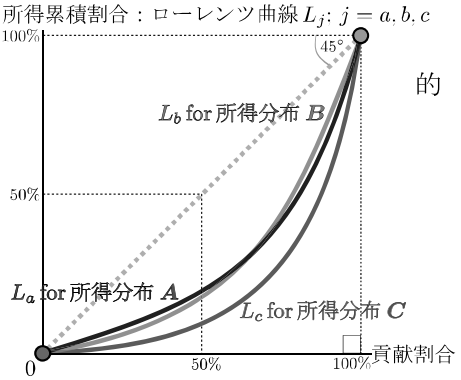

図1：所得分布A,B,C

本稿の主目的は、異ローレンツ曲線の異分布における同ジニ係数値となる例を取り上げ、これに対して別の角度から新たな判断基準を提示することである。そこで、本稿の構成は以下の通りである。第 2 節において、ローレンツ曲線とジニ係数とカルバックライブラー情報量の定義を述べる。続く第 3 節では、二つの所得分布を例に挙げる。そして、ジニ係数が一致している場合を仮定し、カルバックライブラー情報量を利用することで異差を見出す。最後に第 4 節では本稿の例以外の分布のカルバックライブラー情報量を載る。

2　ローレンツ曲線とジニ係数とカルバックライブラー情報量

2.1　ローレンツ曲線の定義

まず、ローレンツ曲線の定義を行う。一般的に所得に関するローレンツ曲線は「横軸に累積世帯数割合」、「縦軸に累積所得割合」を基にして長さ 1 の正方形の座標平面に凸関数として描かれる曲線である。そして、一般的にローレンツ曲線の定義式は$f(y)$を確率密度関数としたとき、$L(x) = \int_{-\infty}^{y(x)} yf(y)\,dy / \int_{-\infty}^{\infty} yf(y)\,dy$で与えられる。ここで、$y(x) = F^{-1}(x)$のことであり累積分布関数$F(y)$の逆関数を意味する。

しかし、我々はローレンツ曲線に関してこのような一本式ではなく、独立変数yを媒介変数とした

(1)　$$L(x) = \begin{cases} x = F(y) \\ L(y) = \frac{1}{\mu}\int_{-\infty}^{y} tf(t)dt \end{cases}$$

で定義を与える。ただし、$\mu = \int_{-\infty}^{\infty} yf(y)dy$のことであり、また、累積分布関数$F(y)$に関して逆関数の存在が保証されるものとする。

2.2　ジニ係数の解釈

ジニ係数 G は図 2 に描かれるように、直角二等辺三角形の面積(面積 A+面積 B)に占める面積 A の割合として定義される。したがって、定義式として$\mathrm{G} = \mathrm{A}/(\mathrm{A}+\mathrm{B})$が与えられ、$2G = \int_0^1 \big(x - L(x)\big)dx$となることがわかる。そして、この式に対して変数変換を行うと

(2)　$$2G = \int_0^1 \big(F(y) - L(y)\big)f(y)dy = E_f[\boldsymbol{F}] - E_f[\boldsymbol{L}]$$

を得る。すなわち、ジニ係数自体が二つの分布関数$F(y)$と$L(y)$の差に関する密度関数$f(y)$の平均になっているのである。また分布関数$L(y)$の確率密度関数を$l(y)$と定義すると、$E_f[\boldsymbol{F}] = 1/2 = E_l[\boldsymbol{L}]$より

(3)　$$2G = E_l[\boldsymbol{L}] - E_f[\boldsymbol{L}] = \int_0^1 \big(l(y) - f(y)\big)\,L(y)dy$$

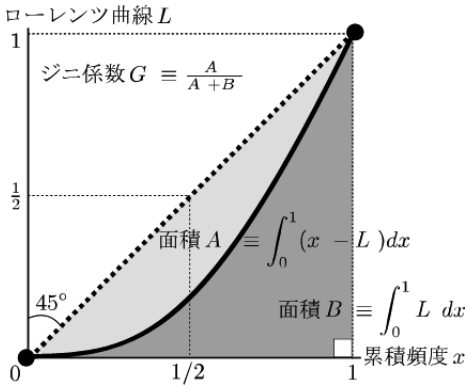

図 2：ローレンツ曲線と面積

と、二つの相異なる密度関数の差で表現される。(2)式と(3)式を見ると、ジニ係数自体は二つの相異なる分布関数間の距離、もしくは、相異なる密度関数間の距離として考えることができる。そして、この相異なる分布関数もしくは密度関数間の距離が小さければ小さいほどジニ係数が小さくなり平等と言える。したがって、確率密度関数間距離を測るために我々はカルバックライブラー情報量を用いる。

2.2　カルバックライブラー情報量

それでは、カルバックライブラー情報量の定義を述べる。

定義．カルバックライブラー情報量(KLD)

真の確率密度関数$f(y)$に対して確率モデル$l(y)$が存在するとき、カルバックライブラー情報量$D(f \parallel l)$は

(4)　$$D(f \parallel l) = \int_{-\infty}^{\infty} f(y)\log\frac{f(y)}{l(y)}dy = \int_{-\infty}^{\infty} f(y)\log f(y)\,dy - \int_{-\infty}^{\infty} f(y)\log l(y)\,dy$$

と表される。ここで、最右辺の第一項をエントロピー、第二項を平均対数度という。

この定義より、我々は$f(y)$を世帯数に関する確率密度関数、$l(y)$を所得に関する確率密度関数として考えこれらの確率密度関数間の距離を考える。$D(f \parallel l)$が小さいほど確率密度関数の差が小さいとみなせ、より平等度が高いと考えられる。ここで、今述べた$f(y)$は世帯数に関する世帯数の累積分布関数$F(y)$を、$l(y)$は所得額に関する所得額の累積分布関数$L(y)$を所得額yでそれぞれ一階微分したものである。

3　二つの所得分布によるカルバックライブラー情報量比較

ジニ係数に代わってカルバックライブラー情報量を用いて所得の不平等度を評価するために、ここでは二つの所得分布を定義する。その二つの所得分布を用いてローレンツ曲線とジニ係数、ジニ係数が一致する条件を求める。そして、カルバックライブラー情報量で評価可能であることを見てゆく。

3.1　パレート分布と逆パレート分布の密度関数と累積分布関数

所得額をy、最低所得額をy_0、パレート指数αを$1 < \alpha < \infty$としたとき、パレートの累積分布関数$F(y)$は$F(y) = 1 - \left(\frac{y_0}{y}\right)^{\alpha}$ $[y_0 \le y < \infty]$と定義される。そして、確率密度関数に関しては下記ように定義される。

(5)　$f_1(y) = \frac{\alpha y_0^{\alpha}}{y^{\alpha+1}}$　$in\ \ y_0 \le y < \infty$

もう一つの所得分布はパレート分布の分母と分子をひっくり返した形で、ここではこれを逆パレート分布と名付けておく。すると、累積分布関数は$F(y) = \left(\frac{y}{y_0}\right)^{\gamma}$ $[0 \le y \le y_0]$と与えられる。すると確率密度関数$f(y)$は

(6)　$f_2(y) = \frac{\gamma y^{\gamma-1}}{y_0^{\gamma}}$　$in\ \ 0 \le y \le y_0$

と定義できる。ただし、$0 \le \gamma < \infty$とする。

3.2　パレート分布と逆パレート分布のローレンツ曲線とジニ係数

パレート分布と逆パレート分布の媒介変数型ローレンツ曲線$L_1(x)$と$L_2(x)$は

(7)　$L_1(x) = \begin{cases} x = 1 - \left(\frac{y_0}{y}\right)^{\alpha} \\ L_1(y) = 1 - \left(\frac{y_0}{y}\right)^{\alpha-1} \end{cases}$　、　$L_2(x) = \begin{cases} x = \left(\frac{y}{y_0}\right)^{\gamma} \\ L_2(y) = \left(\frac{y}{y_0}\right)^{\gamma+1} \end{cases}$

で与えられる。ここで、パレート分布と逆パレート分布の平均は$\mu_1 = \alpha y_0/(\alpha - 1)$と$\mu_2 = \gamma y_0/(\gamma + 1)$である。

ジニ係数 G は図 2 においてG = A/(A + B)で表される。よって、パレート分布に対する面積 A をA_1、逆パレート分布に対する面積 A をA_2とすると、それぞれ$A_1 = \int_0^1 \left(x - L_1(x)\right)dx = \frac{1}{2(2\alpha-1)}$、$A_2 = \int_0^1 \left(x - L_2(x)\right)dx = \frac{1}{2(2\gamma+1)}$で与えられる。このことより、パレート分布のジニ係数G_1と逆パレート分布のジニ係数G_2は

(9)　$G_1 = \frac{1}{2\alpha-1}$　、　$G_2 = \frac{1}{2\gamma+1}$

で表される。そして、パレート分布と逆パレート分布におけるジニ係数が一致する条件は$G_1 = G_2$より

(10)　$\gamma = \alpha - 1$

を得る。これが本稿の分布におけるジニ係数一致条件である。この一致条件を満たしている限りジニ係数は等しいので、どのように平等不平等の判定をすればいいのだろうか。そこでカルバックライブラー情報量を用いて判定可能であることを次節で述べてゆく。

3.3　パレート分布と逆パレート分布のカルバックライブラー情報量

2.2 節を見ると、パレート分布と逆パレート分布に関する世帯数と所得額の確率密度関数$f(y)$、$l(y)$が必要である。世帯数に関する確率密度関数$f(y)$はすでに(5)式と(6)式より得られている。そして、所得に関する確率密度関数$l(y)$は、(7)式の$L_1(y)$と$L_2(y)$を所得額yの一階微分で得られる。ゆえに、

(11)　$l_1 = \frac{(\alpha-1)y_0^{\alpha-1}}{y^{\alpha}}$　、　$l_2(y) = \frac{(\gamma+1)y^{\gamma}}{y_0^{\gamma+1}} = \frac{\alpha y^{\alpha-1}}{y_0^{\alpha}}$

となる。今、所得額の定義域は$y_0 \le y < \infty$であることに注意して、パレート分布と逆パレート分布のカルバ

ックライブラー情報量をそれぞれ$D_1(f_1 \parallel l_2)$、$D_2(f_2 \parallel l_2)$とすると(4)式より

(12)　　パレート分布：$D_1(f_1 \parallel l_2) = \log\frac{\alpha}{\alpha-1} - \frac{1}{\alpha}$

(13)　逆パレート分布：$D_2(f_2 \parallel l_2) = \log\frac{\gamma}{\gamma+1} + \frac{1}{\gamma} = \log\frac{\alpha-1}{\alpha} + \frac{1}{\alpha-1}$

が計算により求められる。そして、これらの所得分布に関するカルバックライブラー情報量とジニ係数に関する数値表を下記にまとめる。すると、同一ジニ係数でもパレート分布と逆パレート分布のカルバックライブラー情報量(KLD)の値に差があることがわかる。例えば表 1 のジニ係数$\mathrm{G} = 0.5$を見ると、$D_1(f_1 \parallel l_2) = 0.43194562$、$D_2(f_2 \parallel l_2) = 0.901387711$となっておりパレート分布の方がより平等度が高いことが伺える。

表 1　同ジニ係数に対するカルバックライブラー情報量(KLD)の違い

パレート		ジニ係数	逆パレート	
α	KLD	G	γ	KLD
1.01	3.625021507	0.980392157	0.01	95.38487948
1.02	2.951433476	0.961538462	0.02	46.06817437
1.03	2.565242913	0.943396226	0.03	29.79721663
1.04	2.296558076	0.925925926	0.04	21.74190346
1.05	2.092141485	0.909090909	0.05	16.95547756
1.06	1.928283398	0.892857143	0.06	13.79498704
1.07	1.792339246	0.877192982	0.07	11.5587956
1.08	1.67676376	0.862068966	0.08	9.897310315
1.09	1.576692112	0.847457627	0.09	8.616987806
1.1	1.488804364	0.833333333	0.1	7.602104727

パレート		ジニ係数	逆パレート	
α	KLD	G	γ	KLD
1.2	0.958426136	0.714285714	0.2	3.208240531
1.3	0.6971063	0.625	0.3	1.866996265
1.4	0.538477254	0.555555556	0.4	1.247237032
1.5	0.431945622	0.5	0.5	0.901387711
1.6	0.355829253	0.454545455	0.6	0.685837414
1.7	0.299067901	0.416666667	0.7	0.541268234
1.8	0.255374661	0.384615385	0.8	0.439069784
1.9	0.220898612	0.357142857	0.9	0.363896709
2	0.193147181	0.333333333	1	0.306852819
2.1	0.170436689	0.3125	1.1	0.262463744

4　おわりに

本稿において、ジニ係数が累積分布関数間の距離、もしくは、確率密度関数間の距離に依存することを示した。そこで、我々は、その確率密度関数間の距離を測るために、分布間の距離測定に用いられるカルバックライブラー情報量の適用を提案する。そして、例として、パレート分布と逆パレート分布とを用い、ジニ係数が一致する場合の平等不平等度の判断基準として、当該情報量が有用であることを述べた。

なお、本稿では、二つの所得分布を例に挙げて、カルバックライブラー情報量を導出のうえ、不平等度を比較したが、これら以外の分布についてカルバックライブラー情報量を求めると、一様分布のカルバックライブラー情報量は、$D(f \parallel l) = 1 + \frac{1}{2}\log(1-a)^{\frac{1}{a}-1}/(1+a)^{\frac{1}{a}+1}$で、対数正規分布のカルバックライブラー情報量は、それぞれ、$D(f \parallel l) = \sigma^2/2$などとなる。ただし、ここでは、一様分布の平均を$\mu$、かつ、端点パラメータを$a(0 \le a \le 1)$としたときの確率密度関数を$f(y) = 1/2\mu a$、また、対数正規分布のパラメータ$\mu$、$\sigma$に関する確率密度関数を$f(y) = \frac{1}{\sqrt{2\pi}\sigma y}e^{(\log y-\mu)^2/2\sigma^2}$と仮定している。これらを各種の所得分布に適宜用いていけば、異分布間、特にローレンツ曲線が交差している場合などで、平等不平等度の測定が容易に可能だと思われる。

参考文献

Gini, C., “Notes and Memoranda: Measurement of the Inequality of Incomes,” *The Economic Journal*, 1921 March, pp. 124-6.

Lorenz, M. O., “Methods of Measuring the Concentration of wealth,” *Publications of American Association*, 1905 June, pp. 209-19.