

ローレンツ曲線とジニ係数

1 ローレンツ曲線

所得分布などの不平等の程度を調べるために便利な道具がローレンツ曲線 (Lorentz curve) と呼ばれる図である。ローレンツ曲線を描くためには、表 1 のような度数分布表を作成する。

表 1: ローレンツ曲線を作図するための度数分布表

階級 番号	階級値	人口			所得計		
		絶対度数	相対度数	累積 相対度数	絶対度数	相対度数	累積 相対度数
1	I_1	n_1	p_1	P_1	y_1	q_1	Q_1
2	I_2	n_2	p_2	P_2	y_2	q_2	Q_2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
j	I_j	n_j	p_j	P_j	y_j	q_j	Q_j
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
m	I_m	n_m	p_m	P_m	y_m	q_m	Q_m
合計		N	1		Y	1	

階級 j には n_j 人が属しており、 y_j はその n_j 人の人たちの所得の合計である。階級値 I_j は平均所得であり、 $I_j = y_j/n_j$, $y_j = I_j \times n_j$ である¹。各階級は平均所得の低い方から順に並べるものとする。すなわち、 $I_1 \leq I_2 \leq \dots \leq I_m$ である。表中の各変数のあいだには以下のような関係があることに注意されたい。

$$N = \sum_{j=1}^m n_j, \quad Y = \sum_{j=1}^m y_j, \quad p_j = \frac{n_j}{N}, \quad q_j = \frac{y_j}{Y}, \quad (j = 1, 2, \dots, m),$$

$$P_j = \sum_{k=1}^j p_k = \sum_{k=1}^{j-1} p_k + p_j = P_{j-1} + p_j, \quad (j = 1, 2, \dots, m),$$

$$Q_j = \sum_{k=1}^j q_k = \sum_{k=1}^{j-1} q_k + q_j = Q_{j-1} + q_j, \quad (j = 1, 2, \dots, m).$$

xy 平面上の x 座標が P_j , y 座標が Q_j である点を点 B_j とする。このことを $B_j(P_j, Q_j)$ と書く。また、原点 (x 座標, y 座標ともに 0 である点) を点 B_0 とする。このことを $B_0(0, 0)$ と書く。ローレンツ曲線は、 $m+1$ 個の点 $B_0(0, 0), B_1(P_1, Q_1), B_2(P_2, Q_2), \dots, B_m(P_m, Q_m)$ を結んだ折れ線である。図 1 は、講

¹各個人のデータから度数分布表を作成する場合、各階級の最小値と上限を定めてから、各階級の階級値を最小値と上限の midpoint (最小値と上限の平均、すなわち (最小値 + 上限) ÷ 2) として求めるのが標準的である。ローレンツ曲線を描くための度数分布表を作成する場合は、階級値を「最小値と上限の平均」ではなく、「階級に属する全個人の平均」とすること。このように階級値を定めることで、元のデータの合計と「階級値 × 人数」として求めた合計 (表中の y_j) とが一致する。あるいは、階級値をとくに定めず、所得の合計を直接 y_j として利用しても良い。

義で提供した 76 カ国の 1 人当たり GDP と人口のデータ (Excel Book countries1.xls) を用いたローレンツ曲線である。

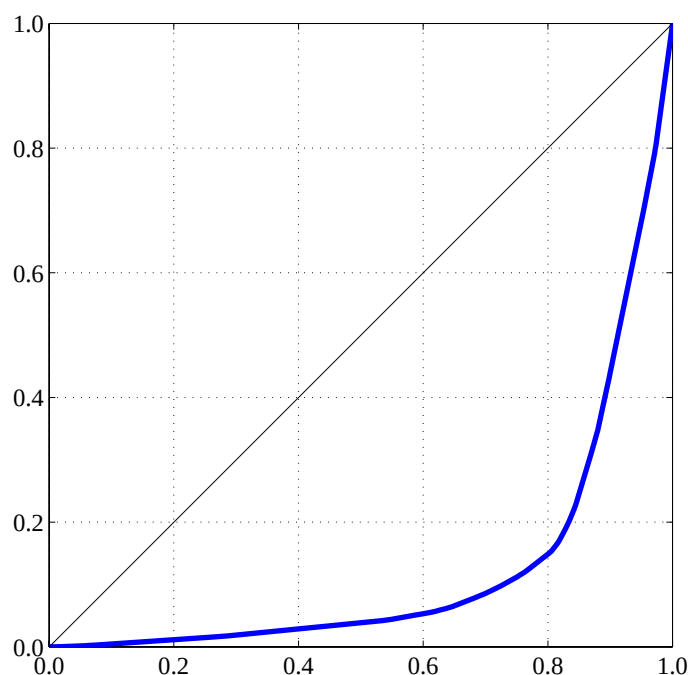


図 1: 76 カ国の所得分布についてのローレンツ曲線

完全平等線と完全不平等線 もし全員の所得が等しく、所得分布が完全に平等な場合には、ローレンツ曲線はどのような形状になるであろうか？ このようなとき、全階級の平均所得 (階級値) はすべて等しく、 $I_1 = I_2 = \dots = I_m = \bar{I}$ である。また、 $y_j = I_j \times n_j$ であるから、

$$Y = \sum_{j=1}^m y_j = \sum_{j=1}^m (I_j \times n_j) = \sum_{j=1}^m (\bar{I} \times n_j) = \bar{I} \times \sum_{j=1}^m n_j = \bar{I} \times N,$$

$$q_j = \frac{y_j}{Y} = \frac{I_j \times n_j}{\bar{I} \times N} = \frac{\bar{I} \times n_j}{\bar{I} \times N} = \frac{n_j}{N} = p_j,$$

となる。すなわち、所得計についての相対度数 q_j は人口についての相対度数 p_j に等しく、それぞれの累積相対度数も等しくなり $P_j = Q_j$ である。このことは、ローレンツ曲線が 2 点 $(0,0)$, $(1,1)$ を結ぶ直線 (45° 線) となることに対応する。このため、 45° 線のことを**完全平等線**と呼ぶ。

逆に、完全に平等な (全員の所得が等しい) 場合の対極としての完全に不平等な場合は、ローレンツ曲線はどのような形状になるだろうか？ 完全に不平等な場合とは、たった 1 人の大富豪がすべての所得を得ており、その他全員の所得がゼロの場合である。図 2 は、完全に不平等なときのローレンツ曲線を、人口が 5 人、10 人、100 人の場合について描いたものである。図からも明らかなように、完全に不平等で人口が十分大きいときのローレンツ曲線は、3 点 $(0,0)$, $(1,0)$, $(1,1)$ を結ぶ折れ線となる。この折れ線のことを**完全不平等線**と呼ぶ。

ローレンツ曲線の性質 縦軸・横軸ともに「累積」相対度数であるので、ローレンツ曲線は右上がりの折れ線である。また、所得の低い方から順に並べてあるので、ローレンツ曲線は右下の方へ凸である (右へ行くほど傾きは急になる)。

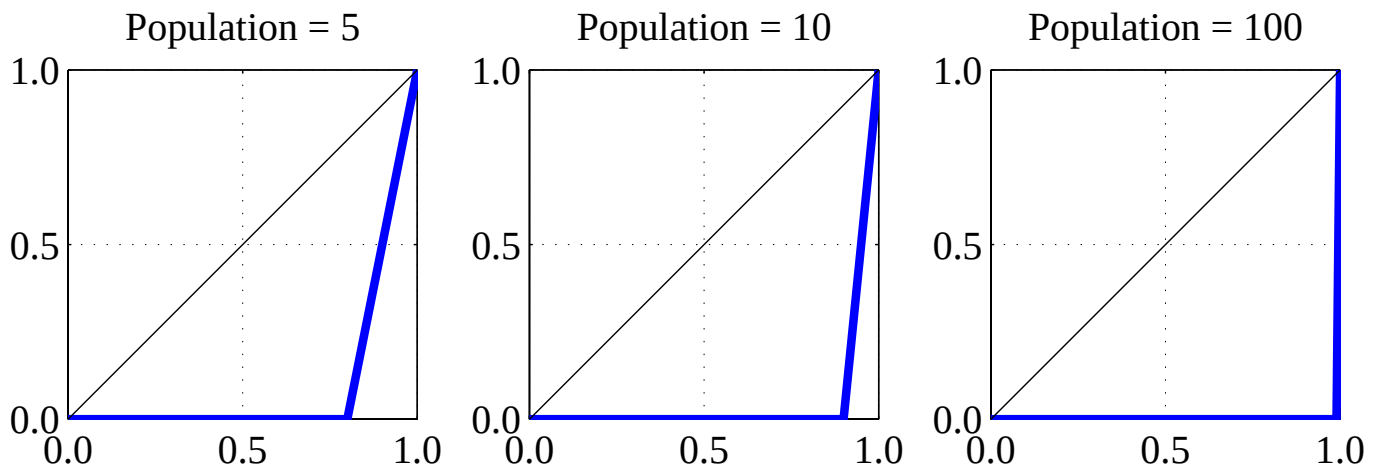


図 2: 完全に不平等な場合のローレンツ曲線：人口が 5 人, 10 人, 100 人の場合

実際のデータを用いて描かれたローレンツ曲線は、完全平等線と完全不平等線のあいだにある。完全平等線に近いほど「より平等」であり、完全不平等線に近いほど「より不平等」である。

2 ジニ係数

所得分布の不平等の程度を図で調べるのがローレンツ曲線であるが、その程度を数値化したものがジニ係数 (Gini index) である。ジニ係数 G は、完全平等線とローレンツ曲線で囲まれたレンズ型の図形の面積を用いて、次のように定義される。

$$G = \frac{\text{データから描かれたレンズ型の面積}}{\text{完全に不平等な場合のレンズ型の面積}}$$

$$= \text{[Lorenz Curve Area]} \div \text{[Perfect Inequality Area]} = \text{[Lorenz Curve Area]} \div \frac{1}{2} = \text{[Lorenz Curve Area]} \times 2$$

定義により、ジニ係数 G は不等式 $0 \leq G \leq 1$ を満たす。完全に平等な場合は $G = 0$ であり、完全に不平等な場合は $G = 1$ である。

ジニ係数の解釈については、次の点に注意する必要がある。

- 2 種類のデータ I, II について所得分布の不平等の程度を比較する場合、I のローレンツ曲線がすべての範囲で II のローレンツ曲線よりも左上にあれば、「I は II より平等である」と解釈できる。また、I のジニ係数の方が II のジニ係数よりも小さくなる。しかし、2 つのローレンツ曲線が互いに交わることもあり、この場合には、一方が他方よりも平等であるか否かの判断は困難である。にもかかわらず、数値としてジニ係数は計算できてしまうので、2 つのジニ係数の大小を比較することは可能である。2 種類のデータについて所得分布の不平等の程度を比較する場合、ローレンツ曲線を描かずにジニ係数の値だけを比較するのは危険である。
- 国・地域別にデータを集計したり、所得階級別にデータを集計したりすると、集計する前と比べて、ローレンツ曲線は必ず (一定であるか、または) 左上方へ移動する。これは、データの集計

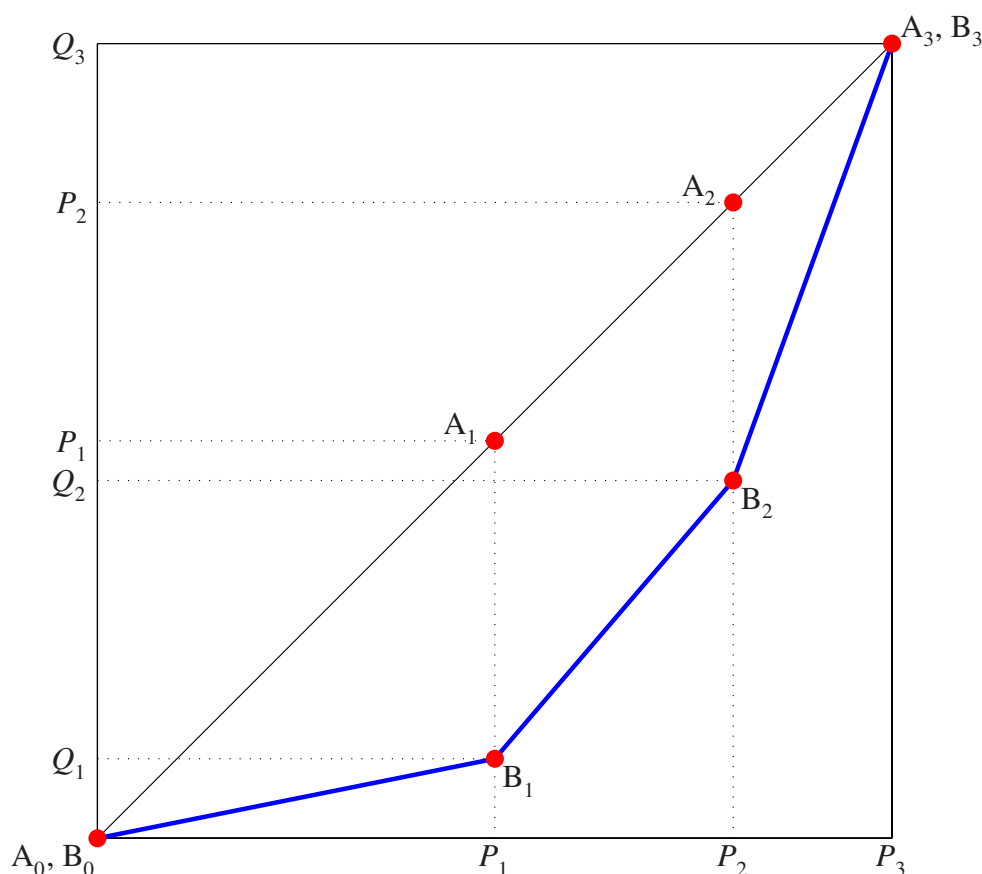


図 3: ジニ係数の計算方法

によって、各階級の中での所得分布は完全に平等であるものと扱われることによる。したがって、もし可能であれば、集計されていないデータを用いるべきである。しかし、集計されていない個人の所得データを入手することは一般には困難である。少なくとも、複数種類のデータ間でローレンツ曲線やジニ係数による比較を行う場合には、集計の程度が同程度であるかどうかには留意すべきである。

ジニ係数の計算方法 ジニ係数はローレンツ曲線を用いて定義されるから、ローレンツ曲線を作図するデータが整えられていれば、ジニ係数も計算可能である。以下では、ジニ係数の計算方法を解説する。

図 3 は階級数が $m = 3$ の場合のローレンツ曲線である。既述のように、ローレンツ曲線は $m+1 = 4$ 個の点 $B_0(0, 0)$, $B_1(P_1, Q_1)$, $B_2(P_2, Q_2)$, $B_3(P_3, Q_3)$ を結ぶ折れ線である。図 3 には、これら 4 個の点の位置が分かるように点の名前を書き入れている。さらに、これら 4 点と x 座標が等しく、 45° 線 (完全平等線) 上にある点を $A_0(0, 0)$, $A_1(P_1, P_1)$, $A_2(P_2, P_2)$, $A_3(P_3, P_3)$ とする。定義により、2 点 A_0, B_0 , および 2 点 A_3, B_3 はそれぞれ同一の点である。

完全平等線とローレンツ曲線で囲まれたレンズ型は、 $m - 1 = 2$ 本の線分 A_1B_1 , A_2B_2 によって $m = 3$ 個の領域に分割される。左右両端の 2 個は三角形であり、中央よりの $m - 2 = 1$ 個は台形である。両端の三角形は台形の特殊な場合と見なせば、分割してできた $m = 3$ 個の領域は、台形 $A_0A_1B_1B_0$, 台形 $A_1A_2B_2B_1$, 台形 $A_2A_3B_3B_2$ である。

各台形の上底、下底、高さなどの対応をあらわしたのが表 2 である。各台形の左右にある向かい合う辺が平行であり、左側を上底、右側を下底と呼ぶこととする。高さは x 軸方向に測る。既述のよう

表 2: ジニ係数の計算方法 ($m = 3$ の場合)

階級	台形の名称	上底	下底	高さ
1	台形 $A_0A_1B_1B_0$	$A_0B_0 = 0$	$A_1B_1 = P_1 - Q_1$	$p_1 = P_1$
2	台形 $A_1A_2B_2B_1$	$A_1B_1 = P_1 - Q_1$	$A_2B_2 = P_2 - Q_2$	$p_2 = P_2 - P_1$
3	台形 $A_2A_3B_3B_2$	$A_2B_2 = P_2 - Q_2$	$A_3B_3 = P_3 - Q_3$	$p_3 = P_3 - P_2$

に, 2 点 A_0, B_0 , および 2 点 A_m, B_m (A_3, B_3) はそれぞれ同一の点であるから, 線分 A_0B_0 と線分 A_mB_m (A_3B_3) の長さは 0 である. また, 線分 A_kB_k ($k = 1, \dots, m$) の長さは, 2 点 A_k, B_k の y 座標の差であるから, $A_kB_k = P_k - Q_k$ として求められる.

以上より, $m = 3$ の場合のジニ係数 G は以下のように求められることが分かる.

$$\begin{aligned}
 G &= 2 \times (\text{台形 } A_0A_1B_1B_0 \text{ の面積}) + 2 \times (\text{台形 } A_1A_2B_2B_1 \text{ の面積}) + 2 \times (\text{台形 } A_2A_3B_3B_2 \text{ の面積}) \\
 &= \{0 + (P_1 - Q_1)\}p_1 + \{(P_1 - Q_1) + (P_2 - Q_2)\}p_2 + \{(P_2 - Q_2) + (P_3 - Q_3)\}p_3
 \end{aligned}$$

同様に, 一般の m については, 以下のように求められる.

$$\begin{aligned}
 G &= 2 \times (\text{台形 } A_0A_1B_1B_0 \text{ の面積}) + 2 \times (\text{台形 } A_1A_2B_2B_1 \text{ の面積}) \\
 &\quad + \dots + 2 \times (\text{台形 } A_{m-2}A_{m-1}B_{m-1}B_{m-2} \text{ の面積}) + 2 \times (\text{台形 } A_{m-1}A_mB_mB_{m-1} \text{ の面積}) \\
 &= \{0 + (P_1 - Q_1)\}p_1 + \{(P_1 - Q_1) + (P_2 - Q_2)\}p_2 \\
 &\quad + \dots + \{(P_{m-2} - Q_{m-2}) + (P_{m-1} - Q_{m-1})\}p_{m-1} + \{(P_{m-1} - Q_{m-1}) + (P_m - Q_m)\}p_m
 \end{aligned}$$