説　苑

等分位データによる所得不平等度推計の一問題点

豊田　　敬

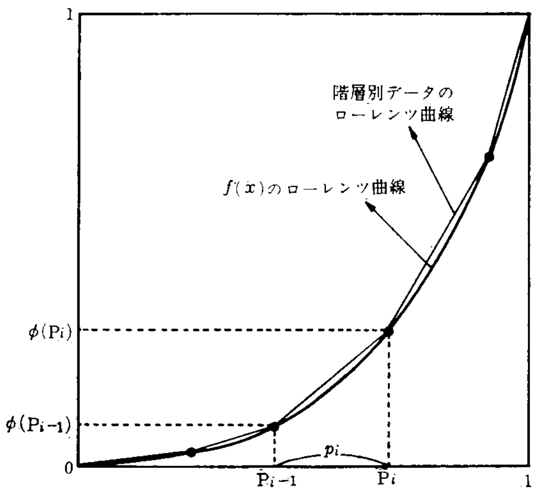
0．所得分布の統計は通常，データの個数が大きいために，各個人の所得額を全て並べた個別データではなく，それを整理要約した階層別データの形で与えられる。階層別データは，同一階層内の不平等を無視したデータであるため，このデータを用いてジニ集中係数などの不平等尺度で不平等度を推計すると，不平等度を過小評価することになる。したがって，不平等度を推計する場合，この過小評価を最小にするような階層区分のなされたデータを用いれば，最良の階層別データを用いた推計であると考えることができる。

過小評価を最小にする階層区分は，一般に所得分布の形に左右され，また使用する尺度により異なる。本稿では，ジニ集中係数・変動係数で不平等度を推計する場合，最良の階層区分が等分位であるような所得分布は一様分布であることを示す。

わが国の代表的な所得（資産）分布統計である「家計調査」や「貯蓄動向調査」の報告書に5分位の等分位階層別データが公表されていることにみられるように，所得分布の統計が5分位・10分位などの等分位データの型式に整理されることは少なくない。等分位の階層別データは，各階層の所得額の占める割合を比較する際に直截簡明であり，しかもローレンツ曲

線による比較が容易であるなど，他の階層別データと較べ，不平等比較に用いる場合，便利なデータである。しかしながら，現実の所得分布が一様分布で近似されるようなことはほとんどありえないので，本稿において得られる結果によれば，ジニ集中係数や変動係数で不平等度を推計する場合，等分位の階層別データは必ずしも良い階層別データではないということになる。実際の不平等度推計の際，このことは相対的には重要ではないかもしれないが，一応，記憶に留めておいてもよい問題点ではあると考えられる。

第1図　ローレンツ曲線



第1表　階層別所得分布統計表

所得階層		相対度数	階層別平均所得
1	$0 \sim y_1$	p_1	μ_1
2	$y_1 \sim y_2$	p_2	μ_2
⋮	⋮	⋮	⋮
i	$y_{i-1} \sim y_i$	p_i	μ_i
⋮	⋮	⋮	⋮
n	$y_{n-1} \sim$	p_n	μ_n

1．所得分布の密度関数を $f(x)$ $(0\leq x<\infty)$，その累積分布関数を $F(x)$，そして平均所得を $\mu(<\infty)$ とする。また，$f(x)$ は連続，そして $dF(x)/dx=f(x)$ とし，$F(\cdot)$ の逆関数を $F^{-1}(\cdot)$ と表わす。

分布 $f(x)$ のローレンツ曲線は

$$\phi(p)=\frac{1}{\mu}\int_0^{F^{-1}(p)}xf(x)\,dx=\frac{1}{\mu}\int_0^p F^{-1}(t)dt,\quad 0\leq p\leq 1$$

で表わされる。$\phi(p)\geq 0$, $\phi(0)=0$, $\phi(1)=1$ であることは明らかである。

また，

$$\phi'(p)=\frac{d\phi(p)}{dp}=\frac{1}{\mu}F^{-1}(p)\geq 0$$

$$\phi''(p)=\frac{d^2\phi(p)}{dp^2}=\frac{1}{\mu}[f(F^{-1}(p))]^{-1}\geq 0$$

である。したがって，分布 $f(x)$ のローレンツ曲線は座標 (0, 0) と座標 (1, 1) とを結ぶ単調増加で凸の曲線（第1図参照）である。

階層別所得分布データが第1表の形式で与えられているとする。分布 $f(x)$ からこの階層別データが作成されたものとすれば，

$$p_i=F(y_i)-F(y_{i-1}),\quad i=1, 2, \ldots, n$$

ただし，$p_i>0$, $F(y_0)=0$, $F(y_n)=1$

で，そして，

$$\mu_i=\mu[\phi(P_i)-\phi(P_{i-1})]/p_i,\quad i=1, 2, \cdots, n$$

ただし，$P_i=p_1+p_2+\cdots+p_i$,

$$\phi(P_0)=\phi(0)=0,\ \phi(P_n)=\phi(1)=1$$

である。この階層別データのローレンツ曲線は単調増加で凸の折線（第1図参照）で表わされる。

2. 分布 $f(x)$ のジニ集中係数は

$$G=1-2\int_0^1\phi(p)dp$$

である。一方，折線下の面積の２倍は，

$$S=\sum_{i=1}^{n}p_i[\phi(P_i)+\phi(P_{i-1})]$$

（第１図参照）であるから，階層別データのジニ集中係数は，

$$G^*=1-\sum_{i=1}^{n}p_i[\phi(P_i)+\phi(P_{i-1})]$$

である。

階層別データの ジニ集中係数 G^* が分布 $f(x)$ の ジニ集中係数 G に最も近くなる ように $\mathbf{P}=(p_1, p_2, \cdots, p_{n-1})$ を定めるという 問題は， G^* が最大になるように $\mathbf{P}$ を定める 問題と 同等である。 これは S を最小にすることである。

$$S=\sum_{i=1}^{n-1}p_i[\phi(P_i)+\phi(P_{i-1})]+(1-P_{n-1})[1+\phi(P_{n-1})]$$

であるから，

$$\frac{\partial S}{\partial p_j}=[\phi(P_j)+\phi(P_{j-1})]+p_j\phi'(P_j)+\sum_{i=1}^{n-1}p_i[\phi'(P_i)+\phi'(P_{i-1})]$$

二三九

$$+p_n\phi'(P_{n-1})-[1+\phi(P_{n-1})],\quad j=1, 2, \cdots, n-1$$

$$\frac{\partial S}{\partial p_j}-\frac{\partial S}{\partial p_{j+1}}=[\phi(P_{j-1})-\phi(P_{j+1})]+(p_j+p_{j+1})\phi'(P_j),$$

$$j=1, 2, \cdots, n-2$$

となる。したがって，S が最小となるための必要条件は，

$$\phi(P_{j+1})-\phi(P_{j-1})=(P_{j+1}-P_{j-1})\phi'(P_j),\quad j=1,\ 2,\ \cdots,\ n-1$$

である。

等分位の階層別データ，すなわち，$p_1=p_2=\cdots=p_n=1/n$ でこの必要条件が成り立つとすると，

$$\phi\left(\frac{j+1}{n}\right)-\phi\left(\frac{j-1}{n}\right)=\frac{2}{n}\phi'\left(\frac{j}{n}\right),\qquad j=1,\ 2,\ \cdots,\ n-1$$

である。ここで，$x=j/n,\ h=1/n$ とおくと，

$$\phi(x+h)-\phi(x-h)=2h\phi'(x) \qquad ①$$

である。階層の個数 n に関係なく①が成り立つためには，区間 $]0,\ 1[$ の任意の有理点 x で①が成り立たなければならない。そこで，①が任意の実数で成り立つとする。①より，

$$[\phi(x+h)-\phi(x-h)]+[\phi(y+h)-\phi(y-h)]=2h[\phi'(x)+\phi'(y)]$$
$$[\phi(x+h)-\phi(y-h)]+[\phi(y+h)-\phi(x-h)]=4h\phi'\left(\frac{x+y}{2}\right)$$

であるから，

$$\phi'\left(\frac{x+y}{2}\right)=\frac{1}{2}[\phi'(x)+\phi'(y)]$$

が得られる。これはイェンセンの方程式であり，その解は， 二二八

$$\phi'(x)=a_1x+a_2,\qquad a_1,\ a_2:\ 任意定数$$

である。したがって，

$$\phi(x)=ax^2+bx+c$$

である。この $\phi(x)$ は明らかに①を満足する。ここで，ローレンツ曲線の条件を考慮に入れると，

$$\phi(p)=ap^2+(1-a)p, \qquad 0<a\leq 1$$

が得られる。

$$\phi'(p)=\frac{1}{\mu}F^{-1}(p)=2ap+(1-a)$$

であるから，

$$F(x)=\frac{1}{2a\mu}x+\frac{a-1}{2a}$$

である。この $F(x)$ は一様分布であり，その密度関数は，

$$f(x)=1/2a\mu, \qquad \mu(1-a)\leq x\leq \mu(1+a)$$

である。すなわち，等分位の階層別データのとき，必要条件を満足する分布は一様分布であることが示された。

つぎに十分条件を調べる。

$$\frac{\partial^2 S}{\partial {p_j}^2}=2\phi'(P_j)+p_j\phi''(P_j)+\sum_{i=j+1}^{n-1}p_i[\phi''(P_i)+\phi''(P_{i-1})]$$
$$+p_n\phi''(P_{n-1})-2\phi'(P_{n-1}), \qquad j=1,\ 2,\ \cdots,\ n-1$$

$$\frac{\partial^2 S}{\partial p_j \partial p_k}=[\phi'(P_k)+\phi'(P_{k-1})]+p_k\phi''(P_k)+\sum_{i=k+1}^{n-1}p_i[\phi''(P_i)$$
$$+\phi''(P_{i-1})]+p_n\phi''(P_{n-1})-2\phi'(P_{n-1}), \quad j<k;\ j,k=1,2,\cdots,n-1$$

であるから，

$$\left.\begin{aligned}\phi(p)&=ap^2+(1-a)p\\ \phi'(p)&=2ap+(1-a)\\ \phi''(p)&=2a\end{aligned}\right\}\qquad ②$$

と $\mathbf{P}_0=(1/n,\ 1/n,\ \cdots,\ 1/n)$ を代入して計算すると，

$$\frac{\partial^2 S}{\partial p_j{}^2}\bigg|\mathbf{P}=\mathbf{P}_0=\frac{4a}{n},\qquad \frac{\partial^2 S}{\partial p_j\partial p_k}\bigg|\mathbf{P}=\mathbf{P}_0=\frac{2a}{n}$$

が得られる。

$$A=\left(\frac{\partial^2 S}{\partial p_j\partial p_k}\bigg|\mathbf{P}=\mathbf{P}_o\right)=\frac{2a}{n}\ \begin{array}{c}\overbrace{\hphantom{2\ \ 1\ \ \cdots\ \ 1}}^{n-1}\\ \begin{pmatrix}2 & 1 & \cdots & 1\\ 1 & 2 & \cdots & 1\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & \cdots & 2\end{pmatrix}\end{array}$$

の主小行列式は，

$$|A_k|=\left(\frac{2a}{n}\right)^k(k+1)>0,\qquad k=1,\ 2,\ \cdots,\ n-1$$

であるから，A は正値定符号である。

よって，一様分布で等分位の階層別データのとき S は最小になる。すなわち，等分位の階層別データのとき，ジニ集中係数を最大にする分布は一様分布であることが示された。

この一様分布のジニ集中係数は，

$$G=1-2\int_0^1\phi(p)dp=\frac{a}{3}$$

で，n 等分位の階層別データのジニ集中係数は，

$$G^*=1-S=\frac{a}{3}-\frac{a}{3n^2}$$

であるから，$(G-G^*)/G=1/n^2$ である。すなわち，一様分布の場合，n 等分位データの ジニ集中係数は 本来の ジニ集中係数より $1/n^2$ の割合で不平等度を過小評価する。

3. 分布 $f(x)$ の変動係数の 2 乗は，

$$V=\frac{1}{\mu^2}\int_o^\infty (x-\mu)^2 f(x)dx=\int_o^\infty\left(\frac{x}{\mu}\right)^2 f(x)dx-1$$

である。階層別データの変動係数の 2 乗は，

$$V^*=\sum_{i=1}^n p_i\left[\frac{\phi(P_i)-\phi(P_{i-1})}{p_i}\right]^2-1=\sum_{i=1}^n\frac{1}{p_i}[\phi(P_i)-\phi(P_{i-1})]^2-1$$

と表現できる。そこで，

$$W=\sum_{i=1}^n\frac{1}{p_i}[\phi(P_i)-\phi(P_{i-1})]^2$$

の最大化問題を考える。

$$I_i=[\phi(P_i)-\phi(P_{i-1})]/p_i,\qquad i=1,\ 2,\ \cdots,\ n$$

とおく。

$$\frac{\partial W}{\partial p_j}=[2\phi'(P_j)-I_j]I_j+2\sum_{i=j+1}^{n-1}[\phi'(P_i)-\phi'(P_{i-1})]I_i$$

$$-[2\phi'(P_{n-1})-I_n]I_n,\qquad j=1,\ 2,\ \cdots,\ n-1$$

$$\frac{\partial W}{\partial p_j}-\frac{\partial W}{\partial p_{j+1}}=[2\phi'(P_j)-I_j]I_j-[2\phi'(P_j)-I_{j+1}]I_{j+1},$$

$$j=1,\ 2,\ \cdots,\ n-2$$

であるから，W が最大となるための必要条件は，

$$(I_{j+1}-I_j)[(I_{j+1}+I_j)-2\phi'(P_j)]=0, \qquad j=1, 2, \cdots, n-1$$

である。$I_{j+1}\fallingdotseq I_j$ であるから，実際の必要条件は，

$$I_{j+1}+I_j=2\phi'(P_j), \qquad j=1, 2, \cdots, n-1$$

である。

$p_1=p_2=\cdots=p_n=1/n$ とし， $x=j/n,\ h=1/n$ とおくと，この必要条件は，

$$\phi(x+h)-\phi(x-h)=2h\phi'(x)$$

となる。これは①と同じであるから，等分位の階層別データのとき，階層の個数 n と関係なくこの必要条件を満足するローレンツ曲線は，

$$\phi(p)=ap^2+(1-a)p, \qquad 0<a\leqq 1$$

である。このローレンツ曲線に対応する分布は一様分布であった。

十分条件を調べる。

$$\frac{\partial^2 W}{\partial p_j^2}=2\phi''(P_j)I_j+\frac{2}{p_j}[\phi'(P_j)-I_j]^2+2\sum_{i=j+1}^{n-1}\{[\phi''(P_i)-\phi''(P_{i-1})]I_i$$
$$+\frac{1}{p_i}[\phi'(P_i)-\phi'(P_{i-1})]^2\}-2\phi''(P_{n-1})I_n+\frac{2}{p_n}[\phi'(P_{n-1})-I_n]^2,$$
$$j=1, 2, \cdots, n-1$$

$$\frac{\partial^2 W}{\partial p_j\partial p_k}=2\phi''(P_k)I_k+\frac{2}{p_k}[\phi'(P_k)-\phi'(P_{k-1})][\phi'(P_k)-I_k]$$
$$+2\sum_{i=k+1}^{n-1}\{[\phi''(P_i)-\phi''(P_{i-1})]I_i+\frac{1}{p_i}[\phi'(P_i)-\phi'(P_{i-1})]^2\}$$
$$-2\phi''(P_{n-1})I_n+\frac{2}{p_n}[\phi'(P_{n-1})-I_n]^2, \qquad j<k;\ i,j=1, 2, \cdots, n-1$$

であるから，②を代入して計算すると，

$$\frac{W^2\partial}{\partial p_j{}^2}=-2a^2(p_j+p_n),\qquad \frac{\partial^2 W}{\partial p_j\partial p_k}=-2a^2p_n$$

となる。したがって，

$$\left.\frac{\partial^2 W}{\partial p_j{}^2}\right|_{\mathbf{P}=\mathbf{P}_0}=-\frac{4a^2}{n},\qquad \frac{\partial^2 W}{\partial p_j\partial p_k}=-\frac{2a^2}{n}$$

で

$$B=\left(\left.\frac{\partial^2 W}{\partial p_j\partial p_k}\right|\mathbf{P}=\mathbf{P}_0\right)=-\frac{2a^2}{n}\overbrace{\begin{pmatrix} 2 & 1 & \cdots & 1\\ 1 & 2 & \cdots & 1\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & \cdots & 2\end{pmatrix}}^{n-1}$$

となる。B の主小行列式は，

$$|B_k|=(-1)^k\left(\frac{2a^2}{n}\right)^k(k+1),\qquad k=1,\ 2,\ \cdots,\ n-1$$

であるから，B は負値定符号である。

よって，一様分布で等分位階層別データのとき W は最大になる すなわち，等分位の階層別データのとき，変動係数が最大になる分布は一様分布であることが示された。

一様分布の変動係数の2乗は $V=a^2/3$ で，n 等分位階層別データの変動係数の2乗は，

$$V^*=\frac{a^2}{3}-\frac{a^2}{3n^2}$$

であるから，過小評価の割合は $(V-V^*)/V=1/n^2$ である。