1995年度日本オペレーションズ・リサーチ学会
春季研究発表会

格差の分析とカイ離係数

01600120　東京理科大学　牧野都治　MAKINO Toji

1．格差の分析に用いられる各種の尺度とカイ離係数

n個のデータt_1，t_2，・・・，t_nのもつ格差を分析するのに，変動係数がよく用いられるが，その他の尺度として，ジニ係数，パレート図での不平等度を表す弓形の面積，相対平均偏差などが用いられている。これと関連してわれわれは，たいへん簡便な乖離係数（以後カイ離係数と書く）という尺度の使用を提唱したい。それは，上の相対平均偏差に類する次のような尺度である。例えば，所得金額T（連続的確率変数として扱う）の分布関数をF（t），平均をνとする。いま金額の大きい方から累積することにして，横軸に累積人数率，縦軸に累積金額率をとって，パレート図を書く。このとき，Tの分布のパレート曲線$y=g(x)$と均等線$y=x$とのふくらみ（カイ離）が最大になる点をP（x_0，y_0）として，

$$\gamma=y_0-x_0$$

のことをカイ離係数と呼ぶ。x_0は

$$x_0=1-F(\nu)$$

により求められるので，カイ離係数の算出は極めて容易である。

2．対数正規分布のカイ離係数

一般に所得金額Tの分布は確率密度関数$f_1(t)$が次式で表される対数正規分布にしたがうとみてよいことが，よくいわれている。

$$f_1(t)$$
$$=1/\{(\sqrt{2\pi}\sigma t)\}\cdot\exp\{-A^2/(2\sigma^2)\}$$

ただし，$A=\log t-\mu$，$t>0$.

上式のσは形のパラメータ，μは尺度パラメータである。パレート曲線は尺度パラメータには依存しないので，$f_1(t)$の代りに次の$f_2(t)$を用いても同じパレート曲線が得られる。

$$f_2(t)$$
$$=1/\{(\sqrt{2\pi}\sigma t)\}\cdot\exp\{-B^2/(2\sigma^2)\}$$

ただし，$B=\log t$，$t>0$.

そこで，この曲線上のふくらみ最大の点の座標を求め，カイ離係数を計算してみると

$$\gamma=1-2\cdot\Phi(-\sigma/2)$$

になる。ただし，$\Phi(t)$は標準正規分布の分布関数である。

3．条件付対数正規分布のカイ離係数

3．1　1000万円以上の条件付分布

毎年5月になると，前年の所得納税金額1000万円以上の人の住所・氏名・納税金額が公示される。これを用いてカイ離係数を計算してみると，例えば東京都京橋税務署で，平

成3年分についてはデータから

$x_0=0.234,\ y_0=0.626$

が得られる。よってカイ離係数は

$\gamma=0.392$

となる。ここで，対数正規分布をあてはめたならばカイ離係数がいくらになるかを計算してみよう。ただし，Tを所得納税金額とすると，いまの場合$T\geq 1$（千万円）という条件付の分布を用いなければならない。そこでまず，このような条件付分布のパレート曲線上でのふくらみ最大の点Pの座標を求めてみると

$x_0=2\cdot\{1-\Phi(\log\nu/\sigma)\}$

$y_0=\{1-\Phi(\log\nu/\sigma-\sigma)\}/\{1-\Phi(-\sigma)\}$

となる。一方，データから計算した平均値は3783万円である。そこで，上のx_0の式に

$\nu=3.783$（千万円）

を代入し

$x_0=0.234$

とおいて，σの値を求めてみると

$\sigma=1.118$

となる。これを上のy_0の式にいれてみると

$y_0=0.543.$

これより，データに条件付対数正規分布をあてはめたときのカイ離係数の値は0.309になり，真の値0.392との相対誤差は0.21程度に上ることがわかる。

3.2 1000万円以上5000万円未満の条件付分布

所得納税金額が1000万円以上ということで，各税務署，各年次毎のパレート図を書いてみると，いろいろ相違がみられる。ところが，例えば1000万円以上～5000万円未満ということにしてパレート図を書くと，それらがほとんど重なってくる。これに関連して$1\leq T<5$（千万円）という条件付対数正規分布のカイ離係数を調べてみた。その方法は前述のものと同様で，まずこの条件付分布のパレート曲線上でのふくらみ最大の点の座標を求める。これは次のようになる。

$x_0=(\alpha_1-\beta_1)/(\alpha_1-1/2)$

$y_0=(\alpha_2-\beta_2)/\{\alpha_2-\Phi(-\sigma)\}.$

ただし，

$\alpha_1=\Phi(\log 5/\sigma),\quad \alpha_2=\Phi(\log 5/\sigma-\sigma),$

$\beta_1=\Phi(\log\nu/\sigma),\quad \beta_2=\Phi(\log\nu/\sigma-\sigma).$

一方，データからは

$\nu=2.028$（千万円）

$x_0=0.377,\ y_0=0.581$

が得られる。これに対し，条件付対数正規分布をあてはめて

$\nu=2.028$（千万円）

$x_0=0.377$

のときのy_0の値を計算してみると

$y_0=0.560$

となり，カイ離係数は0.183になる。これを，データからの真の値0.204と比べてみると，相対誤差は約0.10である。

参考文献
[1] 国土開発ジャーナル；著者寄付.
[2] 牧野都治；格差・パレート図・ABC分析 日本評論社(1984).