

パレート分布とユール分布との対応関係

SUZUKI, Takeshi / 鈴木, 武

(出版者 / Publisher)

法政大学経営学会

(雑誌名 / Journal or Publication Title)

経営志林 / The Hosei journal of business

(巻 / Volume)

44

(号 / Number)

1

(開始ページ / Start Page)

1

(終了ページ / End Page)

15

(発行年 / Year)

2007-04

(URL)

<https://doi.org/10.15002/00007120>

〔論 文〕

パレート分布とユール分布との対応関係

鈴 木 武

本稿では、連続分布であるパレート分布に対応して、離散分布として同じ特性をもつユール分布について記述する。従来の研究では、 x が大きな値をとるときに、ユール分布はパレート分布に近似すると述べられてきた。本稿ではその点を改良し、 x のいかなる値についても、 x のとりうる区間幅を 0 に近づけることによりパレート分布に近似するよう「一般化ユール分布」を提案している。

【1】では、ピアソン分布タイプの一形態であるパレート分布について述べる。【2】パレート分布に類似する離散分布はユール分布であることを示す。そのさい、「類似」の概念としてピアソン分布システムで用いられている「勾配・縦座標比率」を離散型に翻訳して考察している。【3】パレート分布とユール分布のハザード関数が同じ式になることを述べる。逆に、そのハザード関数をもつ連続分布および離散分布を求めると、パレート分布およびユール分布になることをいう。【4】一般化ユール分布を定義し、それが x のとる区間幅を 0 に近づけることによりパレート分布に収束することをいう。【5】一般化ユール分布がパレート分布とどの程度乖離しているかについてシミュレーションをする。【6】従来の論文で「一般化ユール分布」と呼ばれているものについて、その限界を述べる。【7】パレート分布およびユール分布について、従来、 x の値を 0 から始める場合と 1 あるいは α から始める場合とがみられる。それにより分布型が若干異なってくる。ここでは、 x のとる定義域のいかんに関わらず、それらを含むより一般化したユール分布を定義する。【8】より一般化したユール分布とそれに対応するパレート分布の特性のうち、分布関数、ハザード関数、平均、分散について述べる。

【1】ピアソン分布システム

ピアソン分布システムは、 f を確率密度関数とするとき

$$\frac{1}{f} \frac{df}{dx} = \frac{x+a}{b_0+b_1x+b_2x^2} \quad (1)$$

を満たす確率密度関数族である^(注1)。パラメータの値によって、いろいろなタイプの分布に分類される。ここで注目しているのはタイプ VI である。(1)式右辺の分母 $b_0+b_1x+b_2x^2=0$ の2つの根が実根で、かつ同符号の場合である。

2つの実根を $-c_1$ 、 $-c_2$ とし、 $c_2 > c_1 > 0$ とする。そのとき

$$b_0+b_1x+b_2x^2=b_2(x+c_1)(x+c_2)$$

と表すことができる。(1)式から

$$\begin{aligned} \frac{1}{f} \frac{df}{dx} &= \frac{x+a}{b_2(x+c_1)(x+c_2)} \\ &= \frac{1}{b_2} \cdot \frac{a-c_1}{(c_2-c_1)(x+c_1)} \\ &\quad - \frac{1}{b_2} \cdot \frac{a-c_2}{(c_2-c_1)(x+c_2)} \end{aligned}$$

ここで、 $x+c_2 > x+c_1 > 0$ を仮定し

$$\gamma_1 = \frac{1}{b_2} \cdot \frac{a-c_1}{c_2-c_1}, \quad \gamma_2 = \frac{1}{b_2} \cdot \frac{a-c_2}{c_2-c_1}$$

とおき、上式を積分する。

$$\int \frac{1}{f} \frac{df}{dx} = \gamma_1 \int \frac{1}{x+c_1} dx - \gamma_2 \int \frac{1}{x+c_2} dx$$

$$\log f = C + \gamma_1 \log(x+c_1) - \gamma_2 \log(x+c_2)$$

よって

$$f(x) = k(x+c_1)^{\gamma_1} (x+c_2)^{-\gamma_2} \quad (2)$$

が得られる。

【補助定理 1】 (1)式, (2)式において $a=c_1$, $b_2 = -\frac{1}{\beta+1}$, $\alpha=c_2-c_1 > 0$, $\beta > 0$ とおくと $f(x) = \beta\alpha^\beta(x+\alpha)^{-\beta-1}$, $x > 0$ (3) となり, パレート分布になる^(注2).

(証明) $a=c_1$ であるから $\gamma_1=0$, $\gamma_2 = -\frac{1}{b_2} = \beta+1$ になる。したがって(2)式は

$$f(x) = k(x+c_2)^{-\beta-1}$$

となる。

$x+c_1 > 0$ を仮定してるから, $x > -c_1$ である。 k を求めるためには

$$\int_{-c_1}^{\infty} k(x+c_2)^{-\beta-1} dx = 1$$

を解けばよい。

$$\begin{aligned} \int_{-c_1}^{\infty} k(x+c_2)^{-\beta-1} dx &= \left[\frac{k}{-\beta} (x+c_2)^{-\beta} \right]_{-c_1}^{\infty} \\ &= \frac{k}{-\beta} \{ 0 - (c_2-c_1)^{-\beta} \} \\ &= \frac{k}{\beta\alpha^\beta} \end{aligned}$$

よって

$$k = \beta\alpha^\beta$$

したがって

$$f(x) = \beta\alpha^\beta(x+c_2)^{-\beta-1}, \quad x > -c_1$$

ここで x の範囲を 0 以上に調整するために $x+c_1$ をあらためて x と置きなおす。

$$x+c_2-c_1 = x+\alpha$$

であるので

$$f(x) = \beta\alpha^\beta(x+\alpha)^{-\beta-1}, \quad x > 0$$

となり, パレート分布になる。(証明終わり)

【2】 Continuous Analogues

連続分布であるパレート分布に類似する離散分布は何であろうか。「類似」という意味をピアソン分布システムで用いた「勾配・縦座標比率」(slope-ordinate ratio) $\frac{1}{f} \frac{df}{dx}$ で表現してみる。その結果はユール分布になる。以下, それを説明しよう。

いま離散分布が $x=0, 1, 2, 3, \dots$ 上で分布しており, 確率は p_x で表現されるとする。離散分布の勾配・縦座標比率に相当するものを考えよう。点 $(x-1)$ と点 x における $\frac{df}{dx}$ に相当するものは, 1 の変化量に対する確率の変化量であるので $(p_x - p_{x-1})$ と表現される。また, f に相当するものは二点の確率の加重平均と考えられるので, $cp_x + (1-c)p_{x-1}$, ($0 \leq c \leq 1$) である。したがって

$$\frac{p_x - p_{x-1}}{cp_x + (1-c)p_{x-1}}$$

となる^(注3)。

Irwin (1975) によれば, 連続分布であるピアソン分布タイプ VI に類似する離散分布は一般超幾何分布 (generalized hypergeometric distribution) と呼ばれるものが対応する。これは一般ウェアリング分布 (generalized Waring distribution) とも言われている。

一般超幾何分布を表現するために, 上昇階乗ベキという記号を定義しよう。

$$a^{\bar{x}} = a(a+1) \cdots (a+x-1)$$

とする。 $x=0, 1, 2, 3, \dots$ のとき, 一般超幾何分布の確率は

$$p_x = \frac{\Gamma(a+c)\Gamma(b+c)}{\Gamma(a+b+c)\Gamma(c)} \cdot \frac{a^{\bar{x}}b^{\bar{x}}}{(a+b+c)^{\bar{x}}x!}, \quad (a, b, c > 0)$$

で表される。ここで $\Gamma(\cdot)$ はガンマ関数である。

ユール分布は一般超幾何分布の特別なケースで, $a=1$ かつ $b=1$ とした場合である。

$$\begin{aligned} p_x &= \frac{\Gamma(c+1)\Gamma(c+1)}{\Gamma(c+2)\Gamma(c)} \cdot \frac{1^{\bar{x}}1^{\bar{x}}}{(c+2)^{\bar{x}}x!} \\ &= \frac{c}{c+1} \cdot \frac{x!}{(c+2)^{\bar{x}}} \\ &= \frac{cx!}{(c+1)^{\bar{x}+1}} \end{aligned}$$

c を β で置き換えると, ユール分布は

$$p_x = \frac{\beta x!}{(\beta+1)^{\bar{x}+1}} \quad (x=0, 1, 2, \dots) \quad (4)$$

と表現される。

【補助定理 2】 ユール分布に類似する連続分布はパレート分布である。

(証明) ユール分布の勾配・縦座標比率を求めよう。(4)式から

$$\begin{aligned} p_x - p_{x-1} &= p_{x-1} \left(\frac{x}{x+\beta+1} - 1 \right) \\ &= -\frac{p_{x-1}}{x+\beta+1} \cdot (\beta+1) \\ cp_x + (1-c)p_{x-1} &= p_{x-1} \left(\frac{cx}{x+\beta+1} + 1 - c \right) \\ &= \frac{p_{x-1}}{x+\beta+1} \cdot \{cx + (1-c)(x+\beta+1)\} \end{aligned}$$

したがって

$$\frac{p_x - p_{x-1}}{cp_x + (1-c)p_{x-1}} = -\frac{\beta+1}{x+(1-c)(\beta+1)} \quad (x=1, 2, 3, \dots) \quad (5)$$

ここで $\alpha = (1-c)(\beta+1)$ とおく。類似する連続分布を求めるには

$$\frac{1}{f(x)} \frac{df(x)}{x} = -\frac{\beta+1}{x+\alpha}$$

を解けばよい。

微分方程式を解くと

$$\begin{aligned} \log f(x) &= C - (\beta+1) \int \frac{1}{x+\alpha} dx \\ &= C - (\beta+1) \log(x+\alpha) \end{aligned}$$

よって

$$f(x) = k(x+\alpha)^{-\beta-1} \quad (x > 0)$$

k を求めるには

$$\int_0^{\infty} k(x+\alpha)^{-\beta-1} dx = 1$$

$$1 = k \left[-\frac{1}{\beta} (x+\alpha)^{-\beta} \right]_0^{\infty} = \frac{k}{\beta} \alpha^{-\beta}$$

したがって

$$k = \beta \alpha^{\beta}$$

求める確率密度関数は

$$f(x) = \beta \alpha^{\beta} (x+\alpha)^{-\beta-1} \quad (x > 0)$$

よって、ユール分布に類似する連続分布はパレート分布である。(証明終わり)

パラメータ α を 1 にするためには $c = \frac{\beta}{\beta+1}$ である。

【定理 1】 ユール分布の勾配・縦座標比率は(5)式で表される。逆に勾配・縦座標比率が(5)式で表される離散分布はユール分布である。

(証明) 前半は記述したとおりなので、後半を証明しよう。勾配・縦座標比率が(5)式で表されるとする。(5)式を変形して

$$\begin{aligned} (p_x - p_{x-1}) \{x + (1-c)(\beta+1)\} \\ &= -(\beta+1) \{cp_x + (1-c)p_{x-1}\} \\ &\{x + (1-c)(\beta+1) + c(\beta+1)\} p_x \\ &= \{x + (1-c)(\beta+1) - (1-c)(\beta+1)\} p_{x-1} \end{aligned}$$

$$p_x = \frac{x}{x+\beta+1} p_{x-1}$$

p_x を求めるため、以下の変形をする。

$$\begin{aligned} p_x &= \frac{x}{x+\beta+1} p_{x-1} \\ &= \frac{x}{x+\beta+1} \cdot \frac{x-1}{x+\beta} \cdot \frac{x-2}{x+\beta-1} \\ &\quad \cdots \frac{1}{\beta+2} \cdot p_0 \\ &= \frac{x!}{(\beta+2)^{\bar{x}}} \cdot p_0 \end{aligned}$$

ここで p_0 を計算するために、次の差分を計算しよう。

$$\begin{aligned} \Delta \left[\frac{x!}{(\beta+2)^{\bar{x-1}}} \right] &= \frac{(x+1)!}{(\beta+2)^{\bar{x}}} - \frac{x!}{(\beta+2)^{\bar{x-1}}} \\ &= \frac{x!}{(\beta+2)^{\bar{x-1}}} \left(\frac{x+1}{x+\beta+1} - 1 \right) \\ &= -\beta \frac{x!}{(\beta+2)^{\bar{x}}} \end{aligned}$$

$\beta > 0$ だから

$$\begin{aligned} \sum_{x=1}^{\infty} \frac{x!}{(\beta+2)^{\bar{x}}} &= -\frac{1}{\beta} \sum_{x=1}^{\infty} \left(\Delta \left[\frac{x!}{(\beta+2)^{\bar{x-1}}} \right] \right) \\ &= -\frac{1}{\beta} \cdot \left(\frac{x!}{(\beta+2)^{\bar{x-1}}} \Big|_1^{\infty} \right) \end{aligned}$$

$$= -\frac{1}{\beta}(0-1)$$

$$= \frac{1}{\beta}$$

である。

$$1 = \sum_{x=0}^{\infty} p_x = p_0 + p_0 \sum_{x=1}^{\infty} \frac{x!}{(\beta+2)^x}$$

$$= p_0 \left(1 + \frac{1}{\beta}\right)$$

よって

$$p_0 = \frac{\beta}{\beta+1}$$

確率分布は

$$p_x = \frac{x!}{(\beta+2)^x} \cdot \frac{\beta}{\beta+1}$$

$$= \frac{\beta x!}{(\beta+1)^{x+1}}$$

となり、ユール分布になる。

【3】ハザード関数

連続確率変数 X の分布関数を $F(x)$ 、確率密度関数を $f(x)$ とする。生存関数は $S(x) = 1 - F(x)$ 、ハザード関数は

$$h(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{1 - F(x)} \quad (6)$$

と定義される。

離散確率変数 X の場合には

$$h(x) = \frac{p_x}{S(x)} = \frac{p_x}{1 - F(x+0)} = \frac{p_x}{\sum_{k=x+1}^{\infty} p_k} \quad (7)$$

とする^(注4)。

Xekalaki(1983)によれば、確率分布とハザード関数とが次のケースで1対1に対応する。離散分布 $X(x=0, 1, 2, \dots)$ については、ハザード関数を

$$h(x) = \frac{1}{a+bx} \quad (a>0, b \text{ は実数}) \quad (8)$$

とすると、 $b=0$ のとき幾何分布に、 $b>0$ のときウェアリング分布に、 $b<0$ のとき負の超幾何分布に対応する。

連続分布 $X(x>0)$ についても(8)式と同様のハザード関数を仮定すると、確率密度関数が

$$f(x) = \frac{1}{a} \left(1 + \frac{b}{a}x\right)^{-\frac{1}{b}-1}$$

となるピアソン分布族に対応することが言える。

本稿では、ウェアリング分布の特殊ケースであるユール分布のハザード関数と、ピアソン分布タイプVIの特殊ケースであるパレート分布が、同じハザード関数に対応することを述べよう。(8)式で

$$a=b=\frac{1}{\beta}>0 \text{ とおくと}$$

$$h(x) = \frac{\beta}{x+1} \quad (\beta>0) \quad (9)$$

となる。

【定理2】ハザード関数が(9)式となる離散確率変数はユール分布であり、その逆も言える。また、ハザード関数が(9)式となる連続確率変数はパレート分布であり、その逆も成り立つ。

(証明)

(i) 離散ケース

a) (9)式のようなハザード関数を仮定する。

$$P(X \geq x+1) = \frac{x+1}{\beta} p_x$$

同様に

$$P(X \geq x) = \frac{x}{\beta} p_{x-1}$$

であるから

$$p_x = \frac{x}{\beta} p_{x-1} - \frac{x+1}{\beta} p_x$$

変形して

$$p_x = \frac{x}{\beta+1+x} p_{x-1}$$

定理1で証明したように、この式を満たす確率変数はユール分布である。

b) (4)式で表現されるユール分布を仮定する。
ハザード関数を計算するために次式の差分を求めよう。

$$\begin{aligned} \Delta \left[\frac{(x+2)^{\overline{k-1}}}{(x+\beta+2)^{\overline{k-1}}} \right] &= \frac{(x+2)^{\overline{k}}}{(x+\beta+2)^{\overline{k}}} - \frac{(x+2)^{\overline{k-1}}}{(x+\beta+2)^{\overline{k-1}}} \\ &= \frac{(x+2)^{\overline{k-1}}}{(x+\beta+2)^{\overline{k-1}}} \left(\frac{x+k+1}{x+\beta+k+1} - 1 \right) \\ &= \frac{(x+2)^{\overline{k-1}}}{(x+\beta+2)^{\overline{k}}} (-\beta) \end{aligned}$$

よって

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{(x+2)^{\overline{k-1}}}{(x+\beta+2)^{\overline{k}}} &= -\frac{1}{\beta} \sum_{k=1}^{\infty} \left(\Delta \left[\frac{(x+2)^{\overline{k-1}}}{(x+\beta+2)^{\overline{k-1}}} \right] \right) \\ &= -\frac{1}{\beta} \left(\frac{(x+2)^{\overline{k-1}}}{(x+\beta+2)^{\overline{k-1}}} \Big|_1^{\infty} \right) \\ &= -\frac{1}{\beta} (0-1) = \frac{1}{\beta} \end{aligned}$$

生存関数は

$$\begin{aligned} \sum_{k=x+1}^{\infty} p_k &= \sum_{k=x+1}^{\infty} \frac{\beta k!}{(\beta+1)^{\overline{k+1}}} \\ &= \frac{\beta(x+1)!}{(\beta+1)^{\overline{x+1}}} \left\{ \frac{1}{x+\beta+2} + \frac{x+2}{(x+\beta+2)(x+\beta+3)} + \frac{(x+2)(x+3)}{(x+\beta+2)(x+\beta+3)(x+\beta+4)} + \dots \right\} \\ &= \frac{\beta(x+1)!}{(\beta+1)^{\overline{x+1}}} \sum_{k=1}^{\infty} \frac{(x+2)^{\overline{k-1}}}{(x+\beta+2)^{\overline{k}}} \\ &= \frac{\beta(x+1)!}{(\beta+1)^{\overline{x+1}}} \cdot \frac{1}{\beta} \\ &= \frac{(x+1)!}{(\beta+1)^{\overline{x+1}}} \end{aligned}$$

ハザード関数は

$$\begin{aligned} h(x) &= \frac{p_x}{\sum_{k=x+1}^{\infty} p_k} = \frac{\beta x!}{(\beta+1)^{\overline{x+1}}} \cdot \frac{(\beta+1)^{\overline{x+1}}}{(x+1)!} \\ &= \frac{\beta}{x+1} \end{aligned}$$

となり、(9)式に一致する。

(ii) 連続ケース

a) (9)式のようなハザード関数を仮定する。

$$h(x) = \frac{f(x)}{S(x)} = \frac{\beta}{x+1} \quad (\beta > 0)$$

$\frac{dS(x)}{dx} = -f(x)$ であるから、上式より

$$-\frac{1}{S(x)} \frac{dS(x)}{dx} = \frac{\beta}{x+1}$$

微分方程式を解いて

$$S(x) = k(x+1)^{-\beta}$$

$S(0) = 1$ より、 $k=1$ であるから

$$S(x) = (x+1)^{-\beta}$$

したがって確率密度関数は

$$f(x) = -\frac{d}{dx} S(x) = \beta(x+1)^{-\beta-1} \quad (x > 0)$$

となり、(3)式において $\alpha=1$ としたケースに一致し、パレート分布になる。

b) (3)式で表現されるパレート分布を仮定する。ただし $\alpha=1$ とする。生存関数は

$$\begin{aligned} S(x) &= \int_x^{\infty} \beta(t+1)^{-\beta-1} dt \\ &= \beta \left[-\frac{1}{\beta} (t+1)^{-\beta} \right]_x^{\infty} \end{aligned}$$

$\beta > 0$ であるから

$$S(x) = (x+1)^{-\beta}$$

ハザード関数は

$$h(x) = \frac{f(x)}{S(x)} = \frac{\beta}{x+1} \quad (x > 0)$$

となり、(9)式のハザード関数に一致する。

【4】一般化ユール分布

ユール分布がパレート分布に漸近するというのは、従来(4)式において x が大きな値をとるようになった場合が述べられてきた。(4)式は次のように変形される。

$$P_x = \frac{\beta x!}{(\beta+1)^{x+1}} \\ = \frac{\beta \Gamma(x+1) \Gamma(\beta+1)}{\Gamma(x+\beta+2)}$$

ここで $\frac{\Gamma(x+1)}{\Gamma(x+\beta+2)}$ の部分が $x \rightarrow \infty$ のとき $x^{-\beta-1}$ に漸近するからである。

本稿では x が小さな値でもパレート分布に漸近するよう、ユール分布を一般化しよう。

【定義】

$$P(x) = \frac{B\left(\frac{x+\alpha}{h}, \beta+1\right)}{B\left(\frac{\alpha}{h}, \beta\right)} \\ (x=0, h, 2h, 3h, \dots) \quad (10)$$

を「一般化ユール分布」と呼ぶことにする。ここで $B(\cdot, \cdot)$ はベータ関数、区間幅 $h > 0$, $\alpha > 0$, $\beta > 0$ である。□

(10)式のベータ関数を変形すると

$$P(x) = \frac{\Gamma\left(\frac{x+\alpha}{h}\right) \Gamma(\beta+1)}{\Gamma\left(\frac{x+\alpha}{h} + \beta + 1\right)} \cdot \frac{\Gamma\left(\frac{\alpha}{h} + \beta\right)}{\Gamma\left(\frac{\alpha}{h}\right) \Gamma(\beta)} \\ = \beta \frac{\Gamma\left(\frac{x+\alpha}{h}\right)}{\Gamma\left(\frac{x+\alpha}{h} + \beta + 1\right)} \cdot \frac{\Gamma\left(\frac{\alpha}{h} + \beta\right)}{\Gamma\left(\frac{\alpha}{h}\right)}$$

パラメータ α は $(x+\alpha)$ という形のほかに $\frac{\Gamma\left(\frac{\alpha}{h} + \beta\right)}{\Gamma\left(\frac{\alpha}{h}\right)}$ にも関係している。前者は x 軸の下限を α にするという意味であり位置を示している。後者は分布の形に影響を与えている。したがって、パラメータ α は主として位置パラメータとしての性格をもっている。パラメータ β は主として分布の形状を決めている。

(10)式において $h=1$, $\alpha=1$ とすれば

$$P(x) = \frac{B(x+1, \beta+1)}{B(1, \beta)} \\ (x=0, 1, 2, 3, \dots) \quad (11)$$

これを変形すると

$$P(x) = \beta \frac{\Gamma(x+1)}{\Gamma(x+\beta+2)} \cdot \frac{\Gamma(\beta+1)}{\Gamma(1)} \\ = \frac{\beta x!}{(\beta+1)^{x+1}}$$

となり、通常のユール分布の表現になる。

(10)式で定義したものが確率分布になることを証明しよう。

【補助定理3】 $x=0, h, 2h, 3h, \dots$ のとき

$$B\left(\frac{x+\alpha}{h}, \beta+1\right) \\ = -\frac{1}{\beta} \cdot \Delta \left[\frac{\Gamma\left(\frac{x+\alpha}{h}\right) \Gamma(\beta+1)}{\Gamma\left(\frac{x+\alpha}{h} + \beta\right)} \right] \quad (12)$$

(証明)

$$\Delta \left[\frac{\Gamma\left(\frac{x+\alpha}{h}\right) \Gamma(\beta+1)}{\Gamma\left(\frac{x+\alpha}{h} + \beta\right)} \right] \\ = \frac{\Gamma\left(\frac{x+h+\alpha}{h}\right) \Gamma(\beta+1)}{\Gamma\left(\frac{x+h+\alpha}{h} + \beta\right)} - \frac{\Gamma\left(\frac{x+\alpha}{h}\right) \Gamma(\beta+1)}{\Gamma\left(\frac{x+\alpha}{h} + \beta\right)} \\ = \frac{\Gamma\left(\frac{x+\alpha}{h}\right) \Gamma(\beta+1)}{\Gamma\left(\frac{x+\alpha}{h} + \beta\right)} \left(\frac{\frac{x+\alpha}{h}}{\frac{x+\alpha}{h} + \beta} - 1 \right) \\ = \frac{\Gamma\left(\frac{x+\alpha}{h}\right) \Gamma(\beta+1)}{\Gamma\left(\frac{x+h+\alpha}{h} + \beta\right)} (-\beta) \\ = (-\beta) B\left(\frac{x+\alpha}{h}, \beta+1\right)$$

よって(12)式が成り立つ。(証明終わり)

【補助定理4】 $x=0, h, 2h, 3h, \dots$ のとき

$$\sum_{x=0}^{\infty} B\left(\frac{x+\alpha}{h}, \beta+1\right) = B\left(\frac{\alpha}{h}, \beta\right)$$

(証明) (12)式より

$$\sum_{x=0}^{\infty} B\left(\frac{x+\alpha}{h}, \beta+1\right) \\ = -\frac{1}{\beta} \cdot \sum_{x=0}^{\infty} \Delta \left[\frac{\Gamma\left(\frac{x+\alpha}{h}\right) \Gamma(\beta+1)}{\Gamma\left(\frac{x+\alpha}{h} + \beta\right)} \right]$$

$$\begin{aligned}
&= -\frac{\Gamma(\beta+1)}{\beta} \left(\frac{\Gamma(\frac{x+\alpha}{h})}{\Gamma(\frac{x+\alpha}{h} + \beta)} \Big|_0^{\infty} \right) \\
&= -\Gamma(\beta) \left(0 - \frac{\Gamma(\frac{\alpha}{h})}{\Gamma(\frac{\alpha}{h} + \beta)} \right) \\
&= \frac{\Gamma(\frac{\alpha}{h}) \Gamma(\beta)}{\Gamma(\frac{\alpha}{h} + \beta)} \\
&= B\left(\frac{\alpha}{h}, \beta\right) \quad (\text{証明終わり})
\end{aligned}$$

補助定理 4 より (10) 式は確率分布になる。

一般化ユール分布の区間幅 h が 0 に漸近すると、その極限分布はパレート分布になる。その場合、

確率密度関数は $\lim_{h \rightarrow 0} \frac{P(x)}{h}$ である。

【定理 3】

$x=0, h, 2h, 3h, \dots$ のとき

$$\lim_{h \rightarrow 0} \frac{B(\frac{x+\alpha}{h}, \beta+1)}{hB(\frac{\alpha}{h}, \beta)} = \beta \alpha^\beta (x+\alpha)^{-\beta-1}$$

(証明)

$$\begin{aligned}
&\frac{B(\frac{x+\alpha}{h}, \beta+1)}{B(\frac{\alpha}{h}, \beta)} \\
&= \beta \frac{\Gamma(\frac{x+\alpha}{h})}{\Gamma(\frac{x+\alpha}{h} + \beta + 1)} \cdot \frac{\Gamma(\frac{\alpha}{h} + \beta)}{\Gamma(\frac{\alpha}{h})}
\end{aligned}$$

ここで $\frac{x}{h} = 0, 1, 2, 3, \dots$ であるから

$$\begin{aligned}
&\frac{\Gamma(\frac{x+\alpha}{h})}{\Gamma(\frac{x+\alpha}{h} + \beta + 1)} \\
&= \frac{1}{(\frac{x+\alpha}{h})(\frac{x+\alpha}{h} + 1) \cdots (\frac{x+\alpha}{h} + \beta)} \\
&= \frac{1}{h^{\beta+1} (x+\alpha)(x+\alpha+h) \cdots (x+\alpha+h\beta)}
\end{aligned}$$

同様に

$$\frac{\Gamma(\frac{\alpha}{h} + \beta)}{\Gamma(\frac{\alpha}{h})} = \left(\frac{\alpha}{h}\right) \left(\frac{\alpha}{h} + 1\right) \cdots \left(\frac{\alpha}{h} + \beta - 1\right)$$

$$= \frac{(\alpha)(\alpha+h) \cdots (\alpha+h(\beta-1))}{h^\beta}$$

したがって

$$\begin{aligned}
&\frac{B(\frac{x+\alpha}{h}, \beta+1)}{hB(\frac{\alpha}{h}, \beta)} \\
&= \beta \frac{(\alpha)(\alpha+h) \cdots (\alpha+h(\beta-1))}{(x+\alpha)(x+\alpha+h) \cdots (x+\alpha+h\beta)}
\end{aligned}$$

ここで $h \rightarrow 0$ とすると、右辺は $\beta \alpha^\beta (x+\alpha)^{-\beta-1}$ となる。(証明終わり)

【5】パレート分布と一般化ユール分布の乖離度

一般化ユール分布がパレート分布からどの程度乖離しているかをみよう。(3) 式のパレート分布で、 x から $x+h$ までの確率 $\Delta F(x)$ は

$$\begin{aligned}
\Delta F(x) &= F(x+h) - F(x) \\
&= \int_x^{x+h} \beta \alpha^\beta (t+\alpha)^{-\beta-1} dt \\
&= \alpha^\beta \left\{ \frac{1}{(x+\alpha)^\beta} - \frac{1}{(x+\alpha+h)^\beta} \right\}
\end{aligned}$$

これに対応する一般化ユール分布の確率 $P(x)$ は (10) 式から

$$P(x) = \beta \frac{\Gamma(\frac{x+\alpha}{h})}{\Gamma(\frac{x+\alpha}{h} + \beta + 1)} \cdot \frac{\Gamma(\frac{\alpha}{h} + \beta)}{\Gamma(\frac{\alpha}{h})}$$

である。

$\beta=1$ とすると

$$\Delta F(x) = P(x) = \frac{ah}{(x+\alpha)(x+\alpha+h)}$$

であり、一致する。また、 β の値が 1 から離れるほど、両者の確率は乖離する。

乖離度を

$$\left(\frac{\Delta F(x)}{P(x)} - 1 \right) \times 100 \quad (\%)$$

とする。 $\beta=1$ の場合には乖離度は 0 である。

区間幅 $h=1, 0.1, 0.05$ および $\beta=2, 3, \frac{1}{2}, \frac{1}{3}$ のケースについて乖離度を計算した結果が表およびグラフである。エクセルの計算能力の関係で、区間幅 $0.1, 0.05$ の場合には x のすべての値は計算されていない。

(表) パレート分布からの乖離度 (%)

幅 $h = 1$

x	$\beta = 2$	$\beta = 3$	$\beta = 1/2$	$\beta = 1/3$
0	12.5	16.7	-12.1	17.2
1	-16.7	-41.4	-2.7	21.1
2	-27.1	-57.2	1.5	20.3
3	-32.5	-64.4	3.9	18.6
4	-35.8	-68.5	5.5	16.7
5	-38.1	-71.2	6.6	14.9
6	-39.7	-73.1	7.4	13.2
7	-41.0	-74.4	8.0	11.6
8	-41.9	-75.5	8.5	10.2
9	-42.7	-76.3	8.9	8.8
10	-43.4	-77.0	9.2	7.6
11	-43.9	-77.5	9.5	6.4
12	-44.4	-78.0	9.8	5.4
13	-44.8	-78.4	10.0	4.3
14	-45.1	-78.7	10.2	3.4
15	-45.4	-79.0	10.3	2.5
16	-45.7	-79.3	10.5	1.6
17	-45.9	-79.5	10.6	0.8
18	-46.1	-79.7	10.7	0.1
19	-46.3	-79.9	10.8	-0.6
20	-46.5	-80.1	10.9	-1.3
21	-46.6	-80.2	11.0	-2.0
22	-46.8	-80.4	11.1	-2.6
23	-46.9	-80.5	11.1	-3.2
24	-47.0	-80.6	11.2	-3.8
25	-47.2	-80.7	11.3	-4.3
26	-47.3	-80.8	11.3	-4.9
27	-47.4	-80.9	11.4	-5.4
28	-47.4	-81.0	11.4	-5.9
29	-47.5	-81.1	11.5	-6.3
30	-47.6	-81.1	11.5	-6.8

幅 $h = 0.1$

x	$\beta = 2$	$\beta = 3$	$\beta = 1/2$	$\beta = 1/3$
0	4.1	7.8	-2.3	44.3
1	-2.4	-8.7	-0.6	31.7
2	-4.6	-13.9	0.0	24.1
3	-5.7	-16.6	0.3	18.8
4	-6.4	-18.1	0.5	14.8
5	-6.8	-19.1	0.6	11.5
6	-7.2	-19.9	0.7	8.9
7	-7.4	-20.4	0.8	6.6
8	-7.6	-20.9	0.8	4.6
9	-7.7	-21.2	0.9	2.8
10	-7.9	-21.5	0.9	1.2
11	-8.0	-21.7	0.9	-0.2
12	-8.0	-21.9	1.0	-1.5
13	-8.1	-22.1	1.0	-2.7
14	-8.2	-22.2	1.0	-3.7
15	-8.2	-22.3	1.0	-4.8

幅 $h = 0.05$

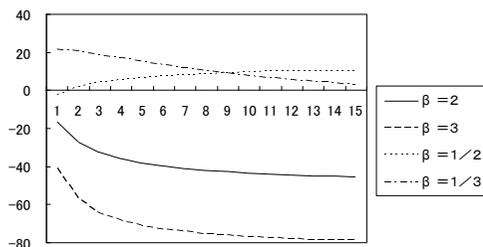
x	$\beta = 2$	$\beta = 3$	$\beta = 1/2$	$\beta = 1/3$
0	2.3	4.4	-1.2	47.0
1	-1.2	-4.6	-0.3	32.6
2	-2.4	-7.6	0.0	24.5
3	-3.0	-9.1	0.2	18.9
4	-3.3	-9.9	0.3	14.7
5	-3.6	-10.5	0.3	11.4
6	-3.7	-10.9	0.4	8.7
7	-3.9	-11.2	0.4	6.3

$$\text{乖離度}(\%) = \left(\frac{\text{パレート分布の確率}}{\text{ユール分布の確率}} - 1 \right) \times 100$$

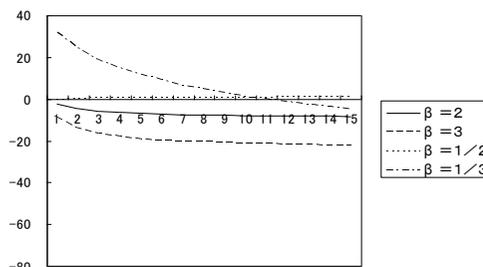
x の値は、たとえば幅 $h = 0.1$ で $x = 1$ のケースでは
1 から 1.1 の区間に対応するもの。

乖離度のグラフ (横軸は x の値, 縦軸は乖離度 %)

幅 $h=1$ のケース



幅 $h=0.1$ のケース



$\beta=1$ のケースはすべての場合において乖離度 0 である。したがって、X軸に一致するので描いていない。

[6] 従来の論文で用いられている “generalized Yule distribution”

“generalized Yule distribution” という用語をネットで検索すると主に 2 つの論文で用いられている。(1) Simon (1955) の 433 ページ, (2) Johnson & Kotz (1989) の 16 ページである。また, (3) Wikipedia の “Yule-Simon distribution” にも掲載されている。それぞれの内容について検討しよう。

(i) Simon の論文 (3・8) 式で

$$\frac{f^*(i)}{f^*(i-1)} = \frac{\lambda(i-1+c)}{\lambda(i+c)+1} = \frac{i-1+c}{i+c+\frac{1}{\lambda}}$$

$$(i = 2, 3, 4, \dots)$$

$$\lambda = \frac{k(1-\alpha)}{k+c}$$

を満たす $f^*(i)$ が generalized Yule distribution であると呼ばれている。ここで $0 \leq \alpha \leq 1$, $0 \leq c$ を想定している。

本稿での一般化ユール分布との関係を見るために, 記号を本稿のものに変更して上式を計算して

みよう。 i のかわりに x で置き換える。本稿で α を別の記号として用いているので, a とする。 f^* のかわりに P とする。上式は

$$\frac{P(x)}{P(x-1)} = \frac{\lambda(x-1+c)}{\lambda(x+c)+1} = \frac{x-1+c}{x+c+\frac{1}{\lambda}}$$

と書き換えられる。以下のように変形して $P(x)$ を求める。

$$P(x) = \frac{x-1+c}{x+c+\frac{1}{\lambda}} P(x-1)$$

$$= \frac{x-1+c}{x+c+\frac{1}{\lambda}} \cdot \frac{x-2+c}{x-1+c+\frac{1}{\lambda}} P(x-2)$$

$$\dots$$

$$= \frac{x-1+c}{x+c+\frac{1}{\lambda}} \cdot \frac{x-2+c}{x-1+c+\frac{1}{\lambda}}$$

$$\dots \frac{1+c}{2+c+\frac{1}{\lambda}} P(1)$$

$$= \frac{\Gamma(x+c)}{\Gamma(1+c)} \cdot \frac{\Gamma(2+c+\frac{1}{\lambda})}{\Gamma(x+1+c+\frac{1}{\lambda})} P(1)$$

$\sum_{x=1}^{\infty} P(x) = 1$ から $P(1)$ の値を求めよう。そのた

めの準備として次式を計算する。

$$\Delta \left[\frac{\Gamma(x+c)}{\Gamma(x+c+\frac{1}{\lambda})} \right]$$

$$= \frac{\Gamma(x+1+c)}{\Gamma(x+1+c+\frac{1}{\lambda})} - \frac{\Gamma(x+c)}{\Gamma(x+c+\frac{1}{\lambda})}$$

$$= \frac{\Gamma(x+c)}{\Gamma(x+c+\frac{1}{\lambda})} \left(\frac{x+c}{x+c+\frac{1}{\lambda}} - 1 \right)$$

$$= \left(-\frac{1}{\lambda} \right) \frac{\Gamma(x+c)}{\Gamma(x+1+c+\frac{1}{\lambda})}$$

したがって

$$\frac{\Gamma(x+c)}{\Gamma(x+1+c+\frac{1}{\lambda})} = (-\lambda) \Delta \left[\frac{\Gamma(x+c)}{\Gamma(x+c+\frac{1}{\lambda})} \right]$$

である。

$$\begin{aligned} \sum_{x=1}^{\infty} P(x) &= P(1) \left(1 - \lambda \cdot \frac{\Gamma(2+c+\frac{1}{\lambda})}{\Gamma(1+c)} \right. \\ &\quad \left. \cdot \sum_{x=2}^{\infty} \Delta \left[\frac{\Gamma(x+c)}{\Gamma(x+c+\frac{1}{\lambda})} \right] \right) \\ &= P(1) \left(1 + \lambda \cdot \frac{\Gamma(2+c+\frac{1}{\lambda})}{\Gamma(1+c)} \cdot \frac{\Gamma(2+c)}{\Gamma(2+c+\frac{1}{\lambda})} \right) \\ &= P(1) \{1 + \lambda(1+c)\} \\ \text{よって } P(1) &= \frac{1}{1 + \lambda(1+c)} \text{ であるから} \end{aligned}$$

$$\begin{aligned} P(x) &= \frac{\Gamma(x+c)}{\Gamma(1+c)} \cdot \frac{\Gamma(2+c+\frac{1}{\lambda})}{\Gamma(x+1+c+\frac{1}{\lambda})} \cdot \frac{1}{1 + \lambda(1+c)} \\ &= \frac{\Gamma(2+c+\frac{1}{\lambda})}{\Gamma(1+c+\frac{1}{\lambda})} \cdot \frac{1}{\Gamma(1+c)} \cdot \frac{\Gamma(x+c)}{\Gamma(x+1+c+\frac{1}{\lambda})} \\ &= \frac{\Gamma(1+c+\frac{1}{\lambda})}{\lambda} \cdot \frac{1}{\Gamma(1+c)\Gamma(1+\frac{1}{\lambda})} \\ &\quad \cdot \frac{\Gamma(x+c)\Gamma(1+\frac{1}{\lambda})}{\Gamma(x+c+1+\frac{1}{\lambda})} \\ &= \frac{\Gamma(1+c+\frac{1}{\lambda})}{\Gamma(1+c)\Gamma(\frac{1}{\lambda})} \cdot \frac{\Gamma(x+c)\Gamma(1+\frac{1}{\lambda})}{\Gamma(x+c+1+\frac{1}{\lambda})} \\ &= \frac{B(x+c, 1+\frac{1}{\lambda})}{B(1+c, \frac{1}{\lambda})} \end{aligned}$$

(10)式の表現と比較すると、 $h=1$ のケースに相当する。ただし、(10)式のとる値は $x=0, 1, 2, \dots$ であるが、上式は $x=1, 2, 3, \dots$ である。上式のとる値を $x=0, 1, 2$ に変更すると、それに対応する式は

$$P(x) = \frac{B(x+1+c, 1+\frac{1}{\lambda})}{B(1+c, \frac{1}{\lambda})} \quad (x=0, 1, 2, \dots)$$

となる。

(10)式との対応では $1+c=\alpha$ 、 $\frac{1}{\lambda}=\beta$ である。 $c=0$ とすれば(11)式で表現される通常のユール分布になる。したがって Simon のいう“generalized”とは、パラメータ α が通常のユール分布の場合1であるが、それが $1+c$ にシフトすることを述べている。

(ii) Johnson & Kotz の論文で用いられている記号を本稿に翻訳して記述しよう。14ページにある(1)式

$$P(X=x|T=t) = g(t)^x \{h(t)\}^x \quad (x=0, 1, 2, \dots)$$

において、 $g(t)=1-h(t)$ 、 $h(t)=\frac{t}{1+t}$ とし、 T の密度関数を

$$f_T(t) = \frac{1}{B(\alpha, \beta)} \cdot \frac{t^{\alpha-1}}{(1+t)^{\alpha+\beta}} \quad (t>0; \alpha, \beta>0)$$

とする。

$$P(X=x) = E_T[P(X=x|T)]$$

であるから

$$\begin{aligned} P(X=x) &= \frac{1}{B(\alpha, \beta)} \int_0^{\infty} \frac{1}{1+t} \\ &\quad \cdot \frac{t^x}{(1+t)^x} \cdot \frac{t^{\alpha-1}}{(1+t)^{\alpha+\beta}} dt \\ &= \frac{1}{B(\alpha, \beta)} \int_0^{\infty} \frac{t^{(x+\alpha)-1}}{(1+t)^{(x+\alpha)+(\beta+1)}} dt \end{aligned}$$

したがって

$$P(x) = \frac{B(x+\alpha, \beta+1)}{B(\alpha, \beta)} \quad (x=0, 1, 2, \dots) \quad (13)$$

これは本稿(10)式において $h=1$ としたケースになる。さらに $\alpha=1$ とすれば(11)式になるので、(13)式はユール分布のパラメータを α に一般化したものといえる。

Johnson & Kotz はさらに $h(t)=\left(\frac{t}{1+t}\right)^\gamma$ のケースも述べている。計算結果は

$$P(x) = \frac{B(x\gamma+\alpha, \beta) - B((x+1)\gamma+\alpha, \beta)}{B(\alpha, \beta)} \quad (x=0, 1, 2, \dots) \quad (14)$$

$\gamma=1$ とすれば(13)式になるので、さらに一般化したものになっているが、式の直感的な意味は分かりにくい。

(iii) Wikipedia の “Yule-Simon distribution” に掲載されている式は

$$f(k; \rho, \alpha) = \frac{\rho}{1 - \alpha^\rho} B_{1-\alpha}(k, \rho + 1) \quad (0 \leq \alpha \leq 1, x = 1, 2, 3, \dots)$$

である。ここで $B_{1-\alpha}$ は不完全ベータ関数を表す。本稿の記号に翻訳すれば

$$P(x) = \frac{\beta}{1 - c^\beta} B_{1-c}(x + 1, \beta + 1) \quad (0 \leq c \leq 1, x = 0, 1, 2, \dots) \quad (15)$$

ここで $c = 0$ とおけば(11)式に一致するので一般化といえる。

【7】定義域を一般化したユール分布

(10)式は定義域を $x = 0, h, 2h, 3h, \dots$ とし, x 軸上でとる値を $\alpha, \alpha + h, \alpha + 2h, \alpha + 3h, \dots$ としている。区間幅 $h = 1$ で, x 軸上の値を 1 から始める場合には, $x = 0, 1, 2, 3, \dots$, x 軸上の値は $1, 2, 3, 4, \dots$ となる。その場合の確率は(11)式あるいは(4)式で示されている。

本稿では x のとる定義域を 0 から始めるよう共通化して議論してきた。しかし, 一般には x のとる定義域と x 軸上でとる値を一致させることが普通である。すなわち, $x = 1, 2, 3, 4, \dots$ とし, x 軸上の値も $1, 2, 3, 4, \dots$ となるケースである。この場合の確率は

$$\begin{aligned} P(x) &= \frac{B(x, \beta + 1)}{B(1, \beta)} \\ &= \frac{\beta(x-1)!}{(\beta+1)^x} \quad (x = 1, 2, 3, \dots) \end{aligned}$$

と表現される。

x の定義域と x 軸上でとる値のいかんに関わらず, それら表現できる一般的なユール分布を定義しよう。

【定義】

定義域を $x = \gamma, \gamma + h, \gamma + 2h, \gamma + 3h, \dots$ とし, x 軸上でとる値を $\alpha, \alpha + h, \alpha + 2h, \alpha + 3h, \dots$ とする。一般化ユール分布を

$$P(x) = \frac{B(\frac{x+\alpha-\gamma}{h}, \beta + 1)}{B(\frac{\alpha}{h}, \beta)} \quad (16)$$

と定義する。 $h, \alpha, \beta, \gamma > 0$ である。□

これに対応するパレート分布は

$$f(x) = \beta \alpha^\beta (x + \alpha - \gamma)^{-\beta - 1} \quad (x > \gamma) \quad (17)$$

である。

通常はユール分布, パレート分布ともに $h = 1, \gamma = \alpha$ で表現する。

【補助定理 5】

(16)式は確率分布である。

(証明) 次の式がなりたつ。

$$\begin{aligned} &\frac{\Gamma(\frac{x+\alpha-\gamma}{h})}{\Gamma(\frac{x+\alpha-\gamma}{h} + \beta + 1)} \\ &= \left(-\frac{1}{\beta}\right) \cdot \Delta \left[\frac{\Gamma(\frac{x+\alpha-\gamma}{h})}{\Gamma(\frac{x+\alpha-\gamma}{h} + \beta)} \right] \end{aligned}$$

これを利用して, 以下の式を計算する。ただし, 区間幅は h である。

$$\begin{aligned} &\sum_{x=\gamma}^{\infty} P(x) \\ &= \frac{\Gamma(\beta + 1)}{B(\frac{\alpha}{h}, \beta)} \cdot \sum_{x=\gamma}^{\infty} \frac{\Gamma(\frac{x+\alpha-\gamma}{h})}{\Gamma(\frac{x+\alpha-\gamma}{h} + \beta + 1)} \\ &= \frac{\Gamma(\beta + 1)}{B(\frac{\alpha}{h}, \beta)} \cdot \left(-\frac{1}{\beta}\right) \cdot \sum_{x=\gamma}^{\infty} \Delta \left[\frac{\Gamma(\frac{x+\alpha-\gamma}{h})}{\Gamma(\frac{x+\alpha-\gamma}{h} + \beta)} \right] \\ &= -\frac{\Gamma(\beta)}{B(\frac{\alpha}{h}, \beta)} \left(0 - \frac{\Gamma(\frac{\alpha}{h})}{\Gamma(\frac{\alpha}{h} + \beta)} \right) \\ &= 1 \end{aligned}$$

したがって(16)式は確率分布である。(証明終わり)

【8】ユール分布とパレート分布の特性

(i) 分布関数とハザード関数

【補助定理6】

(16)式の一般化ユール分布の確率関数を $P(x)$, 分布関数を $F(x)$ とすると

$$F(x) = 1 - \frac{x + \alpha - \gamma}{\beta} \cdot \frac{P(x)}{h} \quad (18)$$

また, (17)式のパレート分布の確率密度関数を $f(x)$, 分布関数を $F(x)$ とすると

$$F(x) = 1 - \frac{x + \alpha - \gamma}{\beta} f(x) \quad (19)$$

(証明)

ユール分布の分布関数を求めるために次の計算をしておこう。

$$\begin{aligned} \Delta \left[\frac{\Gamma\left(\frac{k+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{k+\alpha-\gamma}{h} + \beta\right)} \right] &= \frac{\Gamma\left(\frac{k+h+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{k+h+\alpha-\gamma}{h} + \beta\right)} - \frac{\Gamma\left(\frac{k+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{k+\alpha-\gamma}{h} + \beta\right)} \\ &= \frac{\Gamma\left(\frac{k+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{k+\alpha-\gamma}{h} + \beta\right)} \cdot \frac{(-\beta)}{\frac{k+\alpha-\gamma}{h} + \beta} \end{aligned}$$

よって

$$\frac{\Gamma\left(\frac{k+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{k+\alpha-\gamma}{h} + \beta + 1\right)} = -\frac{1}{\beta} \cdot \Delta \left[\frac{\Gamma\left(\frac{k+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{k+\alpha-\gamma}{h} + \beta\right)} \right]$$

この式を用いて以下の計算をする。

$$\begin{aligned} F(x) &= \sum_{k=\gamma}^x P(k) \\ &= \sum_{k=\gamma}^x \frac{B\left(\frac{k+\alpha-\gamma}{h}, \beta + 1\right)}{B\left(\frac{\alpha}{h}, \beta\right)} \\ &= \frac{\Gamma(\beta + 1)}{B\left(\frac{\alpha}{h}, \beta\right)} \sum_{k=\gamma}^x \frac{\Gamma\left(\frac{k+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{k+\alpha-\gamma}{h} + \beta + 1\right)} \\ &= \frac{\Gamma(\beta + 1)}{B\left(\frac{\alpha}{h}, \beta\right)} \left(-\frac{1}{\beta}\right) \sum_{k=\gamma}^x \Delta \left[\frac{\Gamma\left(\frac{k+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{k+\alpha-\gamma}{h} + \beta\right)} \right] \end{aligned}$$

$$\begin{aligned} &= -\frac{\Gamma\left(\frac{\alpha}{h} + \beta\right)}{\Gamma\left(\frac{\alpha}{h}\right)\Gamma(\beta)} \cdot \frac{\Gamma(\beta + 1)}{\beta} \\ &\quad \cdot \left(\frac{\Gamma\left(\frac{x+h+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+h+\alpha-\gamma}{h} + \beta\right)} - \frac{\Gamma\left(\frac{\alpha}{h}\right)}{\Gamma\left(\frac{\alpha}{h} + \beta\right)} \right) \\ &= 1 - \frac{\Gamma\left(\frac{\alpha}{h} + \beta\right)}{\Gamma\left(\frac{\alpha}{h}\right)} \cdot \frac{\Gamma\left(\frac{x+\alpha-\gamma}{h} + 1\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta + 1\right)} \\ &= 1 - \frac{\Gamma\left(\frac{\alpha}{h} + \beta\right)}{\Gamma\left(\frac{\alpha}{h}\right)\Gamma(\beta)} \cdot \frac{x + \alpha - \gamma}{h} \\ &\quad \cdot \frac{\Gamma\left(\frac{x+\alpha-\gamma}{h}\right)\Gamma(\beta + 1)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta + 1\right)} \cdot \frac{\Gamma(\beta)}{\Gamma(\beta + 1)} \\ &= 1 - \frac{x + \alpha - \gamma}{h\beta} \cdot \frac{B\left(\frac{x+\alpha-\gamma}{h}, \beta + 1\right)}{B\left(\frac{\alpha}{h}, \beta\right)} \\ &= 1 - \frac{x + \alpha - \gamma}{\beta} \cdot \frac{P(x)}{h} \end{aligned}$$

パレート分布の分布関数については以下のようにある。

$$\begin{aligned} F(x) &= \int_{\gamma}^x \beta \alpha^{\beta} (t + \alpha - \gamma)^{-\beta - 1} dt \\ &= 1 - \alpha^{\beta} (x + \alpha - \gamma)^{-\beta} \\ &= 1 - \frac{x + \alpha - \gamma}{\beta} f(x) \quad (\text{証明終わり}) \end{aligned}$$

【補助定理7】

(16)式の一般化ユール分布および(17)式のパレート分布のハザード関数は

$$h(x) = \frac{\beta}{x + \alpha - \gamma} \quad (20)$$

(証明)

(6)式で定義したハザード関数の分子は確率密度関数 $f(x)$ であるので, (7)式における分子は p_x を区間幅で割った値が相当する。通常は区間幅が1であるので分子は p_x になっている。しかし, ここでのケースは区間幅が h であるので, 分子は $\frac{P(x)}{h}$ になる。

$$h(x) = \frac{\frac{P(x)}{h}}{1 - F(x+0)} = \frac{P(x)}{h \sum_{k=x+h}^{\infty} P(k)}$$

$$= \frac{\beta}{x + \alpha - \gamma}$$

パレート分布のハザード関数についても(6)式と(19)式から(20)式が言える。(証明終わり)

(ii) 平均

【補助定理 8】

(16)式的一般化ユール分布および(17)式のパレート分布の平均は

$$\mu = \frac{\beta\gamma + \alpha - \gamma}{\beta - 1} \quad (21)$$

(証明)

一般化ユール分布の平均を求めるさい次の計算を用いる。

$$\frac{x\Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta + 1\right)}$$

$$= \frac{1}{1-\beta} \cdot \Delta \left[\frac{x\Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta\right)} \right]$$

$$- \frac{\alpha - \gamma}{1-\beta} \cdot \frac{\Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta + 1\right)}$$

一般化ユール分布の平均は

$$\sum_{x=\gamma}^{\infty} xP(x)$$

$$= \frac{\Gamma(\beta+1)}{B\left(\frac{\alpha}{h}, \beta\right)} \sum_{x=\gamma}^{\infty} \frac{x\Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta + 1\right)}$$

$$= \frac{\Gamma(\beta+1)}{B\left(\frac{\alpha}{h}, \beta\right)} \left\{ \frac{1}{1-\beta} \sum_{x=\gamma}^{\infty} \Delta \left[\frac{x\Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta\right)} \right] \right.$$

$$\left. - \frac{\alpha - \gamma}{1-\beta} \sum_{x=\gamma}^{\infty} \frac{\Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta + 1\right)} \right\}$$

ここで $\Sigma \Delta []$ の部分が収束するためには $\beta > 1$ が必要である。

$$= \frac{\Gamma(\beta+1)}{B\left(\frac{\alpha}{h}, \beta\right)} \left\{ \frac{1}{1-\beta} \left(-\frac{\gamma \Gamma\left(\frac{\alpha}{h}\right)}{\Gamma\left(\frac{\alpha}{h} + \beta\right)} \right) \right.$$

$$\left. - \frac{\alpha - \gamma}{1-\beta} \sum_{x=\gamma}^{\infty} \frac{\Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta + 1\right)} \right\}$$

$$= \frac{\beta\gamma}{\beta-1} + \frac{\alpha - \gamma}{\beta-1}$$

パレート分布の平均は

$$\int_{\gamma}^{\infty} x\beta\alpha^{\beta}(x+\alpha-\gamma)^{-\beta-1} dx$$

$$= \beta\alpha^{\beta} \left\{ \left[x \left(\frac{1}{-\beta} \right) (x+\alpha-\gamma)^{-\beta} \right]_{\gamma}^{\infty} \right.$$

$$\left. - \int_{\gamma}^{\infty} \left(\frac{1}{-\beta} \right) (x+\alpha-\gamma)^{-\beta} dx \right\}$$

$$= \beta\alpha^{\beta} \left\{ \left[x \left(\frac{1}{-\beta} \right) (x+\alpha-\gamma)^{-\beta} \right]_{\gamma}^{\infty} \right.$$

$$\left. - \left[\left(\frac{1}{-\beta} \cdot \frac{1}{-\beta+1} \right) (x+\alpha-\gamma)^{-\beta+1} \right]_{\gamma}^{\infty} \right\}$$

ここで収束するためには $\beta > 1$ が必要である。

$$= \beta\alpha^{\beta} \left\{ \frac{\gamma\alpha^{-\beta}}{\beta} + \frac{\alpha^{-\beta+1}}{\beta(\beta-1)} \right\}$$

$$= \gamma + \frac{\alpha}{\beta-1} \quad (\text{証明終わり})$$

$\gamma = \alpha$ のときには平均が $\frac{\beta\alpha}{\beta-1}$ になる。また

$\gamma = 0$ のときには $\frac{\alpha}{\beta-1}$ になる。

(iii) 分散

【補助定理 9】

(16)式的一般化ユール分布の分散は

$$\sigma^2 = \frac{\beta\alpha^2 + \beta(\beta-1)ah}{(\beta-1)^2(\beta-2)} \quad (22)$$

(証明)

$\sigma^2 = E(X^2) - \{E(X)\}^2$ であるから、 $E(X^2)$ を計算しよう。そのさい次の計算式を用いる。

$$\begin{aligned} & \frac{x^2 \Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta + 1\right)} \\ &= \frac{1}{2-\beta} \cdot \Delta \left[\frac{x^2 \Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta\right)} \right] \\ & \quad - \frac{2(\alpha-\gamma)+h}{2-\beta} \cdot \frac{x \Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta + 1\right)} \\ & \quad - \frac{(\alpha-\gamma)h}{2-\beta} \cdot \frac{\Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta + 1\right)} \end{aligned}$$

 $E(X^2)$ は

$$\begin{aligned} & \sum_{x=\gamma}^{\infty} x^2 P(x) \\ &= \frac{\Gamma(\beta+1)}{B\left(\frac{\alpha}{h}, \beta\right)} \sum_{x=\gamma}^{\infty} \frac{x^2 \Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta + 1\right)} \\ &= \frac{\Gamma(\beta+1)}{B\left(\frac{\alpha}{h}, \beta\right)} \\ & \quad \left\{ \frac{1}{2-\beta} \sum_{x=\gamma}^{\infty} \Delta \left[\frac{x^2 \Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta\right)} \right] \right. \\ & \quad - \frac{2(\alpha-\gamma)+h}{2-\beta} \sum_{x=\gamma}^{\infty} \frac{x \Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta + 1\right)} \\ & \quad \left. - \frac{(\alpha-\gamma)h}{2-\beta} \sum_{x=\gamma}^{\infty} \frac{\Gamma\left(\frac{x+\alpha-\gamma}{h}\right)}{\Gamma\left(\frac{x+\alpha-\gamma}{h} + \beta + 1\right)} \right\} \end{aligned}$$

ここで $\Sigma \Delta [\]$ の部分が収束するためには $\beta > 2$ が必要である。

$$\begin{aligned} &= \frac{\Gamma(\beta+1)}{B\left(\frac{\alpha}{h}, \beta\right)} \cdot \frac{1}{2-\beta} \left\{ - \frac{\gamma^2 \Gamma\left(\frac{\alpha}{h}\right)}{\Gamma\left(\frac{\alpha}{h} + \beta\right)} \right\} \\ & \quad - \frac{2(\alpha-\gamma)+h}{2-\beta} E(X) \\ & \quad - \frac{(\alpha-\gamma)h}{2-\beta} \sum_{x=\gamma}^{\infty} P(x) \\ &= - \frac{\beta \gamma^2}{2-\beta} - \frac{2(\alpha-\gamma)+h}{2-\beta} \end{aligned}$$

$$\begin{aligned} & \cdot \frac{\beta \gamma + (\alpha - \gamma)}{\beta - 1} - \frac{(\alpha - \gamma)h}{2 - \beta} \\ &= \frac{\beta \gamma^2}{\beta - 2} + \frac{\{2(\alpha - \gamma) + h\} \{\beta \gamma + (\alpha - \gamma)\}}{(\beta - 1)(\beta - 2)} \\ & \quad + \frac{(\alpha - \gamma)h}{\beta - 2} \end{aligned}$$

分散は

$$\begin{aligned} \sigma^2 &= \frac{\beta \gamma^2}{\beta - 2} \\ & \quad + \frac{\{2(\alpha - \gamma) + h\} \{\beta \gamma + (\alpha - \gamma)\}}{(\beta - 1)(\beta - 2)} \\ & \quad + \frac{(\alpha - \gamma)h}{\beta - 2} - \frac{\beta^2 \alpha^2}{(\beta - 1)^2} \\ &= \frac{(\text{分子})}{(\beta - 1)^2 (\beta - 2)} \\ \text{分子} &= \beta \gamma^2 (\beta - 1)^2 + \{2(\alpha - \gamma) + h\} \{\beta \gamma \\ & \quad + (\alpha - \gamma)\} (\beta - 1) + (\alpha - \gamma)h (\beta - 1)^2 \\ & \quad - \beta^2 \alpha^2 (\beta - 2) \\ &= \beta^2 ch + \beta (\alpha^2 - ch) \end{aligned}$$

したがって、分散は(22)式のようになる。(証明終わり)

【補助定理10】

(17)式のパレート分布の分散は

$$\sigma^2 = \frac{\beta \alpha^2}{(\beta - 1)^2 (\beta - 2)} \quad (23)$$

(証明)

$$\begin{aligned} E(X^2) &= \int_{\gamma}^{\infty} x^2 \beta \alpha^{\beta} (x + \alpha - \gamma)^{-\beta - 1} dx \\ &= \beta \alpha^{\beta} \left\{ \left[x^2 \left(\frac{1}{-\beta} \right) (x + \alpha - \gamma)^{-\beta} \right]_{\gamma}^{\infty} \right. \\ & \quad \left. - \int_{\gamma}^{\infty} 2x \left(\frac{1}{-\beta} \right) (x + \alpha - \gamma)^{-\beta} dx \right\} \\ &= \beta \alpha^{\beta} \left(\frac{1}{\beta} \right) \gamma^2 \alpha^{-\beta} + 2\beta \alpha^{\beta} \left(\frac{1}{\beta} \right) \end{aligned}$$

$$\begin{aligned}
& \left[x \left(\frac{1}{-\beta+1} \right) (x+\alpha-\gamma)^{-\beta+1} \right]_{\gamma}^{\infty} \\
& - \int_{\gamma}^{\infty} \left(\frac{1}{-\beta+1} \right) (x+\alpha-\gamma)^{-\beta+1} dx \\
& = \gamma^2 + 2\beta\alpha^{\beta} \\
& \cdot \frac{1}{\beta(1-\beta)} \left[x(x+\alpha-\gamma)^{-\beta+1} \right]_{\gamma}^{\infty} \\
& - 2\beta\alpha^{\beta} \cdot \frac{1}{\beta(1-\beta)} \left[\left(\frac{1}{-\beta+2} \right) (x+\alpha-\gamma)^{-\beta+2} \right]_{\gamma}^{\infty}
\end{aligned}$$

ここで収束するためには $\beta > 2$ が必要である。

$$= \gamma^2 + \frac{2\alpha\gamma}{\beta-1} + \frac{2\alpha^2}{(\beta-1)(\beta-2)}$$

したがって分散は

$$\begin{aligned}
\sigma^2 &= \gamma^2 + \frac{2\alpha\gamma}{\beta-1} + \frac{2\alpha^2}{(\beta-1)(\beta-2)} \\
& - \frac{(\beta\gamma + \alpha - \gamma)^2}{(\beta-1)^2} \\
& = \frac{\beta\alpha^2}{(\beta-1)^2(\beta-2)} \quad (\text{証明終わり})
\end{aligned}$$

一般化ユール分布の分散である(22)式において、区間幅 h が 0 に近づくときパレート分布の分散である(23)式になる。

【注】

- (1) 蓑谷 (2003), Johnson, Kotz, Balakrishnan (1994) を参照
- (2) $f(x) = \beta\alpha^{\beta}x^{-\beta-1}$, ($x > \alpha$) と表現する方が普通である。本稿では x のとりうる範囲を 0 以上に統一して記述し、各式が比較しやすいようにした。そのため、(3)式のようにパラメータ α として式の中に明示した。
- (3) Irwin (1975) では $c = 0.5$ としている。
- (4) 離散確率変数のハザード関数は通常

$$h(x) = \frac{p_x}{S(x)} = \frac{p_x}{1 - F(x-0)} = \frac{p_x}{\sum_{k=x}^{\infty} p_k}$$

のように定義されている (Xekalaki (1983), 蓑谷千風

彦 (2003) の48ページを参照)。本稿では、パレート分布とユール分布がより自然に対応するように(7)式のように定義した。

【参考文献】

- J. O. Irwin (1975) “The Generalized Waring Distribution. Part I, Part II, Part III”, Journal of the Royal Statistical Society. Series A (General), Vol. 138, No. 1
- N. L. Johnson, S. Kotz (1989) “Characterization based on conditional distributions”, Annals of the Institute of Statistical Mathematics, Vol. 41, No. 1, p. 13-17
- N. L. Johnson, S. Kotz, N. Balakrishnan (1994) “Continuous Univariate Distributions, Volume 1, Second Edition”, John Wiley & Sons, Inc.
- H. A. Simon (1955) “On a Class of Skew Distribution Functions”, Biometrika, Vol. 42, No. 3/4, p. 425-440
- E. Xekalaki (1983) “Hazard functions and life distribution in discrete time”, Communication Statistics, Series A, Vol. 12, p. 2503-2509
- 蓑谷千風彦 (2003) 『統計分布ハンドブック』朝倉書店