〔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

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

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

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

ここで，$x+c_2>x+c_1>0$ を仮定し

$$g_1=\frac{1}{b_2}\cdot\frac{a-c_1}{c_2-c_1}\quad,\ g_2=\frac{1}{b_2}\cdot\frac{a-c_2}{c_2-c_1}$$

とおき，上式を積分する。

$$\int\frac{1}{f}\frac{df}{dx}=g_1\int\frac{1}{x+c_1}dx-g_2\int\frac{1}{x+c_2}dx$$
$$\log f=C+g_1\log(x+c_1)-g_2\log(x+c_2)$$

よって

$$f(x)=k(x+c_1)^{g_1}(x+c_2)^{-g_2} \tag{2}$$

が得られる。

【補助定理 1】 (1)式，(2)式において $a=c_1$，$b_2=-\dfrac{1}{b+1}$，$a=c_2-c_1>0$，$b>0$ とおくと

$$f(x)=ba^{b}(x+a)^{-b-1} \quad , \quad x>0 \qquad (3)$$

となり，パレート分布になる(注2)。

(証明)　$a=c_1$ であるから $g_1=0$，$g_2=-\dfrac{1}{b_2}=b+1$ になる。したがって(2)式は

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

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

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

を解けばよい。

$$\int_{-c_1}^{\infty} k(x+c_2)^{-b-1}dx=\left[\frac{k}{-b}(x+c_2)^{-b}\right]_{-c_1}^{\infty}$$

$$=\frac{k}{-b}\left\{0-(c_2-c_1)^{-b}\right\}$$

$$=\frac{k}{ba^{b}}$$

よって

$$k=ba^{b}$$

したがって

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

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。

$$x+c_2-c_1=x+a$$

であるので

$$f(x)=ba^{b}(x+a)^{-b-1} \quad , \quad x>0$$

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) $\dfrac{1}{f}\dfrac{df}{dx}$ 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 $x=0,1,2,3,\cdots$ 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 $\dfrac{df}{dx}$ 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，$cp_x+(1-c)p_{x-1}$，$(0\le c\le 1)$ 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。

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

とする。$x=0$，1，2，3,…のとき，一般超幾何分布の確率は

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

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。

$$p_x=\frac{\Gamma(c+1)\,\Gamma(c+1)}{\Gamma(c+2)\,\Gamma(c)}\cdot\frac{1^{\overline{x}}1^{\overline{x}}}{(c+2)^{\overline{x}}\,x\,!}$$

$$=\frac{c}{c+1}\cdot\frac{x!}{(c+2)^{\overline{x}}}$$

$$=\frac{cx!}{(c+1)^{\overline{x+1}}}$$

c を b で置き換えると，ユール分布は

$$p_x=\frac{b\,x!}{(b+1)^{\overline{x+1}}} \qquad (x=0,1,2,\cdots) \qquad (4)$$

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

$$p_x - p_{x-1} = p_{x-1}\left(\frac{x}{x+b+1}-1\right)$$
$$= -\frac{p_{x-1}}{x+b+1}\cdot(b+1)$$
$$cp_x+(1-c)p_{x-1} = p_{x-1}\left(\frac{cx}{x+b+1}+1-c\right)$$
$$= \frac{p_{x-1}}{x+b+1}\cdot\{cx+(1-c)(x+b+1)\}$$

したがって

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

ここで $a=(1-c)(b+1)$ とおく。類似する連続分布を求めるには

$$\frac{1}{f(x)}\frac{df(x)}{x} = -\frac{b+1}{x+a}$$

を解けばよい。

微分方程式を解くと

$$\log f(x) = C-(b+1)\int\frac{1}{x+a}dx$$
$$= C-(b+1)\ \log(x+a)$$

よって

$$f(x)=k(x+a)^{-b-1} \qquad (x>0)$$

k を求めるには

$$\int_0^\infty k(x+a)^{-b-1}dx=1$$
$$1=k\left[-\frac{1}{b}(x+a)^{-b}\right]_0^\infty=\frac{k}{b}a^{-b}$$

したがって

$$k=ba^b$$

求める確率密度関数は

$$f(x)=ba^b(x+a)^{-b-1} \qquad (x>0)$$

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 a を１にするためには $c=\dfrac{b}{b+1}$ 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

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

p_x を求めるため，以下の変形をする。

$$p_x=\frac{x}{x+b+1}p_{x-1}$$
$$=\frac{x}{x+b+1}\cdot\frac{x-1}{x+b}\cdot\frac{x-2}{x+b-1}$$
$$\cdots\frac{1}{b+2}\cdot p_0$$
$$=\frac{x!}{(b+2)^{\overline{x}}}\cdot p_0$$

ここで p_0 を計算するために，次の差分を計算しよう。

$$\Delta\left[\frac{x!}{(b+2)^{\overline{x-1}}}\right]=\frac{(x+1)!}{(b+2)^{\overline{x}}}-\frac{x!}{(b+2)^{\overline{x-1}}}$$
$$=\frac{x!}{(b+2)^{\overline{x-1}}}\left(\frac{x+1}{x+b+1}-1\right)$$
$$=-b\frac{x!}{(b+2)^{\overline{x}}}$$

$b>0$ だから

$$\sum_{x=1}^{\infty}\frac{x!}{(b+2)^{\overline{x}}}=-\frac{1}{b}\sum_{x=1}^{\infty}\left(\Delta\left[\frac{x!}{(b+2)^{\overline{x-1}}}\right]\right)$$
$$=-\frac{1}{b}\cdot\left(\left.\frac{x!}{(b+2)^{\overline{x-1}}}\right|_1^\infty\right)$$

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

である。

$$1 = \sum_{x=0}^{\infty} p_x = p_0 + p_0 \sum_{x=1}^{\infty} \frac{x!}{(b+2)^{\overline{x}}}$$
$$= p_0\left(1 + \frac{1}{b}\right)$$

よって

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

確率分布は

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

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

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

と定義される。

離散確率変数 X の場合には

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

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 $X(\ x=0,1,2,\cdots)$ については，ハザード関数を

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

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 $a=b=\dfrac{1}{b}>0$ とおくと

$$h(x) = \frac{b}{x+1} \qquad (\ b>0\ ) \tag{9}$$

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。

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

同様に

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

であるから

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

変形して

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

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。

$$\Delta\left[\frac{(x+2)^{\overline{k-1}}}{(x+b+2)^{\overline{k-1}}}\right]$$
$$=\frac{(x+2)^{\overline{k}}}{(x+b+2)^{\overline{k}}}-\frac{(x+2)^{\overline{k-1}}}{(x+b+2)^{\overline{k-1}}}$$
$$=\frac{(x+2)^{\overline{k-1}}}{(x+b+2)^{\overline{k-1}}}\left(\frac{x+k+1}{x+b+k+1}-1\right)$$
$$=\frac{(x+2)^{\overline{k-1}}}{(x+b+2)^{\overline{k}}}(-b)$$

よって

$$\sum_{k=1}^{\infty}\frac{(x+2)^{\overline{k-1}}}{(x+b+2)^{\overline{k}}}$$
$$=-\frac{1}{b}\sum_{k=1}^{\infty}\left(\Delta\left[\frac{(x+2)^{\overline{k-1}}}{(x+b+2)^{\overline{k-1}}}\right]\right)$$
$$=-\frac{1}{b}\left(\left.\frac{(x+2)^{\overline{k-1}}}{(x+b+2)^{\overline{k-1}}}\right|_{1}^{\infty}\right)$$
$$=-\frac{1}{b}(0-1)=\frac{1}{b}$$

生存関数は

$$\sum_{k=x+1}^{\infty}p_k=\sum_{k=x+1}^{\infty}\frac{b\,k!}{(b+1)^{\overline{k+1}}}$$
$$=\frac{b\,(x+1)!}{(b+1)^{\overline{x+1}}}\left\{\frac{1}{x+b+2}+\frac{x+2}{(x+b+2)(x+b+3)}+\frac{(x+2)(x+3)}{(x+b+2)(x+b+3)(x+b+4)}+\cdots\right\}$$
$$=\frac{b\,(x+1)!}{(b+1)^{\overline{x+1}}}\sum_{k=1}^{\infty}\frac{(x+2)^{\overline{k-1}}}{(x+b+2)^{\overline{k}}}$$
$$=\frac{b\,(x+1)!}{(b+1)^{\overline{x+1}}}\cdot\frac{1}{b}$$
$$=\frac{(x+1)!}{(b+1)^{\overline{x+1}}}$$

ハザード関数は

$$h(x)=\frac{p_x}{\sum_{k=x+1}^{\infty}p_k}=\frac{b\,x!}{(b+1)^{\overline{x+1}}}\cdot\frac{(b+1)^{\overline{x+1}}}{(x+1)!}$$
$$=\frac{b}{x+1}$$

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。

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

$\dfrac{dS(x)}{dx}=-f(x)$ であるから，上式より

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

微分方程式を解いて

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

$S(0)=1$ より，$k=1$ であるから

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

したがって確率密度関数は

$$f(x)=-\frac{d}{dx}S(x)=b\,(x+1)^{-b-1}\qquad(x>0)$$

となり，(3)式において $a=1$ 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 $a=1$ とする。生存関数は

$$S(x)=\int_x^{\infty}b\,(t+1)^{-b-1}dt$$
$$=b\left[-\frac{1}{b}(t+1)^{-b}\right]_x^{\infty}$$

$b>0$ であるから

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

ハザード関数は

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

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。

$$p_x = \frac{b\,x!}{(b+1)^{\overline{x+1}}} = \frac{b\,\Gamma(x+1)\,\Gamma(b+1)}{\Gamma(x+b+2)}$$

ここで $\dfrac{\Gamma(x+1)}{\Gamma(x+b+2)}$ の部分が $x\to\infty$ のとき x^{-b-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】

$$P(x) = \frac{B(\frac{x+a}{h}, b+1)}{B(\frac{a}{h}, b)} \qquad (x=0, h, 2h, 3h, \cdots) \tag{10}$$

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 $h>0$，$a>0$，$b>0$ である。□

(10)式のベータ関数を変形すると

$$P(x) = \frac{\Gamma(\frac{x+a}{h})\,\Gamma(b+1)}{\Gamma(\frac{x+a}{h}+b+1)} \cdot \frac{\Gamma(\frac{a}{h}+b)}{\Gamma(\frac{a}{h})\,\Gamma(b)} = b\,\frac{\Gamma(\frac{x+a}{h})}{\Gamma(\frac{x+a}{h}+b+1)} \cdot \frac{\Gamma(\frac{a}{h}+b)}{\Gamma(\frac{a}{h})}$$

パラメータ a は($x+a$)という形のほかに $\dfrac{\Gamma(\frac{a}{h}+b)}{\Gamma(\frac{a}{h})}$ にも関係している。前者は x 軸の下限を a 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 a 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 b は主として分布の形状を決めている。

(10)式において $h=1$，$a=1$ とすれば

$$P(x) = \frac{B(x+1, b+1)}{B(1, b)} \qquad (x=0, 1, 2, 3, \cdots) \tag{11}$$

これを変形すると

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

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】　$x=0, h, 2h, 3h, \cdots$ のとき

$$B\left(\frac{x+a}{h}, b+1\right) = -\frac{1}{b}\cdot\Delta\left[\frac{\Gamma(\frac{x+a}{h})\,\Gamma(b+1)}{\Gamma(\frac{x+a}{h}+b)}\right] \tag{12}$$

(証明)

$$\begin{aligned}
\Delta\left[\frac{\Gamma(\frac{x+a}{h})\,\Gamma(b+1)}{\Gamma(\frac{x+a}{h}+b)}\right]
&= \frac{\Gamma(\frac{x+h+a}{h})\,\Gamma(b+1)}{\Gamma(\frac{x+h+a}{h}+b)} - \frac{\Gamma(\frac{x+a}{h})\,\Gamma(b+1)}{\Gamma(\frac{x+a}{h}+b)} \\
&= \frac{\Gamma(\frac{x+a}{h})\,\Gamma(b+1)}{\Gamma(\frac{x+a}{h}+b)}\left(\frac{\frac{x+a}{h}}{\frac{x+a}{h}+b} - 1\right) \\
&= \frac{\Gamma(\frac{x+a}{h})\,\Gamma(b+1)}{\Gamma(\frac{x+h+a}{h}+b)}(-b) \\
&= (-b)\,B\left(\frac{x+a}{h}, b+1\right)
\end{aligned}$$

よって(12)式が成りたつ。(証明終わり)

【補助定理 4】　$x=0, h, 2h, 3h, \cdots$ のとき

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

(証明)　(12)式より

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

$$=-\frac{\Gamma(b+1)}{b}\left(\left.\frac{\Gamma(\frac{x+a}{h})}{\Gamma(\frac{x+a}{h}+b)}\right|_0^\infty\right)$$

$$=-\Gamma(b)\left(0-\frac{\Gamma(\frac{a}{h})}{\Gamma(\frac{a}{h}+b)}\right)$$

$$=\frac{\Gamma(\frac{a}{h})\Gamma(b)}{\Gamma(\frac{a}{h}+b)}$$

$$=B\left(\frac{a}{h},b\right)$$ （証明終わり）

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 $\lim_{h\to 0}\frac{P(x)}{h}$ である。

【定理 3 】

$x=0,h,2h,3h,\cdots$ のとき

$$\lim_{h\to 0}\frac{B(\frac{x+a}{h},b+1)}{hB(\frac{a}{h},b)}=ba^b(x+a)^{-b-1}$$

（証明）

$$\frac{B(\frac{x+a}{h},b+1)}{B(\frac{a}{h},b)}$$

$$=b\frac{\Gamma(\frac{x+a}{h})}{\Gamma(\frac{x+a}{h}+b+1)}\cdot\frac{\Gamma(\frac{a}{h}+b)}{\Gamma(\frac{a}{h})}$$

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

$$\frac{\Gamma(\frac{x+a}{h})}{\Gamma(\frac{x+a}{h}+b+1)}$$

$$=\frac{1}{(\frac{x+a}{h})(\frac{x+a}{h}+1)\cdots(\frac{x+a}{h}+b)}$$

$$=\frac{h^{b+1}}{(x+a)(x+a+h)\cdots(x+a+hb)}$$

同様に

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

$$=\frac{(a)(a+h)\cdots(a+h(b-1))}{h^b}$$

したがって

$$\frac{B(\frac{x+a}{h},b+1)}{hB(\frac{a}{h},b)}$$

$$=b\frac{(a)(a+h)\cdots(a+h(b-1))}{(x+a)(x+a+h)\cdots(x+a+hb)}$$

ここで $h\to 0$ とすると，右辺は $ba^b(x+a)^{-b-1}$ 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

$$\Delta F(x)=F(x+h)-F(x)$$

$$=\int_x^{x+h}ba^b(t+a)^{-b-1}dt$$

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

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

$$P(x)=b\frac{\Gamma(\frac{x+a}{h})}{\Gamma(\frac{x+a}{h}+b+1)}\cdot\frac{\Gamma(\frac{a}{h}+b)}{\Gamma(\frac{a}{h})}$$

である。

$b=1$ とすると

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

であり，一致する。また，b 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。$b=1$ 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 $b=2,\ 3,\ \frac{1}{2},\ \frac{1}{3}$ 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	β＝2	β＝3	β＝1/2	β＝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	β＝2	β＝3	β＝1/2	β＝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

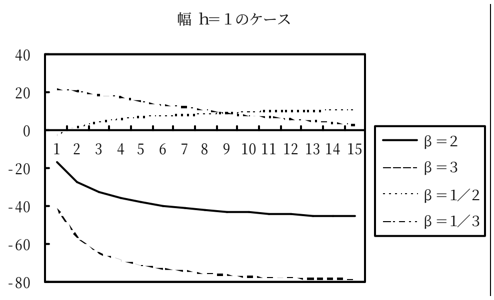
幅 h ＝ 0.05

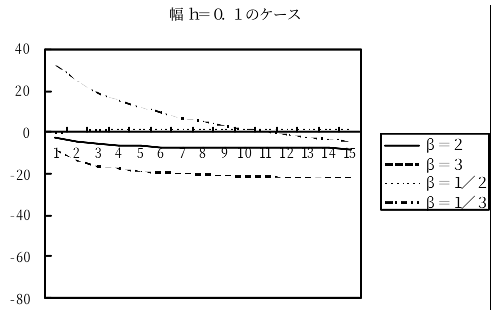
x	β＝2	β＝3	β＝1/2	β＝1/3
0	2.3	4.4	-1.2	47.0
1	-1.2	-4.6	-0.3	32.6
2	-2.4	-7.6	0.0	24.5
3	-3.0	-9.1	0.2	18.9
4	-3.3	-9.9	0.3	14.7
5	-3.6	-10.5	0.3	11.4
6	-3.7	-10.9	0.4	8.7
7	-3.9	-11.2	0.4	6.3

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

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 %）





β＝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

$$\frac{f^*(i)}{f^*(i-1)}=\frac{l(i-1+c)}{l(i+c)+1}=\frac{i-1+c}{i+c+\frac{1}{l}} \qquad (i=2,3,4,\cdots)$$

$$l=\frac{k(1-a)}{k+c}$$

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

本稿での一般化ユール分布との関係をみるために，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 a 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

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

と書き換えられる。以下のように変形して $P(x)$ を求める。

$$\begin{aligned}P(x)&=\frac{x-1+c}{x+c+\frac{1}{l}}P(x-1)\\&=\frac{x-1+c}{x+c+\frac{1}{l}}\cdot\frac{x-2+c}{x-1+c+\frac{1}{l}}P(x-2)\\&\quad\cdot\cdot\cdot\\&=\frac{x-1+c}{x+c+\frac{1}{l}}\cdot\frac{x-2+c}{x-1+c+\frac{1}{l}}\\&\quad\cdots\frac{1+c}{2+c+\frac{1}{l}}P(1)\\&=\frac{\Gamma(x+c)}{\Gamma(1+c)}\cdot\frac{\Gamma(2+c+\frac{1}{l})}{\Gamma(x+1+c+\frac{1}{l})}P(1)\end{aligned}$$

$\sum_{x=1}^{\infty}P(x)=1$ 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。

$$\begin{aligned}&\Delta\left[\frac{\Gamma(x+c)}{\Gamma(x+c+\frac{1}{l})}\right]\\&=\frac{\Gamma(x+1+c)}{\Gamma(x+1+c+\frac{1}{l})}-\frac{\Gamma(x+c)}{\Gamma(x+c+\frac{1}{l})}\\&=\frac{\Gamma(x+c)}{\Gamma(x+c+\frac{1}{l})}\left(\frac{x+c}{x+c+\frac{1}{l}}-1\right)\\&=\left(-\frac{1}{l}\right)\frac{\Gamma(x+c)}{\Gamma(x+1+c+\frac{1}{l})}\end{aligned}$$

したがって

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

である。

$$\sum_{x=1}^{\infty} P(x) = P(1)\left(1 - l \cdot \frac{\Gamma(2+c+\frac{1}{l})}{\Gamma(1+c)} \cdot \sum_{x=2}^{\infty}\Delta\left[\frac{\Gamma(x+c)}{\Gamma(x+c+\frac{1}{l})}\right]\right)$$

$$= P(1)\left(1 + l \cdot \frac{\Gamma(2+c+\frac{1}{l})}{\Gamma(1+c)} \cdot \frac{\Gamma(2+c)}{\Gamma(2+c+\frac{1}{l})}\right)$$

$$= P(1)\{1 + l(1+c)\}$$

よって $P(1) = \dfrac{1}{1+l(1+c)}$ であるから

$$P(x)$$

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

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

$$= \frac{\Gamma(1+c+\frac{1}{l})}{l} \cdot \frac{1}{\Gamma(1+c)\Gamma(1+\frac{1}{l})} \cdot \frac{\Gamma(x+c)\Gamma(1+\frac{1}{l})}{\Gamma(x+c+1+\frac{1}{l})}$$

$$= \frac{\Gamma(1+c+\frac{1}{l})}{\Gamma(1+c)\Gamma(\frac{1}{l})} \cdot \frac{\Gamma(x+c)\,\Gamma(1+\frac{1}{l})}{\Gamma(x+c+1+\frac{1}{l})}$$

$$= \frac{B(x+c, 1+\frac{1}{l})}{B(1+c, \frac{1}{l})}$$

(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$x=0,1,2,\cdots$であるが，上式は$x=1,2,3,\cdots$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

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

となる。

(10)式との対応では$1+c=a$，$\dfrac{1}{l}=b$ 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a 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

$$P(X=x \mid T=t) = g(t)^x\{h(t)\}^x$$

$$(x=0,1,2,\cdots)$$

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

$$f_T(t) = \frac{1}{B(a,b)} \cdot \frac{t^{a-1}}{(1+t)^{a+b}}$$

$$(t>0\,;\ a,b>0)$$

とする。

$$P(X=x) = E_T[P(X=x \mid T)]$$

であるから

$$P(X=x) = \frac{1}{B(a,b)}\int_0^{\infty}\frac{1}{1+t} \cdot \frac{t^x}{(1+t)^x} \cdot \frac{t^{a-1}}{(1+t)^{a+b}}dt$$

$$= \frac{1}{B(a,b)}\int_0^{\infty}\frac{t^{(x+a)-1}}{(1+t)^{(x+a)+(b+1)}}dt$$

したがって

$$P(x) = \frac{B(x+a, b+1)}{B(a,b)}$$

$$(x=0,1,2,\cdots) \qquad (13)$$

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$a=1$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a に一般化したものといえる。

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

$$P(x) = \frac{B(xg+a, b) - B((x+1)g+a, b)}{B(a,b)}$$

$$(x=0,1,2,\cdots) \qquad (14)$$

$g=1$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

$$f(k;\ r,a)=\frac{r}{1-a^{r}}B_{1-a}(k,\ r+1)$$
$$(0\le a\le 1,\quad x=1,2,3,\cdots)$$

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

$$P(x)=\frac{b}{1-c^{b}}B_{1-c}(x+1,\ b+1)$$
$$(0\le c\le 1,\quad x=0,1,2,\cdots)\qquad(15)$$

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 $x=0,h,2h,3h,\cdots$ とし，x 軸上でとる値を $a,\ a+h,\ a+2h,\ a+3h,\cdots$ 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，$x=0,1,2,3,\cdots$，x 軸上の値は $1,2,3,4,\cdots$ 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，$x=1,2,3,4,\cdots$ とし，x 軸上の値も $1,2,3,4,\cdots$ となるケースである。この場合の確率は

$$P(x)=\frac{B(x,\ b+1)}{B(1,\ b)}$$
$$=\frac{b\,(x-1)!}{(b+1)^{\bar{x}}}\qquad(x=1,2,3,\cdots)$$

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，それらを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 $x=g,\ g+h,\ g+2h,\ g+3h,\cdots$ とし，x 軸上でとる値を $a,\ a+h,\ a+2h,\ a+3h,\cdots$ とする。一般化ユール分布を

$$P(x)=\frac{B(\frac{x+a-g}{h},\ b+1)}{B(\frac{a}{h},\ b)}\qquad(16)$$

と定義する。$h,\ a,\ b,\ g>0$ である。□

これに対応するパレート分布は

$$f(x)=ba^{b}(x+a-g)^{-b-1}\qquad(x>g)\ (17)$$

である。

通常はユール分布，パレート分布ともに $h=1$，$g=a$ 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。

$$\frac{\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b+1)}=\left(-\frac{1}{b}\right)\cdot\Delta\left[\frac{\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b)}\right]$$

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。

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

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

$$F(x)=1-\frac{x+a-g}{b}\cdot\frac{P(x)}{h} \qquad (18)$$

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

$$F(x)=1-\frac{x+a-g}{b}f(x) \qquad (19)$$

(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。

$$\Delta\left[\frac{\Gamma(\frac{k+a-g}{h})}{\Gamma(\frac{k+a-g}{h}+b)}\right]$$

$$=\frac{\Gamma(\frac{k+h+a-g}{h})}{\Gamma(\frac{k+h+a-g}{h}+b)}-\frac{\Gamma(\frac{k+a-g}{h})}{\Gamma(\frac{k+a-g}{h}+b)}$$

$$=\frac{\Gamma(\frac{k+a-g}{h})}{\Gamma(\frac{k+a-g}{h}+b)}\cdot\frac{(-b)}{\frac{k+a-g}{h}+b}$$

よって

$$\frac{\Gamma(\frac{k+a-g}{h})}{\Gamma(\frac{k+a-g}{h}+b+1)}=-\frac{1}{b}\cdot\Delta\left[\frac{\Gamma(\frac{k+a-g}{h})}{\Gamma(\frac{k+a-g}{h}+b)}\right]$$

この式を用いて以下の計算をする。

$$F(x)=\sum_{k=g}^{x}P(k)$$

$$=\sum_{k=g}^{x}\frac{B(\frac{k+a-g}{h},b+1)}{B(\frac{a}{h},b)}$$

$$=\frac{\Gamma(b+1)}{B(\frac{a}{h},b)}\sum_{k=g}^{x}\frac{\Gamma(\frac{k+a-g}{h})}{\Gamma(\frac{k+a-g}{h}+b+1)}$$

$$=\frac{\Gamma(b+1)}{B(\frac{a}{h},b)}\left(-\frac{1}{b}\right)\sum_{k=g}^{x}\Delta\left[\frac{\Gamma(\frac{k+a-g}{h})}{\Gamma(\frac{k+a-g}{h}+b)}\right]$$

$$=-\frac{\Gamma(\frac{a}{h}+b)}{\Gamma(\frac{a}{h})\Gamma(b)}\cdot\frac{\Gamma(b+1)}{b}$$

$$\cdot\left(\frac{\Gamma(\frac{x+h+a-g}{h})}{\Gamma(\frac{x+h+a-g}{h}+b)}-\frac{\Gamma(\frac{a}{h})}{\Gamma(\frac{a}{h}+b)}\right)$$

$$=1-\frac{\Gamma(\frac{a}{h}+b)}{\Gamma(\frac{a}{h})}\cdot\frac{\Gamma(\frac{x+a-g}{h}+1)}{\Gamma(\frac{x+a-g}{h}+b+1)}$$

$$=1-\frac{\Gamma(\frac{a}{h}+b)}{\Gamma(\frac{a}{h})\Gamma(b)}\cdot\frac{x+a-g}{h}$$

$$\cdot\frac{\Gamma(\frac{x+a-g}{h})\Gamma(b+1)}{\Gamma(\frac{x+a-g}{h}+b+1)}\cdot\frac{\Gamma(b)}{\Gamma(b+1)}$$

$$=1-\frac{x+a-g}{hb}\cdot\frac{B(\frac{x+a-g}{h},b+1)}{B(\frac{a}{h},b)}$$

$$=1-\frac{x+a-g}{b}\cdot\frac{P(x)}{h}$$

パレート分布の分布関数については以下のようである。

$$F(x)=\int_{g}^{x}ba^{b}(t+a-g)^{-b-1}dt$$

$$=1-a^{b}(x+a-g)^{-b}$$

$$=1-\frac{x+a-g}{b}f(x) \qquad \text{(証明終わり)}$$

【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

$$h(x)=\frac{b}{x+a-g} \qquad (20)$$

(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。

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

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)式の一般化ユール分布および(17)式のパレート分布の平均は

$$m=\frac{bg+a-g}{b-1} \tag{21}$$

(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。

$$\frac{x\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b+1)}$$
$$=\frac{1}{1-b}\cdot\Delta\left[\frac{x\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b)}\right]$$
$$-\frac{a-g}{1-b}\cdot\frac{\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b+1)}$$

一般化ユール分布の平均は

$$\sum_{x=g}^{\infty}xP(x)$$
$$=\frac{\Gamma(b+1)}{B(\frac{a}{h},b)}\sum_{x=g}^{\infty}\frac{x\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b+1)}$$
$$=\frac{\Gamma(b+1)}{B(\frac{a}{h},b)}\left\{\frac{1}{1-b}\sum_{x=g}^{\infty}\Delta\left[\frac{x\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b)}\right]\right.$$
$$\left.-\frac{a-g}{1-b}\sum_{x=g}^{\infty}\frac{\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b+1)}\right\}$$

ここで Σ Δ [ ]の部分が収束するためには $b>1$ が必要である。

$$=\frac{\Gamma(b+1)}{B(\frac{a}{h},b)}\left\{\frac{1}{1-b}\left(-\frac{g\ \Gamma(\frac{a}{h})}{\Gamma(\frac{a}{h}+b)}\right)\right.$$
$$\left.-\frac{a-g}{1-b}\sum_{x=g}^{\infty}\frac{\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b+1)}\right\}$$
$$=\frac{bg}{b-1}+\frac{a-g}{b-1}$$

パレート分布の平均は

$$\int_g^{\infty}xba^b(x+a-g)^{-b-1}dx$$
$$=ba^b\left\{\left[x\left(\frac{1}{-b}\right)(x+a-g)^{-b}\right]_g^{\infty}\right.$$
$$\left.-\int_g^{\infty}\left(\frac{1}{-b}\right)(x+a-g)^{-b}dx\right\}$$
$$=ba^b\left\{\left[x\left(\frac{1}{-b}\right)(x+a-g)^{-b}\right]_g^{\infty}\right.$$
$$\left.-\left[\left(\frac{1}{-b}\cdot\frac{1}{-ß+1}\right)(x+a-g)^{-b+1}\right]_g^{\infty}\right\}$$

ここで収束するためには $b>1$ が必要である。

$$=ba^b\left\{\frac{ga^{-b}}{b}+\frac{a^{-b+1}}{b(b-1)}\right\}$$
$$=g+\frac{a}{b-1}$$

(証明終わり)

$g=a$ のときには平均が $\frac{ba}{b-1}$ になる。また $g=0$ のときには $\frac{a}{b-1}$ になる。

(ⅲ) 分散

【補助定理 9】

(16)式の一般化ユール分布の分散は

$$s^2=\frac{ba^2+b(b-1)ah}{(b-1)^2(b-2)} \tag{22}$$

（証明）

$s^2=E(X^2)-\{E(X)\}^2$ 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。

$$\frac{x^2\Gamma(\frac{k+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b+1)}$$
$$=\frac{1}{2-b}\cdot\Delta\left[\frac{x^2\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b)}\right]$$
$$-\frac{2(a-g)+h}{2-b}\cdot\frac{x\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b+1)}$$
$$-\frac{(a-g)h}{2-b}\cdot\frac{\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b+1)}$$

$E(X^2)$ は

$$\sum_{x=g}^{\infty}x^2P(x)$$
$$=\frac{\Gamma(b+1)}{B(\frac{a}{h},b)}\sum_{x=g}^{\infty}\frac{x^2\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b+1)}$$
$$=\frac{\Gamma(b+1)}{B(\frac{a}{h},b)}$$
$$\left\{\frac{1}{2-b}\sum_{x=g}^{\infty}\Delta\left[\frac{x^2\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b)}\right]\right.$$
$$-\frac{2(a-g)+h}{2-b}\sum_{x=g}^{\infty}\frac{x\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b+1)}$$
$$\left.-\frac{(a-g)h}{2-b}\sum_{x=g}^{\infty}\frac{\Gamma(\frac{x+a-g}{h})}{\Gamma(\frac{x+a-g}{h}+b+1)}\right\}$$

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

$$=\frac{\Gamma(b+1)}{B(\frac{a}{h},b)}\cdot\frac{1}{2-b}\left\{-\frac{g^2\Gamma(\frac{a}{h})}{\Gamma(\frac{a}{h}+b)}\right\}$$
$$-\frac{2(a-g)+h}{2-b}E(X)$$
$$-\frac{(a-g)h}{2-b}\sum_{x=g}^{\infty}P(x)$$
$$=-\frac{bg^2}{2-b}-\frac{2(a-g)+h}{2-b}$$
$$\cdot\frac{bg+(a-g)}{b-1}-\frac{(a-g)h}{2-b}$$
$$=\frac{bg^2}{b-2}+\frac{\{2(a-g)+h\}\{bg+(a-g)\}}{(b-1)(b-2)}$$
$$+\frac{(a-g)h}{b-2}$$

分散は

$$s^2=\frac{bg^2}{b-2}$$
$$+\frac{\{2(a-g)+h\}\{bg+(a-g)\}}{(b-1)(b-2)}$$
$$+\frac{(a-g)h}{b-2}-\frac{b^2a^2}{(b-1)^2}$$
$$=\frac{（分子）}{(b-1)^2(b-2)}$$

$$分子=bg^2(b-1)^2+\{2(a-g)+h\}\{bg+(a-g)\}(b-1)+(a-g)h(b-1)^2-b^2a^2(b-2)$$
$$=b^2ah+b(a^2-ah)$$

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

$$s^2=\frac{ba^2}{(b-1)^2(b-2)} \tag{23}$$

（証明）

$$E(X^2)=\int_g^{\infty}x^2ba^b(x+a-g)^{-b-1}dx$$
$$=ba^b\left\{\left[x^2\left(\frac{1}{-b}\right)(x+a-g)^{-b}\right]_g^{\infty}\right.$$
$$\left.-\int_g^{\infty}2x\left(\frac{1}{-b}\right)(x+a-g)^{-b}\,dx\right\}$$
$$=ba^b\left(\frac{1}{b}\right)g^2a^{-b}+2ba^b\left(\frac{1}{b}\right)$$

$$\left\{\left[x\left(\frac{1}{-b+1}\right)(x+a-g)^{-b+1}\right]_g^\infty - \int_g^\infty \left(\frac{1}{-b+1}\right)(x+a-g)^{-b+1}dx\right\}$$

$$= g^2 + 2ba^b \cdot \frac{1}{b(1-b)}\left[x(x+a-g)^{-b+1}\right]_g^\infty - 2ba^b \cdot \frac{1}{b(1-b)}\left[\left(\frac{1}{-b+2}\right)(x+a-g)^{-b+2}\right]_g^\infty$$

ここで収束するためには $b>2$ が必要である。

$$= g^2 + \frac{2ag}{b-1} + \frac{2a^2}{(b-1)(b-2)}$$

したがって分散は

$$s^2 = g^2 + \frac{2ag}{b-1} + \frac{2a^2}{(b-1)(b-2)} - \frac{(bg+a-g)^2}{(b-1)^2}$$

$$= \frac{ba^2}{(b-1)^2(b-2)}$$

（証明終わり）

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)　$f(x)=ba^bx^{-b-1}$，（$x>a$）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a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），蓑谷千凰彦(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”，Communcation Statistics，Series A，Vol.12，p.2503-2509
- 蓑谷千凰彦(2003)『統計分布ハンドブック』朝倉書店