連続確率変数を用いたエントロピー最大化によるベキ乗則の成立条件

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)
43
(号 / Number)
3
(開始ページ / Start Page)
1
(終了ページ / End Page)
12
(発行年 / Year)
2006-10-30
(URL)
https://doi.org/10.15002/00007245

〔論　文〕

連続確率変数を用いたエントロピー最大化によるベキ乗則の成立条件

鈴　木　　武

本稿では，ベキ乗則が生成するための条件を，連続確率変数のエントロピーを最大化するという観点から考察している。【１】で，ベキ乗則の定義を述べる。【２】で，ベキ乗則生成のモデルについて従来の研究をいくつか概観する。従来の研究は，特定の現象を説明する観点からモデル化しているので，ベキ乗則生成の現象に共通するものではないことを述べる。【３】で，期待値を一定にしてエントロピーを最大にする分布は指数分布であること，および，ベキ乗則の分布であるパレート分布と指数分布との関係を述べる。【４】で，連続確率変数が時間経過にもかかわらず連続性を保つための条件を考察する。【５】で，パレート分布をする確率変数の対数をとり，その変数の期待値が時間経過にもかかわらず一定であるための条件を求める。【６】で，期待値一定のもとでエントロピーを最大にする条件を求める。【７】で，以上の考察をまとめ，ベキ乗則生成の条件を求める。【８】で，ベキ乗係数が１である場合をジップ則というが，その生成条件を述べる。

【１】ベキ乗則とは

都市人口の大きさとその順位には一定の関係が見られる。たとえば日本では，2000年国勢調査における人口第１位は東京特別区で，813万人である。第２位は横浜市343万人であり，第１位の約半分である。また，第３位は大阪市260万人で約１／３，第４位は名古屋市217万人で約１／４である。すなわち，(順位)×(人口)＝(一定）という関係がおおよそ成り立つ。

ある現象において n 個の結果が生じたとする。各結果の数値は正であり，大きさの順序は $x_1 \geq x_2 \geq \cdots \geq x_n > 0$ である。規模 X と順位 N との間に $N=\alpha X^{-\beta}$ (α，$\beta > 0$）の関係があるとき，ベキ乗則（power law）が成り立つという。とくに $\beta = 1$ のとき，ジップ則（Zipf's law）という。都市人口の場合には，ジップ則が成り立つと言われている。

ベキ乗則の成立は都市人口のほかに，ある言語における単語の使用頻度，ウェブページへのリンク数，地震のマグニチュード，月のクレーターの大きさ，名字の使用割合，論文の引用回数など，かなり普遍的にみられる。従来多くの研究は，それぞれ特定の現象についてベキ乗則生成のメカニズムをモデル化してきた。したがって，他の現象については適用しにくい面がある。

本稿では特定の現象を念頭に置かず，ベキ乗則が成立するためにはどのような条件が必要かという観点から考察する。議論の容易さから，連続確率変数に限定している。連続確率変数が連続性を保ちながら時間経過とともに変化していき，その極限としてベキ乗則が出現するという観点から考察している。その場合，規模の順位は変化しない。したがって，ベキ乗則生成の最終段階における限定された議論になる。

都市人口を例にとると，時間経過とともに各都市の人口が変化していくが，その変化の行きつく先にベキ乗則が現れてくると考えられる。とくに都市人口の場合には $\beta = 1$ になるジップ則が出現する。

ある時点における第 i 順位の都市人口を x_i とする。すでにジップ則が成立している場合，もしすべての都市において等しい人口成長率を示すならば，α については変化するが，β については変わらない。

時点 t と時点（$t+1$）の関係がすべての X について $X_{t+1}=kX_t$ であるならば，順位 N は変化しないので

$$\begin{aligned} N &= \alpha X_t^{-\beta} \\ &= \alpha \left(\frac{X_{t+1}}{k} \right)^{-\beta} \\ &= \alpha k^{\beta} X_{t+1}^{-\beta} \end{aligned}$$

αk^{β}を改めてαとおけば

$$N = \alpha X_{t+1}^{-\beta}$$

となる。したがって，ベキ乗則においてはx_iそのものの大きさではなく，他都市の人口と比較したときの相対的な大きさが問題となる。たとえば，全人口に占める第i都市の人口シェアを対象にすればよい。相対的な大きさを改めてx_iとおくことにし，ここでは規模と呼ぶことにする。

各x_iが一つずつではなく，より一般的に，相対度数または確率p_iで表されるケースを考えよう。規模x_iの順位は確率

$$P(X \geq x_i) = \sum_{k=1}^{i} p_k$$

で表現される。

さらにnが大きな値をとるとき，離散的確率変数Xは連続型で近似されるとしよう。それにより，式の展開を容易にすることができる。すなわち，現象は連続的確率変数Xで表現される。その分布関数は$F(x)$，確率密度関数は$f(x)$とする。

値xの順位は，確率$P(X \geq x)$になる。変数Xがベキ乗則に従うなら

$$P(X \geq x) = \alpha x^{-\beta}, \quad x \geq \alpha^{\frac{1}{\beta}} \tag{1}$$

と表される。これはパレート分布と呼ばれるものである。本稿では，前述したようにXの絶対的な大きさではなく，相対的な大きさが問題とされる。したがって，Xの最小値を1と基準化してもさしつかえない。そこでベキ乗則を

$$P(X \geq x) = x^{-\beta}, \quad x \geq 1 \tag{2}$$

と表現しよう。この分布関数は

$$F(x) = P(X \leq x) = 1 - x^{-\beta}, \quad x \geq 1 \tag{3}$$

確率密度関数は

$$f(x) = \beta x^{-(\beta-1)}, \quad x \geq 1$$

である。

【2】ベキ乗則の研究といくつかのモデル

ベキ乗則の研究は1910年代からみられるが，1949年にZipfが論文を著してから注目されるようになった。ベキ乗則の生成メカニズムについては，Simon (1955) やBelevitch (1956) によって当初から研究されてきた。最近では，Barabási et al. (1999)，Gabaix (1999)，Reed (2003) 等によりに行われており，簡単なモデルから複雑なモデルまでいろいろ提案されている。ここでは，そのうちのいくつかについて記述しよう。

(a) 指数的関係の組合せによるモデル[注1]

yが指数分布に従うと仮定する。

$$p(y) \sim e^{ay}$$

aはプラスでもマイナスでもよい。ただし，プラスのときはある値以上の確率は0になる。ここで注目している変数はyではなくxであり，両者は指数的に関係して

$$x \sim e^{by}$$

であるとする。xの確率密度関数は

$$p(x) = p(y)\frac{dy}{dx} \sim \frac{e^{ay}}{be^{by}} = \frac{x^{-1+a/b}}{b}$$

$\beta = -\dfrac{a}{b}$とおけば

$$p(x) \sim x^{-(\beta+1)} \quad \text{あるいは} \quad P(X \geq x) \sim x^{-\beta}$$

となり，xはベキ乗則に従う。

この例として，Miller (1957) は単語の使用頻度について適用している。アルファベットがm文字あり，空白によって単語を識別しているとする。単語はm文字をランダムに用いて作られるとすると，l番目の文字が用いられる確率は$q_l = \dfrac{1-q_s}{m}$である。ただし，q_sは空白の確率である。y文字からなる単語のうち，ある特定の単

語が出現する頻度 x は

$$x=\left(\frac{1-q_s}{m}\right)^y q_s \sim e^{by}$$

である。ここで $b=\log(1-q_s)-\log m$。長さ y 文字から $(y+dy)$ 文字にある単語の確率は $p(y)\sim m^y=e^{ay}$ である。ここで $a=\log m$。したがって

$$\beta=-\frac{a}{b}=\frac{\log m}{\log m-\log(1-q_s)}$$

この場合，q_s の確率が小さければ $\beta\simeq 1$ になり，ジップ則が成立する。

もしこのモデルを都市人口に適用しベキ乗則を導こうとするならば，都市を規模別にランク分けすることになろう。大都市・中都市・小都市・零細都市というランクや，さらに細かいランク分けである。ランクを y とする。ランク y のなかで，ある特定規模の都市が出現する頻度を x とする。y の分布や，x と y の関係が上述のように導かれるためには，かなり主観を交える必要があり，単語の使用頻度のような明確さはない。

また，このモデルをウェブページへのリンク数やその他の現象に適用しようとすると，さらに明確さを欠くことになる。すなわち，モデルはある特定の現象を説明するためのモデルであり，ベキ乗則を生成するすべての現象に適用できるものではない。

(b) Yule 過程を用いたモデル

Yule 過程を用いてベキ乗則を説明したモデルとしては Simon (1955) が知られているが，ここではより単純なモデルでウェブページへのリンク数を議論した Barabási et al. (1999) をとりあげよう。モデルは 2 ステップから成り立つ。

(1) 初期に m_0 個のノードが存在する。各時点で新たに 1 個ずつノードが追加され，それぞれ $m(\leq m_0)$ 個のリンク数を張る。

(2) i 時点で加わったノードが t 時点ではリンク数を $k_i(t)$ にしているとする。t 時点で新たに加わるノードが第 i ノードにリンクする確率は

$$\Pi(k_i(t))=\frac{k_i(t)}{\sum_j k_j(t)}=\frac{k_i(t)}{2mt}$$

であると仮定する。

このモデルで基本的なものは（2）の仮定である。バラバシはこの仮定を「優先的選択」(preferential attachment) と呼んでいる。より一般的には「ジブラ法則」(Gibrat's law) と言われており，「金持ちはもっと金持ちに」(Rich-get-richer) という現象を引き起こす。

k が連続的に変化すると仮定して，t 時点における第 i ノードの変化は

$$\frac{\partial k_i(t)}{\partial t}=m\Pi(k_i(t))$$
$$=m\frac{k_i(t)}{\sum_j k_j(t)}=\frac{k_i(t)}{2t}$$

i 時点を t_i とする。$k_i(t_i)=m$ であるから，微分方程式を解くと

$$k_i(t)=m\left(\frac{t}{t_i}\right)^{0.5}$$

$k_i(t)$ の分布を求めよう。

$$P(k_i(t)<k)=P\left[m\left(\frac{t}{t_i}\right)^{0.5}<k\right]$$
$$=P\left(t_i>\frac{m^2t}{k^2}\right)$$

ノードは等しい時間間隔で加わるので，t_i の確率密度関数は

$$p(t_i)=\frac{1}{m_0+t}$$

したがって

$$P\left(t_i>\frac{m^0t}{k^2}\right)=1-P\left(t_i<\frac{m^2t}{k^2}\right)$$
$$=1-\frac{m^2t}{k^2(t+m_0)}$$

k の確率密度関数は

$$p(k)=\frac{\partial P(k_i(t)<k)}{\partial k}=\frac{2m^2t}{m_0+t}\frac{1}{k^3}$$

よって $p(k)\sim k^{-3}$ である。言い換えれば $P(k_i(t)\geq k)\sim k^{-2}$ であり，ベキ乗則が成り立つ。

Yule 過程を用いたモデルはウェブページへのリンク数のほか，単語の使用頻度，都市人口の規模等，広く適用されている。

(c) 下限のあるブラウン運動を用いたモデル

Gabaix（1999）は幾何ブラウン運動で下限のあるモデルを用いて，都市人口がベキ乗則に従うことを示している。

t 時点における第 i 都市の人口を P_{it} とする。各都市の人口変化率は（1）時間とともに一定の割合 γ_i で変化する部分と，（2）各都市が各時点でそれぞれ独立に，しかし同じ分散 σ^2 で正規分布に従って変化する部分とから構成されていると仮定する。都市人口の規模が連続的に変化し，時間変化も連続的であるとして

$$\frac{dP_{it}}{P_{it}}=\gamma_i dt+\sigma dB_{it}$$

が言える。ここで，B_{it} はブラウン運動である。

規模を比例的に変化させても，ベキ乗則における係数 β は変わらない。したがって，人口そのものではなく，人口シェアで議論する。$S_{it}=P_{it}/$(t 時点における全都市人口の期待値）とおく。シェアで表した関係式は

$$\frac{dS_{it}}{S_{it}}=\mu_i dt+\sigma dB_{it}$$

ここで $\mu_i=\gamma_i-\bar{\gamma}$ である。γ_i は人口シェア S_i の都市の変化率であり，$\bar{\gamma}$ は全都市の平均変化率である。

添字 i を省略して上式を変形すれば

$$dS_t=\mu S_t dt+\sigma S_t dB_t$$

積分の形に書き直して

$$S_t=S_0+\mu\int_0^t S_s ds+\sigma\int_0^t S_s dB_s$$

である。右辺第２項はリーマン＝スチルチェス積分で，第３項は伊藤確率積分で表現される。

これを解くと

$$S_t=S_0 e^{(\mu-0.5\sigma^2)t+\sigma B_t}$$

となり，幾何ブラウン運動になる。

人口シェア S_t はある値以下にはならない下限が存在すると仮定する。すなわち，$S_t>S_{min}$ に対しては上式が適用され，$S_t\leq S_{min}$ の場合には $dS_t=S_t\max(\mu dt+\sigma dB_t,\ 0)$ であるとする。

人口シェアの対数をとり，$s_t=\log S_t$ とおく。

$$s_t=\log S_t=s_0+(\mu-0.5\sigma^2)t+\sigma B_t$$

B_t は標準ブラウン運動なので，正規分布 $N(0,t)$ に従う。よって，下限に制約がなければ

$$s_t\sim N\left[s_0+(\mu-0.5\delta^2)t,\ \sigma^2 t\right]$$

となり，人口シェア S_t は対数正規分布に従う。

しかし，下限を設けているので Harrison（1985）p.15の記述により，$t\to\infty$ のときには以下のようになる。

$s_t>s_{min}$ のとき

$$\begin{aligned}
&P(s_t\leq s)\\
&\to 1-\exp\left[\left(\frac{2(\mu-0.5\sigma^2)}{\sigma^2}\right)(s-s_{min})\right]\\
&=1-\exp\left[-\left(1-\frac{2\mu}{\sigma^2}\right)(s-s_{min})\right]\\
&=1-e^{-\zeta(s-s_{min})}
\end{aligned}$$

ここで $\zeta=1-\dfrac{2\mu}{\sigma^2}$ である。よって，s_t は指数分布に収束する。

$s_t\leq s_{min}$ のときは

$$P(s_t\leq s)\to 0$$

となる。したがって下限のあるモデルでは，$t\to\infty$ の極限において

$$P(s_t\geq s)=e^{-\zeta(s-s_{min})}$$

である。

この対数を戻して S_t でみよう。$S_t=e^{s_t}$，$S_{min}=e^{s_{min}}$ だから

$$P(S_t\geq S)=\left(\frac{S}{S_{min}}\right)^{-\zeta}$$

となり，ベキ乗則が成立する。

【3】エントロピー最大化とベキ乗則

従来の研究では特定の現象を念頭においてモデル化してきたが，本稿では一般的にベキ乗則を生成するための条件という観点から考察してみよう。

確率変数 X がパレート分布に従うならば，$P(X \geq x) \sim x^{-\beta}$ と書けるので，ベキ乗則が生成している。パレート分布と深い関係にあるものとして指数分布がある。$Y=\log X$ とおくと，Y は指数分布に従っている。

X の確率密度関数を

$$f(x)=\beta x^{-(\beta+1)},\quad x \geq 1 \tag{4}$$

とする。$x=e^y$，$dx=xdy$ であるから

$$f(x)dx=\beta x^{-(\beta+1)}xdy=\beta x^{-\beta}dy$$
$$=\beta e^{-\beta y}dy=g(y)dy$$

したがって Y は指数分布に従い，その確率密度関数は

$$g(y)=\beta e^{-by},\quad y \geq 0 \tag{5}$$

である。逆に Y が（5）式のような指数分布で表されているならば，$X=e^Y$ は（4）式のようなパレート分布になる。

X がパレート分布になるためには $\log X$ が指数分布になることが必要十分である。それでは $Y=\log X$ が指数分布になるためにはどのようなことが言えればよいであろうか。エントロピーの概念を用いると，平均が一定という条件の下でエントロピーが最大になるようにすればよい。エントロピーは状態が秩序だっているほど低く，規制をしないで自然の状態にしておくほど高くなる。

Y のエントロピー $H(Y)$ は

$$H(Y)=-\int_0^\infty g(y)\log g(y)dy$$

条件は

$$\int_0^\infty g(y)dy=1 \tag{6}$$

$$\int_0^\infty yg(y)dy=c\ (\text{一定})\quad c>0 \tag{7}$$

である。

エントロピーを最大にする Y の分布を求めよう。(注2) ラグランジュ乗数法を用いて

$$\frac{\partial}{\partial g}(-g\log g)+\kappa\frac{\partial}{\partial g}(g)+\lambda\frac{\partial}{\partial g}(gy)$$
$$=(-1+\log g)+\kappa+\lambda y=0$$

これを解いて

$$g(y)=e^{\kappa-1+\lambda y}$$

κ および λ を求めるために条件式（6），（7）に代入する。

$$e^{\kappa-1}\int_0^\infty e^{\lambda y}dy=1$$

$$e^{\kappa-1}\int_0^\infty ye^{\lambda y}dy=c$$

各定積分を求めるための計算として

$$\int_0^\infty e^{\lambda y}dy=\frac{1}{\lambda}\left[e^{\lambda y}\right]_0^\infty$$

ここで $\lambda<0$ とすると

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

また

$$\int_0^\infty ye^{\lambda y}dy=\frac{1}{\lambda}\left[ye^{\lambda y}\right]_0^\infty-\frac{1}{\lambda}\int_0^\infty e^{\lambda y}dy$$
$$=\frac{1}{\lambda}\left[ye^{\lambda y}\right]_0^\infty-\frac{1}{\lambda^2}\left[e^{\lambda y}\right]_0^\infty$$
$$=\frac{1}{\lambda}(0-0)-\frac{1}{\lambda^2}(0-1)$$
$$=\frac{1}{\lambda^2}$$

したがって条件式（6），（7）は

$$-\frac{e^{\kappa-1}}{\lambda}=1$$

$$\frac{e^{\kappa-1}}{\lambda^2}=c$$

変形して

$$e^{\kappa-1}=-\lambda$$

$$\lambda = -\frac{1}{c}$$

よって

$$g(y) = \frac{1}{c}\, e^{-y/c}$$

となり，$g(y)$ は指数分布の確率密度関数になる。

以上のことから，ベキ乗則が生じるためには $\log X$ の期待値が一定の値をとりつつ，エントロピーが最大になるようにすればよい。

【4】連続的確率変数であるための条件

時点 t における確率変数 X_t が連続であるとしよう。1単位時間経過後に X_{t+1} に変化するとし，それが連続であるための条件を求めよう。ここで，関数 X を

$$X(y) = e^{y},\quad y \geq 0$$

とする。すなわち，$y = \log X(y)$ である。

確率変数 Y の分布関数を F とする。期待値は

$$E(Y) = \int_{-\infty}^{\infty} y dF(y)$$

このスチルチェス積分（リーマン＝スチルチェス積分）を定義しよう。[注3]

まず，閉区間 $I = [a,\ b]$ でのみ定義されているとする。任意の分割 Δ を

$$\Delta : a = r_0 < r_1 < \cdots < r_{n-1} < r_n = b$$

分割 Δ の幅を $d(\Delta) = \max_{1 \leq i \leq n}(r_i - r_{i-1})$ とする。各小区間 $I_i = [r_{i-1},\ r_i]$ の代表元 $y_i \in I_i$ を選んで作った和

$$s(F; \Delta; y) = \sum_{i=1}^{n} y_i \{F(r_i) - F(r_{i-1})\}$$

をスチルチェス和という。$d(\Delta) \to 0$ とするとき，y_i のとり方によらず一定の極限が存在するならば，スチルチェス積分可能である。

積分範囲を $-\infty$ から ∞ までにするためには，広義積分を定義し，範囲を $a \to -\infty$，$b \to \infty$ とすればよい。[注4]

時点 t における確率変数を Y_t とし，その実現値を y_t とする。さらに $X_t = X(Y_t) = e^{Y_t}$ とする。Y_t の分布関数を F_t とし，任意の分割を Δ_t，各小区間 $I_{ti} = [r_{t,i-1},\ r_{ti}]$ の代表元を $y_{ti} \in I_{ti}$ とする。ここで，$x_{ti} = X(y_{ti}) = e^{y_{ti}}$ とする。スチルチェス和は

$$s(F_t; \Delta_t; y_t) = \sum_{i=1}^{n} y_{ti} \{F_t(r_{ti}) - F_t(r_{t,i-1})\}$$

F_t は分布関数なので，単調非減少，有界，右連続という性質をもつ。さらに，ここでは連続的な確率変数の期待値を対象としているので，F_t は C^1 級としてよい。すなわち，F_t の1階の導関数が存在し，F_t およびその導関数は連続である。

y_t は 閉区間 $I = [a,\ b]$ 上で連続なので，有界でリーマン可積分である。[注5] また，F_t は I 上で C^1 級であるので，y_t は F_t に関しスチルチェス積分可能である。[注6]

1単位時間経過後の時点（$t+1$）における確率変数 Y_{t+1} について考察しよう。t から（$t+1$）への経過により，x_{ti} は $R_{t+1}(y_{ti})$ 倍になるとする。ここで，関数 R_{t+1} は正の値をとり，区間 I で C^1 級であると仮定する。ただし，確率変数 X_{t+1} の下限 $r_{t+1,0} = r_{t0} R_{t+1}(r_{t0})$ は1に基準化されるとするから，

$$x_{t+1,i} = x_{ti} \frac{R_{t+1}(y_{ti})}{R_{t+1}(r_{t0})} \tag{8}$$

と定義される。

時点 t と時点（$t+1$）との関係は

$$y_{ti} = \log(x_{ti})$$

$$y_{t+1,i} = \log(x_{t+1,i})$$

$$x_{t+1,i} = x_{ti} \frac{R_{t+1}(y_{ti})}{R_{t+1}(r_{t0})}$$

$$r_{ti} = \log\{X(r_{ti})\}$$

$$r_{t+1,i} = \log\{X(r_{t+1,i})\}$$

$$X(r_{t+1,i}) = X(r_{ti}) \frac{R_{t+1}(r_{ti})}{R_{t+1}(r_{t0})}$$

である。よって

$$y_{t+1,\,i}=y_{ti}+\log\{R_{t+1}(y_{ti})\}-\log\{R_{t+1}(r_{t0})\}$$
$$r_{t+1,\,i}=r_{ti}+\log\{R_{t+1}(r_{ti})\}-\log\{R_{t-1}(r_{t0})\}$$
$$r_{t-1,\,i-1}=r_{t,\,i-1}+\log\{R_{t+1}(r_{t,\,i-1})\}-\log\{R_{t-1}(r_{t0})\}$$

である。

関数 F_{t+1} の性質を調べよう。

(a) F_t の分割 Δ_t における小区間 $I_{ti}=[r_{t,\,i-1},\ r_{ti}]$ の代表元を $y_{ti}\in I_{ti}$ としているから，

$$r_{t,\,i-1}\leq y_{ti}\leq r_{ti}$$

である。F_t は単調非減少関数であるので

$$F_t(r_{ti})-F_t(r_{t,\,i-1})\geq 0$$

となる。このとき，時点 t の分割 Δ_t に対応する時点（t+1）の分割 Δ_{t+1} が

$$r_{t+1,\,i-1}\leq y_{t+1,\,i}\leq r_{t+1,\,i} \tag{9}$$

を満たすであろうか。

（９）式が成立するためには

$$r_{t,\,i-1}+\log\{R_{t+1}(r_{t,\,i-1})\}\leq y_{ti}+\log\{R_{t+1}(y_{ti})\}$$
$$\leq r_{ti}+\log\{R_{t+1}(r_{ti})\}$$

がいえればよい。$y_{ti}=\log(x_{ti})$，$r_{t,\,i-1}=\log\{X(r_{t,\,i-1})\}$ であるから，（９）式は

$$\log\{X(r_{t,\,i-1})R_{t+1}(r_{t,\,i-1})\}\leq\log\{x_{ti}R_{t+1}(y_{ti})\}$$
$$\leq\log\{X(r_{ti})R_{t+1}(r_{ti})\}$$

したがって

$$X(r_{t,\,i-1})R_{t+1}(r_{t,\,i-1})\leq x_{ti}R_{t-1}(y_{ti})\leq X(r_{ti})R_{t+1}(r_{ti})$$

が成り立てばよい。すなわち，時間が経過しても規模の順位に変化がなければよい。

（９）式において $r_{t+1,\,i-1}<r_{t+1,\,i}$ であるから

$$r_{t,\,i-1}+\log\{R_{t+1}(r_{t,\,i-1})\}<r_{ti}+\log\{R_{t+1}(r_{ti})\}$$

である。変形すると，$r_{t,\,i-1}<r_{ti}$ であるから

$$1+\frac{\log\{R_{t+1}(r_{ti})\}-\log\{R_{t+1}(r_{t,\,i-1})\}}{r_{ti}-r_{t,\,i-1}}>0$$

R_{t+1} は微分可能であるから，$\log(R_{t+1})$ も微分可能である。したがって $r_{t,\,i-1}$ を限りなく r_{ti} に近づけると

$$1+[\log\{R_{t+1}(r_{ti})\}]'=1+\frac{[R_{t+1}(r_{ti})]'}{R_{t+1}(r_{ti})}>0$$

$R_{t+1}(r_{ti})>0$ であるから

$$R_{t+1}(r_{ti})+[R_{t-1}(r_{ti})]'>0 \tag{10}$$

がいえる。よって，時間が経過しても規模の順位に変化がないことと，（10）式とは同等である。

(b) F_{t+1} は単調非減少関数であるか。このためには次式が成り立てばよい。

$$F_{t+1}(r_{t+1,\,i})-F_{t+1}(r_{t+1,\,i-1})\geq 0 \tag{11}$$

時点 t における小区間 $I_{ti}=[r_{t,\,i-1},\ r_{ti}]$ が，時点（t+1）においては小区間 $I_{t+1,\,i}=[r_{t+1,\,i-1},\ r_{t+1,\,i}]$ に遷移する。したがって，小区間 $I_{t+1,\,i}$ が生起する確率は小区間 I_{ti} の確率に等しい。

$$P(I_{t+1,\,i})=P(I_{ti})$$

また，（９）式が成立しているので，小区間の両端点の順序は入れ替わらない。よって

$$F_t(r_{ti})=\sum_{h=1}^{i}P(I_{th})=\sum_{h=1}^{i}P(I_{t+1,\,h})=F_{t-1}(r_{t+1,\,i})$$

それゆえ

$$F_{t+1}(r_{t+1,\,i})-F_{t+1}(r_{t+1,\,i-1})=F_t(r_{ti})-F_t(r_{t,\,i-1})\geq 0$$

（11）式が成立するので，F_{t+1} は単調非減少関数である。

(c) 連続性

F_t は連続であるから，任意の $r_t\in I$ について次のことが成り立つ。任意の ε_1 に対し δ_1 が存在して $|y_t-r_t|<\delta_1$ となるすべての $y_t\in I$ に対し $|F_t(y_t)-F_t(r_t)|<\varepsilon_1$ となる。同様に $\log(R_{t+1})$ は連続であるから，任意の $r_t\in I$ について次のことが成り立つ。任意の ε_2 に対し δ_2 が存在して $|y_t-r_t|<\delta_2$ となるすべての $y_t\in I$ に対し $|\log\{R_{t+1}(y_t)\}-\log\{R_{t+1}(r_t)\}|<\varepsilon_2$ となる。

F_{t+1} に関し，任意の $r_{t+1}\in I$ について次のことがいえる。任意の $\varepsilon=\varepsilon_1+\varepsilon_2$ に対して $\delta=\delta_1+\delta_2$ をとる。

$$
\begin{aligned}
&|y_{t+1}-r_{t+1}| \\
&=|y_t-r_t+\log\{R_{t+1}(y_t)\}-\log\{R_{t+1}(r_t)\}| \\
&\quad\leq|y_t-r_t|+|\log\{R_{t+1}(y_t)\}-\log\{R_{t+1}(r_t)\}| \\
&=\delta_1+\delta_2=\delta
\end{aligned}
$$

上式の関係を満たすすべての y_{t+1} に対し

$$|F_{t+1}(y_{t+1})-F_{t+1}(r_{t+1})|=|F_t(y_t)-F_t(r_t)|<\varepsilon_1\leq\varepsilon$$

したがって F_{t+1} は連続である。

(d) 微分可能性と導関数の連続性

F_t は微分可能であるから

$$\lim_{r_{ti}-r_{t,i-1}\to 0}\frac{F_t(r_{ti})-F_t(r_{t,i-1})}{r_{ti}-r_{t,i-1}}=f_t(r_{ti})$$

が存在する。

$$
\begin{aligned}
&\frac{F_{t+1}(r_{t+1,i})-F_{t+1}(r_{t+1,i-1})}{r_{t+1,i}-r_{t+1,i-1}} \\
&=\frac{F_t(r_{ti})-F_t(r_{t,i-1})}{r_{ti}-r_{t,i-1}+\log\{R_{t+1}(r_{ti})\}-\log\{R_{t+1}(r_{t,i-1})\}} \\
&=\frac{F_t(r_{ti})-F_t(r_{t,i-1})}{(r_{ti}-r_{t,i-1})\left[1+\dfrac{\log\{R_{t+1}(r_{ti})\}-\log\{R_{t+1}(r_{t,i-1})\}}{r_{ti}-r_{t,i-1}}\right]}
\end{aligned}
$$

R_{t+1} は正の値をとり C^1 級であるから，$\log(R_{t+1})$ も C^1 級である。したがって

$$
\begin{aligned}
&\lim_{r_{t+1,i}-r_{t+1,i-1}\to 0}\frac{F_{t+1}(r_{t+1,i})-F_{t+1}(r_{t+1,i-1})}{r_{t+1,i}-r_{t+1,i-1}} \\
&=f_t(r_{ti})\frac{1}{1+[\log\{R_{t+1}(r_{ti})\}]'} \\
&=f_t(r_{ti})\frac{1}{1+\dfrac{[R_{t+1}(r_{ti})]'}{R_{t+1}(r_{ti})}}
\end{aligned}
$$

$f_t(r_{ti})$，$R_{t+1}(r_{ti})$，$[R_{t+1}(r_{ti})]'$ がそれぞれ連続であり，(10) 式から

$$1+\frac{[R_{t+1}(r_{ti}))]'}{R_{t+1}(r_{ti})}>0$$

であるので，F_{t+1} は微分可能で，その導関数は連続である。

(a) ～ (d) をふまえて，関数 R_{t+1} が正の値をとり，区間 I で C^1 級であり，(10) 式が成りたてば，F_{t+1} は単調非減少で C^1 級である。したがって，連続的確率変数の分布関数である。

【5】対数の期待値が一定である条件

$Y=\log X$ の期待値が一定であるための条件を求めよう。そのために，時点 t と $(t+1)$ における Y の期待値の関係をみよう。

$$
\begin{aligned}
&E(Y_{t+1}) \\
&=\int_0^\infty ydF_{t+1}(y) \\
&=\lim_{d(\Delta_{t+1})\to 0}\sum_{i=1}^{n}\log(x_{t+1,i})\{F_{t+1}(r_{t+1,i})-F_{t-1}(r_{t+1,i-1})\} \\
&=\lim_{d(\Delta_t)\to 0}\sum_{i=1}^{n}\log\left(x_{ti}\frac{R_{t+1}(r_{ti})}{R_{t-1}(r_{t0})}\right)\{F_t(r_{ti})-F_t(r_{t,i-1})\} \\
&=\int_0^\infty ydF_t(y)+\int_0^\infty\log\left(\frac{R_{t+1}(y)}{R_{t+1}(r_{t0})}\right)dF_t(y)
\end{aligned}
$$

ここで下限の変化倍率を $R_{t+1}(r_{t0})=R_{t+1}^{min}$ と書き換え，下限の変化倍率を基準にした各値の相対変化倍率を

$$\rho_{t+1}(y)=\frac{R_{t+1}(y)}{R_{t+1}^{min}}$$

と定義する。Y_{t+1} の期待値は

$$E(Y_{t+1})=E(Y_t)+\int_0^\infty\log\{\rho_{t+1}(y)\}dF_t(y)$$

となる。期待値が時間経過にかかわらず一定であるためには

$$
\begin{aligned}
&\int_0^\infty\log\{\rho_{t+1}(y)\}dF_t(y) \\
&=\int_0^\infty\log\left(\frac{R_{t+1}(y)}{R_{t+1}^{min}}\right)dF_t(y)=0 \qquad (12)
\end{aligned}
$$

であることが必要十分である。

この意味を考えてみよう。

$$\Delta_{ti}=F_t(r_{ti})-F_t(r_{t,i-1})$$

とおく。

$$\int_0^\infty \log\{\rho_{t+1}(y)\}\,dF_t(y)$$

$$=\lim_{d(\Delta_t)\to 0}\sum_{i=1}^{n}\Delta_{ti}\log\{\rho_{t+1}(y_{ti})\}$$

$$=\lim_{d(\Delta_t)\to 0}\log\left(\prod_{i=1}^{n}\rho_{t+1}(y_{ti})^{\Delta ti}\right)$$

したがって

$$\prod_{i=1}^{n}\rho_t(y_{ti})^{\Delta ti}\to 1$$

ならば（12）式を満たす。すなわち，（12）式は相対変化倍率の相乗平均が1に等しいときに成立する。

【6】エントロピー最大化の条件

一次元の連続確率変数 V と W がある。確率変数 V が1対1の連続な変換によって新たな確率変数 W に変形されるとしよう。V，W の確率密度関数をそれぞれ $f(v)$，$g(w)$ とする。両者の関係は

$$g(w)=f(v)\left|\frac{dv}{dw}\right|$$

である。

$W=h(V)$ が単調で一価の変換であるならば，W のエントロピーは

$$H(W)$$

$$=-\int_{-\infty}^{\infty}g(w)\log g(w)\,dw$$

$$=-\int_{-\infty}^{\infty}\left[f(v)\left|\frac{dv}{dw}\right|\right]\log\left[f(v)\left|\frac{dv}{dw}\right|\right]dw$$

$$=-\int_{-\infty}^{\infty}f(v)\log f(v)\,dv-\int_{-\infty}^{\infty}f(v)\log\left|\frac{dv}{dw}\right|dv$$

$$=H(V)+\int_{-\infty}^{\infty}f(v)\log\left|\frac{dw}{dv}\right|dv$$

となる。[注7]

上式を本論に用いよう。確率変数 X_t と X_{t+1} の関係は（8）式から

$$X_{t+1}=X_t\,\frac{R_{t+1}(\log(X_t))}{R_{t+1}^{min}}$$

である。ただし，$R_{t-1}(\cdot)$ は変化倍率，R_{t+1}^{min} は下限の変化倍率である。

両辺の対数をとって

$$\log(X_{t+1})$$
$$=\log(X_t)+\log\{R_{t+1}(\log(X_t))\}-\log(R_{t+1}^{min})$$

$Y=\log X$ の関係を用いると

$$Y_{t+1}=Y_t+\log\{R_{t+1}(Y_t)\}-\log(R_{t+1}^{min})\quad(13)$$

$W=Y_{t+1}$，$V=Y_t$ とおいて上式の関係を適用する。Y_t の確率密度関数を $f_t(y_t)$ とすると

$$H(Y_{t+1})=H(Y_t)+\int_0^\infty f_t(y_t)\log\left|\frac{dy_{t+1}}{dy_t}\right|dy_t\quad(14)$$

（13）式から

$$\frac{dy_{t+1}}{dy_t}=1+\frac{d\log\{R_{t+1}(y_t)\}}{dy_t}$$
$$=1+\frac{1}{R_{t+1}(y_t)}\,\frac{dR_{t+1}(y_t)}{dy_t}$$
$$=1+\frac{[R_{t+1}(y_t)]'}{R_{t+1}(y_t)}>0$$

（10）式より右辺は正である。したがって（14）式の右辺第2項は

$$\int_0^\infty f_t(y_t)\log\left|\frac{dy_{t+1}}{dy_t}\right|dy_t$$
$$=\int_0^\infty f_t(y_t)\log\left(1+\frac{[R_{t+1}(y_t)]'}{R_{t+1}(y_t)}\right)dy_t$$
$$=\int_0^\infty f_t(y_t)\log\left(\frac{R_{t+1}(y_t)+[R_{t+1}(y_t)]'}{R_{t+1}^{min}}\right)dy_t$$
$$-\int_0^\infty f_t(y_t)\log\left(\frac{R_{t+1}(y_t)}{R_{t+1}^{min}}\right)dy_t$$

（12）式より右辺第2項は0であるから，（14）式は

$$H(Y_{t+1})$$
$$=H(Y_t)+\int_0^\infty f_t(y_t)\log\{R_{t+1}(y_t)$$
$$+[R_{t+1}(y_t)]'\}\,dy_t-\log(R_{t+1}^{min})\qquad(15)$$

となる。

時間経過とともにエントロピーが増大していくためには，(15) 式の右辺第 2 項と第 3 項をあわせた値がプラスになればよい。すなわち

$$\int_0^\infty f_t(y_t)\log\{R_{t+1}(y_t)+[R_{t+1}(y_t)]'\}dy_t-\log(R_{t+1}^{min})>0 \qquad (16)$$

である。左辺第 1 項は，「(変化倍率)＋(変化倍率の傾き)」の相乗平均にたいし，対数をとった値である。この値が下限値の変化倍率よりも大きければエントロピーが増大する。

【7】ベキ乗則の生じる条件

以上のことをまとめると，ベキ乗則が生じるためには

① (10) 式：規模の順位が変化しないこと
② (12) 式：相対変化倍率の相乗平均が 1 になること
③ (16) 式：「(変化倍率)＋(変化倍率の傾き)」の相乗平均＞下限値の変化倍率

が満たされることが必要である。

ただし本論では，連続的確率変数が時間経過によっても連続的確率変数として推移すると前提して論理を組み立てている。それゆえ，①規模の順位が変化しないという条件が必要である，このような状況は実際の場面では最終局面において生じうると考えられる。したがって，ここでの必要条件はダイナミックに順位が変化する初期の現象ではなく，ベキ乗則が生成しつつある最終局面における条件と想定した方がよい。

②の相対変化倍率の相乗平均が 1 という状況の具体的なケースを考えてみよう。たとえば Barabási et al. (1999) が前提している場合である。新たにウェブに参加する者が既存のどのウェブページに接続するかという問題である。Barabási et al. は，個々のウェブページに張られているリンク数に比例して接続するという仮定をもうけた。この場合，本稿の論理に翻訳すれば，個々の変化倍率がすべて等しいということになる。下限値の変化倍率も同様に等しいので，個々の相対変化倍率は 1 となる。したがって，相対変化倍率の相乗平均も 1 になる。

次に，下限値が変化しない $R_{t+1}^{min}=1$ のケースを考えてみよう。この場合には，ここで定義した相対的な規模ではなく，絶対的な規模に下限が存在するケースである。

(12) 式は

$$\int_0^\infty \log\{R_{t+1}(y)\}f_t(y)dy=0 \qquad (17)$$

となる。

下限が存在するケースとしては Gabaix (1999) が議論をしている。ただし，幾何ブラウン運動を用いてモデルを構築しているので，標本経路はいたるところ微分可能ではない。したがって本稿の論理には翻訳しにくい。Gabaix のモデルを離れて本稿の場合で考えれば，個々の変化倍率が 1 に近い収束局面が想定される。

③の「(変化倍率)＋(変化倍率の傾き)」の相乗平均が下限値の変化倍率より大きければエントロピーが増大するという具体的な状況を考えてみよう。$R_{t+1}^{min}=1$ の場合がより単純なので，このケースをみる。(16) 式は

$$\int_0^\infty \log\{R_{t+1}(y)+[R_{t+1}(y)]'\}f_t(y)dy>0 \qquad (18)$$

となる。

$R_{t+1}(y)$ は時点 t における規模に関する変化倍率の関数である。したがって変化倍率の傾きとは，同一時点で規模が変化したとき変化倍率がどのように変化するかを表す。変化倍率の相乗平均が 1 であり，かつ，「(変化倍率)＋(変化倍率の傾き)」の相乗平均が 1 以上という条件から，変化倍率の傾きが全般的にプラスということが想定される。ということは，規模が大きいほど変化倍率もプラスに大きい。すなわち，規模が大きいものはより大きくなる。Yule 過程を用いたモデルでジブラ法則が成り立つ場合に，そのようなことが言えた。そのような状況が③の具体的なケースとして想定される。

もちろん，ベキ乗則が成立する収束局面ではエントロピーが最大になっているので，すべての変化は静止し，変化倍率は 1 になる。したがって変化倍率の傾きは 0 になる。そのとき (18) 式は

(17) 式と同じものになる。

【8】ジップ則の成立条件(注8)

ベキ乗係数βが1になる条件を考察しよう。この問題では変数Xの値は正であることを前提にしている。連続的確率変数で近似したとき，Xの値は限りなく0に近いものまで範囲にしていた。しかし実際には，Xの下限は0よりは大きな値である。その下限をx_Lとしよう。

$Y=\log X$であるから，下限に対応するYの値は$y_L=\log x_L$になる。Yは指数分布をする。範囲は0以上としていたが，実際にはy_L以上になる。したがって（5）式 の確率密度関数は

$$g(y)=\beta e^{-\beta(y-y_L)},\quad y\geq y_L$$

となる。（2）式のベキ乗則の関係になおすと

$$P(X\geq x)=\left(\frac{x}{x_L}\right)^{-\beta},\quad x\geq x_L$$

これは（1）式において$\alpha=(x_L)^{\beta}$とおいたものになる。

（4）式に対応するXの確率密度関数は

$$f(x)=\beta\,(x_L)^{\beta}x^{-(\beta+1)},\quad x\geq x_L$$

になる。

Xの期待値を計算しよう。

$$E(X)=\int_{x_L}^{\infty}x\beta\,(x_L)^{\beta}x^{-(\beta+1)}dx$$

$$=\beta(x_L)^{\beta}\int_{x_L}^{\infty}x^{-\beta}dx$$

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

$\beta>1$であれば

$$E(X)=x_L\frac{\beta}{\beta-1}$$

したがって

$$\beta=\frac{1}{1-\dfrac{x_L}{E(X)}}$$

となる。

期待値$E(X)$が下限x_Lより十分大きければ，βは1に近くなりジップ則がなりたつ。

本稿の主要部分を国友直人・東京大学経済学部教授に検討していただきました。謝意を表します。もちろん誤りがあればすべて筆者の責任です。

〔注〕

（1） ここでの議論はNewman (2005) p.13を参照している。
（2） Reza p.280, Johnson p.13 (Lemma 1.2)
（3） 杉浦 p.351
（4） 杉浦 p.290
（5） 杉浦 p.227 定理4.2
（6） 杉浦 p.357 定理17.9
（7） 以上の計算はReza p.273
（8） Gabaix (1999) の考え方による。

【参考文献】

・Barabási, A.L., R. Albert, H. Jeong (1999) "Mean-field theory for scale-free random networks", Physica A 272 p.173-187
・Belevitch, V. (1956) "Langage des Machines et Langage Humain", Office de Publicite, Pruxelles（佐々木宗雄訳『機械の言葉と人間の言葉』みすず書房，1968）
・Gabaix, X. (1999) "Zipf's law for cities: an explanation", The Quarterly Journal of Economics, p.739-767
・Harrison, J. (1985) "Brownian Motion and Stochastic Flow Systems", John Wiley & Sons
・Johnson, O. (2004) "Information Theory and The Central Limit Theorem", Imperial College Press, London
・Miller, G. A. (1957) "Some effects of intermittent silence", American Journal of Psychology 70, p.311-314.
・Mitzenmacher, M. (2003). "A brief history of generative models for power law and lognormal distributions", Internet Mathematics 1, p.226-

251
・Newman, M. E. J. (2005) "Power laws, Pareto distributions and Zipf's law" Contemporary Physics 46, p.323-351
・Reed, W. J. (2003) "The Pareto law of income-an explanation and an extension", Physica A 319, p.579-597
・Reza, F. M. (1961) "An introduction to information theory" Dover Publication, INc., New York
・Simon, H. A. (1955) "On a class of skew distribution function", Biometrika Vol.42, No. 3/4 p. 425-440
・Zipf, G. K. (1949) "Human Behaviour and the Principle of Least Effort", Addison-Wesley, Reading, MA.
・杉浦光夫 (1980)「解析入門 I」東京大学出版会