〔論　文〕

要素の選択肢数に基づくベキ乗則生成モデル

鈴　木　　武

Abstract

In this paper, we describe that the power law appears when leaving to the natural state under certain conditions. We took the logarithm of the size of the members making up the system and found that the expected value was constant, and we maximized the entropy under that condition. The distribution of the size of the members appearing at that time follows the Pareto distribution. Another law that emerges when leaving to the natural state under certain conditions, that is the central limit theorem. Distribution that maximizes entropy under constant variance condition is normal distribution. The power law is considered to be a universal phenomenon compared with this central limit theorem.

はじめに

ベキ乗則（power law）の生成モデルについて議論する。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。その生成メカニズムについて有力なモデルを提示したのは Simon (1955) である。基本的な仮定は、優先的選択 (preferential attachment) と呼ばれるものである。都市人口の大きさを例にとり、新たに都市に参入する人がどの都市に住もうとするかは、都市人口の大きさに比例して誘引される、というものである。

その後、ベキ乗則の議論はしばらく静かであったが、1990 年代になりネットワークにおけるベキ乗則の生成について Barabási & Albert (1999) が論文を発表してから、研究が盛んになった。その基本的仮定も優先的選択であった。その意味で、数式の展開は少し異なるが、本質的には同じ議論であった。ただし扱う対象が、Simon の場合は既存の都市に人が住むケースを、Barabási & Albert は既存のウェブサイトに新たなウェブサイトが参入するケースを扱っている点が異なっている。すなわち、Simon の場合は扱う対象である都市数が変化しない。それに対し、Barabási & Albert の場合は対象であるウェッブサイト数が変化している。

Simon と Barabási & Albert はメンバーが追加されていくという観点から、「成長するモデル」と呼ばれている。他方、メンバーを固定する「成長しないモデル」もある。各メンバーに重みを確率的に賦与して、それぞれを閾値ルールによって結合させる、閾値モデルである。

本稿では、上記のどの考えとも違う、事象を生成する要素の観点からモデルを構築したい。ベキ乗則生成の誘因として、規制のない自然の状態でエントロピーが増大していく過程を想定している。中心極限定理が分散一定のもとでのエントロピー最大化現象であるのに対して、ベキ乗則は期待値一定のもとでのエントロピー最大化現象であると言えることを示したい。

１では、事象を構成する要素と、似た要素からなる要素群について述べる。一連の要素群から選ばれた要素の集まりが選択肢であり、その選択肢のどれかを選ぶと事象が生起する。本稿

では、生起する事象の大きさ、あるいは生起するまでの期間の長さは選択肢数に比例すると仮定している。また、特定の事象は一度しか生起しないと前提している。2では、生起する事象の大きさ、あるいは生起するまでの期間の長さの対数をとり、その期待値を計算している。3では、対数の期待値が一定の値であることを条件にエントロピーを最大化すると、大きさの分布はパレート分布になることを証明する。4では、事象の生起する過程を、一定の条件のもとでポアソン過程として理解する。5では、優先的選択モデルおよび閾値モデルと比較して、本稿のモデルの特色を述べる。特に、一定の条件のもとで自然状態に任せる結果としてベキ乗則が生成されるのが、中心極限定理と比肩しうるのではないかと述べる。6では、都市人口数、ウェブサイト接続数、名字世帯数、地震発生数のデータを用いて計測されたベキ指数を示し、その解釈を記述する。

1　モデルの前提

1.1　事象を構成する要素

ベキ乗則の生成モデルを議論するとき、例として都市人口の大きさとネットワークへの接続数を取り上げよう。都市人口の大きさでは、対象としている国の都市という「カテゴリー」を前提にしている。また、ネットワークへの接続のケースでは、ネットワークを構成してるウェブサイトというカテゴリーを前提にしている。

都市人口の大きさを決める要因を考えてみよう。人がどの都市に住みたいと思うであろうか。ある人が日常的な買い物だけではなく、博物館など文化的な施設にアクセスしやすいことや、高級なレストランにアクセスしやすいことなどを望んでいるならば、その人は大都会を選択するであろう。あるいは、ごく普通の日常的な買い物だけにアクセスしやすければよいと思っている人ならば、小さな町でもよいであろう。すなわち、人がある特定の都市に住もうと決断するためには、その人が要求する都市の構成要素を満たす必要がある。

ある事象が生起するケースを考えてみよう。その事象が生起するまでには、いくつかの「要素」を満たす必要がある。都市選択の例では、都市を選択する各人の要求を網羅する都市構成要素の集合があると想定できる。各人は、その集合のある部分集合を対象として都市を選択することになる。

ここで、似た性質をもつ要素を集めたものを「要素群」と呼び、全部で n 個あるとする。各人が満たすべき要素は各要素群からそれぞれ1つ選ぶことにする。各要素群には、その要素群から選ばないという空要素も含む。

また、特定のカテゴリーには、他のカテゴリーにはない特有の要素があるとする。都市というカテゴリーの例では、都市が立地する場所という要素は、その都市にしかないものである。このような特有な要素も、要素として含めている。

1.2　モデルの仮定

要素群 i の記号を x_i とする。要素群 i に含まれる要素数を k_i 個とする。要素を x_{ij} で表すならば

$$x_i=\left(x_{i1},\cdots,x_{ik_i}\right)$$

n 個の要素群の全体を $(x_1,\cdots,x_n)$ とする。各要素群から1つずつ選択される要素の組が選択肢となる。したがって、選択肢数は $k_1\times k_2\times\cdots\times k_n$ 個になる。

要素群の記号を以下では単に要素数を表す数字として扱うことにすると、選択肢数は $x_1x_2,\cdots x_n$ と表現される。

都市の例で、大都会を選択する人にとっては、多くの要素群から選択しなければいけないので、選択肢数は大きい。それに対し、小さな町でよい人にとっては、大都会に関わる要素は初めから除外してもよいであろうから、選択肢数は小さい。

ある特定の人にとって、都市の選択は一度限りとする。どの大きさの都市を選択するかは満たすべき要素数に関連していると想定してもよいであろう。また、満たすべき要素数が多ければ、それだけ決定までの時間も長くなると考えられる。そこで、次の仮定を設けよう。

【仮定】事象が生起するまでにかかる期間の長さは、その事象が直面する選択肢数に比例する。

ここで、「期間の長さ」という表現をしたが、それを「事象の大きさ」と解釈することもできるものとする。

事象が生起するまでの期間の長さ、あるいは事象の大きさを X とする。それは選択肢数に比例する。比例定数を c とすると

$$X=cx_1x_2\cdots x_n$$

となる。

いま各要素群に含まれる要素数がすべて同じ s 個であると仮定するなら、選択肢数は s^n であり、$X=cs^n$ となる。また、検討すべき要素数は $x_1+x_2+\cdots+x_n=ns$ である。

事象がいつ生起するかは検討すべき要素数に関連するので、それは n に比例する尺度で議論する必要がある。したがって、事象が生起するまでの期間の尺度を構成するには、X の対数をとればよい。

$$\ln X=\ln c+n\ln s$$

ln は自然対数を表している。n は自然数であり、$\ln c$ と $\ln s$ は定数である。したがって、$\ln X$ は $\ln c$ から始まる等間隔な尺度上の値となっている。ここで、X を「期間」、$\ln X$ を「対数期間」と呼ぶことにする。期間は「要素の組合せ数」に、対数期間は「要素数」に比例している。

1.3　対象期間と生起回数

都市を選択する例でみよう。ある人 k が都市を選択するさいに対象としている要素群がある。その要素群からなる選択肢数があり、それに比例する期間 X_{kT} がある。これを「事象 k の対象期間」とよぶ。その対象期間において事象が生起する回数を $\beta(>0)$ としよう。ある人が都市に所属するケースでは、それによって都市数が変化するわけではない。その人が都市に所属するという事象は一度限りであるので、$\beta=1$ である。しかし、ネットワークに接続するケースでは異なってくる。あるサイトというカテゴリーがネットワークに参加すると、{そのサイトに特有の要素，または空要素} という要素群が追加される。それにより、従来まであったサイトから構成される要素の選択肢にそれぞれ２つの選択肢が加わるので、選択肢数が２倍になる。新たに参加するサイトは、従来まであったサイトの要素から構成される選択肢のどれか一つを選ぶことになるので、対象期間の半分で１回生起することになる。したがって、対象期間では $\beta=2$ になる。

各要素 x_{ij} はさらに細かい要素に分解されうる。各要素をそれぞれ一つの要素群 x_{ij} として、その構成要素を $(x_{ij1},\cdots,x_{ijh})$ とする。このような作業を繰り返して微小な要素にまで分解していく。微小な要素群の構成要素がそれ以上分解できないものになっていると、その微小な構成要素を満たすか、満たさないかの二者択一になり、微小な要素群の構成要素数はほとんどすべて等しい数になるであろう。

はじめに各要素群に含まれる要素数が皆等しいと仮定したが、もし要素数が等しくなくても、微小な要素群に分解していくうちに、各要素群を構成する要素数は皆ほぼ等しくなっていくので、その時点を起点として理論を進めればよい。

いったん定めた期間、これが対象期間である。対象期間において事象が生起する回数 β は、微小な要素群に分解されても不変である。というのは、対象期間における要素を満たすことは、その要素を構成している微小な各要素を満たすことを意味しているからである。

以下の議論で、各要素は互いに独立であると仮定する。微小な要素に分解されても、互いに独立である。もし独立でない要素があるなら、それらは一つのまとまった要素とする。

2　対数期間の期待値

2.1　個別事象の期待値

個別の事象ごとに満たすべき要素の構成は異なるので、はじめに、ある事象 k について議論しよう。事象 k の対象期間を X_{kT} とし、その対数対象期間を $\ln X_{kT}$ とする。対数対象期

間のある時点で事象が生起する、それを対数期間 $\ln X_k$ とよぶ。その期待値 $E(\ln X_k)$ を議論する。対数対象期間を微小で等間隔な対数期間 $\Delta\ln x$ に細分化する。細分化された各対数期間における事象の生起は、互いに独立であると想定する。

対数対象期間で事象が生起する回数は β 回である。細分化された微小な対数期間 $\Delta\ln x$ における生起回数の期待値は皆等しいと仮定して、$\beta\dfrac{\Delta\ln x}{\ln X_{kT}}$ となる。この値は 0 に近い正の値とみなされるので、対数期間 $\Delta\ln x$ における事象の生起確率とみなしてよい。

$\ln X_k$ の期待値を求めよう。$\ln X_{kT}$ を n 等分したうちの i 番目の値は

$$i\Delta\ln x$$

その確率は

$$\beta\frac{\Delta\ln x}{\ln X_{kT}}\left(1-\beta\frac{\Delta\ln x}{\ln X_{kT}}\right)^{i-1}$$

である。したがって、期待値は

$$E(\ln X_k)=\sum_{i=1}^{n} i\Delta\ln x\cdot\beta\frac{\Delta\ln x}{\ln X_{kT}}\left(1-\beta\frac{\Delta\ln x}{\ln X_{kT}}\right)^{i-1}$$

ここで、計算のさいの記号の煩わしさを避けるため

$$p=\beta\frac{\Delta\ln x}{\ln X_{kT}}、n=\frac{\ln X_{kT}}{\Delta\ln x}、np=\beta$$

とおく。

$$E(\ln X_k)=\Delta\ln x\cdot p\sum_{i=1}^{n} i(1-p)^{i-1}$$

さらに

$$S=\sum_{i=1}^{n} i(1-p)^{i-1}$$

とおいて、S を計算すると [1]

$$S=\frac{1}{p^2}\{1-(1+\beta)(1-p)^n\}$$

となるので、期待値は

$$E(\ln X_k)=\frac{\Delta\ln x}{p}\{1-(1+\beta)(1-p)^n\}$$

$$=\frac{\ln X_{kT}}{\beta}\left\{1-(1+\beta)\left(1-\frac{\beta\Delta\ln x}{\ln X_{kT}}\right)^{\frac{\ln X_{kT}}{\Delta\ln x}}\right\}$$

$$=\frac{\ln X_{kT}}{\beta}\left\{1-(1+\beta)\left(1-\frac{1}{\frac{\ln X_{kT}}{\beta\Delta\ln x}}\right)^{\left(-\frac{\ln X_{kT}}{\beta\Delta\ln x}\right)(-\beta)}\right\}$$

ここで、$\Delta\ln x\to 0$ にすると

$$E(\ln X_k)\to\frac{\ln X_{kT}}{\beta}\{1-e^{-\beta}(1+\beta)\}$$

になる。

2.2　平均量の期待値

個別事象を集計したものの性質を議論しているので、その平均量 $\ln X$ について考えよう。

$$\ln X=\sum_{k=1}^{m} w_k\ln X_k\ ,\ \sum_{k=1}^{m} w_k=1$$

とする。対象期間の平均量は

$$\ln X_T=\sum_{k=1}^{m} w_k\ln X_{kT}$$

である。ここで、$\ln X_T$ の値は一定になる。というのは、各 $\ln X_{kT}$ は定まった値であり、各事象は同じウェイトを持つ。ここでは、値が $\ln X_{kT}$ になる事象がいくつかあり、その全体の割合が w_k になっていることを表している。

平均量の期待値は

$$E(\ln X)=E\left(\sum_{k=1}^{m} w_k\ln X_k\right)$$
$$=\sum_{k=1}^{m} w_k E(\ln X_k)$$
$$=\sum_{k=1}^{m} w_k\cdot\frac{\ln X_{kT}}{\beta}\{1-e^{-\beta}(1+\beta)\}$$
$$=\frac{1}{\beta}\ln X_T\{1-e^{-\beta}(1+\beta)\}$$

となる。

β および $\ln X_T$ は一定の値なので、期待値 $E(\ln X)$ は一定の値になる。平均量の期待値を $\ln X_M$ で表し、「平均対数期間」とよぶ。事象が平均 1 回生起する期間が平均対数期間である。事象が平均 β 回生起するのは、平均対数期間の β 倍の期間になる。それを「基準対数期間」として、長さを 1 と定めよう。すなわ

ち、

$$\ln X_T\{1-e^{-\beta}(1+\beta)\}=1$$

になるように定数 c を決める。

このモデルでは、理論上想定される対象期間ではなく、実際に事象が生起する期間の長さが重要である。それが平均対数期間であり、形式的にその β 倍の基準対数期間が基礎となる。したがって、基準対数期間の長さを１とおく。それゆえ、$\ln X$ の期待値は

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

であり、一定の値となる。

３．エントロピー最大化

一定の条件のもとで自然に任せると、状態はランダムに拡散していき、エントロピーは増大していく。中心極限定理は、分散一定のもとでエントロピーを最大化する分布は正規分布になることを示している。それに対しベキ乗則は、対数の期待値が一定のもとでエントロピーを最大化する分布はパレート分布になることを示している。

確率変数 X の期待値が一定の値をとるとき、エントロピーを最大化する分布 X は指数分布である。ここでは、対数期間 $\ln X$ の期待値が一定であるので、エントロピーを最大化する $\ln X$ の分布は指数分布である。したがって、X の分布はパレート分布になる。以下では、直接それを計算しよう。

３．１　パレート分布のエントロピー

確率変数 X がパレート分布に従い、その確率密度関数が

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

であるとする。そのエントロピーは

$$\begin{aligned}H(X)&=-\int_1^{\infty}\beta x^{-(\beta+1)}\ln[\beta x^{-(\beta+1)}]dx\\&=-\ln\beta+\beta(\beta+1)\int_1^{\infty}x^{-(\beta+1)}\ln x dx\\&=1-\ln\beta+\frac{1}{\beta}\end{aligned}$$

というのは

$$\begin{aligned}&\int_1^{\infty}x^{-(\beta+1)}\ln x dx\\&=\left[-\frac{1}{\beta}x^{-\beta}\ln x\right]_1^{\infty}-\int_1^{\infty}\left(-\frac{1}{\beta}x^{-\beta}\right)\frac{1}{x}dx\\&=(0-0)+\frac{1}{\beta}\int_1^{\infty}x^{-(\beta+1)}dx\\&=\frac{1}{\beta}\left[-\frac{1}{\beta}x^{-\beta}\right]_1^{\infty}\\&=-\frac{1}{\beta^2}(0-1)\\&=\frac{1}{\beta^2}\end{aligned}$$

３．２　平均対数期間のエントロピー最大化

次の定理が成り立つ。

【定理】$E[\ln X]=\dfrac{1}{\beta}$; $(\beta>0)$ のとき、エントロピーを最大にする分布の確率密度関数は

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

である。

（証明）

条件を満たす任意の確率密度関数を $g(x)$ とすると、

$$E_g[\ln X]=\frac{1}{\beta}$$

エントロピーの記号を $H(\cdot)$ とする。カルバック・ライブラー情報量は負にはならないので

$$H(X_g)=-\int_1^{\infty}g(x)\ln g(x)dx\leq-\int_1^{\infty}g(x)\ln f(x)dx$$

である。$f(x)$ にパレート分布の確率密度関数を代入する。

$$\begin{aligned}&-\int_1^{\infty}g(x)\ln f(x)dx\\&=-\int_1^{\infty}g(x)\ln\left(\beta x^{-(\beta+1)}\right)dx\\&=-\ln\beta+(\beta+1)\int_1^{\infty}g(x)\ln x dx\\&=-\ln\beta+(\beta+1)E_g[\ln X]\\&=-\ln\beta+(\beta+1)\frac{1}{\beta}\\&=1-\ln\beta+\frac{1}{\beta}=H(X_f)\end{aligned}$$

したがって、条件のもとでエントロピーを最大にするのはパレート分布である。

4　ポアソン過程から導くベキ乗則

基準対数期間の長さを1として、対数対象期間を微小で等間隔な対数期間 $\Delta\ln x$ に細分化する。細分化された各対数期間における事象の生起は、互いに独立であると想定する。基準対数期間で事象が生起する回数は平均 β 回であるので、微小な対数期間 $\Delta\ln x$ において事象が生起する回数の期待値は $\beta\Delta\ln x$ である。この値は0に近い正の値とみなされるので、対数期間 $\Delta\ln x$ における事象の生起確率とみなしてよい。

事象が生起するまでの対数期間を $\ln X$ とする。$\ln X$ が $\ln x$ から $(\ln x+\Delta\ln x)$ の対数期間で初めて生起する確率を求めよう。

$$P(\ln x \leq \ln X \leq \ln x+\Delta\ln x)$$
$$=\beta\Delta\ln x(1-\beta\Delta\ln x)^{\frac{\ln x}{\Delta\ln x}}$$
$$=\beta\Delta\ln x\left(1-\frac{1}{\frac{1}{\beta\Delta\ln x}}\right)^{\left(-\frac{1}{\beta\Delta\ln x}\right)(-\beta\ln x)}$$

ここで $\Delta\ln x\to 0$ ならば

$$P(\ln x \leq \ln X \leq \ln x+\Delta\ln x)$$
$$\to\beta d\ln x\times e^{-\beta\ln x}$$
$$=\left(\beta\cdot\frac{1}{x}\,dx\right)x^{-\beta}$$
$$=\beta x^{-(\beta+1)}dx$$

となる。したがって、確率変数 X の密度関数は

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

となる。これはパレート分布であり、ベキ乗則が生成する。

基準対数期間を1として、対数対象期間を微小な対数期間 $\Delta\ln x$ に区分した。それぞれの対数期間において事象が生起する確率は $\beta\Delta\ln x$ で一定で、独立であるとした。したがって、$\ln X$ はポアソン過程に従っていると言える。その場合、事象が生起するまでの期間の長さ $\ln X$ は指数分布に従う。それゆえ、X はパレート分布に従うと言える。

5　他モデルとの比較

5.1　優先的選択モデル

優先的選択とは、システムに参入するメンバーが既存のどのグループに所属するかは、各グループの大きさに誘引される、というものである。このモデルでは、メンバーが順次システムに参入するという観点から、成長するモデルとも言われている。最初にこのモデルを提示したのは Simon (1955) である。1990年代になりネットワークを題材にして Barabási & Albert (1999) が論文を発表している[2]。

成長するモデルであるので、各メンバーが参入するごとに確率分布の変化を記述し、その極限としてベキ乗則を導いている。本稿では、各メンバーがシステムに一度だけ参入するという点は同じであるが、確率分布の変化を記述するのではなく、一定の条件のもとで自然に任せたときの極限をエントロピー最大化という観点から記述している。その意味で、成長するモデルとは言いがたい。

Simon (1955) は、新たに参入するメンバーが既存のどのグループにも所属しない割合を α と明示して、式の中に取り入れている。それに対し Barabási & Albert (1999) では、α を明示せず、それを $\frac{1}{2}$ として式の展開を行っている。本稿では、{新規サイトに特有の要素，または空要素} という要素群を導入することで対処している。

優先的選択では、新たに参入するメンバーは各グループの大きさに誘因されると仮定している。本稿では、検討した要素数から構成される選択肢数がグループの大きさに対応する。

5.2　閾値モデル

閾値モデルでは、各メンバーが既にシステムの中にいるのか、新たにシステムに参入してくるのかは問わない。その意味で成長しないモデ

ルである。閾値モデルを提案したのはM.Granovetter (1983) である。各メンバーは固有の閾値をもっており、その行動基準は閾値に達しているか、いないかの二者択一である。また、各メンバーの閾値はシステム全体である確率分布をもっていて、時間的に一定である。

例えば、ある集団内で交友関係を結ぶか結ばないかという場合、各カップルの行動基準の合計が閾値を超えるか超えないかで、行動を決定する。モデルでは、各メンバーの行動基準に重みを確率的に賦与して、閾値ルールによって結合させるかどうかを決めている。Masuda-Miwa-Konno (2004) では、重みを与える確率分布がいくつか提案されている。

本稿では、各メンバーが要素を満たした時点で閾値を超えたことになる。

５.３　本稿モデルの特色

本稿のモデルでは、一定の条件のもとで自然状態に任せたときに出現する確率分布を求めている。すなわち、システムを構成するメンバーの大きさの対数をとり、その期待値が一定であることを求めて、その条件のもとでエントロピーを最大化した。そのときに出現するメンバーの大きさの分布はパレート分布に従っている、というものである。これがベキ乗則である。

一定の条件のもとで自然状態に任せたときに出現する確率分布としては、中心極限定理がある。すなわち、分散一定の条件のもとでエントロピーを最大化する分布は正規分布である、というものである。

ベキ乗則は、この中心極限定理とならぶ普遍的な現象であると考えられる。

６　いくつかの事例

６.１　都市の大きさと順位

日本とアメリカ合衆国について、都市人口のベキ指数を求めよう。一般的に、人口分布についてはベキ乗則が成立すると言われている。それが成り立つなら、新たに追加される人口は既存の市町村以外に属することはないので、$\beta=1$ となる。

データから β を推定するために、都市人口を大きい方から小さい方に並べ、順位をつけ

$$\log(\text{都市順位})=(\text{定数})-\beta\log(\text{都市人口})$$

として、回帰分析から β を求めることができる。[3]

2010年センサスから、日本とアメリカ合衆国の β を計算した。結果は人口順位グループによって異なっている。日本のデータは2010年国勢調査による。市町村人口は平成大合併以前の区分が実情をより反映していると考えられるので、それを用いた。人口集中地区データは旧区分が明瞭でないので、2010年時点の区分を用いている。

アメリカ合衆国のデータは2010年Census

表1　日本の2010年国勢調査による都市人口順位別の β 値

市町村人口		人口集中地区	
都市順位	β	都市順位	β
1 ～ 14	1.20	1 ～ 15	1.19
15 ～ 35	1.99	16 ～ 40	1.57
36 ～ 100	1.45	41 ～ 100	1.38
101 ～ 440	1.06	101 ～ 130	1.00
441 ～ 2500	0.71	131 ～ 300	0.89
2501 ～ 3233	0.11	301 ～ 829	0.43

（注）市町村人口は2010年の値であるが、平成大合併前の旧区分を用いた。人口集中地区の値は2010年の区分を用いた。

表２　アメリア合衆国2010年都市圏人口順位別の β 値

都市圏順位	β
1 ～ 15	1.64
16 ～ 40	1.21
41 ～ 100	0.97
101 ～ 250	0.75
251 ～ 700	0.68
701 ～ 917	0.27

（出所）Census 2010 of the United States of America
The 917 Core Based Statistical Areas

である。都市圏データの方がより適切であると考えられるので、Core Based Statistical Areas の人口を用いた。

表1と表2から言えることは、大都市あるいはそれに準ずる都市群ではベキ指数が1より大きい。中都市では1であり、小都市あるいは零細都市では1以下である。

この現象は、人口が小さな都市から大きな都市に移動している過程であり、都市人口の分布が定常状態に達していないから、と解釈できる。大都市カテゴリーに新たな都市の参入があり、選択肢数が増えている。ただし、ウェブサイトの例とは異なり、大都市カテゴリーへの参入はある日突然ではなく、徐々に人口が増えて参入しているので、βは2ではなく、それより小さくなる。逆に小都市の場合には、カテゴリーのメンバーが減少しているので、βは1より小さくなる。中都市では退出と参入が同時に行われるので、βは1に近い。

6.2　ネットワークの大きさと順位

ネットワーク接続における計測例をみよう[4]。表3における out は自分からつなぐネット数であり、in は相手からつながれるネット数である。自分からつなぐ場合は、自分は小さなサイトであるケースが想定され、参入することによって選択肢数が増加することが予想される。したがって、βは1より大きい。それに対し、相手からつながれる場合は、自分は大きなサイトであるケースが想定され、相手が参入しても選択肢数が増加したとは思われないと予想される。したがって、βは1に近い。

理論的にはβ=2である。ネットワークに新たに参加するサイトが、すでにある選択肢のすべてと、{新規サイトに特有の要素，または空要素}の2つに枝分かれするので、選択肢数が2倍になるからである。しかし実際には、自分からつなぐケースでも、選択肢数の増加は1.5倍前後である。既存のすべての選択肢につなぐ必要がないということである。

6.3　名字数の大きさと順位

ウェブサイトから入手した「名字ランキング5000」のデータ（出所：http://www2s.biglobe.ne.jp/~suzakihp/ju0001.html）から、

log(名字順位)＝(定数)－β log(名字世帯数)

として、回帰分析からβを求めた。それが表4である。

新しい名字が追加されることはないと想定するなら、β=1になる。実際には0.86となり、1より小さかった。順位区分別にみると、1～500位までのβは1.27であり、501以下の順位では1より小さい。ここから推測されることは、世帯数の多い名字の割合が増加しており、世帯数の少ない名字の割合が減少している、ということである。

6.4　地震の大きさと順位

地震の規模と頻度についてはグーテンベルク・リヒター則（1944年）が知られている。

表4　名字世帯数とその順位別のβ値

	名字順位	β
全名字世帯	1～5000	0.86
順位区分別	1～500	1.27
	501～1000	0.88
	1001～5000	0.75

表3　ネットワークにおける接続数とβ値

ネットワーク	標本数	β out	β in	出典論文
www	325,729	1.45	1.1	Albert, Jeong & Barabási 1999
www	4×10^7	1.38	1.1	Kumar et al. 1999
www	2×10^8	1.72	1.1	Broder et al. 2000

（出所）Albert & Barabási (2002) TABLE Ⅱ。
原典では、βの値ではなく、$\beta+1$の値が掲載されている。

地震が発するエネルギーの大きさを対数で表したマグニチュード（M）と、そのマグニチュード以上の地震の発生回数（N）との関係について

$$\log_{10}(N)=\alpha-\beta M$$

というベキ乗則が成り立つ、というものである。β は１を中心に 0.7~1.4 の範囲にあるとされている。

表５　マグニチュード別の地震回数

マグニチュード	発生度数
1	847,609
2	332,633
3	92,351
4	16,780
5	1,938
6	274
7	29
8	1

（出所）http://www.hinet.bosai.go.jp/about_earthquake/sec1.2.html

本稿の主旨から、個々の地震はマグニチュードというカテゴリーにはいるので、新たに追加されるカテゴリーはない。したがって、$\beta=1$ になる。

実際に報告された β をみよう。梅田康宏「地震の規模別頻度分布」（2012 年１月報告 http://www.eonet.ne.jp/~kansai-catfish/hindobumpu.pdf ）では、期間 1950 年１月１日から 2004 年５月１日まで、深さ 30 ～ 80 ｋｍにおける南関東の地震分布について、$\beta=0.96$ と報告している。

宇野徳治「地震学（第２版）」（1984 年、共立出版、pp.310）では、1965 年～ 1974 年の日本付近で発生した地震分布から、$\beta=0.936$ と記述している。

本稿で計測した数値は、表５の 1997 年～ 2016 年に日本周辺で発生した地震回数である。ここでは、$\beta=0.83$ になった。

以上のことから分かることは、広い範囲で長期間のデータをとれば、β は１に近い値をとるが、狭い範囲や短期間のデータでは、β は１より小さくなる。というのは、狭い範囲や短期間ではマグニチュードの大きな地震が抜けてしまうからと想定できる。

注

（１）$S=\sum_{i=1}^{n} i\,(1-p)^{i-1}$ を計算する。

$$S=1+2(1-p)+3(1-p)^2+\cdots+n(1-p)^{n-1}$$

$$(1-p)S=(1-p)+2(1-p)^2+\cdots+(n-1)(1-p)^{n-1}+n(1-p)^n$$

上式－下式

$$S-(1-p)S=pS$$

$$=1+(1-p)+(1-p)^2+\cdots+(1-p)^{n-1}-n(1-p)^n$$

ここで

$$(1-p)pS=(1-p)+(1-p)^2+\cdots+(1-\mathrm{p})^{n-1}+(1-p)^n-n(1-p)^{n+1}$$

上式－下式

$$pS-(1-p)pS=p^2S$$

$$=1-(n+1)(1-p)^n+n(1-\mathrm{p})^{n+1}$$

$$=1-(1-p)^n(n+1-n+np)$$

$$=1-(1+\beta)(1-p)^n$$

よって

$$S=\frac{1}{p^2}\{1-(1+\beta)(1-p)^n\}$$

（２）　サイモン・モデルの精緻化については、鈴木武 (2007) を参照。サイモン・モデルとバラバシ・モデルの比較については鈴木武 (2016b) を参照。

（３）　鈴木武 (2016a) を参照。

（４）　鈴木武 (2016a) を参照。

参考文献

H.A.Simon (1955) “On a Class of Skew Distribution Functions”, *Biometrika*, Vol.42, No.3/4, pp.425-440.

A.L.Barabási, R.Albert (1999) “Emergence of Scaling in Random Networks”, *Science*, Vol.286, pp.509-512.

R.Albert, A.L.Barabási(2002) “Statistical Mechanics of Complex Networks”, *Review of Modern Physics*, Vol.74, pp.47-97.

M.Granovetter (1978) “Threshold Models of Collective

Behavior", *The American Journal of Sociology*, Vol.83, No.6, pp.1420-1443 .
N.Masuda, H.Miwa, N.Konno (2004) "Analysis of Scale-free Networks based on a Threshold Graph with Intrinsic Vertex Weights", *Physical Review E*, **70**, 036124.
鈴木武 (2007)「参入下限値を単位としたベキ乗則生成モデル」、『経営志林』第 44 巻 2 号、pp.1-13.
鈴木武 (2016a)「超優先的選択および非定常状態におけるベキ乗則生成モデル」、『経営志林』、第 52 巻 4 号、pp.1-13.
鈴木武 (2016b)「ベキ乗則生成に関するサイモン・モデルとバラバシ・モデル」、『経営志林』第 53 巻 1 号、pp.1-10.