

(総説) 「ベキ分布」の特徴と数理

梅野 善雄*

The Feature and Mathematical Principle of the Powerlaw Distribution

UMENO Yoshio*

Abstract: The powerlaw distribution appears in the various scenes in the world. However, it is rarely mentioned in lectures because the conventional statistics education is based on normal distribution. The powerlaw distribution has no mean or variance depending on exponent values, in that case we cannot apply the central limit theorem. Here, We introduce features and practical examples of the powerlaw distribution, then we shall comment on one end of mathematical principles.

Keywords: powerlaw distribution, scale-free, longtail, self-similar, self-organized criticality

1 はじめに

従来の統計教育では、正規分布が重要な位置を占めている。特に、どのような確率分布であるかが不明でも、標本数が大きければ標本平均は正規分布で近似できるとする中心極限定理がある。そのため、検定などでは標本数が大きい場合は正規分布に従っているとみなして差し支えないとして議論が進められることが多い。ただし、中心極限定理は平均や分散が存在することを前提とするので、それらが存在しない確率分布では中心極限定理は成立しない。そのような確率分布としてはコーシー分布が知られているが、統計の授業で取り上げられることはほとんどないといってよい。

正規分布や中心極限定理を重視した授業を受けると、平均や分散が存在しない確率分布があるという発想は思考回路に浮かんでこないのではないだろうか。それらが存在しない確率分布としてはコーシー分布の他に「ベキ分布」があり、近年ベキ分布に従う現象が山のよう

に報告されている。そのような現象の分析を卒業後の学生が任された場合、正規分布や中心極限定理を念頭に置きすぎると判断を誤る可能性があるのではないかと危惧される。

そこで、ここでは、教科書で取り上げられることは少ないものの、マクロからミクロまで、多方面の場面で頻繁に出現する「ベキ分布」について、その概要を紹介する。この確率分布は、ベキ指数の値によっては平均や分散が存在しない場合がある。

* 一関工業高等専門学校名誉教授, National Institute of Technology, Ichinoseki College,
Professor Emeritus, E-mail: umenoy@yunavi.lsv.jp, [URL] <https://yunavi.lsv.jp/>

2 「ベキ分布」の定義と乱数の標本平均

2.1 「ベキ分布」の定義

ベキ分布の確率密度関数は次の式で定義される。

$$f(x) = \frac{ab^a}{x^{a+1}} \quad (a > 0, b > 0, x \geq b)$$

この分布は 19 世紀のイタリアの経済学者ヴィルフレド・パレートが見いだしたものでパレート分布とも呼ばれ $\text{Par}(a, b)$ で表される。ちなみに、「全体の富の 80% は人口の 20% が所有している」という「80 : 20 の法則」は「パレートの法則」とも呼ばれ、ビジネス社会ではいろいろな場合に当てはまるとされる有名な法則である。

この確率分布の平均と分散は、次のように計算される。

$$\begin{aligned} E[X] &= \int_b^\infty \frac{ab^a}{x^a} dx = \left[\frac{ab^a}{(1-a)x^{a-1}} \right]_b^\infty = \frac{ab}{a-1} \quad (a > 1 \text{ のとき}) \\ E[X^2] &= \int_b^\infty \frac{ab^a}{x^{a-1}} dx = \left[\frac{ab^a}{(2-a)x^{a-2}} \right]_b^\infty = \frac{ab^2}{a-2} \quad (a > 2 \text{ のとき}) \\ V[X] &= E[X^2] - E[X]^2 = \frac{ab^2}{(a-1)^2(a-2)} \end{aligned}$$

この計算から、 $a \leq 1$ のときは平均が、 $a \leq 2$ のときは分散が、いずれも広義積分が収束しないので存在しないことが分かる。

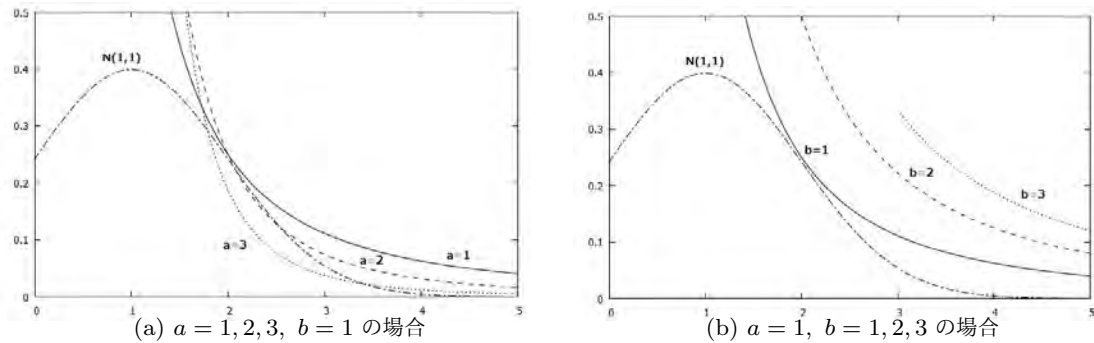


図 1: $\text{Par}(a, b)$ のグラフ

図 1 はベキ分布の確率密度関数のグラフである。比較のために標準正規分布を 1 だけ平行移動した $N(1, 1)$ のグラフも描画した。図 1(a) は $b = 1$ として $a = 1, 2, 3$ の場合であり、(b) は $a = 1$ として $b = 1, 2, 3$ とした場合である。ベキ関数であるので、指数関数と比較すると減衰の仕方が緩やかであることが分かる。

また、 $f(x) = ab^a x^{-(a+1)}$ の対数をとると

$$\log f(x) = \log ab^a - (a+1) \log x$$

となるので、両対数グラフでは右下がりの直線になり、ベキ指数が傾きを表す。

2.2 「ベキ分布」に従う乱数 10 個の標本平均

統計の授業で平均が存在する確率分布ばかり扱っていると「平均が存在しない」ということは感覚的に理解しにくいところがある。そこで次に、平均が存在しない場合について、乱数を利用して標本平均の変化を調べてみる。

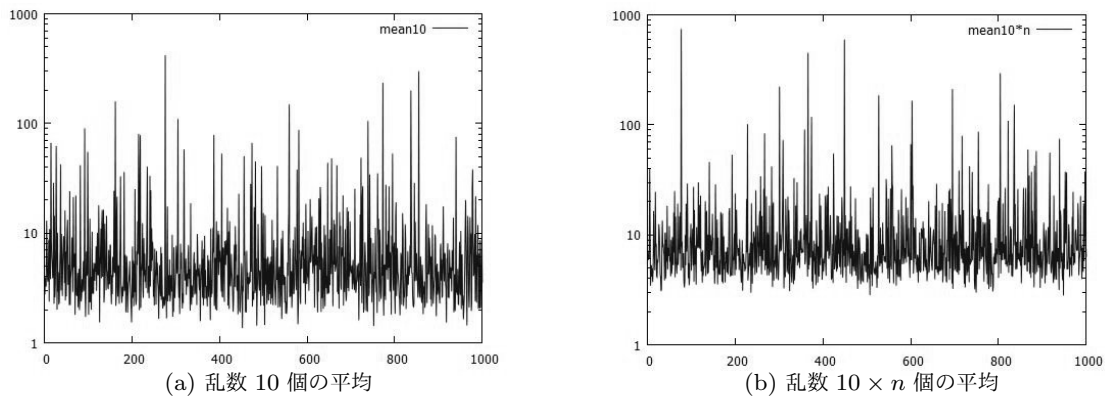


図 2: Par(1,1) にしたがう乱数の平均

図 1(a) は、平均も分散も存在しない Par(1,1) にしたがう乱数を用いて、10 個の標本平均を 1000 回計算して描画したものである。計算するたびに標本平均の値が大きく変動する。その変動の様子を示すため折れ線グラフで示し、変動幅が大きいため縦軸は対数軸にとった。横軸は何回目に計算した値であるかを示す。何度か試すと標本平均が 1000 を超える場合も出てくる。図 1(b) は、 n 回目の標本数を $10 \times n$ ($1 \leq n \leq 1000$) として計算した。計算するごとに標本数が増えるが、標本数を幾ら増やしても標本平均が落ち着く様子は全く見られない。この場合も標本平均が 1000 を超える場合がある。ちなみに、標準正規分布 $N(0,1)$ で同様のことを行くと、10 個の標本平均の値はおおむね $[-1, 1]$ の範囲内にあり、標本の大きさを増やしていくと $[-0.1, 0.1]$ の範囲内に納まっていく。

このように、Par(1,1) の場合は、標本平均の値を何回計算しても、標本数を幾ら大きくしても、(そもそも平均も分散も存在しないので) 平均がどのような値になりそうなのか、また、どのような散らばり具合をするのかが全く分からず、時折とんでもなく大きな値になることがある。たとえば、乱数 10 個の標本平均が 1000 を超える場合にどのような値が生成されたかをみると、標本平均が 1253.14 となったときの値は

$$1, 2, 12503, 2, 1, 3, 1, 10, 1, 2$$

が生成されていた。ただし、整数部分を示した。標本平均が 1000 を超えたのは 12503 という値が 1 個含まれていたことによる。この場合の標本標準偏差は 3750 である。

図 2 を見ても分かるように、標本数を増やしながらか計算を繰り返しても標本平均の値は大きく変動する。そこで次に、計算回数を 1 万回に増やして、 $k+1$ 回目の標本平均と k 回目の標本平均との差の絶対値の変化を調べてみる。

図 3 は、10 個の標本平均を 10001 回計算して、 k 回目の値を $\text{par10}(k)$ とするとき、

$$|\text{par10}(k+1) - \text{par10}(k)| \quad (1 \leq k \leq 10000)$$

の変化をみたものである．計算するたびに差は激しく変動していることが分かる．

図 4(a) は，この変動値から 1 刻みの度数分布表を作成してヒストグラムを描画したものである．図 4(b) はヒストグラムを折れ線にして両対数グラフで示した．出だしの部分を除くとほぼ直線状になっている．直線は $\text{Par}(1,1)$ のグラフである．10 個の標本平均の変動の絶対値は，変動がある程度以上大きい部分は同じベキ指数を持つべき分布に従っていることが読み取れる．乱数平均が図 2 や図 3 のように激しく変動するように見えても，その変動の仕方はもとの確率分布に依存しているように思われる．

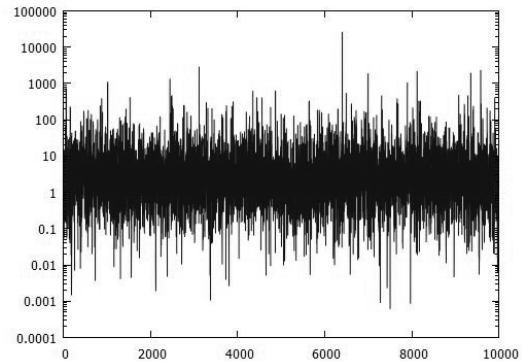
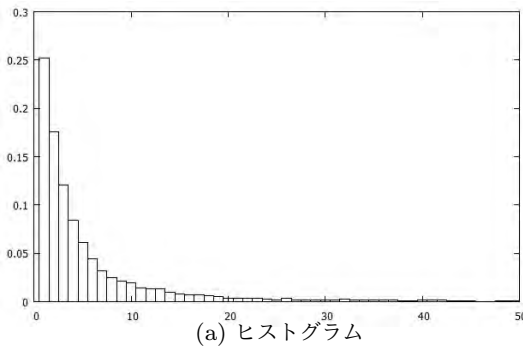
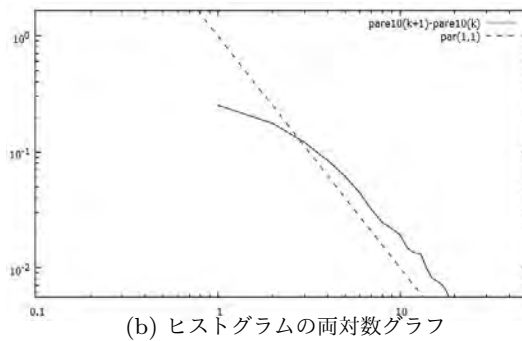


図 3: 10 個の標本平均の変動の絶対値



(a) ヒストグラム



(b) ヒストグラムの両対数グラフ

図 4: $\text{Par}(1,1)$ に従う乱数 10 個の標本平均の変動の絶対値のヒストグラム

3 いろいろな場面で現れる「ベキ分布」

ベキ分布 $f(x) = Cx^{-(a+1)}$ ($C = ab^a$) の累積分布関数は，

$$P(X \leq x) = \int_b^x f(t) dt = 1 - \frac{b^a}{x^a}$$

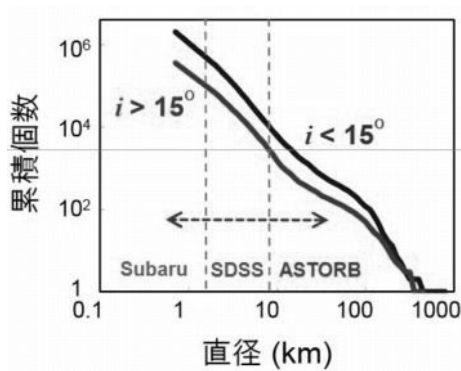
であるので，相補累積分布関数 $P(X \geq x)$ は

$$P(X \geq x) = 1 - P(X \leq x) = \frac{b^a}{x^a}$$

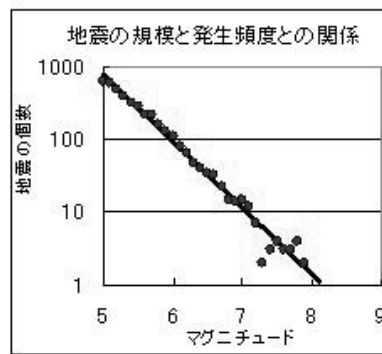
となる．したがって，相補累積分布関数の両対数グラフも直線である．

両対数グラフが直線になるような分布は，自然現象から経済現象・社会現象までのあらゆるところに現れている．図 5 に詳細を Web などから参照できる幾つかの例を示す．すべて両対数グラフであり，相補累積分布の場合も混在している．

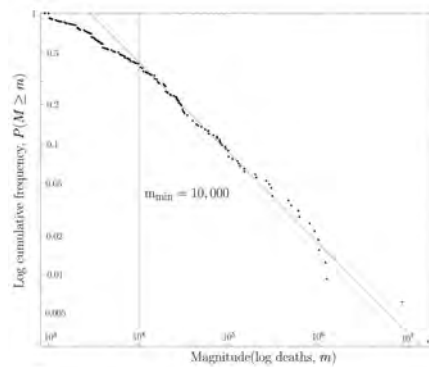
図 5(a) は，火星と木星の間にある小惑星帯を軌道傾斜角 15° を境にして 2 つに分割し，それぞれの惑星の直径 (横軸) と累積個数 (縦軸) を示したものである．小惑星の個数は，直径



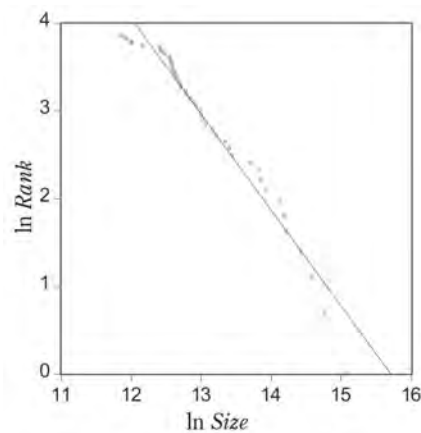
(a) 火星と木星の間にある小惑星の直径と個数 ([2], p.11 より転載)



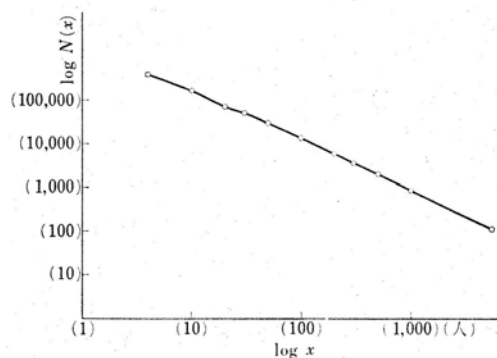
(b) 地震の規模と発生頻度 ([3] より転載)



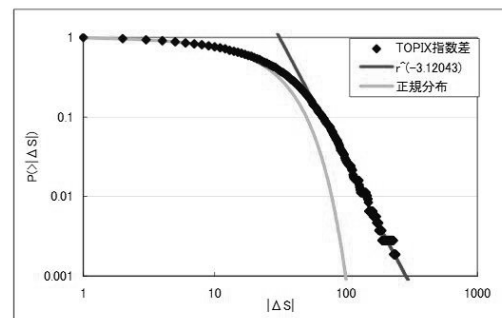
(c) 戦死者の数と戦争の数 ([4], p.4 より転載
目盛りのフォントと配置等を一部修正)



(d) 県庁所在地の人口と順位 (2003 年)
([5], p.34 より転載)



(e) 従業員数と x 人以上の企業の数 (1970 年)
([6], p.2 より転載)



(f) TOPIX(東証株価指数) 変化の累積分布
([7], p.5 より転載)

図 5: ベキ分布の具体例 (両対数グラフ)

を横軸としたベキ分布に従っていることが分かる。図 5(b) は、理科年表 (2005 年) のデータをもとに、1661 年から 1999 年の間に日本で起きたマグニチュード 5 以上の地震の、マグニチュードと発生頻度の関係を示したものである。地震の規模と頻度の間にこのような直線的な関係があることは「ゲーテンベルク・リヒターの法則」と呼ばれている。

図 5(c) は、1816 年から 2007 年までの全ての戦争で、戦死者の数を横軸として、戦死者 x 人以上の戦争の数を縦軸とした場合のグラフである。図 5(d) は、2003 年の県庁所在地の

人口を人口順に並べて、縦軸に順位、横軸には人口をとったものである。この例に限らず、いろいろな事象の出現頻度を順位順に並べて両対数グラフで表示すると直線になることが多い。これは、言語学者 G.K.Zipf が小説における英単語の出現頻度を調べて、(出現頻度) $\times$ (順位) = (定数) という関係があることを見いだしたことによる (1936 年)。この関係は「ジップ (Zipf) の法則」と呼ばれ、分野を問わず非常に多数の場面で見いだされている。

図 5(e) は、1970 年 12 月末現在の日本の製造業において、従業員 4 人以上の全法人企業の従業員規模の分布である。横軸は従業員数 (x)、縦軸は従業員 x 人以上の企業数を表す。図 5(f) は、TOPIX(東証株価指数) の 1986 年 5 月 8 日から 2006 年 11 月 16 日までの木曜日の週次データをもとにしたもので、横軸は 1 週間の値の変化を Δx とするとき $|\Delta x|$ の値、縦軸は変化が $|\Delta x|$ 以上の相補累積分布を表す。

特に、近年、ネットワークの分析でベキ分布の重要性が指摘されるようになってきた。神経回路網・タンパク質間の相互作用・航空経路網・企業の取引関係など、多数のものをネットワークとして捉えることができる。人と人との関係では、論文の共著者、映画の共演者、あるいは感染症の感染経路などもネットワークとして捉えられる。ネットワークは点と点を辺で結んで構成されるので、要するにグラフ理論の世界になる。特定の点 A と辺で結ばれる点が多いほど、点 A は重要な点であることを示している。航空空路網でいえば、そのような点はハブ空港に相当する。

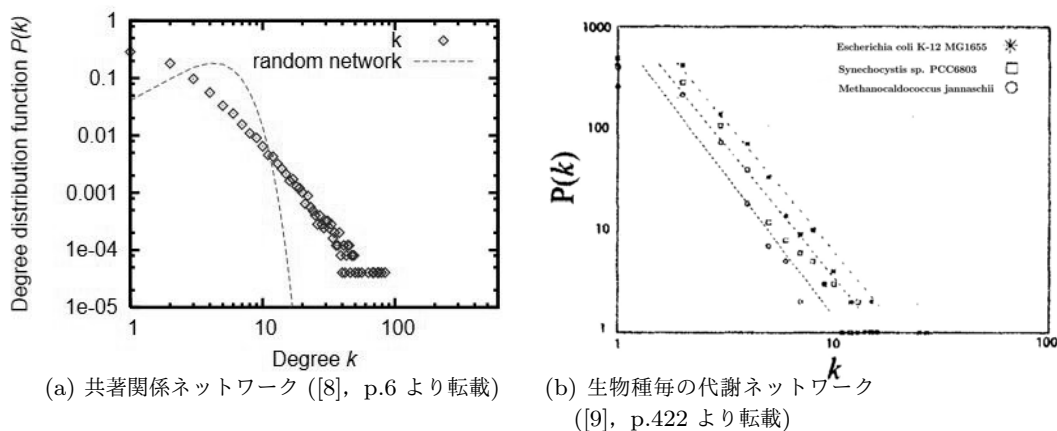


図 6: ネットワークの次数分布

図 6(a) は、1952~2004 年の IEEE の論文誌・国際学会予稿集における論文の著者を点で表し、共著関係にある著者どうしを辺で結んで得られるネットワークにおいて、一つの点と結ばれている辺の本数 (これを「点の次数」という) を横軸とし、その次数を持つ点の個数を縦軸にとったときのヒストグラムを両対数グラフで表したものである。これは、著者の数は共著者数を横軸としてベキ分布にしたがっていることを示している。(b) は、生体内に存在する多くの化合物を酵素反応による代謝ネットワークとみたとき、その次数分布を表したものである。異なる生物種 (大腸菌 K-12 株 MG1655, シアノバクテリア, 超好熱性のメタン菌) において、その次数分布がいずれもベキ分布にしたがうことが示されている。

これらの例に限らず、分野を問わず非常に多くのネットワークで次数分布がベキ分布にしたがうことが報告されている。Newman[10, p.6] には、もっと多数の例が紹介されている。

4 ベキ分布の性質

ここでは、ベキ分布の性質について取りまとめる。ベキ分布は他の確率分布にはないいくつかの特徴がある。

■スケールフリー性

ベキ分布 $f(x) = Cx^{-(a+1)}$ ($C = ab^a$) では、 a の値によっては平均や分散の値が存在しない。分布を特徴づける値が存在しないことから、この性質はスケールフリー性と呼ばれる。実際の例では a は 1 ~ 2 の値を取ることが多いようであるが、その場合は平均が存在しても分散が存在しない。このことは、乱数を発生させて標本平均を取るとき、「時折、平均と比較するととんでもなく大きな値をとることがある」という形で現れてくる。

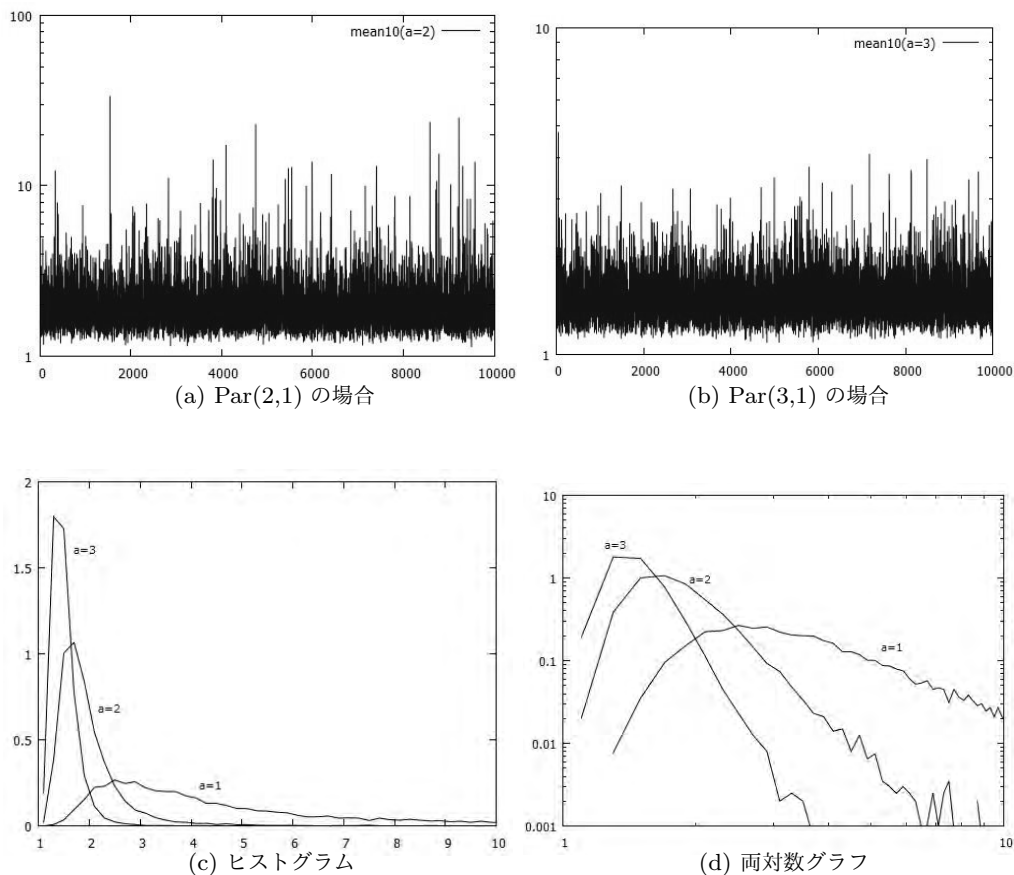


図 7: $\text{Par}(a, 1)$ にしたがう乱数 10 個の標本平均 (1 万回)

図 2(a) では、 $\text{Par}(1,1)$ に従う乱数 10 個の標本平均 1000 個の変化を折れ線で示した。この場合は平均も分散も存在しないので標本平均の値は大きく変動する。図 7(a) は $\text{Par}(2,1)$ での乱数 10 個の標本平均、図 7(b) は $\text{Par}(3,1)$ での乱数 10 個の標本平均を、いずれも 1 万回計算して変化が分かるように折れ線で示したものである。(a) では平均は 2 で分散は存在しない。(b) では平均は 1.5 で分散は 0.75 である。同じようなグラフに見えるが、平均や分散の存在の有無により縦軸のスケールが変わっていることが分かる。(a) では平均は 2 の

に、10 個の標本平均の値が 20 以上の場合が何度か出ている。

図 7(c) は、 $a = 1, 2, 3$ の場合の乱数 10 個の標本平均を 1 万回求めたときのヒストグラムを 0.2 刻みで作成して折れ線グラフで描画したものである。平均が存在すると、その箇所での盛り上がりができ、分散が存在すると平均の近くへの集中が高まっていることが分かる。(d) では、そのヒストグラムを両対数グラフで表示した。いずれも、ある箇所から先は直線状になっており、標本平均がある程度大きい箇所ではベキ分布に従っていることが分かる。

■自己相似性

ベキ分布の重要な性質に自己相似性がある。それは、確率密度関数を $f(x)$ とするとき、

$$f(cx) \propto f(x) \quad (c \text{ は正の定数})$$

という性質である。記号 $\propto$ は比例していることを表す。実際、 x を c 倍した場合は、 $f(x)$ を c^{a+1} 倍すると元の関数と一致する。このことが x の値の大小によらず成立する。この性質は部分が全体と比例することを意味することから、ベキ分布はフラクタル分布と呼ばれることもある。たとえば、図 7(a) の一部を切り出して縦と横のスケールを変えても、図 7(a) とほとんど同じようなグラフになる。

この自己相似性を満たす関数は、微分可能性を仮定するとベキ関数に限る [19, p.10]。

実際、2 つの関数 $f(x), g(x)$ が任意の正数 c に対して $f(cx) = g(c)f(x)$ を満たすとする。 $x = 1$ のときを考えると $g(c) = f(c)/f(1)$ であるので、この関係は $f(x)$ だけを用いて

$$f(cx) = \frac{f(c)}{f(1)} f(x)$$

と表せる。これを c で微分して $c = 1$ の場合を考えると

$$xf'(x) = \frac{f'(1)}{f(1)} f(x)$$

という 1 階線形微分方程式が得られる。この微分方程式を解くと

$$f(x) = Cx^\gamma, \quad \gamma = \frac{f'(1)}{f(1)} \quad (C \text{ は任意定数})$$

が得られ、解はベキ関数に限られる。

この自己相似性を別な観点から眺めてみる。今、簡単のため $\text{Par}(a, b)$ において $b = 1$ の場合を考えて区間 $[1, c]$ での積分を $\int_1^c f(x) dx = p$ とする。実際に計算すると

$$p = \int_1^c \frac{a}{x^{a+1}} dx = 1 - \frac{1}{c^a} \quad \therefore \quad \frac{1}{c^a} = 1 - p$$

である。ここで、 $[c, \infty)$ の中で考えたときに同じ割合 p を持つような区間を考えると、その区間は全体で考えたときの割合が $(1 - p)p = p/c^a$ となるような区間である。この値は、 $x/c = t$ という変数変換をして同じ積分範囲 $[1, c]$ を計算すると得られる。実際、

$$\int_1^c f(ct) c dt = \int_1^c \frac{a}{(ct)^{a+1}} c dt = \frac{1}{c^a} \int_1^c \frac{a}{t^{a+1}} dt = \frac{p}{c^a}$$

となる。つまり、 $[c, c^2]$ で考えれば良いということである。

■ロングテール

ベキ分布の減衰の仕方は、正規分布における指数関数とは異なりベキ乗で緩やかに減衰する。この性質はロングテールと呼ばれている。

表 1: $P(X \geq x) = p$ を満たす x の値

p	平均	0.1	0.01	0.001	0.0001	0.00001
$N(0, 1)$	0	1.2816	2.3263	3.0902	3.7190	4.2649
$\text{Par}(1, 1)$	–	10.000	100.00	1000.0	10000.	100000.
$\text{Par}(1.5, 1)$	3	4.6416	21.544	100.00	464.159	2154.4
$\text{Par}(2, 1)$	2	3.1623	10.000	31.623	100.00	316.23

減衰の程度を、たとえば標準正規分布 $N(0, 1)$ とベキ分布 $\text{Par}(a, 1)$ において、上位確率 p を与えて $P(X \geq x) = p$ となる x の値で比較したのが表 1 である。小さな確率で考えると、ベキ分布ではかなり大きな値まで考慮しなければならない。そのため、乱数を発生させると (平均が存在しても) 平均と比べてとんでもなく大きな値が発生することがある。

■安定分布の漸近近似

通常の統計の教科書では触れられていないが、「安定分布」と呼ばれる確率分布がある。以下、[11] をもとに概略を述べる。安定分布は、ある確率分布にしたがう確率変数 X と、その独立なコピー $X_1, X_2, \dots, X_n$ に対して

$$X_1 + X_2 + \dots + X_n \stackrel{d}{=} a_n + b_n X \quad (a_n \in \mathbb{R}, b_n > 0)$$

を満たす定数 a_n, b_n が存在するような確率分布として定義される [11, def.1]。記号「 $\stackrel{d}{=}$ 」は、両辺の分布が一致することを表す。この定義の中に平均や分散は含まれていない。安定分布は、平均や分散が存在する正規分布や、それらがいずれも存在しないコーシー分布をも含む確率分布になっており、その確率密度関数 $f(x)$ をフーリエ変換した式 (特性関数) で特徴づけられる。なお、フーリエ変換と逆変換は

$$\hat{f}(t) = \int_{-\infty}^{\infty} f(x) e^{itx} dx, \quad f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(t) e^{-itx} dt$$

として考えるものとする。

確率分布が安定分布であるための条件は、その特性関数が次の式で表されることである。

$$\hat{f}(t) = \exp \{ i\delta t - \gamma^\alpha |t|^\alpha \{ 1 + i\beta \operatorname{sgn}(t) \omega(t, \alpha) \} \}$$

ここで、各パラメーターは、次の条件を満たすものとする [11, Prop.11]。

$$0 < \alpha \leq 2, \quad -1 \leq \beta \leq 1, \quad 0 < \gamma < \infty, \quad -\infty < \delta < \infty$$

$\operatorname{sgn}(t)$ は符号関数で、 $\omega(t, \alpha)$ は次式で定義される。

$$\omega(t, \alpha) = \begin{cases} \tan \frac{\pi\alpha}{2} & (\alpha \neq 1) \\ \frac{2}{\pi} \log(|t|) & (\alpha = 1) \end{cases}$$

α は特性指数と呼ばれ分布の厚みを定め (小さいと裾が広い), $\alpha < 2$ のときは分散が, $\alpha \leq 1$ のときは平均が存在しない. β は歪度指数と呼ばれ分布の非対称性を定め, $\beta = 0$ のときは左右対称である. γ は分布の規模を, δ は分布全体の平行移動を定める.

安定分布は 4 つのパラメーターで特徴づけられるので, それを $S(\alpha, \beta, \gamma, \delta)$ で表すことにする. 安定分布の特性関数から確率密度関数を求めることができるのは, 正規分布, コーシー分布, そしてレビ分布の 3 つに限られる. 正規分布 $N(\mu, \sigma^2)$ は $S(2, 0, \sigma/\sqrt{2}, \mu)$ と, コーシー分布 $\text{Cauchy}(\gamma, \delta)$ は $S(1, 0, \gamma, \delta)$ と, そしてレビ分布 $\text{Levy}(\gamma, \delta)$ は $S(0.5, 1, \gamma, \gamma + \delta)$ と対応する [11, Prop.13~17]. なお, レビ分布の確率密度関数は次の式で定義される.

$$\sqrt{\frac{\gamma}{2\pi}} \frac{1}{(x - \delta)^{\frac{3}{2}}} \exp \left\{ -\frac{\gamma}{2(x - \delta)} \right\} \quad (\delta < x)$$

ここで, $0 < \alpha < 2, -1 < \beta \leq 1$ のときを考え, 安定分布 $S(\alpha, \beta, \gamma, 0)$ の確率密度関数を $f(x)$ とすると, $x \rightarrow \infty$ のとき次のことが成り立つ [12, p.24].

$$P(X > x) \sim \frac{\gamma^\alpha c_\alpha (1 + \beta)}{x^\alpha}, \quad f(x) \sim \alpha \frac{\gamma^\alpha c_\alpha (1 + \beta)}{x^{\alpha+1}} \quad \left(c_\alpha = \Gamma(\alpha) \sin \frac{\pi\alpha}{2} \right)$$

つまり, 平均や分散が存在しないような分布も含むかなり広い範疇の安定分布に対して, その相補累積分布関数や確率密度関数を, ベキ関数で漸近的に近似することができる. したがって, 安定分布のロングテール部分の解析においてベキ分布は重要な位置を占める.

5 なぜ「ベキ分布」なのか？

第 3 節で紹介したように, 「ベキ分布」は天文のようなマクロの世界から細胞内部のミクロの世界まで, 分野の異なるいろいろな場面で現れてくる. 人間の行動に関わる社会現象や経済現象にも現れており, この世界には何か共通の原理が働いていることが示唆される. このような「ベキ分布」の重要性が研究者の間ではっきりと認識されるようになってきたのは, 近年の僅かに半世紀以内のことである.

一体, 分野を問わず多くの事象が「ベキ分布」に従うのは何故なのだろうか? その理由はまだ未解明と思われるが, このことに関しては幾つかのキーワードがある. 「自己相似」「相転移」「臨界現象」「自己組織化」「相互作用」「成長と優先的選択」などが挙げられる.

これらに関して, 以下に幾つか簡単な解説を試みる.

5.1 相転移と臨界現象

相転移は, 鉄のような強磁性体に熱を加えていくと磁性が失われたり, 気体であるヘリウムの温度を下げると液体になったり, 寒天を溶かした液体状のもの (ゾル) が冷えると固体状のもの (ゲル) になるなど, 物質の状態が大きく変化する現象である. 相転移の臨界点の近くでは, 幾つかの物理量がベキ乗の関係 (ベキ乗則) を満たし, 臨界点に近づくと発散など大きな揺らぎを生じる. この現象は臨界現象と呼ばれる. 以下は [14][15][16] をもとにする.

強磁性体の場合でいうと, 臨界温度を T_c として $t = (T - T_c)/T_c$ とおくと, 外部磁場が $h = 0$ のとき, 比熱 C は $C \propto |t|^{-\alpha}$, 自発磁化 M_0 は $M_0 \propto (-t)^\beta$, 帯磁率 χ は $\chi \propto |t|^{-\gamma}$, 相関距離 ξ は $\xi \propto |t|^{-\nu}$ などのことが成り立つ. したがって, 臨界温度では比熱, 帯磁率,

そして相関距離は発散する．また，気体-液体の相転移では，臨界圧力 ρ_c での比熱 C_V は $C_V \propto |t|^{-\alpha}$ ，液相と気相の密度差は $\rho_L - \rho_G \propto (-t)^\beta$ ，等温圧縮率 χ_T は $\chi_T \propto |t|^{-\gamma}$ ，相関距離 ξ は $\xi \propto |t|^{-\nu}$ であり，これらの指数 $\alpha, \beta, \gamma, \nu$ は強磁性体の場合の値と一致する．さらに他の同種の相転移でも，物理量は異なっても同様の関係式が成り立ち同じベキ指数を持つ．特に，相関距離が発散することは，たとえば隣合う分子どおしの相互作用が遠く離れた箇所まで影響して系全体が大きく様変わりする場合があることを意味する．これらの指数は臨界指数と呼ばれ， $\alpha + 2\beta + \gamma = 2$ を含む幾つかの関係式が導かれ，物質の個々の性質にはよらず空間の次元や系の対称性に依存するとされる [16, p.1~11]．

そして，臨界点ではある量がベキ分布にしたがっている．たとえば，ゾル-ゲル相転移での臨界点では，固体状になったゲル部分のサイズ (s) が大小さまざまであり，そのサイズ分布 (実際には相補累積分布 $P(X > s)$) はベキ分布にしたがっている [14, p.58]．

以上で述べた相転移は，温度などの外部パラメーターを制御する必要があるが，何の制御をしなくてもベキ分布を示す状態になる場合がある．そのことを示す簡単なモデルとして「砂山崩し」と呼ばれるモデルがある．

平地の 1 点に砂粒を落とし続けると，一定の山の高さと形状が自然に形成される．その後も砂粒を落とし続けると，傾斜に耐え切れなくなった砂粒が崩落してなだれが起きる．なだれに規則性はなく，数粒崩れるときもあれば山が崩れるほど崩落するときもある．精密な装置で調べると実際には砂粒の大きさや受け皿の大きさに依存するが，なだれの規模と頻度はベキ分布に従っている [17, p.676]．特に何かの操作をしたわけでもなく，単に砂粒を落とし続けるだけで一定の形状になり，なだれが起きて崩れたとしても再び同じ形状に戻っていく．この現象は，自己組織化臨界現象と呼ばれている．

この現象をコンピューター上でシミュレートするモデルがある．発案者 3 名 (Bak, Tang, Wiesenfeld) の頭文字を取って BTW モデルと呼ばれ，1987 年に発表された．次のような手順でシミュレートされる [15, p.171]．

- (1) 正形状の平面を格子で区切り，ランダムに格子を選んで格子と同じ大きさのブロックを落とす操作を繰り返す．
- (2) ブロックが 4 個重なったら，それを前後左右の格子に 1 個ずつ振り分ける．この操作を 1 回の「なだれ」とする．振り分けられた箇所が 4 個目の場合は，そこでもなだれが起きる．1 回のなだれで引き起こされるなだれの回数をなだれの規模とする．ただし，なだれで正方形の外側にはみ出るブロックは無視する．
- (3) なだれが収まったら，つまりすべての格子のブロック数が 3 個以下になったら，再度格子をランダムに選んでブロックを落とす．

単純なゲームであるが，ブロックを落とし続けてなだれの規模と頻度を調べると，それはベキ分布の形状になっていく．さらに， k 個が積み上がっている格子の数の割合を $p(k)$ とすると，ブロックを落とす回数がある程度以上大きくなると， $p(3) : p(2) : p(1) : p(0)$ の値はほぼ一定な値を保ち続ける．たとえば，このシミュレーションを行うプログラムを利用して，なだれが千回になったときの値をみると

$$p(3) : p(2) : p(1) : p(0) = 0.432 : 0.307 : 0.188 : 0.073$$

という数値であった。ブロックを落とし続けていくとベキ分布に従うなだれを起こしながらも、積み上がっている個数による格子の割合はほぼ一定の比率を保ち続けていくのである。

5.2 パーコレーション

砂山くずしも BTW モデルも外部からの入力になされ続けるが、そのような外部入力がないくてもベキ分布を生じる場合がある。「パーコレーション」と呼ばれるモデルで、伝染病の感染や森林火災など、様々な伝搬現象や繋がりの様子を表すモデルである。

今、BTW モデルと同様に正方格子があったとし、乱数を利用して格子点に確率 p で黒石を、確率 $1 - p$ で白石を置く。黒石が隣り合うときは病気が感染する、あるいは山火事が燃え移ると考える。そして、黒石が連続的に繋がるのはどのような場合であるかを考える。

明らかに、 $p = 0$ のときは全ての格子点に白石が配置され、 $p = 1$ のときは全ての点に黒石が配置される。下図は、 10×10 の正方格子で $p = 0.4, 0.5, 0.6$ の場合である。

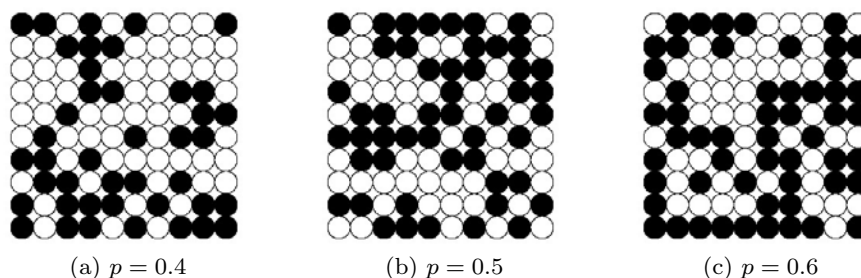


図 8: 10×10 の正方格子でのパーコレーション

黒石が連続的に繋がっている部分をクラスターというが、図 8(c) では黒石が途中で途切れないで上から下まで繋がっているクラスターがある。一般には無限格子を考えて、無限に繋がるクラスターが現れるのは p がどのような場合であるかを考えることになる。つまり、クラスターに含まれる黒石の個数を s として

$$\theta(p) = P(s = \infty)$$

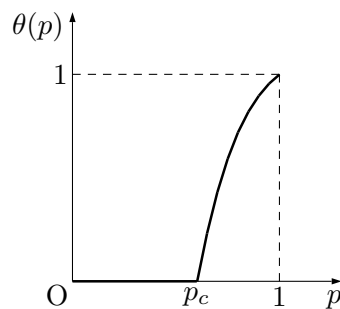


図 9: 臨界確率

を考えることになるが、 $\theta(p)$ を解析的に求めることは難しい問題である。有限格子の場合は数値的な解析から図 9 の

ようなグラフとなることが分かっている。つまり、 $p \leq p_c$ のときは巨大クラスターは現れず、 $p > p_c$ では巨大クラスターの存在確率は正の値を取る。突然繋がるようになるのでパーコレーション相転移と呼ばれ、 p_c を臨界確率という。そして、 $p = p_c$ でのクラスターの個数分布はベキ分布に従っている [18, p.184(6)]。

ここでは 2 次元正方格子で説明したが、三角格子やネットワーク、さらにはフラクタル図形の場合、あるいは次元を上げていった場合など、数学的にみても多くの問題が残されている。2006 年と 2010 年のフィールズ賞は、パーコレーションの研究に与えられている。

5.3 ネットワークの次数分布

ネットワークの点から出ている辺の数 (次数) の分布もベキ分布になることが多い。どのような仕組みでベキ分布になるのか、Barabási-Albert モデル (BA モデル) と呼ばれるモデルがあり (1999 年), 次のような手順でネットワークが構成される [13, p.8].

1. 最初に m 個の点を用意して全ての 2 点間を辺で結ぶ. m は数個程度とする.
2. 次に, 新たな点を追加して, 以下のような手順で既存の点と辺で結ぶ.
 - (a) 既存の点を $i = 1, 2, \dots, N$ と番号づけて, 点 i から出ている辺の本数 (次数) を k_i とする. このとき, N 個の点から m 個の点を次の確率 p_i で選択して, 新たな点と辺で結ぶ.

$$p_i = \frac{k_i}{\sum_{j=1}^N k_j}$$

したがって, 最初に追加する点は m 個の全ての点と辺で結ばれることになる.

- (b) 新しく追加した点を既存の点に含め, N を 1 だけ増加させて (a) と同様のことを繰り返す.

このときのポイントは, 点が常に増え続けていくこと, そして辺の結ばれ方がランダムではなく, 次数の多い点ほど新たな点と結ばれやすいということである. この「成長」と「優先的選択」がキーワードになる. 以上を繰り返して点を増やしていくと, 次数がベキ分布にしたがうことが次のようにして分かる.

まず, 繰り返す回数を時間 t にとり, 毎秒 1 個ずつ点が増えていくものとする. t 秒後の点の数を $N(t)$ とすると $N(t) = m + t$ である. 最初は m 個の点全てが辺で結ばれているので, 辺の本数の初期値は $mC_2 = m(m-1)/2$ 本であり, 1 秒ごとに m 本ずつ辺が増えるので, t 秒後の辺の本数 $E(t)$ は

$$E(t) = \frac{m(m-1)}{2} + mt$$

である. また, t 秒後の点 $i = 1, 2, \dots, N(t)$ の次数の合計は辺の本数 $E(t)$ の 2 倍である. m は小さい数であることを考えると, 十分に時間が経った状態では

$$\sum_{j=1}^{N(t)} k_j(t) = 2 \left\{ \frac{m(m-1)}{2} + mt \right\} \simeq 2mt$$

とみることができるので, 手順 2(a) において t 秒後に m 個の点を選択する確率は

$$p_i(t) \simeq \frac{k_i(t)}{2mt}$$

により近似できる. したがって, 点 i が時刻 t で新たに追加される点と辺で結ばれる期待値は $mp_i(t)$ とみれる. t や $k_i(t)$ を連続変数として考えると, この値は次数 k_i の t に関する変化率とみることができるので,

$$\frac{dk_i}{dt} = mp_i(t) \simeq \frac{k_i(t)}{2t}$$

となる。これを解いて

$$k_i(t) = C\sqrt{t} \quad (C \text{ は任意定数})$$

が得られる。ただし、 i 番目の点が追加される時間を t_i とすると、新たに追加される点は最初は m 本の辺を持つので、初期条件として $k_i(t_i) = m$ を満たす。この条件で考えると $C = m/\sqrt{t_i}$ となるので、解 $k_i(t)$ は

$$k_i(t) = m\sqrt{\frac{t}{t_i}}$$

と表される。これが時刻 t における点 i の次数である。したがって、次数 $k_i(t)$ は $\sqrt{t}$ に比例して増加する。また、 t_i は単調に増えるので、後から追加される頂点ほど次数が小さい。また、時刻 t における点 i の次数を k としてこの式を書き換えると

$$t_i = t \left(\frac{m}{k} \right)^2$$

となり、時刻 t の時点で次数 k を持つ点がいづつ追加されたのかが分かる。

以上をもとにして次数の累積分布を求める。次数が k よりも小さい点の数は、時刻 $t_i = t(m/k)^2$ 以降から現時点 t までに追加された点であるので、

$$N(t) - N(t_i) = (m + t) - (m + t_i) = t - t_i = t \left(1 - \frac{m^2}{k^2} \right)$$

と表される。つまり、次数 k の確率密度関数を $\varphi(k)$ とすると、上式は点の数について

$$N(t) \cdot P(k_i < k) = N(t) \int_m^k \varphi(s) ds = t \left(1 - \frac{m^2}{k^2} \right)$$

を表しており、十分に大きな t については $N(t) \simeq t$ とみると、

$$\int_m^k \varphi(s) ds = 1 - \frac{m^2}{k^2}$$

が成り立つことを示している。そこで、この式を k で微分することにより、次数 k の確率密度関数として次の式が得られ、次数がべき分布にしていることが分かる。

$$\varphi(k) = \frac{2m^2}{k^3} \propto k^{-3}$$

かなり粗い議論になるが、このようにしてべき指数 3 が導かれる。点を選択する確率を

$$p_i(t) = \frac{k_i + a}{\sum_{j=1}^N (k_j + a)}$$

として考えると、

$$\varphi(k) \propto k^{-(3+a/m)}$$

が導かれる。実際のスケールフリー・ネットワークではべき指数は 2 ~ 3 の場合が多いようなので、それは $-m \leq a \leq 0$ の場合を考えることになる。

6 「ベキ分布」の数理

この「ベキ分布」を基礎においた数学的な理論付けがなされつつある．須鎗 [19] をもとにすると，その議論は次のように進められる．

まず，自然現象の表現では指数関数が頻出する．その関数は線形微分方程式 $y' = y$ の解として得られる．そこで，この微分方程式を拡張して，非線形微分方程式

$$\frac{dy}{dx} = y^q \quad (q > 0) \quad \dots\dots\dots \textcircled{1}$$

を考える．この微分方程式は $q = 1$ の場合を含んでいる．次に， $1/y^q$ の積分を

$$\int \frac{1}{y^q} dy = \begin{cases} \ln y + C_1 & (q = 1 \text{ のとき}) \\ \frac{y^{1-q}-1}{1-q} + C_q & (q \neq 1 \text{ のとき}) \end{cases}$$

の形で表すと，

$$\lim_{q \rightarrow 1} \frac{y^{1-q} - 1}{1 - q} = \ln y$$

が成り立つ．そこで，関数 $\ln_q y$ を

$$\ln_q y = \frac{y^{1-q} - 1}{1 - q}$$

と定めると，その定め方から

$$\lim_{q \rightarrow 1} \ln_q y = \ln y, \quad \frac{d}{dy} \ln_q y = \frac{1}{y^q}$$

が成り立ち， $\textcircled{1}$ の一般解は

$$\ln_q y = x + C \quad (C \text{ は任意定数}) \quad \dots\dots\dots \textcircled{2}$$

と表される．この関数 $\ln_q y$ を q -対数関数という．ただし，定義域は $\mathbb{R}^+$ で考えるものとする．値域は q の値により異なる．

次に， q -対数関数 $\ln_q x$ の逆関数を $\exp_q x$ で表して q -指数関数という．具体的には次の式で表される．また，定義域と値域は q -対数関数の場合と逆になる．

$$\exp_q x = \{1 + (1 - q)x\}^{\frac{1}{1-q}}$$

q -指数関数 $\exp_q x$ は

$$\lim_{q \rightarrow 1} \exp_q x = \exp x$$

を満たすことから指数関数 $\exp x$ の拡張とみることができ， $\textcircled{1}$ の一般解として

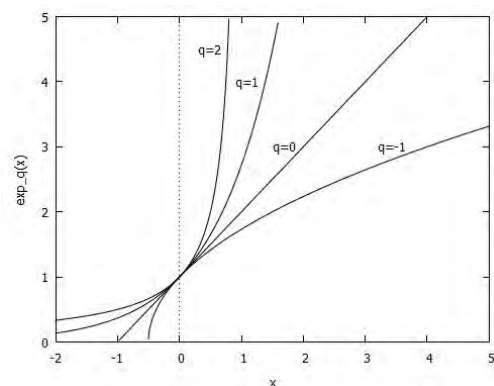


図 10: q -指数関数 $\exp_q x$ のグラフ

$$y = \exp_q(x + C) \quad \dots\dots\dots \textcircled{3}$$

が得られる。ただし、 $y > 0$ であることから、

$$1 + (1 - q)(x + C) > 0$$

を満たすものとする。

以上のもとで、 q -指数関数の定義より解③は

$$\begin{aligned} y &= \{1 + (1 - q)(x + C)\}^{\frac{1}{1-q}} \\ &= \{1 + (1 - q)C\}^{\frac{1}{1-q}} \left\{1 + (1 - q)\frac{x}{1 + (1 - q)C}\right\}^{\frac{1}{1-q}} \\ &= A \exp_q \frac{x}{A^{1-q}} \quad (A = \exp_q C > 0) \end{aligned}$$

と表すことができるので、

$$\frac{y}{A} = \exp_q \frac{x}{A^{1-q}}$$

と変形することができる。ここで、 $\tilde{x} = x/A^{1-q}$ 、 $\tilde{y} = y/A$ において①の関係を利用すると

$$\frac{d\tilde{y}}{d\tilde{x}} = \tilde{y}^q$$

となる。つまり、解に対してこのようなスケール変換をしても同じ非線形微分方程式②を満たす。ただし、これは $q \neq 1$ のときであり、 $q = 1$ のときにスケール不変性は成り立たない。

この q -指数関数 $\exp_q x$ は、 $(1 - q)x \gg 1$ の場合は

$$\exp_q x = \{1 + (1 - q)x\}^{\frac{1}{1-q}} \simeq \{(1 - q)x\}^{\frac{1}{1-q}} \propto x^{\frac{1}{1-q}}$$

となり、ベキ関数で漸近近似できる。確率密度関数が $\exp_q(-\beta x)$ ($\beta > 0, q > 1$) の形で現れる場合は

$$\exp_q(-\beta x) = \{1 + (q - 1)\beta x\}^{\frac{1}{1-q}} \simeq \{(q - 1)\beta x\}^{\frac{1}{1-q}} \propto x^{\frac{1}{1-q}}$$

となる。ベキ分布 $x^{-\gamma}$ との関係では $\gamma = 1/(q - 1)$ である。実際の例では γ は $2 \sim 3$ の場合が多いようなので、対応する q は $4/3 \sim 3/2$ の範囲になる。 $\exp_q(-\beta x^2)$ の形のときは、 $\gamma = 2/(q - 1)$ の関係となる。

この q -指数関数や q -対数関数に対して、通常の指数法則や対数関数の性質が成り立つような積を定義することができる。その積は q -積と呼ばれ、 $x^{1-q} + y^{1-q} - 1 > 0$ を満たす任意の $x, y > 0$ に対して、

$$x \otimes_q y = (x^{1-q} + y^{1-q} - 1)^{\frac{1}{1-q}}$$

で定義される． $q \rightarrow 1$ のときは通常の積と一致して $\lim_{q \rightarrow 1} x \otimes_q y = xy$ が成り立つ．そして，容易に分かるように，この定義により次のことが成り立つ．

$$\begin{aligned}\exp_q x \otimes_q \exp_q y &= \exp_q(x + y) \\ \ln_q(x \otimes_q y) &= \ln_q x + \ln_q y\end{aligned}$$

さらには，比を考えることもできる． $x^{1-q} - y^{1-q} + 1 > 0$ を満たす $x, y > 0$ に対して

$$x \oslash_q y = (x^{1-q} - y^{1-q} + 1)^{\frac{1}{1-q}}$$

と定め q -比という．これにより， q -積の逆演算として次のことが成り立つ．

$$\begin{aligned}\exp_q x \oslash_q \exp_q y &= \exp_q(x - y) \\ \ln_q(x \oslash_q y) &= \ln_q x - \ln_q y\end{aligned}$$

これらの性質を利用すると，指数関数や対数関数に対して成立するいろいろな性質をことごとく q -積や q -比を利用した性質に書きかえることが可能になり，階乗・多項係数・スターリングの公式なども q -積を用いて定式化される．さらには，ガウスの誤差法則やエントロピー，そして情報理論との繋がりも出てくる．いずれにおいても，非線形微分方程式から出発しているので，解の重ね合わせの原理は利用できない．

これらの議論は，もともとは指数関数が中心となる Boltzmann-Gibbs 統計力学を，ベキ関数が頻出するマルチフラクタル系への拡張を考える中で見いだされてきたもので，Tsallis 統計力学と呼ばれている [19]．Tsallis 統計は裾の長い分布をする現象に対して当てはまりがよいとされ，物理系に留まらず情報数理と関連する多方面の分野と繋がりがながら，現在も発展途上にある．

7 おわりに

「ベキ分布」の特徴や具体例，そして数学的取り扱いの一端について紹介してきた．この確率分布が統計教育で取り上げられることは少ないが，分野を問わずにいろいろな場面で現れてくることを考えると，何らかの形で統計教育でも触れる必要があると思われる．

少なくとも，「ベキ分布」という平均や分散が存在しない確率分布があり，この確率分布に従う現象がマクロからミクロまでの様々な分野で山のように存在することは，学生に周知しておく必要があるのではないだろうか．特に，平均や分散が存在しないときに標本平均が激しく乱高下する様子を見せて，一見すると規則性のないような変動でも統計解析により特徴を抽出できる場合があることは伝えておきたい．また，ネットワーク社会で生きていく上では，いろいろな状況をネットワークとして捉えることができる場合があることも伝えておくべきであろう．ベキ分布は，このようなネットワークの中にも含まれている．そして，ネットワークを解析する数学理論としてグラフ理論があることも紹介しておきたい．

「ベキ分布」の重要性が認識されるようになってきたのは，僅かにここ半世紀程度でのことである．これまでの統計教育で取り上げられてこなかったのはやむを得ない部分もあるが，今後は教授すべき項目の一つとして検討する必要があるのではないかとと思われる．

参考文献

- [1] マーク・ブキャナン：歴史は『べき乗則』で動く，早川書房，2009
- [2] 寺居剛・伊藤洋一：高軌道傾斜角メインベルト小惑星を対象とする広域サーベイ観測
[URL] http://gopira.jp/sym2010/proceedings/2/gopira2010-2_terai.pdf
- [3] マグニチュードと発生頻度の関係， [URL] http://www5d.biglobe.ne.jp/~kabataf/yougo/E_jisin/jisin4_magnitude_1.htm
- [4] 伊藤岳：マクロレベルにおける戦争の動態，
[URL] http://yamakage-ken.com/citrus/mas_ir/ito_paper.pdf
- [5] 釜国男：人口分布の数量分析，季刊創価経済論集，Vol.XLVI, No.1~4, pp.33-44, 2017
- [6] 岩田暁一：経済現象における分布，行動計量学，3 巻, 1 号, 1975
[URL] https://www.jstage.jst.go.jp/article/jbhmk1974/3/1/3_1_35/_pdf
- [7] 齋藤健：企業所得成長モデルと投資戦略
[URL] <http://www.f.waseda.jp/watada/News2006/SL20061129.pdf>
- [8] 杉山浩平・大崎博之・今瀬眞：論文の引用・共著関係から何が分かるか？，
[URL] http://www.lsnl.jp/~ohsaki/papers/Sugiyama06_IN5.pdf
- [9] 山田沢司・五斗進：生体分子ネットワークのトポロジー解析，物性研究，Vol.87(3), 2006
[URL] <https://repository.kulib.kyoto-u.ac.jp/dspace/bitstream/2433/110693/1/KJ00004454661.pdf>
- [10] M. E. J. Newman：Power laws, Pareto distribution and Zipf's law, Contemporary Physics 46, pp.323-351, 2005
[URL] <https://arxiv.org/pdf/cond-mat/0412004.pdf>
- [11] Stable Distributions
[URL] <http://www.randomservices.org/random/special/Stable.html>
- [12] John P. Nolan：Stable Distributions – Models for Heavy Tailed Data
[URL] <http://fs2.american.edu/jpnolan/www/stable/chap1.pdf>
- [13] 波多野直道：複雑ネットワーク～統計物理学の視点，物性研究，Vol.3, No.1, 2013
[URL] https://repository.kulib.kyoto-u.ac.jp/dspace/bitstream/2433/182061/1/bussei_el_031210.pdf
- [14] 高安秀樹・高安美佐子：経済・情報・生命の臨界ゆらぎ，ダイヤモンド社，2000
- [15] 香取眞理：複雑系を解く確率モデル，講談社，1997
- [16] 野村清英：相転移の統計力学
[URL] <http://maya.phys.kyushu-u.ac.jp/~knomura/education/Undergraduate/Phase-Transition/Phase-Transition.pdf>
- [17] 吉岡直人：砂山崩し実験から地震予測法を探る試み，材料，Vol.61, No.8, 2012
[URL] https://www.jstage.jst.go.jp/article/jsms/61/8/61_675/_pdf
- [18] 能川友昭・根本幸児・長谷川雄央：ネットワーク構造がもたらすロバストな臨界性
[URL] https://www.jstage.jst.go.jp/article/jnns/21/4/21_182/_pdf/-char/ja
- [19] 須鎗弘樹：複雑系のための基礎数理，牧野書店，2010 (概要は下記参照)
[URL] http://www.ne.jp/asahi/hiroki/suyari/suyari_jps08.pdf