

数理モデルに関する科研費研究会（東京都立産技高専）

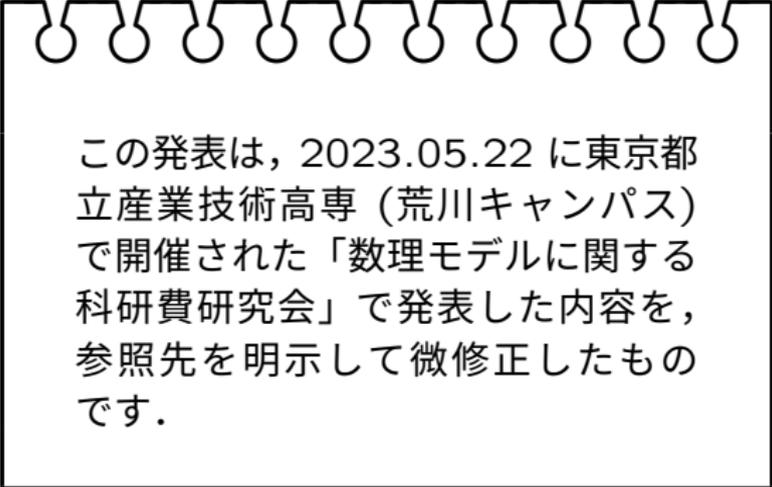
べき分布のデータ

梅野 善雄

元一関工業高等専門学校

2023年6月17日

- 1 べき分布
- 2 べき分布に従う現象例
- 3 ジップの法則とべき分布
- 4 べき分布に従う確率変数の最大値



この発表は、2023.05.22 に東京都立産業技術高専（荒川キャンパス）で開催された「数理モデルに関する科研費研究会」で発表した内容を、参照先を明示して微修正したものです。

べき分布の定義

- べき分布の確率密度関数

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

パレート分布と呼ばれ、 $\text{Pareto}(a, b)$ などで表される。

- 両対数グラフでは、右下がりの直線になる。

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

- 累積分布関数は

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

- 相補累積分布関数はべき関数になる。

$$P(X > x) = \frac{b^a}{x^a}$$

べき分布の平均と分散

■ べき分布の平均

$$\begin{aligned} E(X) &= \int_b^{\infty} x \cdot \frac{ab^a}{x^{a+1}} dx = \int_b^{\infty} \frac{ab^a}{x^a} dx \\ &= \begin{cases} \infty & (0 < a \leq 1) \\ \frac{ab}{a-1} & (a > 1) \end{cases} \end{aligned}$$

■ べき分布の分散

$$V(X) = \frac{ab^2}{(a-1)^2(a-2)} \quad (a > 2)$$

■ べき分布の n 次モーメント

$$E(X^n) = \begin{cases} \infty & (0 < a \leq n) \\ \frac{ab^n}{a-n} & (a > n) \end{cases}$$

べき分布 Pareto(a, b) の性質

- 確率密度関数に自己相似性がある。これは、べき関数に限る。

$$f(cx) = \frac{ab^a}{(cx)^{a+1}} = \frac{1}{c^{a+1}} f(x) \propto f(x)$$

- cX ($c > 0$) は Pareto(a, bc) にしたがう。

$$\begin{aligned} P(cX \leq x) &= P\left(X \leq \frac{x}{c}\right) \\ &= 1 - \frac{b^a}{(x/c)^a} = 1 - \frac{(bc)^a}{x^a} \end{aligned}$$

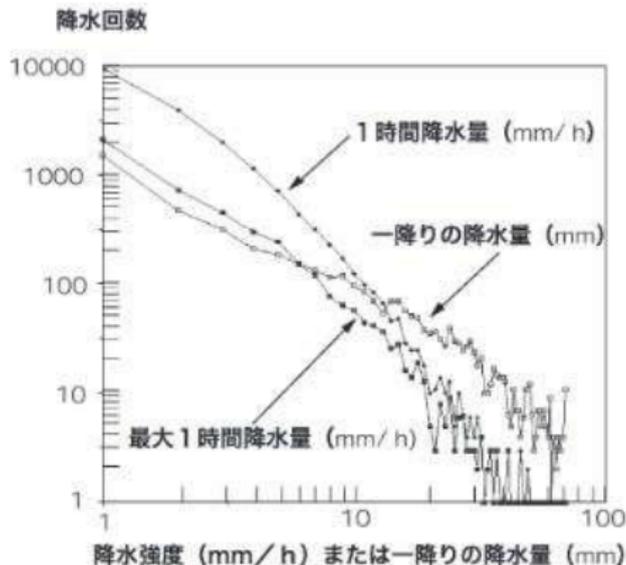
- X^n ($n > 0$) は Pareto($\frac{a}{n}, b^n$) に従う。

$$\begin{aligned} P(X^n \leq x) &= P\left(X \leq x^{\frac{1}{n}}\right) \\ &= 1 - \frac{b^a}{\left(x^{\frac{1}{n}}\right)^a} = 1 - \frac{(b^n)^{\frac{a}{n}}}{x^{\frac{a}{n}}} \end{aligned}$$

- べき分布は、定数倍してもべき乗してもべき分布に従う。

べき分布に従う現象例 (1)

■ 東京における降水量と降水回数 (1976～2009 年)

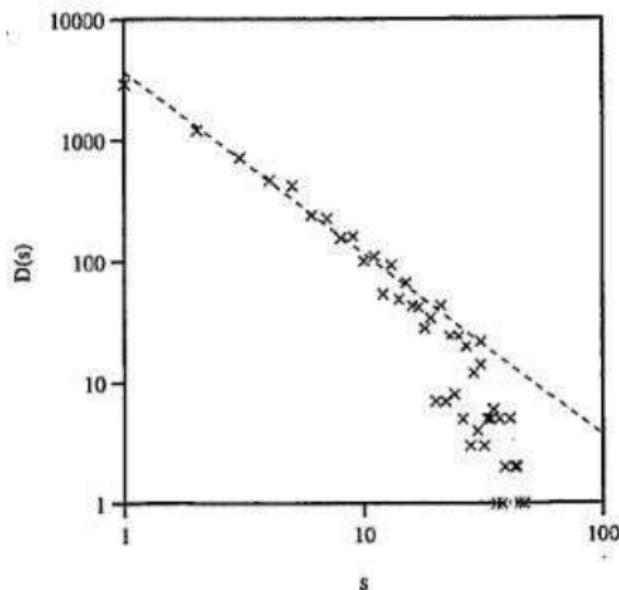


- アメダスデータでの 0 mm には 1 mm 未満の降水も含まれるので、0 mm の前後で 1 mm 以上の降水があるときは「降水あり」と判断し、降水が連続する部分を「一降りの雨」としてデータを修正。

[Ref] https://www.metsoc.jp/tenki/pdf/2018/2018_03_0017.pdf

べき分布に従う現象例 (2)

■ 熱帯マグロの漁獲高と頻度 (1976–1982)：ロペス岬付近

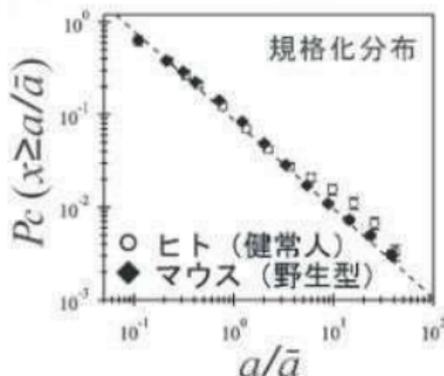
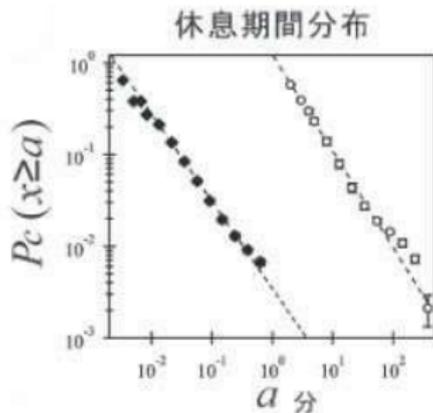


- アフリカ大陸中部大西洋側にあるガボン共和国ロペス岬付近での、周囲 2 km の網を用いた商業漁業による、熱帯マグロの 7 年間の漁獲高のデータ。横軸は 1 回あたりの漁獲高 (トン)、縦軸はその頻度。

[Ref] <https://arxiv.org/pdf/1310.5527>

べき分布に従う現象例 (3)

- ヒトの身体活動を腕時計型加速度計で計測し、1分でゼロレベルを交差する回数で活動量を定義する。
- 野生マウスの活動をケージ下の圧電シートセンサーで計測し、0.1秒内の局所分散値で活動量を定義する。
- 活動データの平均を連続して下回る時間を「休息时间」として、その相補累積分布を $P_c(X \geq x)$ を求める。



- 平均で割って規格化すると、分布がほぼ一致する。

[Ref] https://www.jstage.jst.go.jp/article/sicejl/49/12/49_844/

ジップの法則

- ジップの法則は、あらゆる箇所で見られる経験則である。たとえば、単語の出現頻度、都市の人口、個体が割れたときの破片の大きさ、...
- ある事象を、頻度 f の多い順 R に並べた度数分布表

R	1	2	...	N	All
f	f_1	f_2	...	f_N	F

- ジップの法則は、 C を正定数として

$$f \times R \approx C \quad \therefore f \approx \frac{C}{R}$$

が成り立ち、頻度が順位と反比例の関係にある。

- 別な見方をすると、頻度が f_1 以上のものは 1 つしかなく、 f_2 以上のものは 2 つ、 f_3 以上のものは 3 つある。確率でいうと、頻度がある値以上である確率、つまり、相補累積分布が順位と比例しているとみることができる。
- 一般には、 $\alpha > 0$ として、 $f = \frac{C}{R^\alpha}$ もジップの法則という。

ジップの法則の別な見方

- 合計 F の中から 1 個を選ぶとき、それが k 位である f_k の事象である確率を p_k とすると、 $p_k = \frac{f_k}{F}$.
- 一般のジップの法則に従っていれば、 $f_k = \frac{f_1}{k^\alpha}$ なので、

$$F = \sum_{k=1}^N f_k = f_1 \sum_{k=1}^N \frac{1}{k^\alpha}$$

- したがって、 $p_k = \frac{\frac{1}{k^\alpha}}{\sum_{k=1}^N \frac{1}{k^\alpha}} = \frac{1}{k^\alpha \cdot \sum_{k=1}^N \frac{1}{k^\alpha}}$
- 順位が無限に続くときは、分母にゼータ関数が含まれることになる。
- $P(X = k) = \frac{1}{k^\alpha \zeta(\alpha)}$ ($\alpha > 1$) で表される離散分布はゼータ分布と呼ばれる。

ジップの法則とべき分布

- ジップの法則 $f = \frac{C}{R^\alpha}$ を R について解くと, $C^{\frac{1}{\alpha}} f^{-\frac{1}{\alpha}} = R$
- 最終順位 N で割り $r = \frac{R}{N}$, $A = \frac{C^{\frac{1}{\alpha}}}{N}$ とおくと, $Af^{-\frac{1}{\alpha}} = r$
- r ($0 < r \leq 1$) は順位に関する比率で, r の値が小さいほど頻度 f は大きいので, この式は相補累積分布を表すとみることができる.

$$P(X > f) = Af^{-\frac{1}{\alpha}}$$

- したがって, 累積分布 $P(X \leq f)$ は

$$1 - P(X > f) = 1 - Af^{-\frac{1}{\alpha}}$$

- f で微分すると確率密度関数が得られ, べき分布の形になる.

$$\frac{d}{df} \left(1 - Af^{-\frac{1}{\alpha}} \right) = \frac{A}{\alpha} f^{-\frac{1}{\alpha}-1} = \frac{A/\alpha}{f^{\frac{1}{\alpha}+1}}$$

- ジップの法則の指数 α と Pareto(a, b) の指数 a とは, $a = \frac{1}{\alpha}$ の関係にある.

べき分布に従う確率変数の最大値

- Pareto(a, b) に従う n 個の確率変数の最大値を X_{\max} とする.
- X_{\max} が小区間 $(x, x + \Delta x]$ に含まれる確率は、確率密度関数を $f(x)$ とすると $f(x) \Delta x$ である.
- 他の $n - 1$ 個は $P(X \leq x)$ を満たすので、求める確率は

$$\{P(X \leq x)\}^{n-1} \cdot f(x) \Delta x$$

- n 個の中の 1 個の選び方は n 通りあるので、

$$n \cdot \{P(X \leq x)\}^{n-1} \cdot f(x) \Delta x$$

を考えることになる.

- そこで、関数 $\pi(x)$ を

$$\pi(x) = n \cdot \frac{ab^a}{x^{a+1}} \cdot \left(1 - \frac{b^a}{x^a}\right)^{n-1}$$

とおくと、この関数は確率密度関数になる.

$\pi(x)$ は確率密度関数

- $\pi(x) = n \cdot \frac{ab^a}{x^{a+1}} \cdot \left(1 - \frac{b^a}{x^a}\right)^{n-1}$
- $\frac{b^a}{x^a} = t$ とおくと, $-\frac{ab^a}{x^{a+1}} dx = dt$
- x が b から ∞ まで変化すると, t は 1 から 0 まで変化する.
- したがって, $[b, \infty)$ における積分は

$$\begin{aligned}\int_b^{\infty} \pi(x) dx &= \int_b^{\infty} n \left(1 - \frac{b^a}{x^a}\right)^{n-1} \frac{ab^a}{x^{a+1}} dx \\ &= n \int_0^1 (1-t)^{n-1} dt \\ &= -[(1-t)^n]_0^1 = 1\end{aligned}$$

- したがって, $\pi(x)$ は確率密度関数である.

最大値の平均

- 最大値 X_{\max} の平均は

$$\begin{aligned} E(X_{\max}) &= \int_b^{\infty} x \pi(x) dx \\ &= n \int_b^{\infty} \frac{ab^a}{x^a} \left(1 - \frac{b^a}{x^a}\right)^{n-1} dx \end{aligned}$$

- $1 - \frac{b^a}{x^a} = t$ とおくと, $x = \frac{b}{(1-t)^{\frac{1}{a}}}$ となり, x が b から ∞ まで変化すると, t は 0 から 1 まで変化する.

- 微分は $a \cdot \frac{b^a}{x^a} \cdot \frac{1}{x} dx = dt$ より, $\frac{ab^a}{x^a} dx = \frac{b}{(1-t)^{\frac{1}{a}}} dt$

- したがって,

$$E(X_{\max}) = nb \int_0^1 \frac{t^{n-1}}{(1-t)^{\frac{1}{a}}} dt = nb \int_0^1 t^{n-1} (1-t)^{-\frac{1}{a}} dt$$

最大値の平均とベータ関数

- $B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$ は、ベータ関数と呼ばれる。
- $y > 0$, $y - 1 = -\frac{1}{a}$ となる y は $y = 1 - \frac{1}{a}$, $a > 1$.
- このとき、 X_{\max} の平均は、ベータ関数を用いて

$$E(X_{\max}) = nb \cdot B\left(n, 1 - \frac{1}{a}\right)$$

と表され、 $E(X_{\max}) \sim n^{\frac{1}{a}}$ (漸近近似) が成り立つ。

- 証明は、ベータ関数とガンマ関数との関係式を用いる。

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

- x が十分に大きいとき、スターリングの公式により

$$\Gamma(x) \approx \sqrt{\frac{2\pi}{x}} \left(\frac{x}{e}\right)^x = \sqrt{2\pi} x^{x-\frac{1}{2}} e^{-x}$$

最大値の平均とべき関数

- $\alpha = 1 - \frac{1}{a}$ においてスターリングの公式を利用すると,

$$\begin{aligned}\frac{\Gamma(n)}{\Gamma(n+\alpha)} &\approx \frac{\sqrt{2\pi}n^{n-\frac{1}{2}}e^{-n}}{\sqrt{2\pi}(n+\alpha)^{n+\alpha-\frac{1}{2}}e^{-(n+\alpha)}} \\ &= \left(\frac{n}{n+\alpha}\right)^{n-\frac{1}{2}} \frac{1}{(n+\alpha)^\alpha e^{-\alpha}} \\ &= \left(\frac{1}{1+\frac{\alpha}{n}}\right)^n \left(\frac{1}{1+\frac{\alpha}{n}}\right)^{-\frac{1}{2}} \frac{1}{n^\alpha \left(1+\frac{\alpha}{n}\right)^\alpha e^{-\alpha}}\end{aligned}$$

- α が定数で n が十分に大きいときは $1 + \frac{\alpha}{n} \approx 1$ とみれる。
- テイラー展開すると,

$$\frac{1}{1+\frac{\alpha}{n}} = 1 - \frac{\alpha}{n} + \left(\frac{\alpha}{n}\right)^2 - \dots \approx 1 - \frac{\alpha}{n}$$

最大値の平均とべき関数 2

$$\blacksquare \lim_{n \rightarrow \infty} \left(1 - \frac{\alpha}{n}\right)^n = \lim_{n \rightarrow \infty} \left(1 - \frac{1}{\frac{n}{\alpha}}\right)^{-\frac{n}{\alpha} \cdot (-\alpha)} = e^{-\alpha}$$

であるから、 n が十分大であれば

$$\left(\frac{1}{1 + \frac{\alpha}{n}}\right)^n \approx \left(1 - \frac{\alpha}{n}\right)^n \approx e^{-\alpha}$$

■ 以上を利用すると、

$$\frac{\Gamma(n)}{\Gamma(n + \alpha)} \approx e^{-\alpha} \cdot 1 \cdot \frac{1}{n^\alpha \cdot 1 \cdot e^{-\alpha}} = n^{-\alpha}$$

■ $\alpha = 1 - \frac{1}{a}$ ($a > 1$) であるので、

$$E(X_{\max}) \approx nb \cdot n^{-1 + \frac{1}{a}} \cdot \Gamma\left(1 - \frac{1}{a}\right) = bn^{\frac{1}{a}} \cdot \Gamma\left(1 - \frac{1}{a}\right)$$

したがって、漸近近似では、 $E(X_{\max}) \sim n^{\frac{1}{a}}$ である。

Pareto(2,1) の最大値の平均

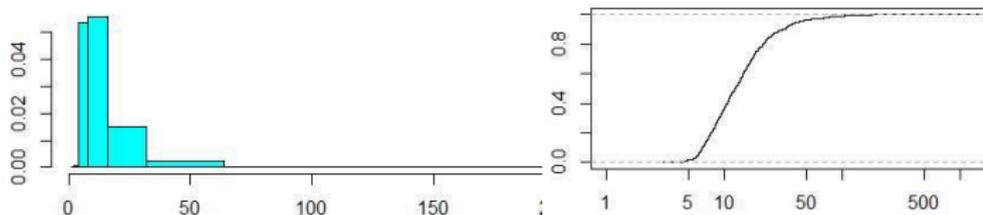
- Pareto(2,1) では, $E(X_{\max}) = n \cdot B\left(n, \frac{1}{2}\right) \approx \sqrt{n\pi}$

n	10	100	1000	10000
$nB(n, 1/2)$	5.675	17.74	56.05	177.24
$\sqrt{n\pi}$	5.604	17.72	56.04	177.24

- 具体的に乱数を発生させ, 最大値 1000 個で確認する.

n	10	100	1000	10000
Mean	5.962	18.477	59.79	169.50
SD	8.220	42.80	134.8	264.1
Min	1.362	3.026	11.86	34.91
Max	123.81	1253.8	3357.5	6192.5

- $n = 100$ の場合の, 最大値のヒストグラムと累積分布



グラフ電卓絡みで開設した Web サイトであるが、いろいろなページで構成される。

- グラフ電卓：活用事例 (数学・工学)・操作マニュアル等
- 数学関係：TeX, emath, MePoTeX, gnuplot, Maxima, R
- リンク集：数学学習・学習総合・べき分布・悩みごと等

下記は、「べき分布」に関するリンク集のメニューである。

■ 「べき分布」：関連リンク集 [\[Map\]](#)

[はじめに\(概説\)](#)

[「べき分布」の解説](#)

[「べき分布」の具体例](#)

[「べき分布」と自然現象](#)

[「べき分布」と生物現象](#)

[「べき分布」と社会現象](#)

[「べき分布」と経済現象](#)

[「べき分布」とネットワーク](#)

[「べき分布」のメカニズム](#)

[相転移・臨界現象](#)

[パーコレーション\(浸透現象\)](#)

[複雑系・フラクタル](#)

[自己組織化臨界現象](#)

[シミュレーション](#)

[参考文献](#)

[Ref] <http://www.cs.cornell.edu/courses/cs6241/2019sp/readings/Newman-2005-distributions.pdf>