

## 大数（たいすう）の法則

平均値がある一定に近づくこと

### コラム：べき乗分布

発生した地震の大きさを表すエネルギーなどの測定データの値を  $x$  で表し、 $x$  がどれくらいの頻度で一定の期間中に起こるかを示す分布関数を  $f(x)$  で表すと、べき乗分布は次のような、比較的単純な式で表される。

$$f(x) = Ax^{-\alpha} \quad (2.1)$$

## ジップ則

べき乗的な振る舞い

~~~~~

ジップの法則（ジップのほうそく、Zipf's law）あるいはジフの法則とは、出現頻度が  $k$  番目に大きい要素が全体に占める割合が  $1/k$  に比例するという経験則である。Zipf は「ジフ」と読まれることもある

~~~~~

地球表面が生きている証しに関しては、地球科学の主要な理論であるプレートテクトニクスが説明してくれる。

これは 1912 年のウェゲナーの**大陸移動説**から始まるといい。この大陸移動説は、はじめくっついていた大陸が何らかの原因によって分離し、互いに移動したとする仮説である。

ウェゲナーは、南アメリカ大陸とアフリカ大陸の形をおおむねそのままにして切り離し、地球表面で自由に動かしてみると、ジグソーパズルのピースのように一つの大きな大陸にまとまることを指摘した。**南アメリカ大陸の東岸とアフリカ大陸の西岸がぴったり合う**ことは、地球儀を見れば明らかであろう。

ウェゲナーはそれだけでなく、現在は離れている両大陸間の地質構造がよく似ていること、古生物化石の分布に連続性が見られること、海を渡ることができないはずの生物種が両大陸にまたがって分布することなどを指摘して、大陸移動説を唱えたのである。



図2-7 サンアンドレアス断層 (写真・フット)



サンアンドレアス断層の動きは比較的単純な横ずれである。それでも地震の予知ができなかったのは、単に偶然のせいであろうか。二つのプレートが相対的に横ずれで動いている場合には、その境界のあるところでは、引っ張られては外れ、また引っ張られては外れ、ということがほぼ周期的に繰り返され、そのたびに同じ程度の強さの地震が起こることも考えられる。

このとき、いったん地震が起こったとして、次にいつ起こるかは正確にはわからないという意味で、地震の発生は偶然である。

## $\sqrt{1000}$

ただ、地震のエネルギーをべき乗に表すときの底を歴史的に $\sqrt{1000} \cong 31.6$  (≒は右辺が左辺にだいたい等しいことを表す記号) とする約束のため、少々ややこしく、マグニチュードが1増えると、地震のエネルギーは約31.6倍に、2増えるとちょうど1000 (≒ $31.6^2$ ) 倍になることになる。したがって、最近ときどきニュースになるM4.0程度の茨城県沖地震とM9.0の東北地方太平洋沖地震とでは、マグニチュードが5ちがうので地震のエネルギーに約3000万 (≒ $31.6^5$ ) 倍もの差があることになる。

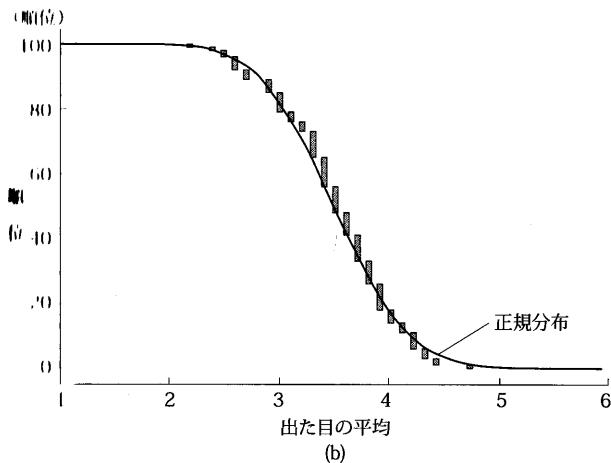
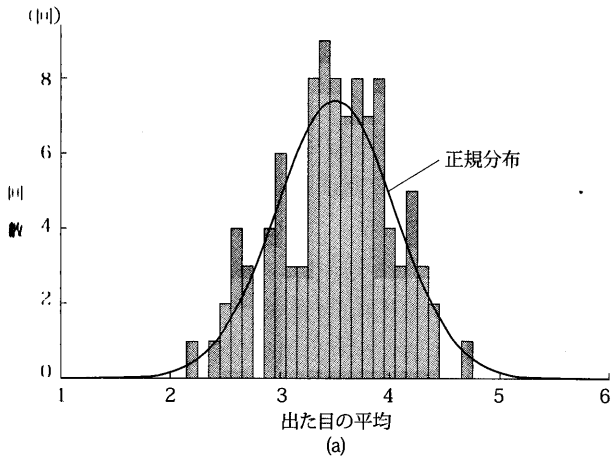


図3-3 サイコロ投げによる100個のデータの個数分布(a)  
 ランキングプロット(b)

実をいうと、ランキングプロット（累積個数分布）の場合にはデータの数を順々に加えていくというのがみそなのである。個数分布そのものはデータ数が少ないと、図3-3 (a) のように、棒の高さがデコボコになってしまうが、ランキングプロットではそれを順々に加えていくので、途中のデコとボコがお互いに相殺して滑らかになるのである。

もし統計性を生み出す原因が乗算過程であれば、生じる結果はいろいろな要因の掛け算で表されるが、対数をとった後で見ると統計性を生み出すばらつきの原因は足し算となり、加算過程とみなされる。したがって、対数をとった後に正規分布が得られることになり、対数正規分布になるというわけである。

こうして、複雑系の共通した特徴が歴史性にあり、歴史性が乗算過程的であるとすれば、複雑系ではその統計性を特徴づける最も自然な分布関数は対数正規分布だといえることができる。このようなわけで、対数正規分布

こそが複雑な系全体の統計性を見わたす際の基準としてふさわしいと考えられる。すなわち、複雑系の自然な分布は対数正規分布だということができる。

### 対数正規分布

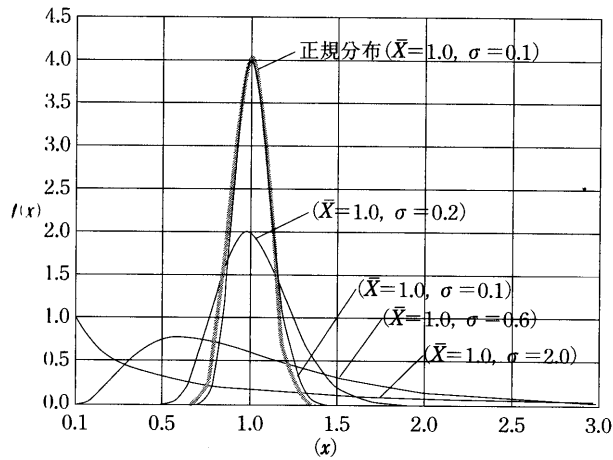


図4-2 対数正規分布

#### コラム：対数正規分布

測定データの値を  $x$  で表し、 $x$  がどのように分布するかを示す分布関数を  $f(x)$  で表すと、対数正規分布は次のような式で表される：

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2x}} \exp\left\{-\frac{[\log(x/\bar{X})]^2}{2\sigma^2}\right\} \quad (4.1)$$

正規分布の場合の (1.1) と比べてみると、よく似ている部分と違う部分があることがわかるであろう。対数正規分布にも、正規分布の場合の平均値  $\mu$  と、分布の幅を与える標準偏差  $\sigma$  に相当する二つのパラメータ  $\bar{X}$  と  $\sigma$  が含まれている。図4-2には、(4.1) で  $\bar{X}=1.0$  と固定して  $\sigma$  の値をいくつか変えたときの対数正規分布の様子を示した。

実際にデータ解析する場合にはデータ数は有限なので、上の分布関数そのものより、データのある値  $x$  以上の値をもつデータ数がいくつあるかを示す累積個数分布（ランキングプロット）が便利である。対数正規分布の場合の累積個数分布  $N(x)$  は、

$$N(x) = N_T \int_x^\infty f(x') dx' \\ = \frac{N_T}{2} \left\{ 1 - \operatorname{erf} \left[ \frac{\log(x/\bar{X})}{\sqrt{2}\sigma} \right] \right\} \quad (4.2)$$

と表される。ここで  $N_T$  はデータの総数であり、 $\operatorname{erf}(x)$  は、

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-y^2) dy \quad (4.3)$$

と定義される誤差関数である。

前章でも強調したように、有限個数のデータがあるとき、累積個数分布あるいはランキングプロットを作るのは非常に簡単であるとともに有用でもある。それが対数正規分布であるかどうかを確かめるには、(4.2) で表される曲線がランキングプロットしたデータにちょうど合うように、(4.2) に含まれる二つのパラメータ  $\bar{X}$  と  $\sigma$  の値を調節する。これをデータフィッティングといい、試みの中で最良のものをベストフィットという。実際に世界のGDPのランキングプロットに対して対数正規分布でベストフィットを求めたのが、図3-4に示した曲線だったのである。

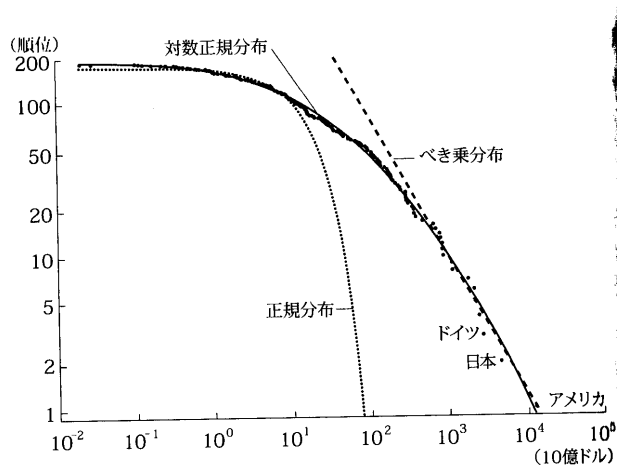


図4-3 世界各国のGDP(2005年)のランクを各種の分布関数でフィット (IMF World Economic Outlook Database, April 2019)

統計の世界で最も普通に使われる釣鐘状の正規分布は、点線で示されているように、右裾の順位の上位部分で曲線が鋭く落ち込み、データにまったくフィットできない。これは、対数正規分布では右裾が尾を引くのが特徴であるのに対して、正規分布ではそのようなことがないからである。

また、複雑系でよく使われるべき乗分布(破線)は両対数グラフ上で直線になるので、べき乗分布もこの場合には右裾のごく一部を除いてまったく不適格である。この図を見て明らかなように、正規分布やべき乗分布に比べて、対数正規分布は裾の部分だけでなく全体にわたって見事にフィットしている。

超大企業や超高額所得者は周囲から富を吸い上げて成長し、その結果、いっそう吸い上げる力をもつようになるのであって、大多数の中小企業や普通の人々の場合とは異なる。このような傾向を標語的に「富めるものはより豊かになる (“The rich get richer”）」という。

社会福祉政策が貧弱な自由主義経済の下では

「貧しきものはより貧困になる (“The poor get poorer”）」という現象が現実には起き始めている

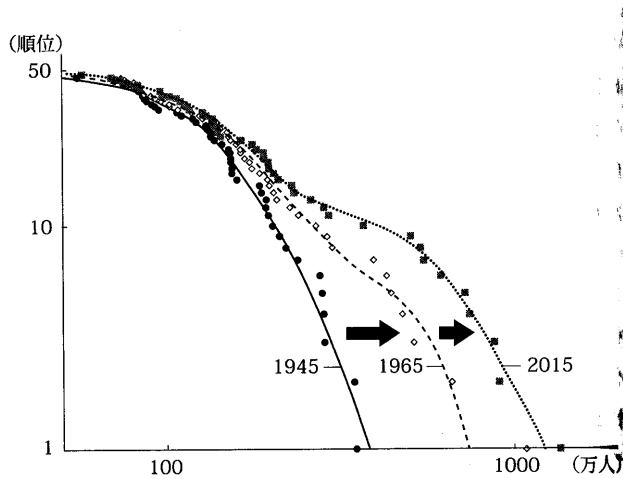


図6-4 都道府県の人口ランキングの推移（『国勢調査』及び「人口推計」）

1位の東京への極端な人口集中が始まり、  
2位大阪、3位神奈川、4位愛知、5位埼玉、6位千葉、  
などと推移。

さらに図からもわかるように、ランク  
じ位の12ほどの都道府県がそれ以下の県からくっきり  
と系統的にずれ始める。そして、年の経過とともにこの  
二つのグループの差はどんどん広がっていくのである。

トップ10への人口集中を表現するグラフだが、地方の過疎化は左上の領域に埋没されて表現できない。

地方行政の効率化ばかりを考慮して道州制を導  
入したりすると、道州より小さな地方自治体組織が抱え  
るいろいろな問題や矛盾がいつそう見えなくなり、無視  
されるようになることはほぼ間違いない。安易に道州制  
に賛成すべきでない理由の一つがここにある。

~~~~~

道州制（どうしゅうせい）とは、行政区画として道と州を置く地方行政制度である。府県制、市制、町村制  
などになった用語である。日本では、現行の都道府県よりも広域な行政区分として道と州を新たに設ける  
構想を指す。道州には現在の都道府県よりもレベルの高い地方自治権を与える。

~~~~~

べき乗分布が厄介なのは、正規分布と違って平均値が存在せず、標準的な大きさがわからないことである。

富めるものがより豊かになれば、いずれ貧しきもの  
にも富がしたり落ちてくる（トリクルダウン）というも  
っともらしい発言が、一部の経済学者や政治家から聞こ  
えてくる。もしこれが事実なら、もうとっくに格差の縮

小しつある社会が実現しているはずであるが、貧富の格差はいつそう広がっているのが現実である。