対数正規分布

～lognormal distribution～

平成 20 年 3 月

中央大学理工学部物理学科４年

香取研究室

浅野　翔　　　金田　佐和子

目次

1.対数正規分布とその性質

1.1　はじめに

私達の周りの身近な現象や事柄（例；体重や身長の分布、成績の分布、人口の分布など）は、正規分布、べき分布、対数正規分布に従うことが多い。私達はこの対数正規分布について研究した。

まず、対数正規分布とは、ある確率変数xの対数をとったもの$\log x$が、正規分布に従うことをいう。分布関数は、正規分布の確率密度関数

$$f_N(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

において$x \to \log x$に変数変換をすれば、

$$f_\Lambda(x) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left[-\frac{(\log x-\mu)^2}{2\sigma^2}\right] \qquad (1)$$

となる。これが対数正規分布の確率密度関数である。（ただし対数をとっているので$x \geq 0$）

1.2 モーメント

分布の特性を表すのに、モーメントが用いられる。例えば、1 次のモーメントは平均値を示し、2 次のモーメントを用いて分散を示すことが出来、3 次は歪度（分布の非対称性を示す量）、4 次は尖度（分布の尖り具合を示す量）を示す。ここで r 次のモーメントを示しておく。

$$\mu_r' = E(X^r) = \int_0^\infty x^r f_\Lambda(x)dx$$

$$= \exp(r\mu + \frac{1}{2}r^2\sigma^2)$$

これより、1 次のモーメント（平均値）は

$$\mu_1' = \exp(\mu + \frac{1}{2}\sigma^2)$$

となり、次に 2 次のモーメントを用いて分散を計算すると、

$$\mu_2 = E(X^2) - E(X)^2 = \mu_2{}' - \mu_1{}'$$

$$= \exp(2\mu + \sigma^2)\{\exp(\sigma^2) - 1\}$$

となる。

また、分布の特性を表す量として、Med（中央値）と Mode（最頻値）がある。これも示しておくと、

$$Med(X) = \exp(\mu)$$

$$Mode(X) = \exp(\mu - \sigma^2)$$

平均値、中央値、最頻値の間で成り立つ関係式は、

$$Mode(X) < Med(X) < E(X)$$

である。正規分布の場合、中央値と最頻値は共に μ で等しい。

1.3 グラフの概形

(1)式が描くグラフはどのような概形になるのだろうか。増減表を作成して考えてみる。(1)式の両辺を 1 階微分、2 階微分して、

$\dfrac{df(x)}{dx} = 0$ となる値は、

$$x_1 = \exp(\mu - \sigma^2) \quad \text{このとき} \quad f(x_1) = \frac{1}{\sqrt{2\pi}\sigma}\exp(\frac{\sigma^2}{2} - \mu)$$

$\dfrac{d^2 f(x)}{dx^2} = 0$ となる値は、

$$x_{2+} = \exp(\mu + \frac{-3\sigma^2 + \sigma\sqrt{\sigma^2 + 4}}{2})$$

$$x_{2-} = \exp(\mu + \frac{-3\sigma^2 - \sigma\sqrt{\sigma^2 + 4}}{2})$$

$x \to 0_+$ では、

$$\lim_{x \to 0_+} \frac{1}{x}\exp\left[-\frac{(\log x - \mu^2)}{2\sigma^2}\right] = \lim_{x \to 0_+} y$$

とおき、ここで対数をとると、

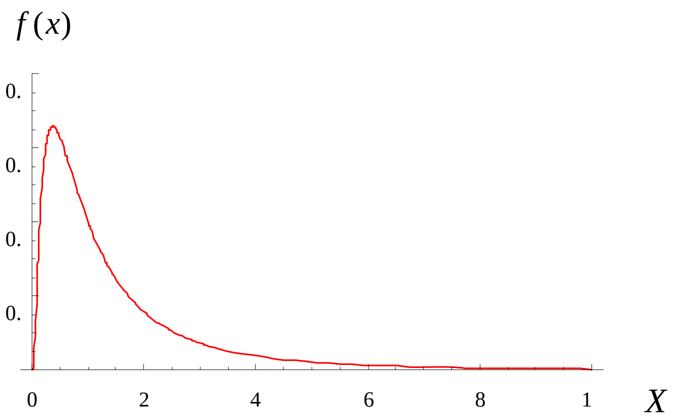
$$\lim_{x \to 0_+} \log y = -\infty \quad \text{より} \quad \lim_{x \to 0_+} y = 0$$

(1) 式が描くグラフは、$x \to 0_+$ で 0 に近づく。

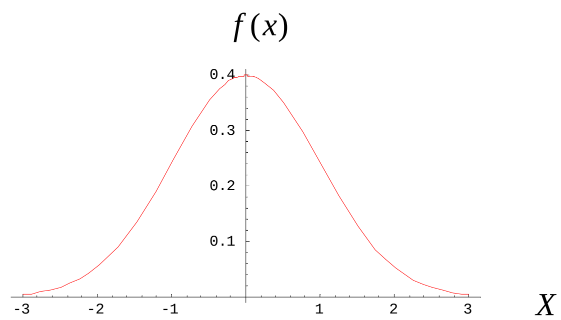
以上より増減表は、

x	0		x_{2-}		x_1		x_{2+}		∞
$\dfrac{df}{dx}$	／	$+$	$+$	$+$	0	$-$	$-$	$-$	$-$
$\dfrac{d^2 f}{dx^2}$	／	$+$	0	$-$	$-$	$-$	0	$+$	$+$
$f(x)$	0	$\nearrow$	$f(x_{2-})$	$\nearrow$	$f(x_1)$	$\searrow$	$f(x_{2+})$	$\searrow$	0 に近づく
			(変曲点)		(極値)		(変曲点)		

この増減表をもとにグラフの概形を考えると、



このようなグラフを描く事がわかる。非対称なグラフである。これに比べて、正規分布のグラフは対称なグラフを描く。



1.4　比例効果の法則と中心極限定理

　対数正規分布の特性として、比例効果の法則（Gibrat 1930,1931）というものがある。 j ステップ目のある正の変数 X_j (初期値 X_0>0)は、

$$X \quad = X_{j-1}(1+\varepsilon_j)$$

と乗法の式で表すことが出来る。（$\{X_j\}$は一様に分布した確率変数であり、ε_j はそれぞれ独立している集合で、かつ$\{X_j\}$とも独立している。）これを比例効果の法則という。

　次に中心極限定理というものがある。これは、確率変数 $x_1, x_2, x_3,,,,$ について

$X = x_1 + x_2 + x_3 + \cdots + x_N$ とおくと、

$$\frac{(x_1 - \langle x \rangle) + (x_2 - \langle x \rangle) + \cdots + (x_N - \langle x \rangle)}{\sqrt{N}} = \frac{X - N\langle x \rangle}{\sqrt{N}}$$

N→∞の極限で平均０、分散 σ^2 の正規分布に従う。

　これが対数正規分布になると、$Y = x_1 \times x_2 \times x_3 \times \cdots \times x_N$ と乗法の式でおく。

$\log Y = \log x_1 + \log x_2 + \log x_3 + \cdots + \log x_N$ となり、これが正規分布に従う。

　比例効果の法則と中心極限定理から分かるように、対数正規分布に従う現象が過去の影響を乗法的に受けているような現象である。

　そして私達は、実際に身の周りにある現象や事柄のデータをグラフにしてみた。

2.グラフ

2.1.累積分布

　グラフを作成する際、縦軸を順位や度数にした。（累積分布）
累積分布関数をここに定義する。

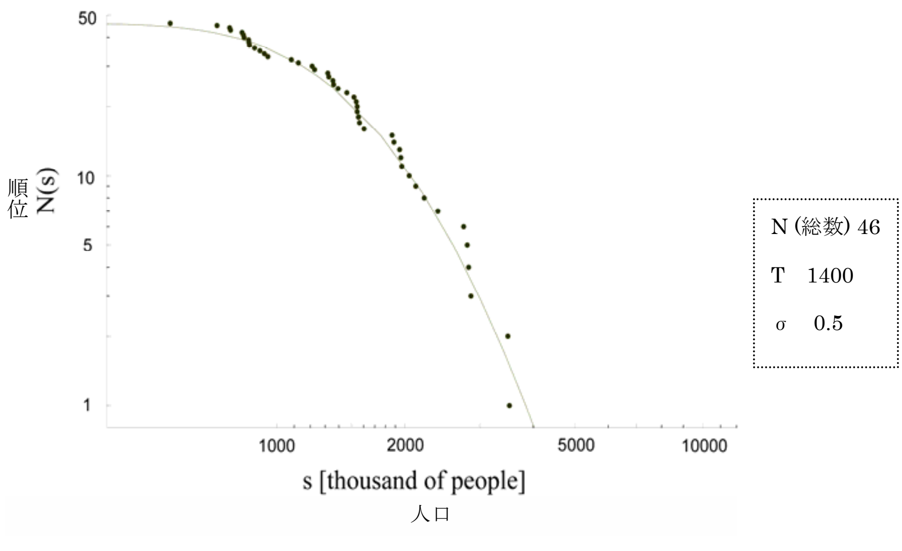
$$N(x) = \frac{N_T}{2}\left\{1 - erf\left[\frac{\log\left(x/T\right)}{\sqrt{2}\sigma}\right]\right\} \qquad (\mu = \log T)$$

ただし、$erf(x) = \dfrac{2}{\sqrt{\pi}}\int_0^x \exp(-y^2)dy$　　（誤差関数）

また、データフィッティングは非線形最小二乗法で行った。

2.2 日本の都道府県の人口分布(1945 年度)　Kobayashi *et* al. , (2006)

データ出展：*Japan statistical year book 2005*, edited by the Statistical Training Institute
(the Statistics Bureau, both under the Ministry of Internal Affairs and Communications, Japan, 2004).

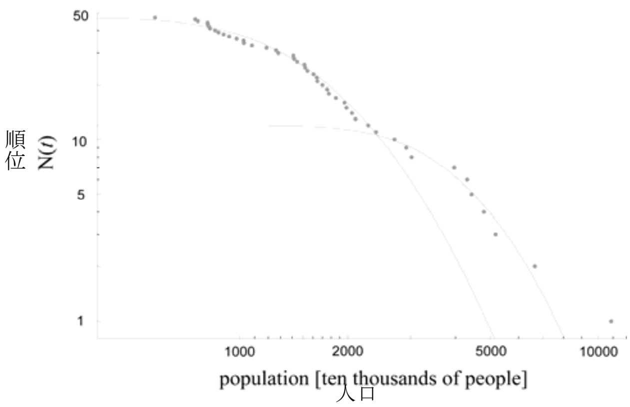


　これは、都道府県別に人口の多い順に 1 から番号をふって、その順位を縦軸に、人口を横軸にとったものである。1945 年の人口分布は、第 2 次世界大戦により都市部から地方部への疎開や戦争により戦死した人などの影響を受けている。

2.3 日本の都道府県の人口分布(1965,1985,2003 年) Kobayashi *et al.* , (2006)

データ出展：*Japan statistical year book 2005*, edited by the Statistical Training Institute (the Statistics Bureau, both under the Ministry of Internal Affairs and Communications, Japan, 2004).

1965年



N_{large} 35

T_{large} 1300

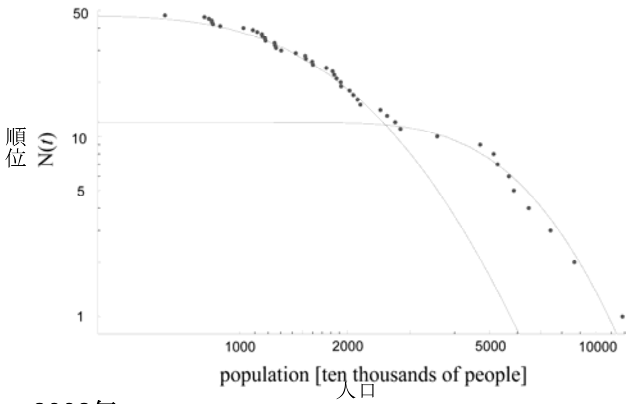
σ_{large} 0.35

N_{small} 12

T_{small} 3900

σ_{small} 0.49

1985年



N_{large} 35

T_{large} 1430

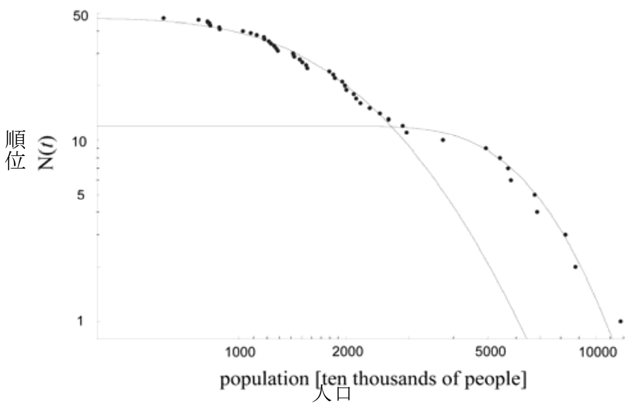
σ_{large} 0.34

N_{small} 12

T_{small} 5800

σ_{small} 0.45

2003年



N_{large} 35

T_{large} 1450

σ_{large} 0.34

N_{small} 12

T_{small} 6100

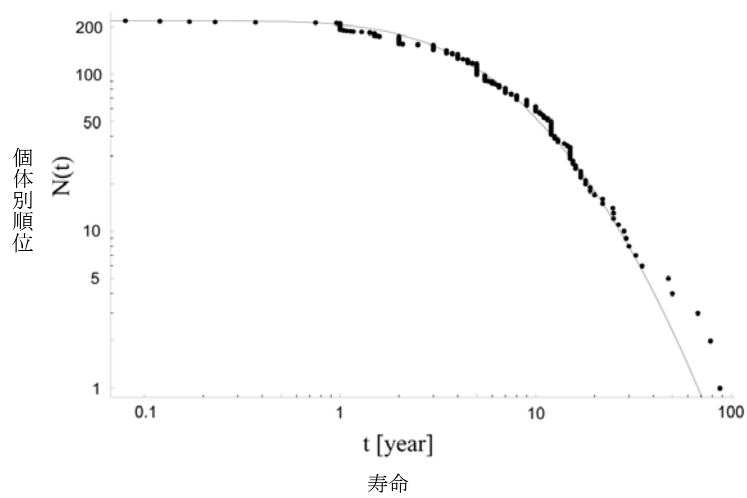
σ_{small} 0.42

これらのグラフの特徴は、2つの対数正規分布が見えるところである。時代の背景として、高度経済成長による都市部の経済の発展と、それによる人口の集

中での過密化、地方への過疎化が影響を受けている。二極化が2つの対数正規分布で現れているようである。

2.4　無脊椎動物の平均寿命の分布　　Kobayashi *et al.* , (2006)

データ出展：A. Comfort, *The Biology of Senescence* (Churchill Livingstone, Edinburgh, 1979).

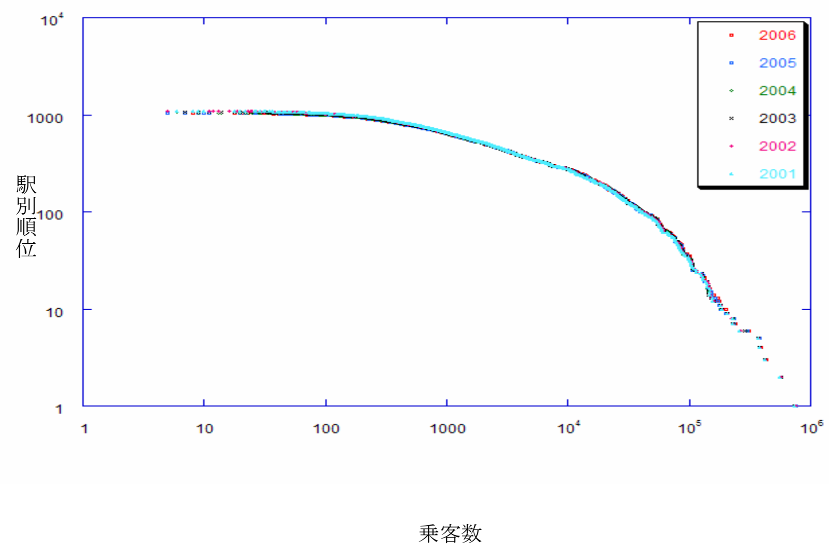


N	218
T	5.0
σ	1.0

無脊椎動物の寿命の分布は、自然界と関係がある。過去の影響を受けているのと照らし合わせて考えると、進化の過程において、様々な原因や事象の影響を受けていると考えられる。

2.5　ＪＲ東日本の乗客数分布(2001～2006 年)

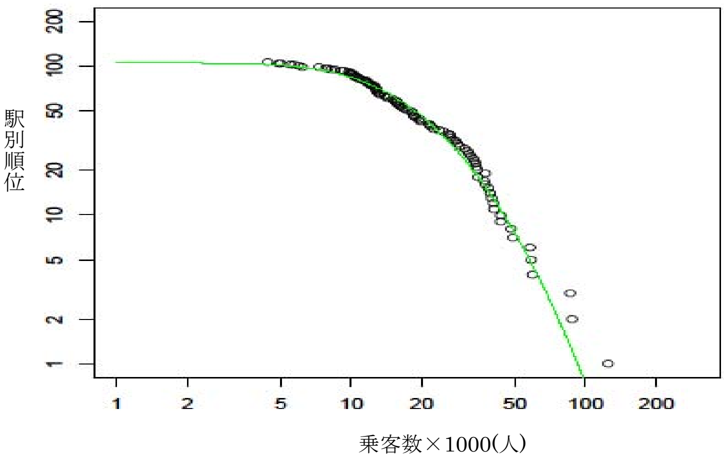
データ出展：JR 東日本ホームページ　<http://www.jreast.co.jp/passenger/>



	N	T	σ
2001	1095	1906	2.15
2002	1090	1895	2.17
2003	1066	1980	2.17
2004	1052	2023	2.18
2005	1048	2033	2.20
2006	1045	2049	2.20

2.6　都営地下鉄の各駅の乗客数分布(2006 年)

データ出展：東京都交通局 http://www.kotsu.metro.tokyo.jp/



N	105
T	17.85
σ	0.70

2.5 JR の乗客数分布と 2.6 地下鉄の乗客数分布については、乗り換え、快速、特急が止まる駅、その駅の周辺の様子（会社や学校、住宅など）が関わっている。急行の止まる駅や他社線への乗り換えなどの影響も受けていると考えられる。（都営地下鉄については、これはあと 2,3 年のデータを調査してみればもう少し何かが見えてくるかもしれない。）

3.グラフの考察

1945 年の人口分布、無脊椎動物の平均寿命の分布、都営地下鉄の乗客数分布の 3 つのグラフについて、実測してデータとグラフから読み取ったデータをそれぞれ表にまとめて比較してみる。

日本の都道府県の人口分布(1945)

	実測データ	グラフデータ
平均値	1565	1586
中央値	1428	1400
最頻値	1546	1090
分散	534482	714805
標準偏差	845.5	731.1

無脊椎動物の平均寿命の分布

	実測データ	グラフデータ
平均値	8.34	8.24
中央値	5.00	5.00
最頻値	1.00	1.84
分散	128	118
標準偏差	11.3	10.8

都営地下鉄の各駅の乗客数の分布(2006 年)

	実測データ	グラフデータ
平均値	23.06	22.83
中央値	16.40	17.85
最頻値	18.60	10.90
分散	343.2	328.9
標準偏差	18.52	18.13

比較してみると、平均値、中央値、標準偏差については各データそれぞれ誤差は少ないが。最頻値の誤差が目立つのがわかる。

そもそも最頻値はグラフで見ると、点が１番混みあっている部分である。どのデータもその混みあっている部分が一つではなく、いくつかあるので、グラフのデータを読み取る時に誤差が生じたのではないかと考えられる。

また、都営地下鉄の乗客数の分布については、グラフを眺めていると 2 つの対数正規分布が見えてくる。これも誤差の原因のひとつではないかと考えられる。

参考文献

１. Lognormal distributions: theory and applications, edited by Edwin L. Crow and Knio Shimizu (Marcel Dekker, New York, 1988).

２.　Japan statistical year book 2005, edited by the Statistical Training Institute (the Statistics Bureau,both under the Ministry of Internal Affairs and Communications, Japan, 2004).

３. A. Comfort, The Biology of Senescence (Churchill Livingstone, Edinburgh, 1979).

４.　N,Kobayashi, Y.Sasaki, O. Moriyama, S. Matsushita and M.Matsushita , Empirical Studies of Random Multiplicative Stochastic Processes: Revisit to Lognormals Nonlinear Phenomena in Complex System, Vol.9,No.3 (2006) pp.276-282, 2006

５. JR 東日本ホームページ<http://www.kotsu.metro.tokyo.jp/>