

対数正規分布

~ lognormal distribution ~

H20.1.16

浅野 翔

金田 佐和子

対数正規分布とは

ある確率変数 X の対数をとったもの $\log x$ が正規分布する

「対数正規分布に従う」 という

(ただし、 $x \geq 0$)

確率密度関数

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

正規分布の場合

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

分布の特徴を表すのにモーメントが用いられる

r次モーメント

$$\begin{aligned}\mu'_r &= E(X^r) = \int_0^\infty x^r f_\Lambda(x) dx \\ &= \exp\left(r\mu + \frac{1}{2}r^2\sigma^2\right)\end{aligned}$$

1次のモーメント(平均値)

$$\mu'_1 = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

また、2次のモーメントを用いて分散を計算

$$\begin{aligned}\mu_2 &= E(X^2) - E(X)^2 = \mu'_2 - \mu_1'^2 \\ &= \exp(2\mu + \sigma^2) \left\{ \exp(\sigma^2) - 1 \right\}\end{aligned}$$

グラフの概形を考える

$$\frac{df_{\Lambda}(x)}{dx} = 0 \quad \text{となる値 } x_1 = \exp(\mu - \sigma^2), \quad f_{\Lambda}(x_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{\sigma^2}{2} - \mu\right)$$

$$\frac{d^2 f_{\Lambda}(x)}{dx^2} = 0 \quad \text{となる値 } x_{2+} = \exp\left(\mu + \frac{-3\sigma^2 + \sigma\sqrt{\sigma^2 + 4}}{2}\right)$$

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

また、 $x \rightarrow 0+$ では





$$\lim_{x \rightarrow 0+} \frac{1}{x} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right] = \lim_{x \rightarrow 0+} y \quad \text{とおく}$$

ここで対数をとると、

$$\lim_{x \rightarrow 0^+} \log y = -\infty \quad \text{なので、} \quad \lim_{x \rightarrow 0^+} y = 0$$

$x \rightarrow 0^+$ で 0 に近づく

以上より、増減表は

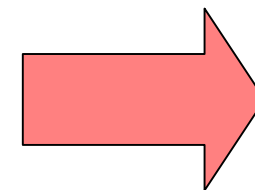
x	0		x_{2-}		x_1		x_{2+}		∞
$\frac{df}{dx}$	/	+	+	+	0	-	-	-	-
$\frac{d^2f}{dx^2}$	/	+	0	-	-	-	0	+	+
$f(x)$	0		$f(x_{2-})$		$f(x_1)$		$f(x_{2+})$		0 に近づく

変曲点

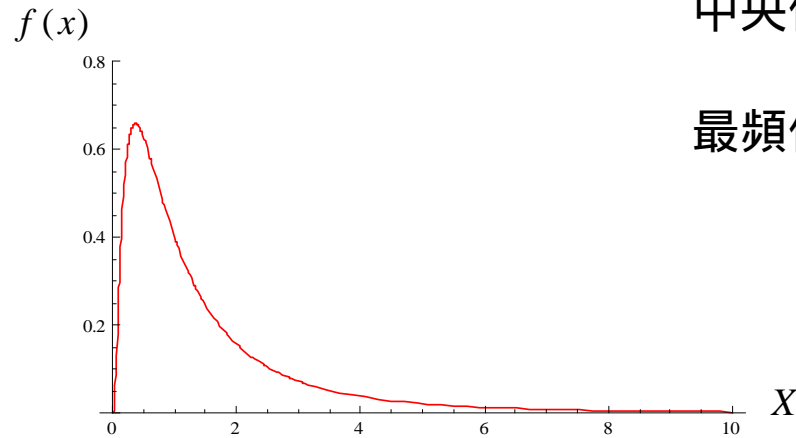
極値

変曲点

どのようなグラフを描くのか？



グラフ



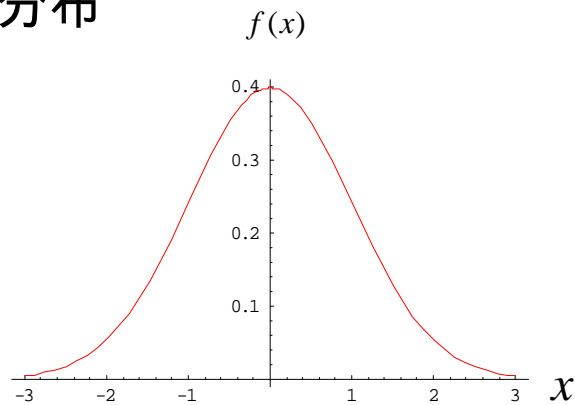
中央値Med(x) $e^{-\mu}$

最頻値Mode(x) $\exp(-\frac{\mu}{\sigma^2})$

Mode(x) < Med(x) < E(x)

非対称なグラフ

正規分布



中央値、最頻値ともに μ

対称なグラフ

比例効果の法則~The Law of Proportionate Effect~

(Gibrat 1930,1931)

jステップ目のある変数 X_j (初期値 X_0)が

$$X_j = X_{j-1} (1 + \varepsilon_j) \quad \text{乗法で表せる}$$

と表せるとき、 $\{X_j\}$ のプロセスは比例効果の法則に従う

中心極限定理

確率変数 $x_1, x_2, x_3, \dots, x_N$ について、 $X = x_1 + x_2 + x_3 + \dots + x_N$ とおくと

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

N の極限で平均0、分散 σ^2 の正規分布に従う

加法的

対数正規の場合

$$Y = x_1 \times x_2 \times x_3 \times \cdots \times x_N \quad \text{とおくと}$$

$$\log Y = \log x_1 + \log x_2 + \cdots + \log x_N$$

乗法的!

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わかるとおり、

対数正規分布に従う現象は

過去の影響を乗法的に受けているような現象である。

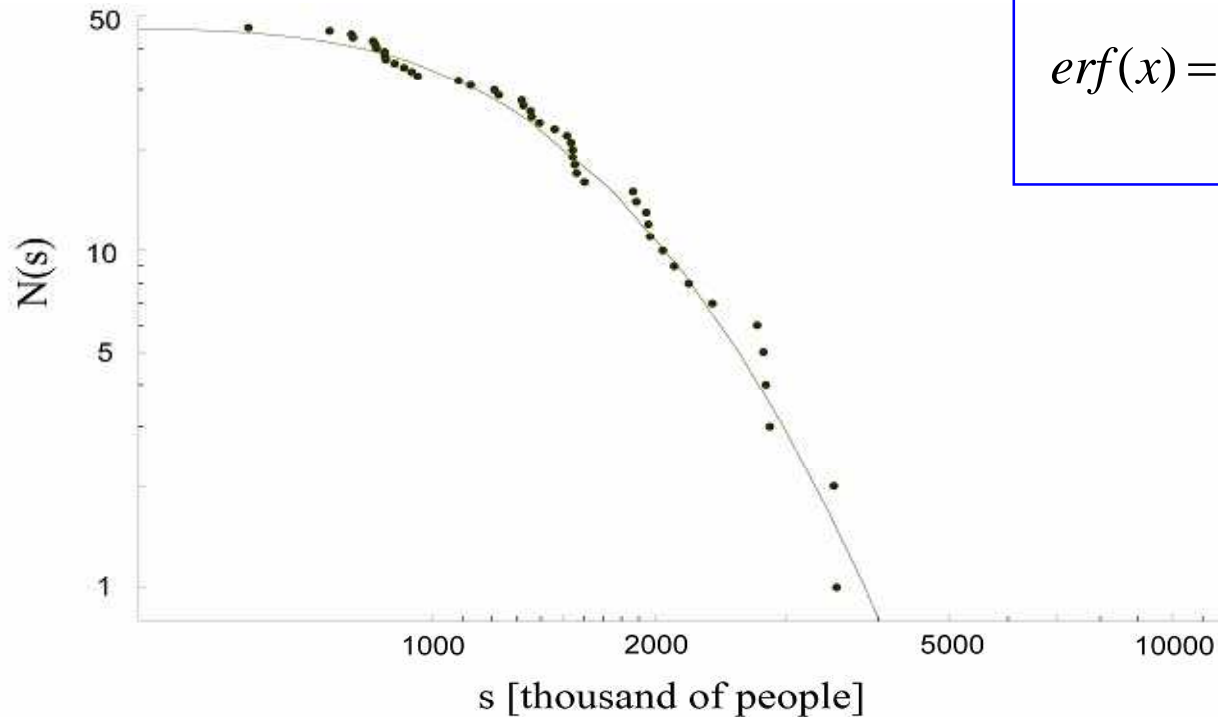
グラフ

日本の都道府県人口分布 (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).



累積分布関数

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

ただし、 $(\mu = \log T)$

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

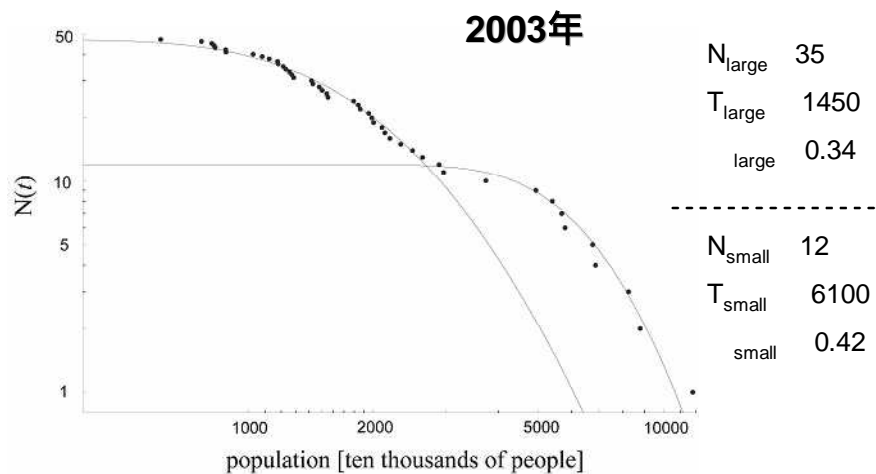
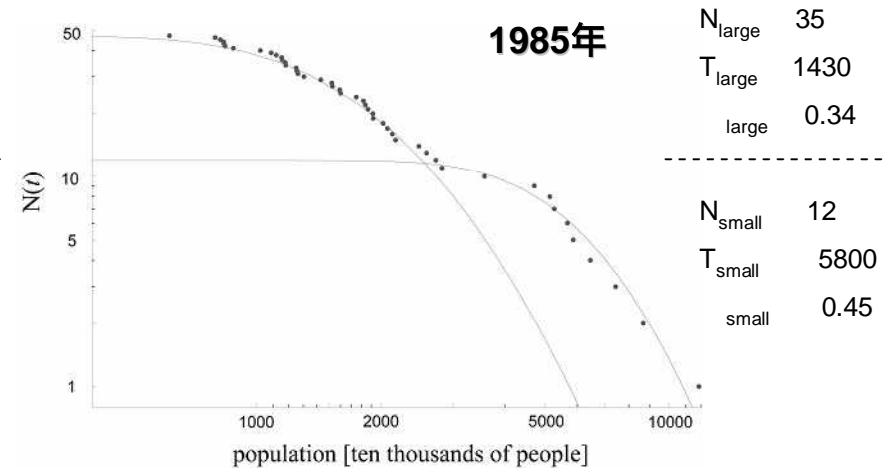
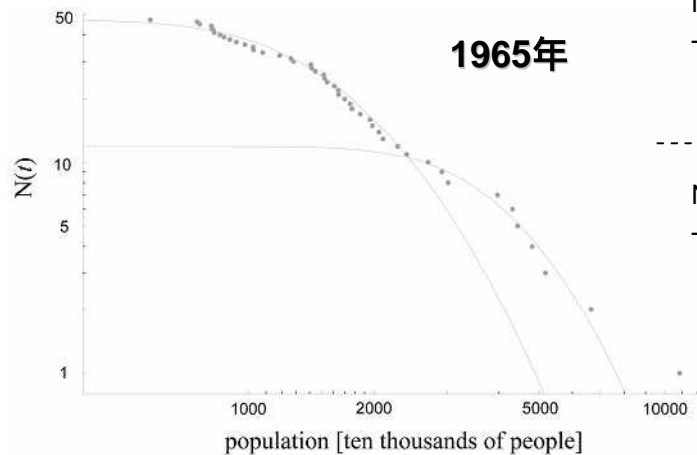
N	46
T	1400
	0.50

日本の都道府県人口分布

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).



2つの対数正規分布が見える

2 lognormal distribution

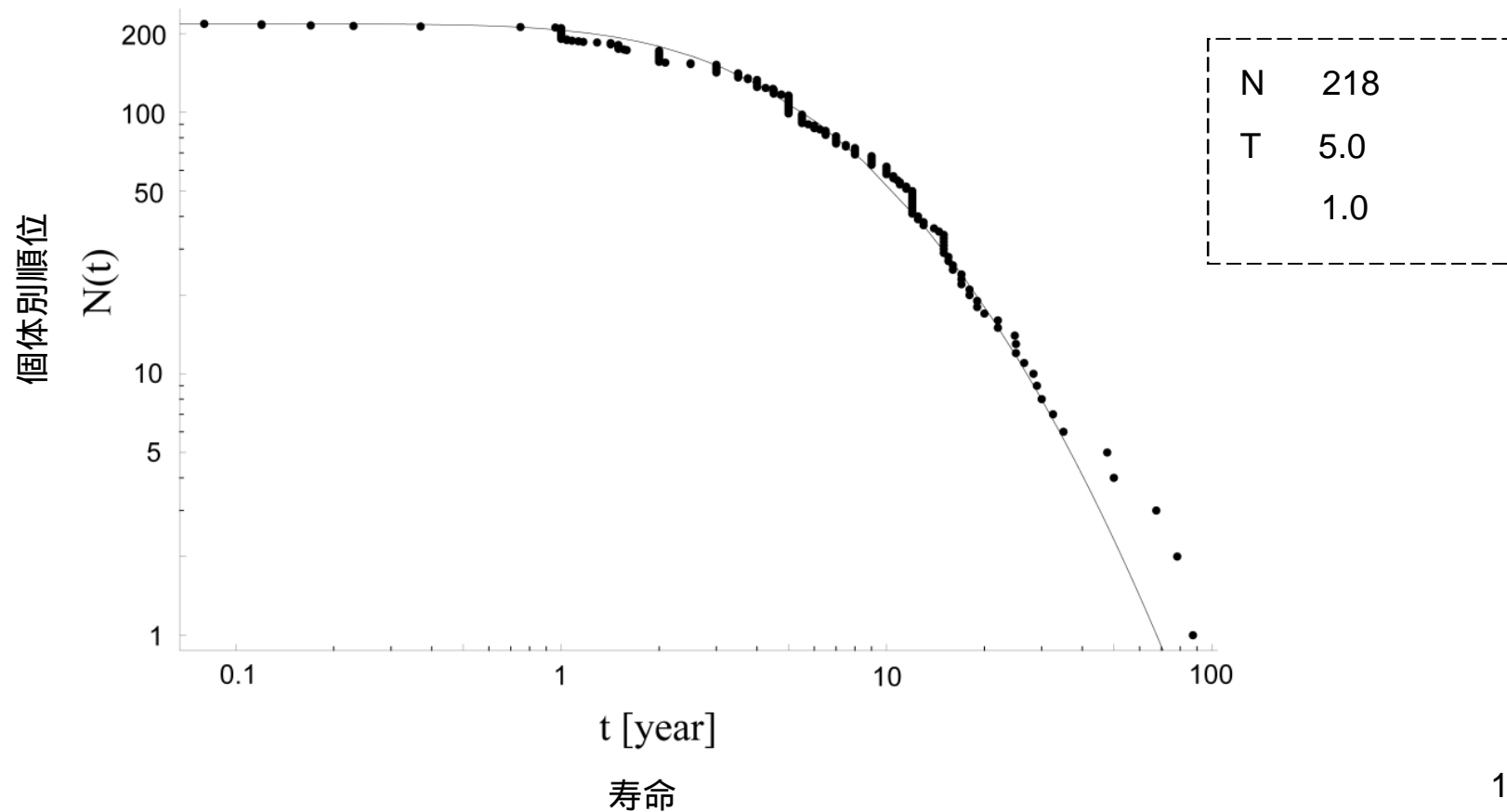
身近な現象とそのグラフ(1)

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).



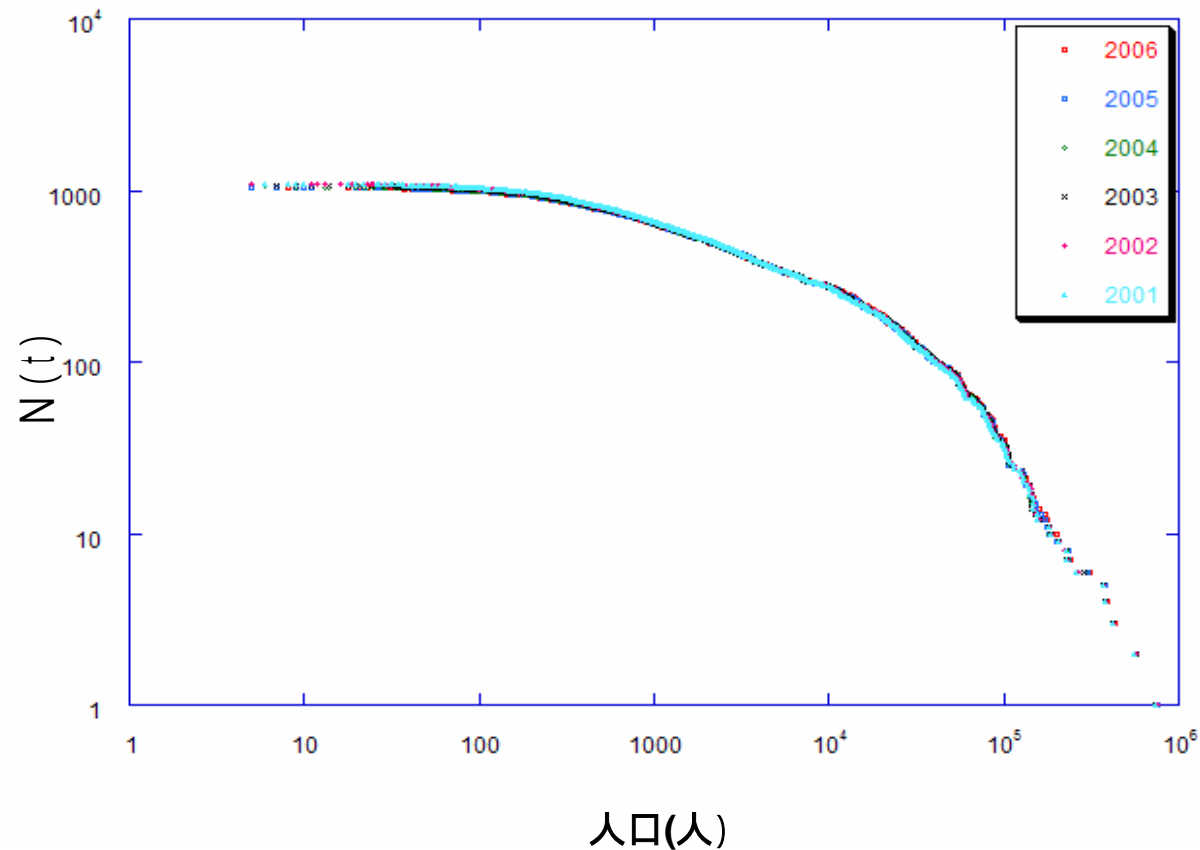
身近な現象とそのグラフ(2)

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	
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

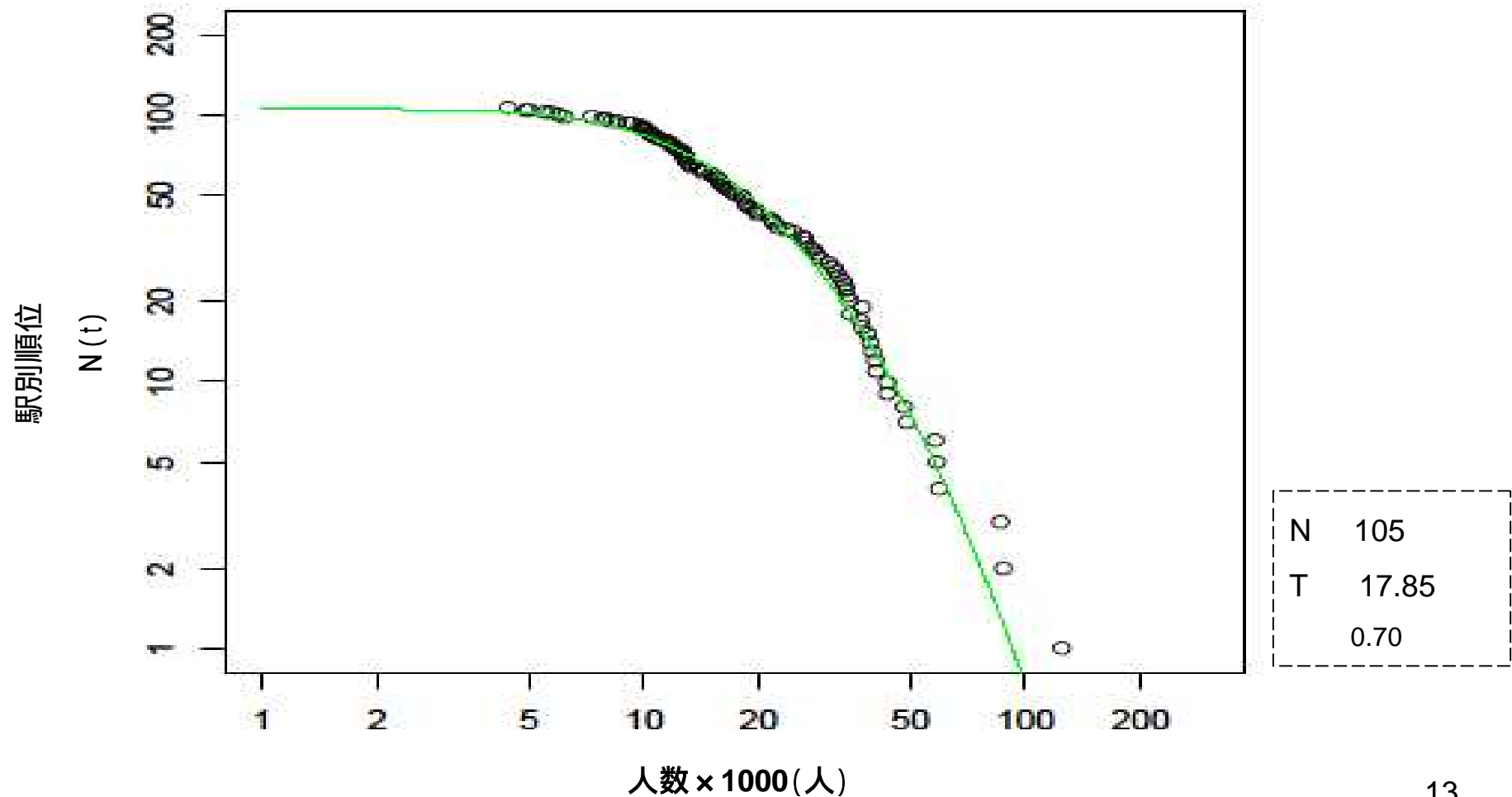
身近な現象とそのグラフ (3)

都営地下鉄の各駅の乗客数分布 (2006年度)

<データ出典>

東京都交通局ホームページ

<<http://www.kotsu.metro.tokyo.jp/>>



グラフの考察

日本の都道府県の人口分布(1945年)

	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

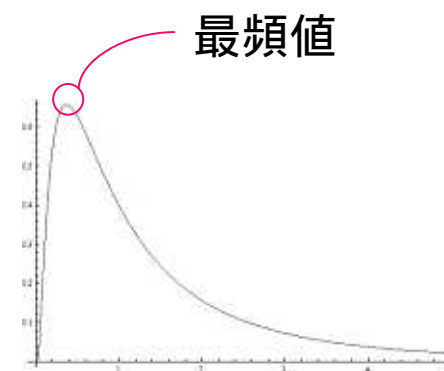
比較してみると・・・

- ・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。
- ・最頻値の誤差が目立つ。

なぜ最頻値の誤差が大きいのだろうか？

考えられる理由

最頻値は右図のとおりグラフの山の値である。しかし、実際の現象を対数正規分布でプロットすると山がハッキリとはわからない事が多い。従って、誤差が生じやすいのではと考えられる。



さらに・・・

都営地下鉄の各駅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一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 Kunio 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/>>