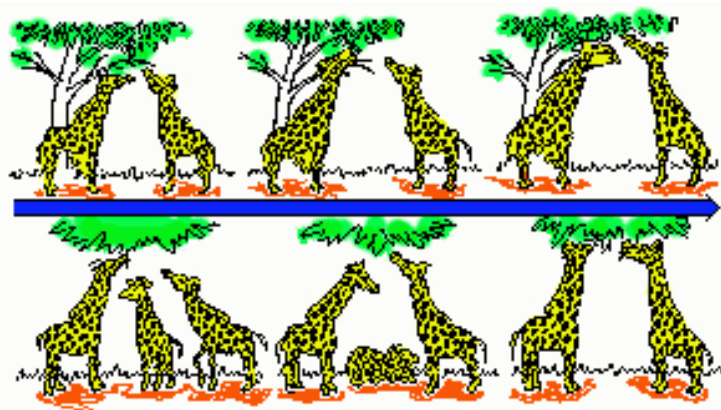


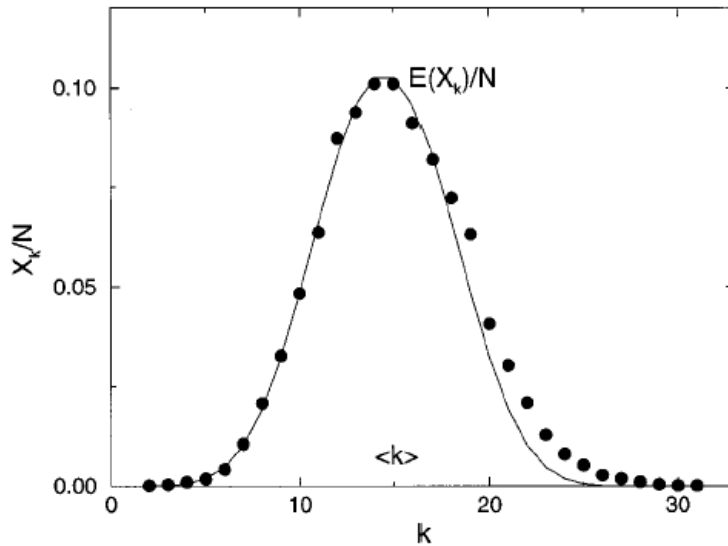
# 複雑系というシステム

## 複雑系という知能

東京大学大学院  
情報理工学系科学研究科  
電気情報学専攻  
伊庭齐志

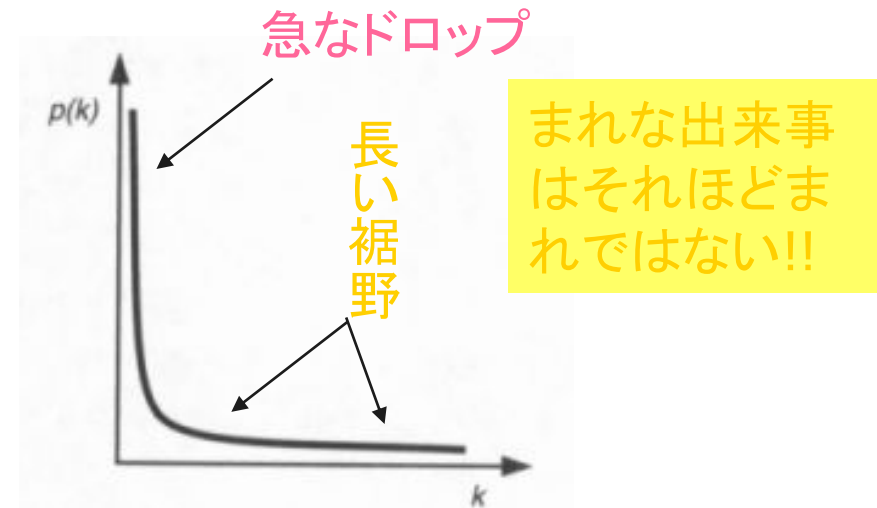


# 次数分布とべき乗法則



ランダムグラフの次数分布,  
 $N = 10,000$   $p = 0.0015$   $\langle k \rangle = 15$ .  
(比較のため、ポアソン曲線を示している)

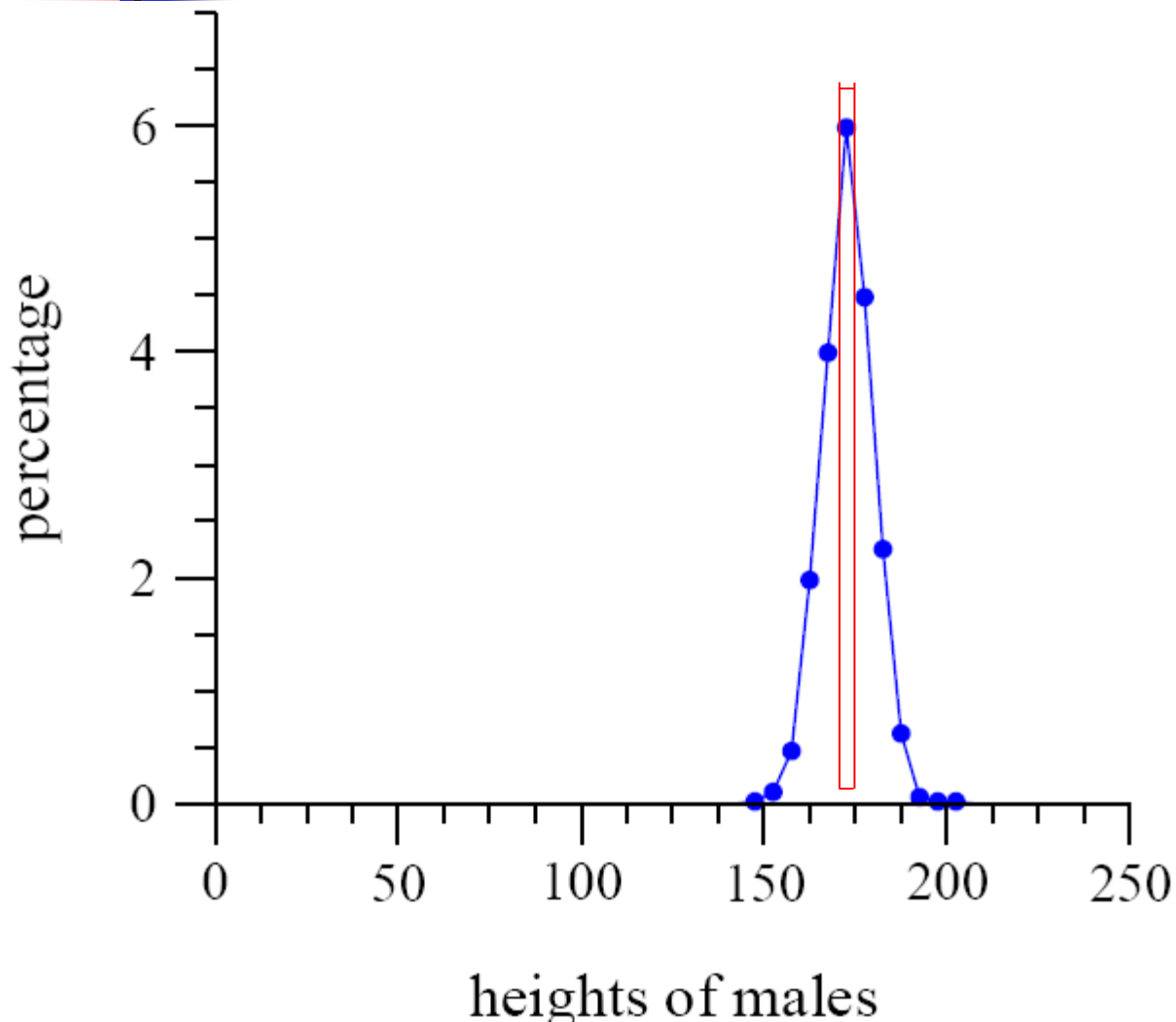
ポアソン分布は指数的に減衰するが、  
べき乗則は多項式的なことに注意



しかし多くの実世界のネットワークはべき乗則分布を示す  
→これはまた「重い裾野 (heavy tailed) のをもつ」分布とも呼ばれる

典型的には  $2 < r < 3$  である。  
Webのグラフでは、  
入次数で  $r \sim 2.1$   
出次数で  $2.7$

# スケールフリー性



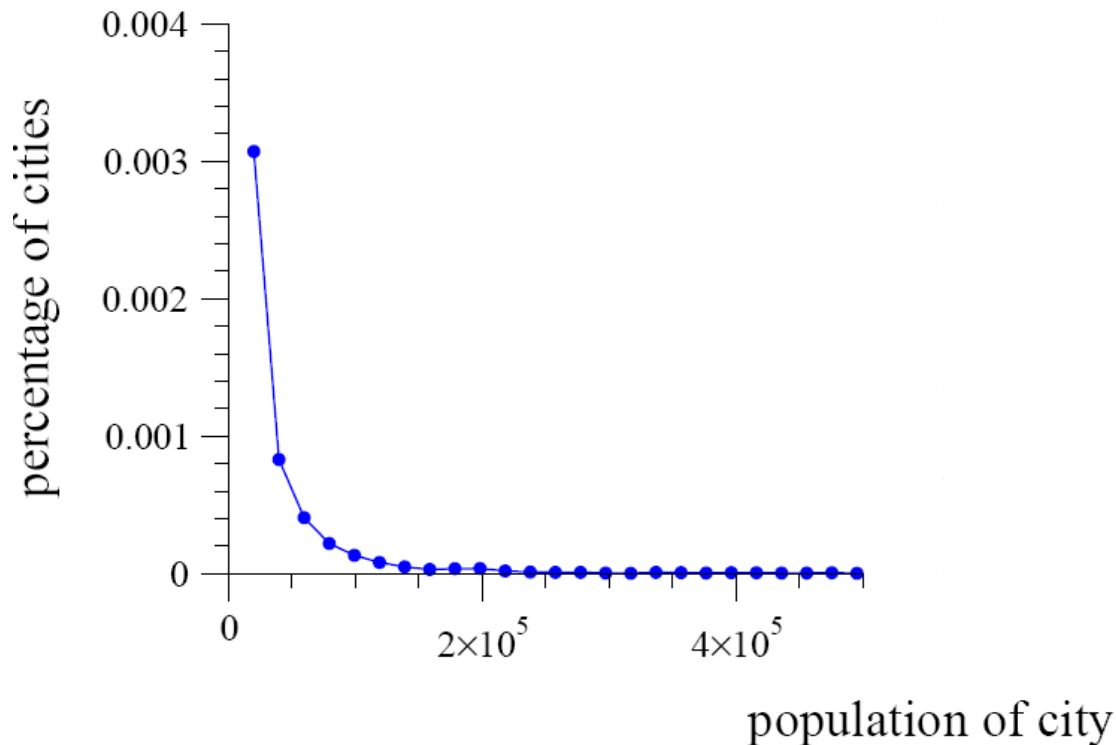
われわれの測定する多くのものはある特定の値の中心に集まる

この値は典型的なサイズである

# スケールフリー性

しかし、分布に巨大な**変異**があるものもある。

もしこのヒストグラムをログスケールでプロットすると**直線**パターンが明確にあらわれる





# スケールフリー性

---

しかし、分布に巨大な**変異**があるものもある。

もしこのヒストグラムをログスケールでプロットすると**直線**パターンが明確にあらわれる

通常のヒストグラムでは、この直線は $p(x) = -ax + c$ である。ここでlog-logなので:

$$\ln p(x) = -a \ln x + c$$

↓  
指数 $e$ を適用する

$$p(x) = e^c x^{-a}$$

この分布は指数 $a$ の**べき乗則**に従うという

べき乗則はどんなスケールで見ても同じである唯一の分布である。  
つまり、 $p(bx) = g(b)p(x)$  となる。したがって**スケールフリー**と呼ばれている。

# スケールフリー性

人口はスケールフリー性を持っていることはすでに知られていた!

1955年に、Herbert Simonはすでに多くのシステムがべき乗則の分布に従うことを示していた。そのためこの結果は新規でも唯一でもない。

- 地震の大きさ
- 月のクレータ
- 太陽フレア
- コンピュータのファイル



- 戦争
- 論文の引用回数
- Webページのヒット数
- 人々の年間収入

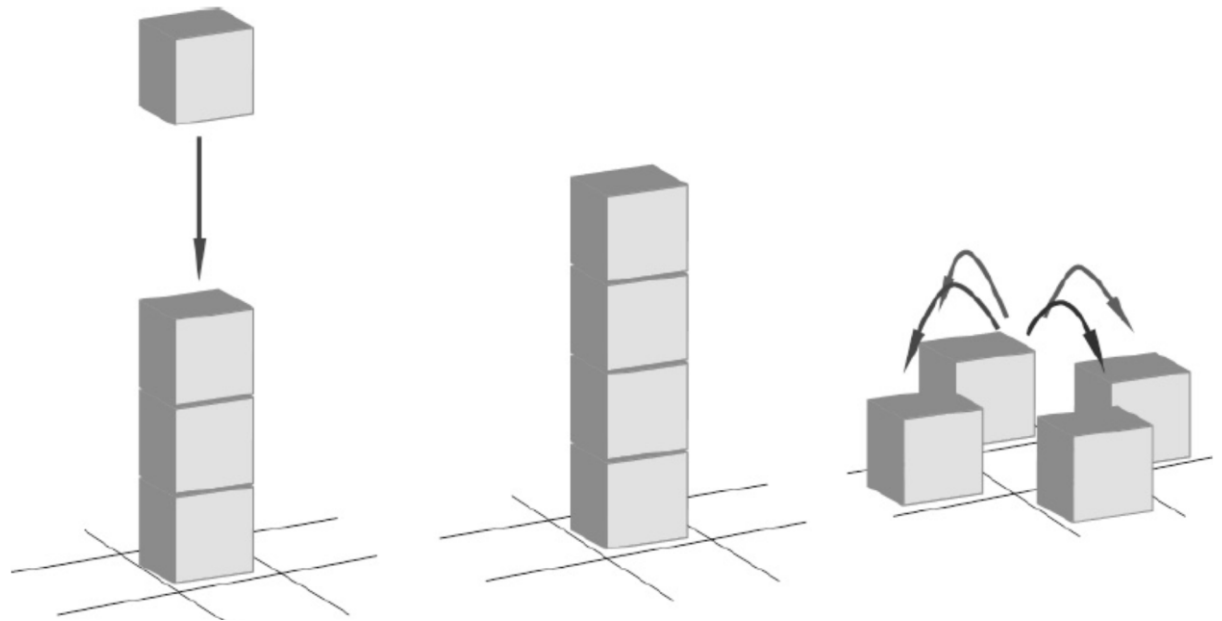


ノードの次数分布は多くのネットワークでべき乗則に従うことがわかっている。つまり多くのネットワークはスケールフリーである。

重要なことはべき乗則を見出すことではない(むしろありふれている)  
なぜそうなるのか、他のどんな構造的パラメータがあるのかを理解することである。

# 砂山モナ

- 1987年に唱されたC
- 砂粒を地
- 砂粒は段、ある程度

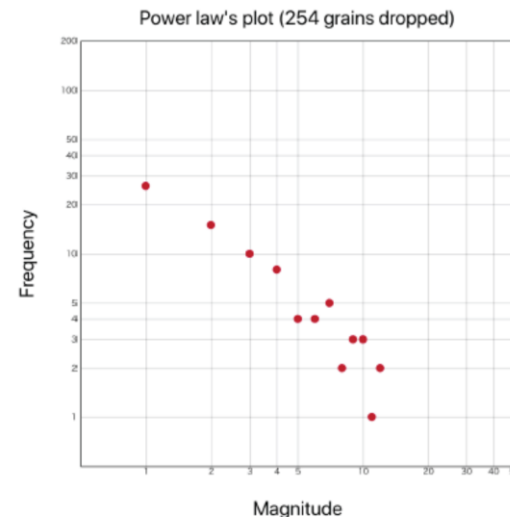
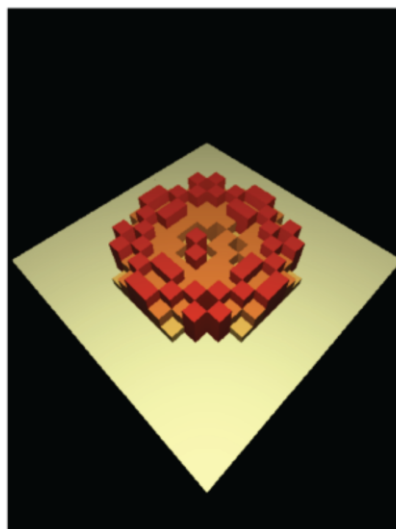


- ブロックが4個以上積み上がると,4個のブロックが上下左右のセルに飛散する
- 飛散した先でまたブロックが4個以上積み上がると,再帰的にブロックの飛散が繰り返される

# 砂山モデル

- 3D CAによるシミュレーション
- 積み上がった砂粒の高さに応じて砂粒の色を濃くする
- 砂山の崩壊の規模を対数グラフにプロット
- 崩壊規模の定義 = 「レーム数」
- 同時に複数個所で

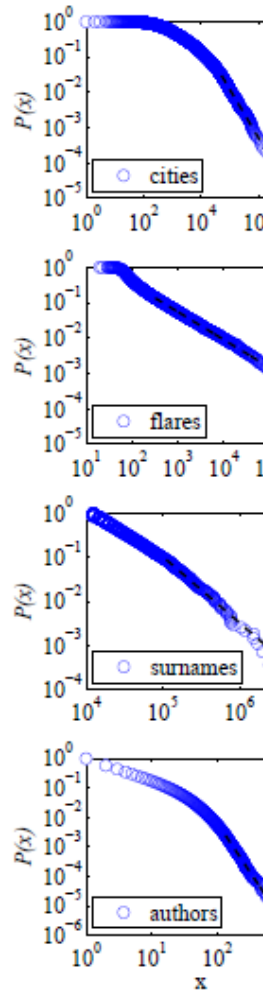
Sandpile model simulator  
The scale-free property of avalanches





# スケールフリー性

"POWER-LAW DISTRIBUTIONS IN EMPIRICAL DATA"  
 AARON CLAUSET, COSMA ROHILLA SHALIZI, AND M.E.J.  
 NEWMAN



| quantity                               | $n$         | $\langle x \rangle$ | $\sigma$   | $x_{\max}$ | $\hat{x}_{\min}$   | $\hat{\alpha}$ | $n_{\text{tail}}$     | $p$         |
|--|-------------|---------------------|------------|------------|--------------------|----------------|-----------------------|-------------|
| count of word use                      | 18 855      | 11.14               | 148.33     | 14 086     | $7 \pm 2$          | 1.95(2)        | $2958 \pm 987$        | <b>0.49</b> |
| protein interaction degree             | 1846        | 2.34                | 3.05       | 56         | $5 \pm 2$          | 3.1(3)         | $204 \pm 263$         | <b>0.31</b> |
| metabolic degree                       | 1641        | 5.68                | 17.81      | 468        | $4 \pm 1$          | 2.8(1)         | $748 \pm 136$         | 0.00        |
| Internet degree                        | 22 688      | 5.63                | 37.83      | 2583       | $21 \pm 9$         | 2.12(9)        | $770 \pm 1124$        | <b>0.29</b> |
| telephone calls received               | 51 360 423  | 3.88                | 179.09     | 375 746    | $120 \pm 49$       | 2.09(1)        | $102 592 \pm 210 147$ | <b>0.63</b> |
| intensity of wars                      | 115         | 15.70               | 49.97      | 382        | $2.1 \pm 3.5$      | 1.7(2)         | $70 \pm 14$           | <b>0.20</b> |
| terrorist attack severity              | 9101        | 4.35                | 31.58      | 2749       | $12 \pm 4$         | 2.4(2)         | $547 \pm 1663$        | <b>0.68</b> |
| HTTP size (kilobytes)                  | 226 386     | 7.36                | 57.94      | 10 971     | $36.25 \pm 22.74$  | 2.48(5)        | $6794 \pm 2232$       | 0.00        |
| species per genus                      | 509         | 5.59                | 6.94       | 56         | $4 \pm 2$          | 2.4(2)         | $233 \pm 138$         | <b>0.10</b> |
| bird species sightings                 | 591         | 3384.36             | 10 952.34  | 138 705    | $6679 \pm 2463$    | 2.1(2)         | $66 \pm 41$           | <b>0.55</b> |
| blackouts ( $\times 10^3$ )            | 211         | 253.87              | 610.31     | 7500       | $230 \pm 90$       | 2.3(3)         | $59 \pm 35$           | <b>0.62</b> |
| sales of books ( $\times 10^3$ )       | 633         | 1986.67             | 1396.60    | 19 077     | $2400 \pm 430$     | 3.7(3)         | $139 \pm 115$         | <b>0.66</b> |
| population of cities ( $\times 10^3$ ) | 19 447      | 9.00                | 77.83      | 8 009      | $52.46 \pm 11.88$  | 2.37(8)        | $580 \pm 177$         | <b>0.76</b> |
| email address books size               | 4581        | 12.45               | 21.49      | 333        | $57 \pm 21$        | 3.5(6)         | $196 \pm 449$         | <b>0.16</b> |
| forest fire size (acres)               | 203 785     | 0.90                | 20.99      | 4121       | $6324 \pm 3487$    | 2.2(3)         | $521 \pm 6801$        | 0.05        |
| solar flare intensity                  | 12 773      | 689.41              | 6520.59    | 231 300    | $323 \pm 89$       | 1.79(2)        | $1711 \pm 384$        | <b>1.00</b> |
| quake intensity ( $\times 10^3$ )      | 19 302      | 24.54               | 563.83     | 63 096     | $0.794 \pm 80.198$ | 1.64(4)        | $11 697 \pm 2159$     | 0.00        |
| religious followers ( $\times 10^6$ )  | 103         | 27.36               | 136.64     | 1050       | $3.85 \pm 1.60$    | 1.8(1)         | $39 \pm 26$           | <b>0.42</b> |
| freq. of surnames ( $\times 10^3$ )    | 2753        | 50.59               | 113.99     | 2502       | $111.92 \pm 40.67$ | 2.5(2)         | $239 \pm 215$         | <b>0.20</b> |
| net worth (mil. USD)                   | 400         | 2388.69             | 4 167.35   | 46 000     | $900 \pm 364$      | 2.3(1)         | $302 \pm 77$          | 0.00        |
| citations to papers                    | 415 229     | 16.17               | 44.02      | 8904       | $160 \pm 35$       | 3.16(6)        | $3455 \pm 1859$       | <b>0.20</b> |
| papers authored                        | 401 445     | 7.21                | 16.52      | 1416       | $133 \pm 13$       | 4.3(1)         | $988 \pm 377$         | <b>0.90</b> |
| hits to web sites                      | 119 724     | 9.83                | 392.52     | 129 641    | $2 \pm 13$         | 1.81(8)        | $50 981 \pm 16 898$   | 0.00        |
| links to web sites                     | 241 428 853 | 9.15                | 106 871.65 | 1 199 466  | $3684 \pm 151$     | 2.336(9)       | $28 986 \pm 1560$     | 0.00        |

TABLE 6.1

Basic parameters of the data sets described in this section, along with their power-law fits and the corresponding  $p$ -value (statistically significant values are denoted in bold).

# スケールフリー性

"POWER-LAW DISTRIBUTIONS IN EMPIRICAL DATA"  
AARON CLAUSET, COSMA ROHILLA SHALIZI, AND M.E.J.  
NEWMAN

| data set   | power law   | log-normal |             | exponential |             | stretched exp. |             | power law + cut-off |             | support for power law |
|------------|-------------|------------|-------------|-------------|-------------|----------------|-------------|---------------------|-------------|-----------------------|
|            | $p$         | LR         | $p$         | LR          | $p$         | LR             | $p$         | LR                  | $p$         |                       |
| birds      | <b>0.55</b> | -0.850     | 0.40        | 1.87        | <b>0.06</b> | -0.882         | 0.38        | -1.24               | 0.12        | moderate              |
| blackouts  | <b>0.62</b> | -0.412     | 0.68        | 1.21        | 0.23        | -0.417         | 0.68        | -0.382              | 0.38        | moderate              |
| book sales | <b>0.66</b> | -0.267     | 0.79        | 2.70        | <b>0.01</b> | 3.885          | <b>0.00</b> | -0.140              | 0.60        | moderate              |
| cities     | <b>0.76</b> | -0.090     | 0.93        | 3.65        | <b>0.00</b> | 0.204          | 0.84        | -0.123              | 0.62        | moderate              |
| fires      | 0.05        | -1.78      | <b>0.08</b> | 4.00        | <b>0.00</b> | -1.82          | <b>0.07</b> | -5.02               | <b>0.00</b> | with cut-off          |
| flares     | <b>1.00</b> | -0.803     | 0.42        | 13.7        | <b>0.00</b> | -0.546         | 0.59        | -4.52               | <b>0.00</b> | with cut-off          |
| HTTP       | 0.00        | 1.77       | <b>0.08</b> | 11.8        | <b>0.00</b> | 2.65           | <b>0.01</b> | 0.000               | 1.00        | none                  |
| quakes     | 0.00        | -7.14      | <b>0.00</b> | 11.6        | <b>0.00</b> | -7.09          | <b>0.00</b> | -24.4               | <b>0.00</b> | with cut-off          |
| religions  | <b>0.42</b> | -0.073     | 0.94        | 1.59        | 0.11        | 1.75           | <b>0.08</b> | -0.167              | 0.56        | moderate              |
| surnames   | <b>0.20</b> | -0.836     | 0.40        | 2.89        | <b>0.00</b> | -0.844         | 0.40        | -1.36               | <b>0.10</b> | with cut-off          |
| wars       | <b>0.20</b> | -0.737     | 0.46        | 3.68        | <b>0.00</b> | -0.767         | 0.44        | -0.847              | 0.19        | moderate              |
| wealth     | 0.00        | 0.249      | 0.80        | 6.20        | <b>0.00</b> | 8.05           | <b>0.00</b> | -0.142              | 0.59        | none                  |
| web hits   | 0.00        | -10.21     | <b>0.00</b> | 8.55        | <b>0.00</b> | 10.94          | <b>0.00</b> | -74.66              | <b>0.00</b> | with cut-off          |
| web links  | 0.00        | -2.24      | <b>0.03</b> | 25.3        | <b>0.00</b> | -1.08          | 0.28        | -21.2               | <b>0.00</b> | with cut-off          |

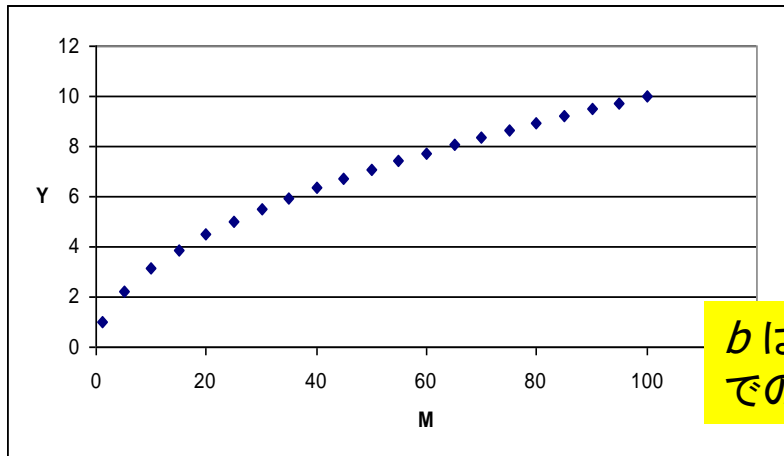
| data set  | $p$         | Poisson |             | log-normal |             | exponential |             | stretched exp. |             | power law + cut-off |             | support for power law |
|-----------|-------------|---------|-------------|------------|-------------|-------------|-------------|----------------|-------------|---------------------|-------------|-----------------------|
|           |             | LR      | $p$         | LR         | $p$         | LR          | $p$         | LR             | $p$         | LR                  | $p$         |                       |
| Internet  | <b>0.29</b> | 5.31    | <b>0.00</b> | -0.807     | 0.42        | 6.49        | <b>0.00</b> | 0.493          | 0.62        | -1.97               | <b>0.05</b> | with cut-off          |
| calls     | <b>0.63</b> | 17.9    | <b>0.00</b> | -2.03      | <b>0.04</b> | 35.0        | <b>0.00</b> | 14.3           | <b>0.00</b> | -30.2               | <b>0.00</b> | with cut-off          |
| citations | <b>0.20</b> | 6.54    | <b>0.00</b> | -0.141     | 0.89        | 5.91        | <b>0.00</b> | 1.72           | <b>0.09</b> | -0.007              | 0.91        | moderate              |
| email     | <b>0.16</b> | 4.65    | <b>0.00</b> | -1.10      | 0.27        | 0.639       | 0.52        | -1.13          | 0.26        | -1.89               | <b>0.05</b> | with cut-off          |
| metabolic | 0.00        | 3.53    | <b>0.00</b> | -1.05      | 0.29        | 5.59        | <b>0.00</b> | 3.66           | <b>0.00</b> | 0.000               | 1.00        | none                  |
| papers    | <b>0.90</b> | 5.71    | <b>0.00</b> | -0.091     | 0.93        | 3.08        | <b>0.00</b> | 0.709          | 0.48        | -0.016              | 0.86        | moderate              |
| proteins  | <b>0.31</b> | 3.05    | <b>0.00</b> | -0.456     | 0.65        | 2.21        | <b>0.03</b> | 0.055          | 0.96        | -0.414              | 0.36        | moderate              |
| species   | <b>0.10</b> | 5.04    | <b>0.00</b> | -1.63      | 0.10        | 2.39        | <b>0.02</b> | -1.59          | 0.11        | -3.80               | <b>0.01</b> | with cut-off          |
| terrorism | <b>0.68</b> | 1.81    | <b>0.07</b> | -0.278     | 0.78        | 2.457       | <b>0.01</b> | 0.772          | 0.44        | -0.077              | 0.70        | moderate              |
| words     | <b>0.49</b> | 4.43    | <b>0.00</b> | 0.395      | 0.69        | 9.09        | <b>0.00</b> | 4.13           | <b>0.00</b> | -0.899              | 0.18        | good                  |

# スケールフリー性

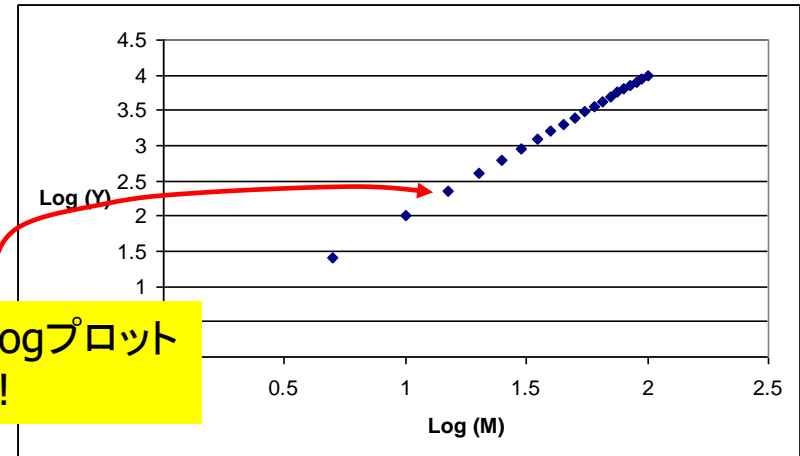
- Log-logプロットと通常のプロット

$$Y = Y_0 M^b$$

$$\text{Log } Y = \text{Log } Y_0 + b \text{ log } M$$



$b$  はlog-logプロットでの傾き!



# Benford's Law



- 膨大な数字が並んでいる冊子(例えば会計報告書)には、ランダムに数値が並んでいるように見えるが、意外なことに1ではじまる数が多い。
- **先頭の数字**がどのような確率で出現するかを考える。単純に各数字(0~9)の出現確率が同じと考えれば、同じ確率 $1/9$ で現れるはずである。
- 実際には**1から始まる数値が圧倒的に多く30%くらいもある**。逆に、**9から始まる数値は4.5%程度まで落ちる**



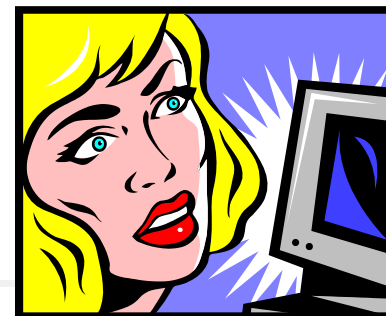
# ベンフォードの法則

---

先頭の数字が  $k$  である確率  $p(k) = \log_{10}\left(\frac{k+1}{k}\right)$

- 1から始まる確率  $\rightarrow \log_{10}(2) = 0.3010,$
- 2から始まる確率  $\rightarrow \log_{10}(3/2) = 0.1761,$
- .....,
- 9から始まる確率  $\rightarrow \log_{10}(10/9) = 0.0458$

# ベンフォードの法則



会計監査で事務処理の不正をこの法則を利用して見破ったというニュースもある

## Science

The New York Times  
ON THE WEB

Home Site Index Site Search Forums Archives Marketplace

August 4, 1998

## Following Benford's Law, or Looking Out for No. 1

By MALCOLM W. BROWNE

**D**r. Theodore P. Hill asks his mathematics students at the Georgia Institute of Technology to go home and either flip a coin 200 times and record the results, or merely pretend to flip a coin and fake 200 results. The following day he runs his eye over the homework data, and to the students' amazement, he easily finds nearly all those who faked their tosses.

"The truth is," he said in an interview, "most people don't know the real odds of such an exercise, so they can't fake data convincingly."

There is more to this than a classroom trick.

Dr. Hill is one of a growing number of statisticians, accountants and mathematicians who are convinced that an astonishing mathematical theorem known as **Benford's Law** is a powerful and relatively simple tool for pointing suspicion at frauds, embezzlers, tax evaders, sloppy accountants and even computer bugs.

The income tax agencies of several nations and several states, including California, are using detection software based on



### Dow Illustrates Benford's Law

To illustrate **Benford's Law**, D. J. Nigrini offered this example:

"If we think of the Dow Jones average as 1,000, our first digit will be 1."

# 数学でも法則を満たす

## ■ フィボナッチ数列

$$f_0 = 1,$$

$$f_1 = 1,$$

$$f_n = f_{n-1} + f_{n-2} (n > 1)$$

| 先頭の数字 | 頻度 (%) |
|-------|--------|
| 1     | 30     |
| 2     | 18     |
| 3     | 13     |
| 4     | 9      |
| 5     | 8      |
| 6     | 6      |
| 7     | 5      |
| 8     | 7      |
| 9     | 4      |

最初の100個について

このほかに2のべき乗 (N乗)  
もベンフォード則を満たす

再帰  
例)

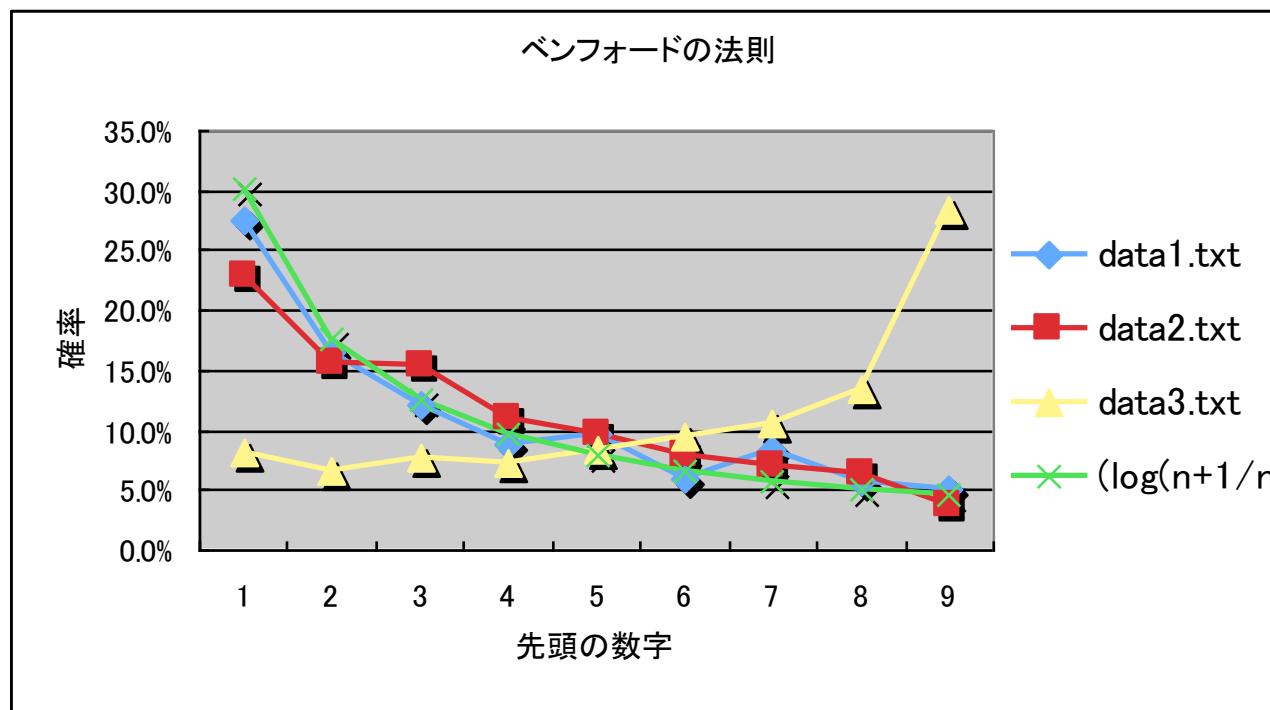
1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ....

# 東大の情報データ

**data1:** <http://www.u-tokyo.ac.jp/gen03/kouhou/1325/no1325.pdf>  
東京大学学内広報No.1325(学生生活実態調査の結果)より、表1-1から1-4のデータを抽出

**data2:** <http://www.u-tokyo.ac.jp/fin01/pdf/H17zaimusyohyou.pdf>  
平成17年度東京大学財務諸表より、P1からP7までを抽出

**data3:** 正弦の値を0度から180度まで、0.5度おきに出力したもの。sin.cにより作成。





# プロ野球のデータなど



\* 環境パフォーマンス・データ:第1次基本集計結果(01北海道)統計表  
第1表 口, 人口増減(平成12年~17年), 面積及び人口密度  
ダウンロード先:<http://www.tokyo-gas.co.jp/csr/report/other/data.pdf>

\*  $2^n$ の上位桁

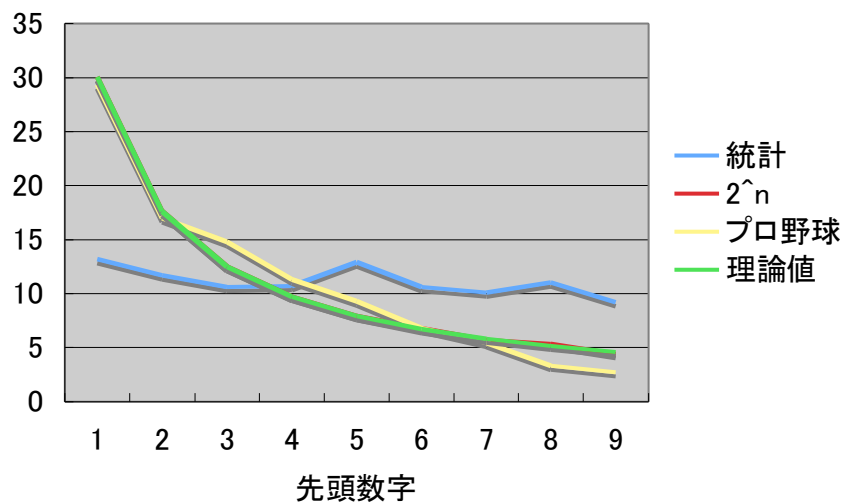
\* プロ野球選手のシーズン成績

プロ野球の2006年度シーズン成績(パリーグとセリーグの投手と打者の成績)で、投手は防御率順にソートして17or18位まで、打者は打率でソートして29位までの一覧の数値部分を用いた。

※ 参照先

セリーグ打者／セリーグ投手／パリーグ打者／パリーグ投手  
[http://baseball.yahoo.co.jp/npb/stats/stats\\_c\\_bat\\_avg.html](http://baseball.yahoo.co.jp/npb/stats/stats_c_bat_avg.html)  
[http://baseball.yahoo.co.jp/npb/stats/stats\\_c\\_pit\\_era.html](http://baseball.yahoo.co.jp/npb/stats/stats_c_pit_era.html)  
[http://baseball.yahoo.co.jp/npb/stats/stats\\_p\\_bat\\_avg.html](http://baseball.yahoo.co.jp/npb/stats/stats_p_bat_avg.html)  
[http://baseball.yahoo.co.jp/npb/stats/stats\\_p\\_pit\\_era.html](http://baseball.yahoo.co.jp/npb/stats/stats_p_pit_era.html)

ベンフォードの法則との比較

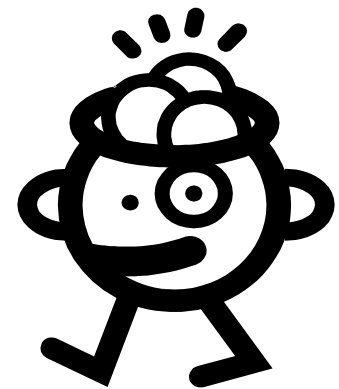
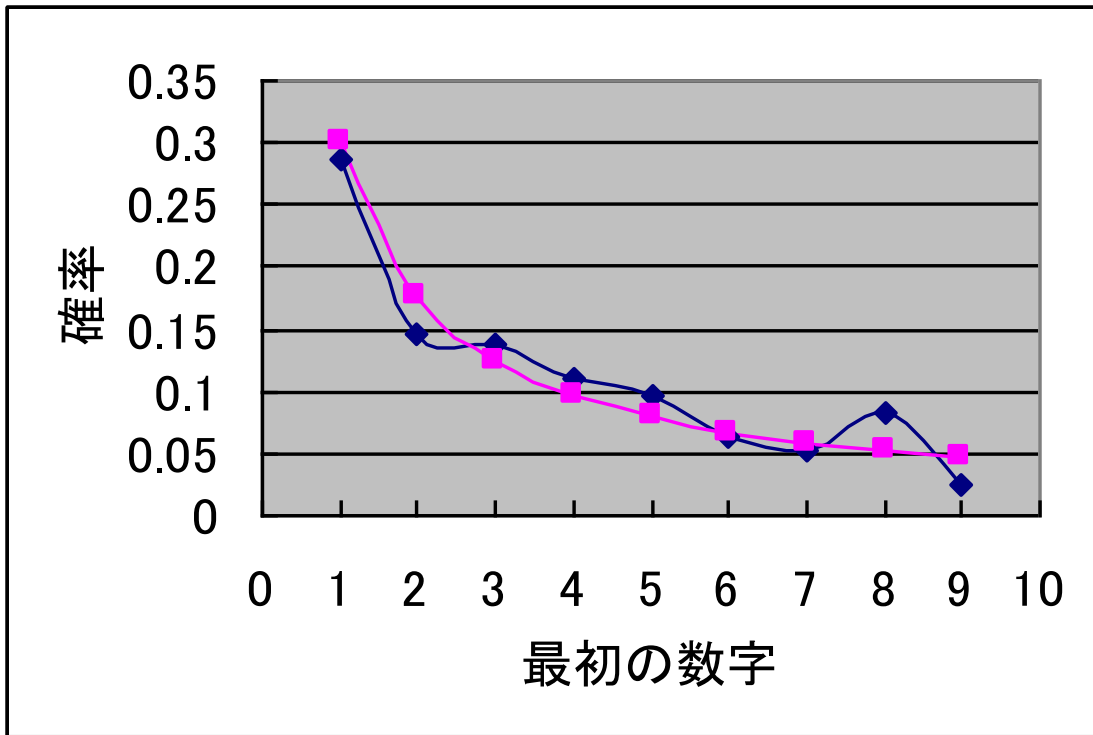


赤線で示した $2^n$ のグラフ  
はほとんど理論値と重なっ  
て見えない



# ビール暗号

<http://rogergrambihler.tripod.com/BealeHoax.htm>



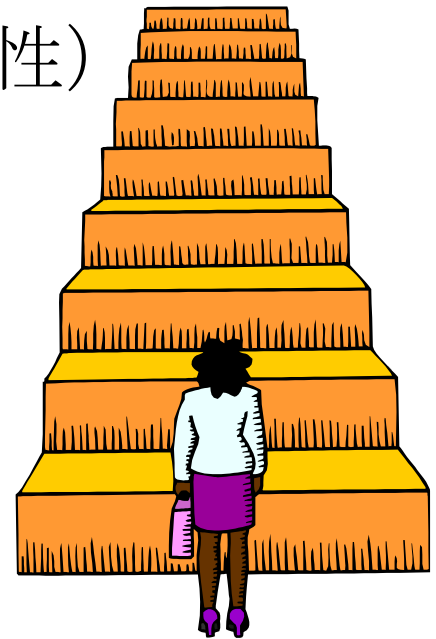
これは何を意味  
するのか？



# For challengers!!

---

- ベンフォードの法則の説明を考えよ。
  - もっともらしい説明があるか？
  - ヒントはスケールフリー(スケール不変性)



# 2のべき乗のベンフォード則

- ワイルの均等分布定理
  - 任意の無理数 $a$ について、  
数列 $\{na\} = na$ の小数部分  
は、1を法として均等に分布する。
  - 1を法として均等に分布 = 割合が部分区間の長さとなる
  - 例: 区間 $(0.6, 0.8)$ は区間 $(0, 1)$ の20%なので割合も20%となる



ヘルマン・ワイル(1885-1955) 20世紀において最も影響力のあるドイツの数学者。数論を含む純粋数学と理論物理学の双方の分野で顕著な業績を残した。



# 2のべき乗のベンフォード則

- 2のN乗の最初の桁がdとなる条件

$$d \times 10^n \leq 2^N < (d + 1) \times 10^n$$



$$\log_{10} d \leq \{N \log_{10} 2\} < \log_{10} (d + 1)$$

- $\{N \log_{10} 2\}$  は1を法として均等分布
- ワイルの均等分布定理から

$$P[\log_{10} d \leq \{N \log_{10} 2\} < \log_{10} (d + 1)]$$

$$= \log_{10} (d + 1) - \log_{10} d = \log_{10} \left(1 + \frac{1}{d}\right)$$

# 法則はどうして成り立つか？

- スケール不変性 (スケールフリー)

$$P(\alpha < X < \beta) = P(a \cdot \alpha < X < a \cdot \beta)$$

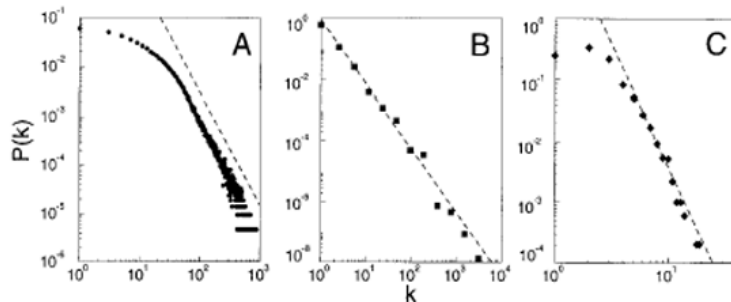


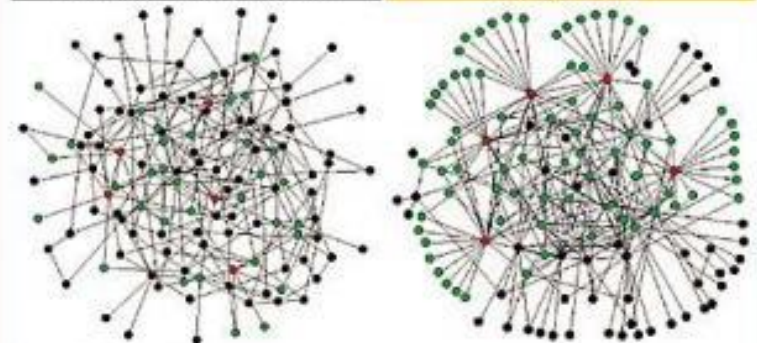
Fig. 1. The distribution function of connectivities for various large networks. (A) Actor collaboration graph with  $N = 212,250$  vertices and average connectivity  $\langle k \rangle = 28.78$ . (B) WWW,  $N = 325,729$ ,  $\langle k \rangle = 5.46$  (6). (C) Power grid data,  $N = 4941$ ,  $\langle k \rangle = 2.67$ . The dashed lines have slopes (A)  $\gamma_{\text{actor}} = 2.3$ , (B)  $\gamma_{\text{www}} = 2.1$  and (C)  $\gamma_{\text{power}} = 4$ .

## Comparing Random and Scale-Free Distribution

In the random network, the five nodes with the most links (in red) are connected to only 27% of all nodes (green). In the scale-free network, the five most connected nodes (red) are connected to 60% of all nodes (green).

RANDOM/EXPONENTIAL

SCALE-FREE



Source: the journal Nature



# スケールフリー性による証明

教科書 140ページ

べき乗の係数(スケール係数)が1のべき乗分布  $P(x)=1/x$  を考える。

この分布に従うデータについて、先頭の数字  $n$  が出現する確率は、

$$p(n) = \int_k^{k+1} P(x)dx = \int_k^{k+1} x^{-1}dx = \int_k^{k+1} d(\log x) = \log \left( \frac{n+1}{n} \right)$$

となる。



# パレートの法則

教科書 140ページ

## 定理 4.4 パレートの法則

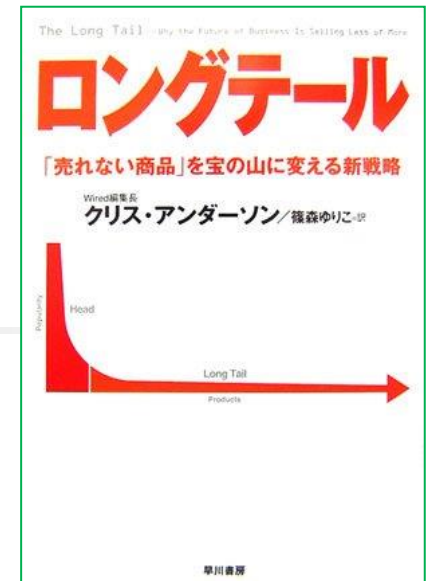
全体の数値の大部分は一部の要素が生み出している。とくに、20対80の法則が有名で、「売上の80%は全商品の20%が作る」「社会の富の80%は全体の20%の富裕層が独占している」などと解釈される。

**BUT 最近は違ってきた!!**



# 98%の法則

- インターネットが世界を変えた
- 販売機会の少ない商品でもアイテム数を幅広く取り揃えたり、顧客の総数を増やすこと(オタクをターゲットにすること)で、総体としての売上げを大きくできる。
- amazon, google, などの成功例
  - 検索エンジンの必要性
  - 在庫の保管費がゼロに近い



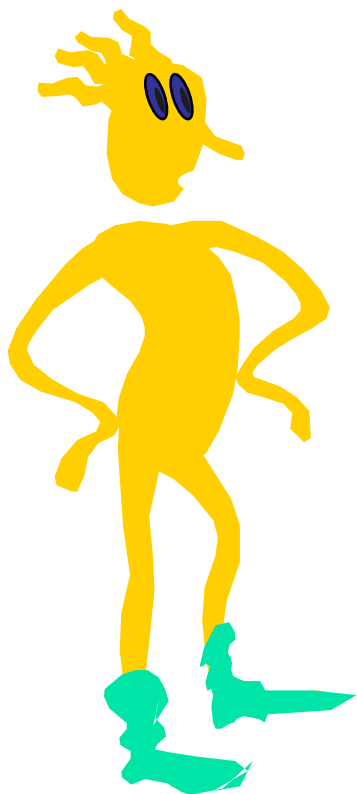
# チャイティンの $\Omega$

- $\Omega$  は“停止確率”のことである
- あるプログラムを表す列が得られるまで公正なコインを投げる。そしてそのプログラムを実行する。
- $\Omega$  はこのプログラムが停止する確率である。

$$\Omega = \sum_{\text{halting programs } p} 2^{-\text{length of } p}$$



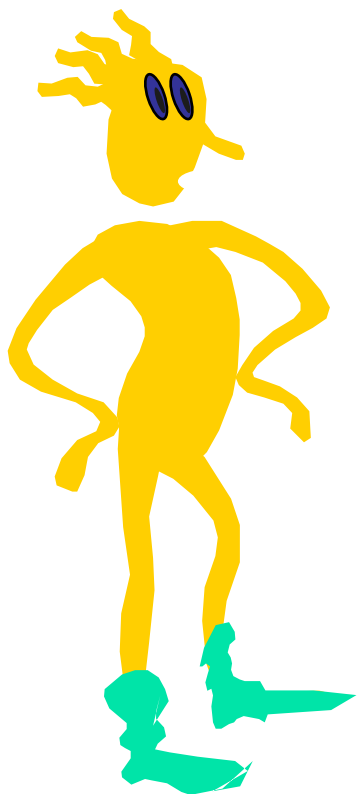
# チャイティンの $\Omega$



$\Omega$  は最大限に  
未知の数である



# チャイティンの $\Omega$



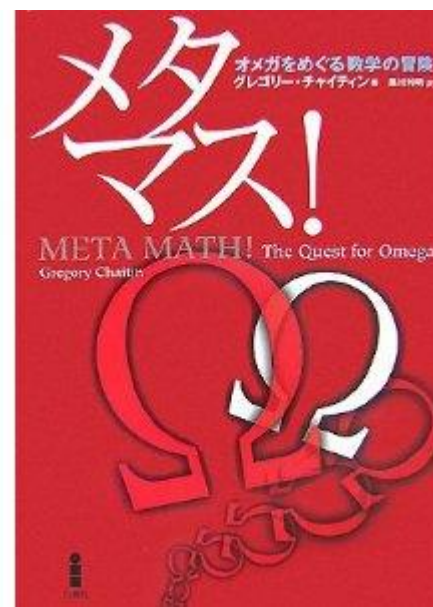
$\Omega$ は  
停止性のオラクル  
の最適な圧縮系で  
ある

# コンピュータ科学者にとっての $\Omega$



知の限界

グレゴリー チャイティン,  
エスアイビー・アクセス



メタマス!—オメガをめぐる数学  
の冒険 グレゴリー チャイティン,  
白揚社



# 複雑性の定義：非圧縮

- アルゴリズム的な情報量
- データの複雑性=そのデータを生成するのに必要なコンピュータプログラムのサイズ.
  - そのプログラムを記述するために必要なビット数(0, 1文字列のサイズ)の最小値.
- E.g. 1,2,3,4,5 ... → ほとんどアルゴリズム的な情報はない
  - とても短いプログラムでこれらの数を生成できる.
- E.g. 3.141592... → ほとんどアルゴリズム的な情報はない
  - 比較的短いプログラムで円周率は生成できる.



# 複雑性の定義：非圧縮

- 1.341285... 64
  - より大きなアルゴリズム的情報量を持つ
- なぜなら、この数は明確なパターンを持たないの。
- よって、出力するための最小のプログラムはその数自身とほぼ同じ長さとなるだろう。
- つまり、以下のようなプログラム：

```
Begin
Print "1.341285... 64"
End
```

# チャイティンの $\Omega$

- 0と1の間の実数値
- もしそれが計算できれば、停止問題を解決することができる。
- $\Omega$ は「ランダム」である。nビットの $\Omega$ を記述するにはnビットのプログラムを必要とする。 $\Omega$ には単純なパターンがない。

$$\Omega = \sum_{p \in P} 2^{-|p|}$$

- ただし  $P$  はすべてのチューリングマシンの集合であり、 $|p|$  は  $p$  をエンコードするビット長。
- $P_n$  を  $n$ ステップ以下で停止するすべてのチューリングマシンの集合とする。

$$\Omega_n = \sum_{p \in P_n} 2^{-|p|}$$

すると $\Omega_n$  は  $n$  が  $\infty$  に近づけば  $\Omega$  に近づく。





# Ω をどう定義するか？

- いま考えているコンピュータが停止するプログラムが3つだけであり、それらが0, 1文字列で110, 11100 と 11110だとしよう。これらのプログラムはそれぞれ3, 5 と 5のビットサイズである。
- もしも各ビットの0, 1をコインを投げてランダムにプログラムを選んでいるとすれば、偶然にこれらの停止するプログラムの1つを得る確率は正確に $1/2^3$ ,  $1/2^5$  および  $1/2^5$ となる。
- よって,  $\Omega = 1/2^3 + 1/2^5 + 1/2^5$   
 $= .001 + .00001 + .00001 = .00110$
- ただし、プログラム110は停止するので110で始まるどのようなプログラム(たとえば1100 や 1101)も考える必要がないことに注意



# チャイティンの $\Omega$

---

- $\Omega_n$  はどの精度でも計算できる.
- $\Omega$ を用いて停止問題を解決するには次のようにする:
  - もし  $M$  が  $n$ ステップで停止するなら“yes”と答える.
  - もし  $\Omega - \Omega_n < 2^{-|M|}$  なら“no”と答える.
  - このことを  $n = 1, 2, 3, \dots$  と答えが得られるまで繰り返す.
  - 最終的には“yes”または“no”の答えが必ず得られることになる.
- したがって  $\Omega$  は決して知られない. お告げ以外には知ることはできない.



# チャイティンの $\Omega$

---

- $\Omega$  の最初の  $n$  ビットは、bits of 長さが  $n$  で制限されるすべてのプログラムに対して停止問題を解決するのに十分な情報を与える
  - $L := 0$
  - 各プログラムを時分割する
  - 長さ  $k$  のプログラムが停止するとき,  $2^{-k}$  を  $L$  に加える.
  - $L$  の最初の  $n$  ビットは  $\Omega$  の最初の  $n$  ビットに等しい
  - 長さが  $n$  以下のプログラムで停止するはずのものは停止しているだろう.



# $\Omega$ は圧縮可能でない

---

- もしも $\Omega$ の最初の $N$ ビットを、 $N$ ビットより短いプログラムで計算できると仮定しよう.
- すると $\Omega$ の最初の $N$ ビットを用いて、サイズが $N$ ビット以下のすべてのプログラムに対して停止問題を解くことができる.
- 一方、長さ $N$ ビットまでのプログラムの停止問題は $N$ ビットより短いプログラムでは解けないことが知られている.
- したがって矛盾となる.
- よって、 $\Omega$ は圧縮可能でないことがわかる.



# 未解決問題の解決

---

- 数学上の未解決問題は、チャイティンの定数を使って証明(または反証)できる。
  - リーマン予想、ゴールドバッハ予想、。。。。
- アイディア
  - 予想の反例を見つけ続けるプログラムを考える。反例が見つかった場合、プログラムは停止し、数学的予想は間違いだったことが示される。
  - このプログラムの長さを  $N$  ビットとする。
  - チャイティンの定数の先頭から  $N + 1$  ビットまでの値がわかるとする。
- どうやって証明するのか？



# 意識のパラドックス

---

- $H$ をチューリングマシンのループ<sup>o</sup>についての人間の思考をエンコードするチューリングマシンとしよう.
- 十分な時間があれば人間が $T_i$ が入力 $i$ でループすると数学的に演繹するならば $H$ は入力 $i$ で停止し、さもなければ $H$ はループすると仮定する
- すると、 $H$ は部分的なループテスターとなる。よって $H$ は入力 $j$ でループする。ただし $H$ は $T_j$ である。



# 意識のパラドックス

---

- Hは入力jに対してループするので、人間は $T_j$ が入力jに対してループすると演繹することができない。
- $T_j$  はHである。したがって、人間はHが入力jでループすると演繹できない。
- しかしわれわれはHが入力jでループすることを示した。
- われわれは人間だ(マトリックスでなければ)。
- したがって、矛盾する。

# 意識

われわれは自分がどのように思考しているかを決して知ることができなくなる

われわれは数学的に推論しているときでさえ正確に思考していないことになる

- 可能性:
  - H は部分的なループテスターである; 人間の論理は欠陥がある.
  - 人間は決してHを知ることができない.
  - 人間の思考はチューリングマシンでは表現できない.
- このずれの可能性も問題である

われわれの思考はチューリングマシンの容量を超えているところになる