数理解析研究所講究録
第1710巻 2010年 142-153



# Benford's Law and Distribution Functions

Vladimir Baláž スロヴァキア工科大
長坂建二 法政大理工
Oto Strauch スロヴァキアアカデミー

## 1. 一様分布

この小論は，数論の分野の一つである一様分布論に分類されるので，一様分布の基本事項を復習する。

数列は，$\{x_n\}$ や $(x_n)_{n=1}^{\infty}$ のように表現されることが多いが，一般項（第 $n$ 項）を指定することにより特定することができる。したがって，数列を $x_n$, $n=1,2,\cdots$，または単に $x_n$ と書いても，それが数列全体であるか，数列の一般項を指しているのかは，理解されることであろう。ここでは，数列をこのように略記することを，まず断っておく。

一様分布論で取り扱うのは，無限個の項を持つ実数列 $x_n$, $n=1,2,\cdots$ である。ここで，数列 $x_n$ の小数部分の数列を，$x_n \bmod 1$ で表す。つまり，数列 $x_n \bmod 1$ の一般項 $= y_n$, $n=1,2,\cdots$ は

$$y_n = x_n - [x_n] = x_n - \lfloor x_n \rfloor = \{x_n\}$$

であり，$[\cdot]$ はガウス記号，$\lfloor\ \rfloor$ は床関数，$\{\cdot\}$ は小数部分である。

さて，$\theta$ を無理数（$\theta \notin \mathbb{Q}$）とするとき，数列 $n\theta$, $n=1,2,\cdots$ の小数部分の数列 $n\theta \bmod 1$ は，右半開区間 $[0,1)$ 上でどのように分布しているのであろうか。$\theta$ が有理数ならば $n\theta \bmod 1$ はいずれ周期的となるが，無理数のときには $[0,1)$ 上稠密（dense）であることが Sierpinsky により見出

された。しかし，もっと強い性質をもつことが Hermann Weyl により示された。彼の論文

[1] Weyl, H.: Über die Gleichverteilung von Zahlen mod Eins, Math. Ann. 77, 313-352 (1916)

ことが一様分布の出発点である。

実数列 $x_n$ $n=1,2,\cdots$ に対して，その小数部分の数列 $x_n \bmod 1$ が任意の右半開区間 $[a,b)$，$0 \leq a < b < 1$ に対して

$$\lim_{N\to\infty} \frac{\#\{i \leq N;\ x_i \bmod 1 \in [a,b)\}}{N} = b-a$$

が成り立つとき，$x_n$ $n=1,2,\cdots$ は mod 1 で一様分布するという。略して，u.d. mod 1 という。

この概念を用いれば，$\theta \notin \mathbb{Q}$ ならば $n\theta$ $n=1,2,\cdots$ は mod 1 で一様分布する（u.d. mod 1）ということになる。

実際，mod 1 で一様分布するような実数列は沢山あり，十分研究対象となることが，次の判定条件（criterion）からわかる。

・[判定条件] 数列 $x_n$ $n=1,2,\cdots$ が mod 1 で一様分布するための必要十分条件は，任意の周期 1 の連続関数 $f(x)$ に対して

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(x_i \bmod 1) = \int_0^1 f(x)\,dx$$

が成り立つことである。もちろん，$f(x_i \bmod 1) = f(\{x_i\})$ の意味である。

[Weyl 判定条件] 数列 $x_n$ $n=1,2,\cdots$ が mod 1 で一様分布するための必要十分条件は，0 でないすべての整数 $h$ に対して

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} e^{2\pi i h x_i} = 0$$

が成り立つことである。

これは、判定条件の周期1の連続関数として $e^{2\pi i h x}$ を取ったと考えればすぐにわかるであろう。

一様分布論は Weyl の論文 [1] に始まったが、その後

[2] Kuipers, L. & Niederreiter, H.: Uniform Distribution of Sequences, John Wiley & Sons, 1974

はこの分野の The book とも言える大著である。絶版となっていたが、reprint 版が Dover 社から 2006 年に出版されたので、手許に置いておけば便利であろう。

## 2. ベンフォードの法則

フランク・ベンフォードは天文学者であったが、多くの数表等を調べると、最初の数字の分布が一様ではなく、1が最も多く9に向けて次第に減少していくことを見出した。これを

[3] Benford, F.: The law of anomalous numbers, Proc. Amer. Phil. Soc. 78 (1938), 551-572

に発表した。諸君も身の回りの数字について最初の(有効)数字の分布がこの法則に従うことをチェックすることをお勧めする。たとえば、Abramowitz & Stegun の数表中の定数を抜き出して見ると、ベンフォードの法則が成り立つことがわかる。

ベンフォードの法則は、1881年に既に発見されて Simon Newcomb により報告されている。また、特定の数列がベンフォードの法則を満たしている報告は、Fibonacci Quarterly などには繰り返し掲載されている。

ベンフォードの法則の総合報告は

[4] Raimi, R. A.: The first digit problem, Amer. Math. Monthly 83 (1976), 521-538

がそれ以前の結果をほとんど含んでいて便利で読み易い。

数学的にベンフォードの法則はもう少し一般的に定義することができる。

$b \geq 2$ を実数表現の基底とする。通常は $b=10$ であり，実数は10進法で表現されている。$b$ はもちろん整数（自然数）に限定しておく。次に，正数 $x>0$ に対して，$x$ のマンティッサ $M_b(x)$ を $x = M_b(x) \times b^{n(x)}$ で定める。ここで，$1 \leq M_b(x) < b$ を満たすとしておくと，$n(x)$ は一意的に決る。たとえば，$125 = 1.25 \times 10^2$，$0.003 = 3 \times 10^{-3}$ であり，125の基底10に対するマンティッサ $M_{10}(125) = 1.25$ となる。コンピュータ科学の用語では，浮動小数点による表示で用いられていると考えればよい。ここで，正の実数に限定しているが，負の場合にも $|x| = M_b(x) \times b^{n(x)}$ とすれば $x<0$ のマンティッサが定められるのは明らかであるが，ここでは正の場合に限定しておく。

さて，正の実数列 $x_n$ $n=1,2,\cdots$ が基底 $b$ についてオーダー $r$ のベンフォードの法則を満たすことは次のように定義される。$r$ は自然数とし，$K = k_1 k_2 \cdots k_r$ を $b$ 進法により表される自然数とする。つまり

$$K = k_1 \times b^{r-1} + k_2 \times b^{r-2} + \cdots + k_{r-1} \times b + k_r$$

であり，$k_1 \neq 0$，$0 \leq k_i < b$，$k_i \in \mathbb{Z}$ を満たしている。

また，$K = k_1 k_2 \cdots k_r$ は，$r$ 個のブロック（block, string）とも考えることにする。この約束の下で，すべての $r$-digits の整数 $K = k_1 k_2 \cdots k_r$ に対して

$$\lim_{N\to\infty} \frac{\#\{ i \leq N ;\ M_b(x_i) \text{ の最初の } r\text{-digits が } K \text{ に等しい} \}}{N}$$

$$= \log_b (K+1) - \log_b K$$

が成り立つとき，$x_n$ $n=1,2,\cdots$ は基底 $b$ についてオーダー $r$ のベンフォードの法則を満たすという。

$b=10$，$r=1$ の場合が元々のベンフォードの法則である。

このとき

$$\lim_{N\to\infty} \frac{\#\{ i \leq N \ ;\ M_b(x_i) \text{の最初の数字が} k \}}{N}$$

$$= \log_b (k+1) - \log_b k$$

となり，$b=10$ ならば最初の(有効)数字が $k$ $(k=1, 2, \cdots, 9)$ である確率(相対頻度の極限)が $\log_{10}(k+1) - \log_{10} k$ となる。$k=1$ ならば $\log_{10} 2 = 0.3010\cdots$，$k=2$ ならば $\log_{10} 3 - \log_{10} 2 < \log_{10} 2$，$\cdots$ となり，最初の数字が $1, 2, 3, \cdots$ となる確率が次第に減少していくことになる。

$r=1$ の場合に，$x>0$ の最初の(有効)数字が $k$ であるような条件を考えてみると

$$k \leq M_b(x) < k+1$$

となる。$M_b(x)$ を使わなければ，$x = M_b(x) \times b^{n(x)}$ であるから

$$k \times b^{n(x)} \leq x < (k+1) b^{n(x)}$$

とも表されることになる。$k = 1, 2, \cdots, b-1$ であった。そこで上の不等式で $b$ を底とする対数を取ると

$$\log_b k + n(x) \leq \log_b x < \log_b (k+1) + n(x)$$

となる。$n(x) \in \mathbb{Z}$ であるから，上の不等式を $\bmod 1$ で考えれば

$$\log_b k \leq \log_b x \bmod 1 < \log_b (k+1)$$

となる。

これから，数列 $\log_b x_n$ $n=1, 2, \cdots$ が $\bmod 1$ で一様分布すれば，$x_n$ $n=1, 2, \cdots$ は(基底 $b$ において)(オーダー1で)ベンフォードの法則に従うことがわかる。

Persi Diaconis は，ベンフォードの法則よりも強い概念を

[5] Diaconis, P. : The distribution of leading digits and uniform distribution mod 1, Ann. Prob. 5 (1977)

72–81 において導入した。つまり，数列 $x_n$ $n=1,2,\cdots$ が強い意味でベンフォードの法則に従うことを，すべてのオーダー $r=1,2,\cdots$ に対して $x_n$ $n=1,2,\cdots$ が（基底 b について）オーダー r に対してベンフォードの法則を満たしていると定義する。そして，$x_n$ $n=1,2,\cdots$ $(x_n>0)$ が強い意味でベンフォードの法則に従うための必要十分条件が，$\log_b x_n$ が mod 1 で一様分布することであることを示した。

この Diaconis の定理（強い意味でのベンフォードの法則の判定条件）と，一様分布論の結果を適用することにより，（強い意味での）ベンフォードの法則に関する結果を得ることができる。Nagasaka - Kanemitsu - Shiue - Rauzy，Nagasaka - Shiue，Nagasaka，—— の結果はそうである。（Google 等で検索すれば簡単に入手可能であろう。）

## 3. Distribution Functions of Sequences

分布関数は，確率変数 $X$ に対して

$$F(x)=P(X\le x)$$

のように定められる。$-\infty<x<\infty$ であり，分布関数 $F(x)$ は $F(-\infty)=0$，$F(+\infty)=1$，$x\le y\Rightarrow F(x)\le F(y)$，左連続，不連続点は高々可算個である等の性質をもっている。

一方，データ $x_1,x_2,\cdots,x_N$ に対しては，経験的分布関数を

$$F_N(x)=\frac{1}{N}\sum_{i=1}^{N}\#\{i\le N;\ x_i\le x\}$$

のように定められる。$x_1,x_2,\cdots,x_N$ が分布関数 $F(x)$ をもつ母集団からの標本とすれば，経験的分布関数 $F_N(x)$ は $N\to\infty$ のときに $F(x)$ に収束することが期待されることになる。詳しくいえば，もっと厳密な議論が可能であるが，ここでは経験的分布関数のアイディアを基にして，数列 $x_n$ $n=1,2,\cdots$ の階段分布関数 $F_N(x)$ を

$$F_N(x) = \frac{\#\{i \leq N\ ;\ x_i \bmod 1 \in [0, x)\}}{N}$$

として，$x_1, x_2, \cdots, x_N$ に対して定義する。$x_n \bmod 1$ を対象として定義されているので，$0 \leq x \leq 1$ に対して $F_N(x)$ が定められ，$F_N(0)=0$，$F_N(1)=1$，単調増加の階段関数となる。

次に，自然数の単調増加列 $N_k$，$k=1, 2, \cdots$ が存在して

$$\lim_{k \to \infty} F_{N_k}(x) = g(x)$$

がすべての $x \in [0, 1)$ に対して成り立つとき，$g(x)$ を数列 $x_n$ $n=1, 2, \cdots$ の分布関数（d.f.）という。ここで，上の極限をすべての $g(x)$ の連続点 $\in [0, 1]$ に対して成り立つという形の d.f. の定義もあるが，Kuipers-Niederreiter [1] や
[6] Strauch, O. & Porubský, Š. : Distribution of Sequences : A Sampler, Peter Lang, 2005
を参照されたい。

数列 $x_n$ $n=1, 2, \cdots$ に対して，その分布関数 d.f. は一意的とは限らないので，$x_n \bmod 1$ の数列に対するすべての分布関数の集合を $G(x_n \bmod 1)$ と書くことにする。

$g(x) \in G(x_n \bmod 1)$ のとき

$$\lim_{k \to \infty} \frac{1}{N_k} \sum_{i=1}^{N_k} f(\{x_i\}) = \int_0^1 f(x)\, dg(x)$$

がすべての連続関数 $f : [0, 1] \to \mathbb{R}$ に対して成り立つことになり，1. 一様分布における判定条件を一般化したものである。

分布関数の集合 $G(x_n \bmod 1)$ に対しては，次のような性質が成り立つ。（[6] 参照）

- $G(x_n \bmod 1)$ は空集合ではなく，唯一つの元をもつか，無限個の元をもつ。（2個以上の有限個の元をもつことはない。）

//

- $G(x_n \bmod 1)$ は弱収束による位相に関して閉集合，かつ連結である。
- 空でない分布関数の集合 $H$ に対して，ある数列 $x_n$ $n=1,2,\cdots$ ($x_n \in [0,1)$) が存在して，$G(x_n) = H$ となるのは $H$ が閉かつ連結。

ここで，d.f. を用いると，数列 $x_n$ $n=1,2,\cdots$ が mod 1 で一様分布するための必要十分条件は $G(x_n \bmod 1) = \{g(x)\}$，つまり，唯一つの元 $g(x)$ から $G(x_n \bmod 1)$ は成っており，さらに，$g(x) = x$ がすべての $x \in [0,1]$ に対して成り立つことである。

さて，正の数列 $x_n$ $n=1,2,\cdots$ が（基底 $b$ について）オーダー $r$ に対してベンフォードの法則を満たすための条件を分布関数の形で考えよう。

$x>0$ の最初の $k$ 個の block が $K = k_1k_2\cdots k_r$ に一致するための条件は

$$K \leq M_b(x) \times b^{r-1} < K+1$$

であり，対数（底 $b$）を取り，mod 1 で考えると

$$\log_b\left(\frac{K}{b^{r-1}}\right) \leq \log_b x \bmod 1 < \log_b\left(\frac{K+1}{b^{r-1}}\right)$$

と同値になる。したがって，数列 $x_n$ $n=1,2,\cdots$ の最初の $r$ 個の digits が（基底 $b$ として）$K = k_1k_2\cdots k_r$ 以下の digits と一致するのは，任意の $g \in G(x_n \bmod 1)$ に対して

$$g\left(\log_b\left(\frac{K}{b^{r-1}}\right)\right) = \log_b\left(\frac{K}{b^{r-1}}\right)$$

が成り立つことになる。つまり，d.f. $g(x)$ に対して $g(x) = x$ が $\log_b\left(\frac{K}{b^{r-1}}\right)$ に対して成り立っていることである。

## 4. 数列の分布関数の効用

上述のオーダー $r$ のベンフォードの法則の判定条件は，定義を分布関数を用いて書き直しただけであるが，次の結果を

を導くことができる。

[定理A] すべての $r=1,2,\cdots$ に対して，基底 $b$ について
オーダー $r$ でベンフォードの法則を満たすがオーダー $r+1$ で
はベンフォードの法則を満たさないような数列 $x_n\ n=1,2,\cdots$
が存在する。

<証明> $A_r=\{\log_b(k_1.k_2\cdots k_r);\ k_i\in\{0,1,\cdots,b-1\},\ k_1\neq 0\}$
と定義する。$K=k_1k_2\cdots k_r$ とすると，$k_1.k_2\cdots k_r=K/b^{r-1}$
になっている。ここで，分布関数 $g(x)$ を $A_r$ 上では $g(x)=x$
であり，ある $x\in A_{r+1}$ では $g(x)\neq x$ となるようにとる。
分布関数の一般論から，すべての分布関数 $g(x)$ に
対して，$g(x)$ を漸近的分布関数にもつような数列 $y_n$
が存在する。ここで，数列 $y_n$ の漸近的分布関数(a.d.f)
とは，自然数の部分列ではなく

$$g(x)=\lim_{N\to\infty}F_N(y_n \bmod 1)$$

である。今，$x_n=b^{y_n}$ とおく。つまり $\log_b x_n=y_n$ である。
すると，前節の判定条件から，オーダー $r$ に対しては

$$g\left(\log_b\left(\frac{K}{b^{r-1}}\right)\right)=\log_b\left(\frac{K}{b^{r-1}}\right)$$

が成り立っているが，$r+1$ の場合には $g(x)\neq x$ が $A_{r+1}$ 上の
どこかで成り立っている。つまり，$x_n \bmod 1$ はオーダー $r+1$
ではベンフォードの法則を満たさないことがわかる。□

この定理Aは，分布関数を用いて示したが，コンビナトリアル
に考えても簡単にわかる。一様分布論より，強い意味で
ベンフォードの法則を満たす数列 $x_n\ n=1,2,\cdots$ が存在
することを示すのは容易である。この数列 $x_n$ に対して，
最初の $r+1$ 個のブロック $k_1k_2\cdots k_{r+1}$ で $\underset{r+1\text{番目}}{~~k_{r+1}~~}$ を変えてやると，
この変更により

$(1/N)\times\underbrace{\#\{i\leq N;\ 最初の\ r+1\ 個の\ \mathrm{digit}\ が\ k_1k_2\cdots k_{r+1}\ に一致\}}_{\cancel{\#\#}}$

$$\longrightarrow \log_b\left(\frac{K+1}{b^r}\right) - \log_b\left(\frac{K}{b^r}\right) \quad (N \to \infty)$$

$K = k_1 k_2 \cdots k_{r+1}$

となるから，オーダー$r$でベンフォードの法則を満たすが，$r+1$ではベンフォードの法則を満たさない数列が構成できる。説明がわかりにくいかも知れないが，その数列の最初の$r+1$桁のブロックを考えるときに，$r+1$個のdigit（数字）をたとえば1にすべて変えてしまうとしよう。すると，最初の$r+1$個のブロックが$k_1 k_2 \cdots k_r 2$に等しいような項（変更後）は無いので，そのような項数の相対頻度の極限は0となりオーダー$r+1$でベンフォードの法則に従わないことになる。最初の$r$個のブロックについてはいじらないので

オーダー$r$でベンフォードの法則は明らかに成り立っている。

今までは，実数を表現する基底$b$については，固定して扱ってきた。たとえば，$b=2$（2進法）の場合を考えてみよう。$10=(1010)_2$であるが，これを$4=2^2$進法で考えると，$10=(22)_4$となる。つまり，2つの異る基底$b_1$，$b_2$で$\log b_1 / \log b_2 \in \mathbb{Q}$，つまり$b_1{}^r = b_2{}^s$となるような自然数$r, s$が存在する場合には，$b_1$進法の表現を$b_2$進法の表現に簡単に置き換えることができる。これから，何らかの結果が得られるかを試みてみた。オーダー$r$と基底$b$，$b^r$については関係があることが（更に条件を付加して）得られるが，あまり面白い結果ではない。これらの結果を含む論文は分布関数を用いない形でまとめて投稿予定である。

ここでは，強い意味でのベンフォードの法則と基底についての関係を分布関数を用いて議論しよう。

［定理B］ $x_n \in (0,1)$の数列，$G(x_n)$は$x_n$ $n=1,2,\cdots$の分布関数の集合とし，すべてのd.f. $g(x) \in G(x_n)$は$x=0$で

で連続であるとする。このとき，数列 $x_n$ $n=1,2,\cdots$ が基底 $b$ について強い意味でベンフォードの法則を満たすための必要十分条件は，すべての d.f. $g(x)\in G(x_n)$，すべての $x\in[0,1)$ に対して

$$x=\sum_{i=0}^{\infty}\left(g\left(\frac{1}{b^i}\right)-g\left(\frac{1}{b^{i+x}}\right)\right)$$

が成り立つことである。

この定理Bの上の（関数）等式を満たすような分布関数 $g(x)$ の例としては，

$$g(x)=\begin{cases} x & ,\ x\in[0,1/b] \\ 1+\log_b(x)+(1-x)\dfrac{1}{b-1} & ,\ x\in[1/b,1] \end{cases}$$

がある。$x=0$ で連続，$g(0)=0$，$g(1)=1$，単調増加になっているのは明らかである。

ここで，$(0,1)$ の数列 $x_n$ $n=1,2,\cdots$ の漸近的分布関数 a.d.f. として上の $g(x)$ をもつような数列とする。$b_1<b_2$ を基底とすると，$x_n$ $n=1,2,\cdots$ が基底 $b_1$ についても，基底 $b_2$ についても強い意味でのベンフォードの法則を満たすとすると，必然的に

$$\frac{\log b_1}{\log b_2}=\frac{(b_1-1)\,b_2}{(b_2-1)\,b_1}$$

でなければならないという結果が導びかれる。

一様分布論と組み合せ論（初等的 elementary な議論では，$\log b_1/\log b_2\in\mathbb{Q}$ が得られるが，分布関数の議論を用いると，具体的な d.f. の形からより精密な結果が得られることを意味している。

また，$b_1^k=b_2$ $(k>0)$ に対して考えると一見基底 $b_1$ と基底 $b_2$ についてのベンフォードの法則は同様に成り立ちそうな気がするが，上の $g(x)$ の性質より，

$$b_1^k=1+k(b_1-1)b_1^{k-1}$$

が成り立たなければならなくなり，これは不可能，つまり基底 $b_1$，

および基底 $b_1^k$ に対して強い意味でベンフォードの法則が成り立つことはないことがわかる。

また，基底 $b_1$ と基底 $b_2$ の両方に対してベンフォードの法則を満たさないような数列 $x_n$ を次のような形で存在することを示すことができる。

[定理C] すべての基底 $b_1$ に対して，基底 $b_1$ に対して強い意味でのベンフォード~~の法則~~の法則を満たしているが，どんな $b_2 \geq 2$, $b_2^j \neq b_1$, $j=1,2,\cdots$ についても，基底 $b_2$ に対して強い意味でのベンフォードの法則を満たさない数列 $x_n$ $n=1,2,\cdots$ ($x_n \in [0,1)$) が存在する。

この証明には，前述の $g(x)$ を用いている。

以上のように，異なる基底に対する強い意味でのベンフォードの法則を満たす数列 $x_n$ $n=1,2,\cdots$，($x_n \in [0,1)$) について，分布関数を用いて細かな議論が可能であることを感じてもらえたら幸である。

ベンフォードの法則，オーダー，異なる基底については，他にもいろいろと性質がわかっているが，いずれまとめて投稿予定である。

また，本稿の Full Paper は Karatsuba 記念号(?)に投稿中であり，興味ある方々はそちらを参照されたい。

(大洲中央病院の病室にて脱稿)