因数pの出現頻度について

久米(くめ)　秀夫(ひでお)

§1. はじめに

10年ほど前にコラッツ予想(＊1)について考えているときに，自然数の中に2および3の因数がどれくらい分布しているのかを知る必要があったので，そのとき考えたことについて報告します。直感的には明らかな内容なのですが，証明してみました。またその結果がジップの法則(＊2)そのものであったので大変興味を引かれました。

§2. 自然数における因数pの出現頻度について

nを自然数とする。1からnまでの各自然数に含まれる因数$p(p \geqq 2)$の個数の総和を記号$\pi_p(n)$で表すことにする。つまり$n!$に含まれる因数pの個数が$\pi_p(n)$である。例えば，$\pi_2(10)$は2，2^2，$2 \cdot 3$，2^3，$2 \cdot 5$だから $\pi_2(10)=8$ であり，$\pi_3(10)=4$ である。

また，$\dfrac{\pi_p(n)}{n}$を因数pの1からnまでの「出現頻度」と呼ぶことにすると次の等式が成り立つ。

> 出現頻度について
>
> $$\lim_{n \to \infty} \frac{\pi_p(n)}{n} = \frac{1}{p-1}$$

(証明)　1からnまでの正の整数に含まれる因数pの1次の因数の総数は$\left[\dfrac{n}{p}\right]$個($[x]$はガウス記号で$x$を超えない最大の整数を表す。)

2次の因数の総数は$\left[\dfrac{n}{p^2}\right]$個，3次の因数の総数は$\left[\dfrac{n}{p^3}\right]$個，…であるから $n \geqq p^k$ を満たすkの最大値をmとすると

$$\pi_p(n) = \left[\frac{n}{p}\right] + \left[\frac{n}{p^2}\right] + \left[\frac{n}{p^3}\right] + \cdots + \left[\frac{n}{p^m}\right] = \sum_{i=1}^{m} \left[\frac{n}{p^i}\right]$$

また，$\left[\dfrac{n}{p^i}\right] \leqq \dfrac{n}{p^i} < \left[\dfrac{n}{p^i}\right] + 1$ であるから

$$\frac{n}{p^i} - 1 < \left[\frac{n}{p^i}\right] \leqq \frac{n}{p^i}$$

したがって　$\dfrac{1}{n}\displaystyle\sum_{i=1}^{m}\left(\frac{n}{p^i}-1\right) < \frac{1}{n}\sum_{i=1}^{m}\left[\frac{n}{p^i}\right] \leqq \frac{1}{n}\sum_{i=1}^{m}\frac{n}{p^i}$

$$\sum_{i=1}^{m}\left(\frac{1}{p^i}-\frac{1}{n}\right) < \frac{1}{n}\sum_{i=1}^{m}\left[\frac{n}{p^i}\right] \leqq \sum_{i=1}^{m}\frac{1}{p^i}$$

$$\frac{\frac{1}{p}\left(1-\left(\frac{1}{p}\right)^m\right)}{1-\frac{1}{p}} - \frac{m}{n} < \frac{1}{n}\sum_{i=1}^{m}\left[\frac{n}{p^i}\right] \leqq \frac{\frac{1}{p}\left(1-\left(\frac{1}{p}\right)^m\right)}{1-\frac{1}{p}}$$

$$\frac{1-\left(\frac{1}{p}\right)^m}{p-1} - \frac{m}{n} < \frac{1}{n}\sum_{i=1}^{m}\left[\frac{n}{p^i}\right] \leqq \frac{1-\left(\frac{1}{p}\right)^m}{p-1} \quad \cdots\cdots①$$

ここで，$n \geqq p^m$ であるから $\dfrac{m}{n} \leqq \dfrac{m}{p^m}$

また，$n \to \infty$ のとき $m \to \infty$ であるから

$n \to \infty$ のとき $\dfrac{m}{p^m} \to 0$

すなわち　$n \to \infty$ のとき $\dfrac{m}{n} \to 0$

したがって，①で $n \to \infty$ とすると両端の式はともに $\dfrac{1}{p-1}$ に収束し

$$\lim_{n \to \infty} \frac{1}{n}\sum_{i=1}^{m}\left[\frac{n}{p^i}\right] = \frac{1}{p-1}$$

すなわち　$\displaystyle\lim_{n \to \infty}\frac{\pi_p(n)}{n} = \frac{1}{p-1}$ が成り立つ。

(証明終)　■

なお，$p=2$ のとき $\displaystyle\lim_{n \to \infty}\frac{\pi_2(n)}{n}=1$ であるから

$\pi_2(n) \fallingdotseq n$

すなわち十分大きいnに対して，nまでの正の整数に含まれる2の因数の個数はほぼnに等しい。

また，同様に3の因数の個数はほぼ$\dfrac{n}{2}$に等しいことなどがわかります。

因みに，この問題を改めて考えていた年の京大の入試文系5(2009年度)に次のような問題が出題され，偶然の一致に驚きました。

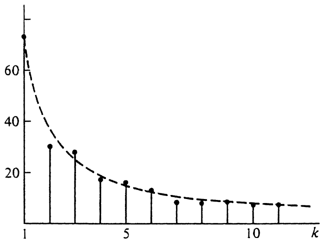
「pを素数，nを正の整数とするとき，$(p^n)!$はpで何回割り切れるか。」

§3. 上の結果とジップの法則

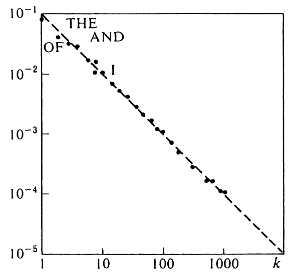
ジップの法則は20世紀半ばにジップという人によって発見された経験則です。

文章における英単語の使用頻度とその順位の関係として発見されました。つまり出現頻度がk番目に多い要素が全体に占める割合は$\dfrac{1}{k}$に比例するという法則です。

その他にもアメリカの都市を人口の多い順に並べた(横軸が順位，縦軸が人口)ものや世界の川の長さと順位，日本の湖の面積と順位，わが国の年間輸入額と国別順位，個人所得と順位，生物の種の個体数と順位などにも同様の分布を見ることができるそうです。



〔図１〕



〔図２〕

ところで，１からnまでのn個の自然数に含まれる因数の個数は多い順に２，３，４，５，… ですが，２で証明したことがらより

順　位	1，2，3，4，5，…
因　数(p)	2，3，4，5，6，…
出現頻度	$\frac{1}{1}$，$\frac{1}{2}$，$\frac{1}{3}$，$\frac{1}{4}$，$\frac{1}{5}$，…

であるので，まさにジップの法則そのものです。

自然界のいろいろな分布と自然数には何らかのつながりがあることの１つの証左なのかもしれません。

(１図はアメリカの都市の人口と順位，参考文献〔1〕のp.93 の上段の図を引用

２図は英語の単語の出現頻度と順位(２図は対数目盛り)，参考文献〔1〕のp.97 の上段の図を引用)

(＊1)　コラッツ予想

任意の１でない自然数nに対して，

⑴　nが偶数ならば，２で割る

⑵　nが奇数ならば，３倍して１を加える

という操作を連続して行う。

このとき，どのような自然数nから出発しても，有限回の操作で１に到達する。

この予想は「角谷の予想」とも呼ばれ，まだ解決されていない数論の問題です。

(＊2)　本文参照

《参考文献》

〔1〕　無限・カオス・ゆらぎ
寺本英・広田良吾・武者利光・山口昌哉
培風館

(大阪府　大阪国際大和田高等学校)