令和４年度大阪大学未来基金「学部学生による自主研究奨励事業」研究成果報告書

ふりがな 氏名	みき　たかや 三木　隆哉	学部 学科	工学部 電子情報工学科	学年	２年
ふりがな 共同 研究者氏名		学部 学科		学年	年
					年
					年
アドバイザー教員 氏名	中村　博昭　教授	所属	大阪大学 大学院理学研究科 数学専攻		
研究課題名	桁数と個数の相関におけるハミング数の特徴について				
研究成果の概要	研究目的、研究計画、研究方法、研究経過、研究成果等について記述すること。必要に応じて用紙を追加してもよい。（先行する研究を引用する場合は、「阪大生のためのアカデミックライティング入門」に従い、盗作剽窃にならないように引用部分を明示し文末に参考文献リストをつけること。）				

1．研究の動機

高校生の時に，ハミング数の個数と桁数の関係について研究を行った．ハミング数とは,

$$2^p \cdot 3^q \cdot 5^r \ (p,q,r \in S, S = \{ n \mid n \in \mathbb{Z}, n \geq 0 \})$$

と表される数のことである．高校では，参考文献[1]を用いて 17 桁までのハミング数の個数を得ることができた．そのため，17 桁までの考察に留まっていた.

17 桁以下のハミング数の個数と桁数の関係について述べる．1 桁のハミング数は，7 以外の自然数すべてであるため，8 個ある．2 桁以上のハミング数の個数については，一つずつ数え上げると時間がかかるため，場合分けをして個数を求める．次のような手順でハミング数の個数を求めることができた.

手順１　変数 p,q,r のうち，2 つを 0 にする．残った変数の範囲を求め，ハミング数の個数を求める．これをすべての変数について行う.

例　2 桁のハミング数の個数を求めるとき

q,r を 0 とすると

$$10^1 \leq 2^p < 10^2$$

となる．これを満たす非負整数 p は，4, 5, 6 の 3 つである.

手順２　変数 p,q,r のうち，1 つを 0 にする．残った 2 つの変数は 1 以上であることから，次の例のようにしてハミング数の個数を求めることができる.

例　2 桁のハミング数の個数を求めるとき

$r = 0$ とすると

$$10^1 \leq 2^p \cdot 3^q < 10^2$$

となる．$p,q \geq 1$ であることから,

$$\frac{10^1}{2 \cdot 3} \leq 2^{p-1} \cdot 3^{q-1} < \frac{10^2}{2 \cdot 3}$$

となる．$2^{p-1} \cdot 3^{q-1}$ は自然数であることから,

$$2 \le 2^{p-1} \cdot 3^{q-1} \le 16$$

となる．これを満たす２変数 p, q の組 $(p-1, q-1)$ を求めると，

$$(p-1, q-1) = (4,0), (3,0), (2,0), (1,0), (0,1), (0,2), (1,1), (2,1)$$

の８通りがあることがわかる．

手順３　r に１以上の自然数を代入していき，手順２と同じ要領で２変数 p, q の組を求める．

$$5^{\alpha+1} \ge 10^n$$

となるとき，α を r の最大値 r_{max} とし，r には α 以下の自然数を代入する．

手順４　手順 1～3 で求めたハミング数の個数をすべて足し，n 桁のハミング数の個数を導出する．

以上の手順で２桁と３桁のハミング数の個数を求めると，それぞれ 25，52 となることがわかった．

$1 \le n \le 3$ において n 桁のハミング数の個数を a_n とおくと，

$$a_2 - a_1 = 25 - 8 = 17$$

$$a_3 - a_2 = 52 - 25 = 27$$

より，

$$a_{n+1} - a_n = 17 + 10(n-1) = 10n + 7$$

となると推測される．

$$a_{n+1} = a_n + 10n + 7$$

であるから，$n \ge 2$ において推測値は

$$a_n = a_1 + \sum_{i=1}^{n-1} (10i + 7)$$

$$= 8 + 10 \cdot \frac{1}{2} n(n-1) + 7(n-1)$$

$$= 5n^2 + 2n + 1$$

となる．a_n に $n = 1$ を代入すると，

$$a_1 = 8$$

となるから，$n = 1$ でも成り立つことがわかる．この式は，$1 \le n \le 3$ においては成り立つが，$n \ge 4$ においても成り立つとは限らない. そこで a_n に $n = 4, 5, \cdots, 17$ を代入して得られた値と実際のハミング数の個数との差を求めると，差の絶対値は２以下であることがわかった．しかし，$n > 17$ においてもこの式が成り立つか高校生の時の研究では明らかにすることができなかった．そこで，大学で $n > 17$ の場合について考えることにした．

2．研究結果

ここでの計算は，Maple*で行った．

文献[2]を調査したところ，$N (\in \mathbb{N})$ までのハミング数の個数を表す式

$$\frac{\left(\ln N\sqrt{30}\right)^3}{6(\ln 2)(\ln 3)(\ln 5)} + O(\log N)$$

がわかった．$N = 10^n$ とすると，$n (\in \mathbb{N})$ 桁までのハミング数の個数は

$$\frac{\left(\ln 10^n\sqrt{30}\right)^3}{6(\ln 2)(\ln 3)(\ln 5)} + O(\log 10^n) = \frac{\left(\ln 10^n\sqrt{30}\right)^3}{6(\ln 2)(\ln 3)(\ln 5)} + O(n) \qquad \mathbf{(1)}$$

と表すことができる．ここで，

$$g(n) = \left\lfloor \frac{\left(\ln 10^n \sqrt{30}\right)^3}{6(\ln 2)(\ln 3)(\ln 5)} \right\rfloor$$

とおく．これは，誤差 $O(n)$ を無視し，小数点以下を切り捨てたものであり，n 桁までのハミング数の個数と近い値をとる式である．さらに，

$$g(0) = 0$$

であることから，n 桁のハミング数の個数と近い値をとる式を $h(n)$ とおくと，

$$h(n) = \left\lfloor \frac{\left(\ln 10^n \sqrt{30}\right)^3}{6(\ln 2)(\ln 3)(\ln 5)} \right\rfloor - \left\lfloor \frac{\left(\ln 10^{n-1} \sqrt{30}\right)^3}{6(\ln 2)(\ln 3)(\ln 5)} \right\rfloor$$

となる．また，高校生の時に導いた n 桁のハミング数の個数の近似式

$$a(n) = 5n^2 + 2n + 1$$

との差

$$d(n) = a(n) - h(n)$$

も定義する．$d(n)$ に 1 から 200 までの自然数を代入すると，$n = 17$ までは $h(n)$ の値が $a(n)$ の値を少し（最大でも 2）上回ることがわかった．しかし，$n \geq 20$ においては，$a(n)$ の値が $h(n)$ の値を上回ることがわかった．$d(n)$ がどの程度の割合で増加するのかを調べるために，$h(n)$ において小数点以下を切り捨てずに n 桁のハミング数の個数を表した式

$$b(n) = \frac{\left(\ln 10^n \sqrt{30}\right)^3}{6(\ln 2)(\ln 3)(\ln 5)} - \frac{\left(\ln 10^{n-1} \sqrt{30}\right)^3}{6(\ln 2)(\ln 3)(\ln 5)}$$

を定義する．$b(n)$ は整理すると，

$$b(n) = \frac{(\ln 10)^3}{2(\ln 2)(\ln 3)(\ln 5)} n^2 - \frac{(\ln 10)^3}{2(\ln 2)(\ln 3)(\ln 5)} n + \frac{(\ln 10)^2 (\ln 30)}{2(\ln 2)(\ln 3)(\ln 5)} n + \frac{(\ln 10)^3}{6(\ln 2)(\ln 3)(\ln 5)}$$
$$- \frac{(\ln 10)^2 (\ln 30)}{4(\ln 2)(\ln 3)(\ln 5)} + \frac{(\ln 10)(\ln 30)^2}{8(\ln 2)(\ln 3)(\ln 5)}$$

となり，n の整式となる．各係数を小数第九位まで表したものを $c(n)$ とすると，

$$c(n) = 4.980499826n^2 + 2.376302329n + 0.698487739$$

となる．

$$a(n) - c(n) = 0.019500174n^2 - 0.376302329n + 0.301512261 \qquad (\mathbf{2})$$

となり，$a(n) - c(n)$ は n の 2 次関数であり，

$$\frac{d}{dn}\{a(n) - c(n)\} = 0.039000348n - 0.376302329$$

となり，

$$\frac{d}{dn}\{a(n) - c(n)\} = 0$$

のとき

$$n = 9.648691571$$

であることから，$a(n)$ の値が $h(n)$ の値を上回る $n \geq 20$ において $d(n)$ が n^2 のオーダーで単調増加することがわかった．

$a(n)$ と $b(n)$ は共に 2 次式であるが，n^2 の係数が完全に一致していないため，その差は $n \to \infty$ のとき無限大に発散する．すなわち

$$\lim_{n \to \infty} \{a(n) - c(n)\} = \infty$$

である．ここで，方程式

$$a(n) - c(n) = 0$$

を解くと，

$$n = 0.8376065203, 18.45977662$$

となることから，19 桁で $a(n)$ の値が $c(n)$ の値を上回ることがわかった．

$(\mathbf{1})$ より，

$$\frac{\left(\ln 10^n \sqrt{30}\right)^3}{6(\ln 2)(\ln 3)(\ln 5)}$$

と実際のハミング数の個数との誤差は

$$O(n)$$

と表すことができることがわかる．$b(n)$ は小数点以下を切り捨てていないだけであるため，$b(n)$ や $c(n)$ と実際のハミング数の個数との誤差も

$$O(n)$$

である．また，

$$\lim_{n\to\infty}\frac{a(n)}{c(n)} = \lim_{n\to\infty}\frac{5+\frac{2}{n}+\frac{1}{n^2}}{4.980499826+\frac{2.376302329}{n}+\frac{0.301512261}{n^2}} \fallingdotseq \frac{5}{4.98} \fallingdotseq 1.004016$$

であるから，n が十分大きな値をとるとき，$a(n)$ と実際のハミング数の個数の違いは，比率として 0.4% 程度になる．

3．考察

研究結果より，$a(n)$ は桁数が大きくなると $b(n)$ との差の絶対値が大きくなることがわかった．しかし，$(\mathbf{2})$ 式からわかるように，$a(n)$ と $b(n)$ や $c(n)$ の各項の係数を比較すると，$a(n)$ は整数係数の n の 2 次式としては $b(n)$ や $c(n)$ との係数の差が最も小さいことがわかる．このことから，$a(n)$ はハミング数の個数と桁数の関係を高精度で近似できている式であると考える．

$a(n)$ を導出した時の情報，すなわち 1 桁から 3 桁までのハミング数の個数のみから，これ以上正確な式を求めることは困難であったと考える．しかし，参考文献より得られた $n \leq 17$ におけるハミング数の個数と桁数の関係を用いると，$a(n)$ よりも正確な式を導くことができた可能性がある．

4．参考文献

［1］T.D. Noe and Reinhard Zumkeller, Table of n, a(n) for n = 1.. 10000,
First 1000 terms from T.D. Noe, https://oeis.org/A051037/b051037.txt
［2］THE ON-LINE ENCYCLOPEDIA OF INTEGER SEQUENCES ,A051037 5-smooth numbers, i.e., numbers whose prime divisors are all <= 5 ,https://oeis.org/A051037,(参照 2022-10-15).

*Maple： 数式処理ソフトウェア(Maplesoft Japan 株式会社)