JOURNAL OF NUMBER THEORY 40, 1–11 (1992)

# The Error Term in Golomb's Sequence

ILAN VARDI

*Wolfram Research Inc., and the Department of Mathematics, Stanford University, Stanford, California 94305*

*Communicated by Hans Zassenhaus*

Received August 22, 1990

S. Golomb discovered a self describing sequence of integers with a simple asymptotic behavior. This paper examines how close the sequence is to the asymptotic estimate. I give an upper bound for the error term and give strong evidence that this upper bound is actually the best possible. The evidence consists of a formal solution to a recurrence relation, as well as numerical evidence. I also present an efficient method for computing Golomb's sequence for large values. This method relies on the enumeration of a special kind of tree. 

## 1. INTRODUCTION

In [2] S. Golomb defined a sequence $F(n)$ of integers

$$1, 2, 2, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 6, 6, 7, \ldots$$

defined by $F(1)=1$, $F(n)=\#\{m : F(m)=n\}$, and $F$ is nondecreasing. Golomb asked for an asymptotic formula for $F(n)$. The answer appeared in [3], where a solution by N. J. Fine was given. The question has also appeared as an exercise in [1, p. 246] and as an open problem in [5, p. 126]. The sequence is #91 in [6, p. 39].

The asymptotic formula for $F(n)$ that Golomb discovered is

$$F(n) \sim cn^{\phi-1}, \qquad c=\phi^{2-\phi}, \quad \text{where} \quad \phi=\frac{\sqrt{5}+1}{2}=1.6180339887498949\ldots. \tag{1}$$

D. E. Knuth asked the question of how good the asymptotic formula (1) is. Write $F(n)=H(n)+E(n)$, where $H(n)=cn^{\phi-1}$. The asymptotic formula says that $E(n)=o(n^{\phi-1})$. Professor Knuth computed $E(n)$ for $n \leqslant 400$ and, surprisingly, found that $E(n) \leqslant 1/2$ except for $n=273$ when $E(n) \approx 0.5003$ [4, p. 577]. In this paper I will show



0022-314X/92 $3.00

THEOREM 1.

$$E(n) = O\left(\frac{n^{\phi-1}}{\log n}\right). \tag{2}$$

In view of Knuth's computation, Theorem 1 seems like a rather weak result. However, I believe that the following is true

*Conjecture* 1.

$$E(n) = \Omega_{\pm}\left(\frac{n^{\phi-1}}{\log n}\right), \tag{3}$$

*where $\Omega_{\pm}(f) = g$, $g > 0$, means that there are constants $A$ and $B$ such that $f(n) > Ag(n)$ for infinitely many $n$ and $f(n) < -Bg(n)$ for infinitely many $n$.*

This conjecture follows from the stronger statement

*Conjecture* 2.

$$E(n) = \frac{n^{\phi-1} h(\log\log n/\log\phi)}{\log n} + O\left(\frac{n^{\phi-1}}{\log^2 n}\right), \tag{4}$$

*where $h(x)$ satisfies*

$$h(x+1) = -h(x), \qquad |h(x)| > 0, \quad x \in (0, 1).$$

*Remark.* I have not even been able to show that $\limsup_{n\to\infty} |E(n)| = \infty$.

I thank Professor Knuth for suggesting this problem to me. The results of this paper are communicated in [4].

## 2. FACTS ABOUT THE ASYMPTOTIC FUNCTION

We use the following result [7, p. 13]

LEMMA 1. *For any continuously differentiable function $\alpha(x)$ on $[a, b]$*

$$\sum_{a<n<b} \alpha(n) = \int_a^b \alpha(x)\,dx + \int_a^b ((x))\alpha'(x)\,dx + ((a))\alpha(a) - ((b))\alpha(b), \tag{5}$$

*where $((x)) = x - \lfloor x \rfloor - 1/2$.*

LEMMA 2.

$$\sum_{k=1}^{n} \frac{k^{\phi-1}}{\log k} = \frac{1}{\phi}\left(1 + \frac{O(1)}{\log n}\right)\frac{n^{\phi}}{\log n}.$$

*Proof.* We have that

$$\int \frac{x^{\phi-1}}{\log x}\,dx = \frac{1}{\phi}\frac{x^{\phi}}{\log x} + \frac{1}{\phi}\int \frac{x^{\phi-1}}{\log^2 x}\,dx$$
$$= \frac{1}{\phi}\frac{x^{\phi}}{\log x} + O\left(\frac{x^{\phi}}{\log^2 x}\right) = \frac{1}{\phi}\left(1 + \frac{O(1)}{\log x}\right)\frac{x^{\phi}}{\log x}.$$

The lemma follows from Lemma 1 and the fact that all but the first term on the right side of (5) are $\ll n^{\phi-1}/\log^2 n$.

As a further observation we note that one can write $H(n)$ as

$$\frac{H(n)}{\phi} = \left(\frac{n}{\phi}\right)^{\phi-1}$$

which clearly shows that

$$F_k(n) \sim \phi\left(\frac{n}{\phi}\right)^{\phi-k},$$

where $F_k(n) = F(F_{k-1}(n))$, etc. In particular,

$$F(F(n)) \sim c^{\phi} n^{2-\phi}.$$

## 3. Proof of Theorem 1

The key idea of the proof is the heuristic formula

$$E(n) \approx -\frac{\sum_{k=1}^{cn^{\phi-1}} E(k)}{c^{\phi} n^{2-\phi}}. \tag{6}$$

*Remark.* Assuming that this formula holds in some sense, then Conjecture 2 is plausible since

$$\frac{n^{\phi-1} h(\log\log n/\log\phi)}{\log n}$$

is an asymptotic solution to (6) when $h(x+1) = -h(x)$. We also note that a variant to the argument below might establish Conjecture 2.

*Proof of Theorem* 1. First define $G(n)$ as the last time that $n$ appears in the sequence, i.e., that $F(G(n)) = n$. Since $n$ has occurred $F(n)$ times, $n-1$ has occurred $F(n-1)$ times, and so on, it is clear that

$$G(n) = \sum_{k=1}^{n} F(k). \tag{7}$$

As in [3] an inductive argument will be used. Let $N_1, N_2, \ldots, N_k$ be a sequence such that $N_{k+1} = G(N_k)$. From (7) it is easy to show that

$$G(n) > \tfrac{2}{3} n^{3/2}$$

so $N_k \to \infty$ and in fact

$$N_k > 3^{(1.1)^{k-1}}.$$

For convenience, we drop the $k$'s and working with a fixed $n$ assume the induction for $k < n$ holds. We then show that the induction hypothesis also holds for $k < G(n)$. What we want to show is that

$$|E(k)| < A(k) \frac{k^{\phi - 1}}{\log k}$$

holds for $k < n$, where the $A(k)$ are bounded above.

The way the proof works is to first use (7) to get an estimate on $G(n)$. The main term will come from Euler–MacLaurin summation, and since the error is a sum of lower order error terms, it will be bounded using the induction assumption. The next step is to feed this estimate back into

$$F(G(n)) = n = c[G(n)]^{\phi - 1} + E(G(n)).$$

From this one derives an estimate of the error term at $G(n)$ .

To be precise about how the induction is carried out, note that for each $N_k$ one chooses an $A(N_k)$ and it must be shown that the $A(N_k)$ are bounded above.

I will assume the asympotic formula (1), i.e., that $F(n) \sim cn^{\phi - 1}$ (see [3] for a proof). This implies that one can choose $A(n) = o(\log n)$.

Writing out what the asymptotic formula corresponds to for $G(n)$ we get

$$\begin{aligned} G(n) &= \sum_{k=1}^{n} F(k) = \sum_{k=1}^{n} cn^{\phi - 1} + \sum_{k=1}^{n} E(k) \\ &= \frac{cn^{\phi}}{\phi} + \frac{cn^{\phi - 1}}{2} + c\zeta(1 - \phi) + \frac{c(\phi - 1)n^{\phi - 2}}{12} + \cdots + R(n), \qquad (8) \end{aligned}$$

where $R(n) = \sum_{k=1}^{n} E(k)$. From Lemma 2 and the induction assumption, we have that

$$|R(n)| < \frac{A(n)}{\phi} \left(1 + \frac{O(1)}{\log n}\right) \frac{n^{\phi}}{\log n}. \qquad (9)$$

Let $0 \leqslant r < F(n)$, then

$$\begin{aligned} n &= F(G(n)-r) \\ &= c\left(\frac{c}{\phi}n^{\phi} + \frac{cn^{\phi-1}}{2} + c\zeta(1-\phi) + \frac{c(\phi-1)n^{\phi-2}}{12} + \cdots + R(n) - r\right)^{\phi-1} \\ &\quad + E(G(n)-r) \\ &= n + \frac{1}{2} - \frac{r}{cn^{\phi-1}} + \frac{R(n)}{cn^{\phi-1}} - \frac{[R(n)]^2}{2c^2\phi n^{2\phi-1}} + E(G(n)-r) + \cdots, \end{aligned} \tag{10}$$

so that

$$E(G(n)) = -\frac{R(n)}{cn^{\phi-1}} + \frac{[R(n)]^2}{2c^2\phi n^{2\phi-1}} + O\left(\frac{n}{\log^2 n}\right), \tag{11}$$

where the variation due to $r = O(n^{\phi-1})$ is small enough that it has been absorbed in the error term. Note that (6) states that $-R(n)/c \log n$ is the leading term of the right hand side of (11), since (6) corresponds to the case where $n \Rightarrow F(n)$.

Using the bound of Eq. (9) in Eq. (11) gives

$$|E(G(n))| < \left(1 + \frac{O(1)}{\log n}\right) A(n) \frac{n}{c\phi \log n},$$

since all other terms are $O(n/\log^2 n)$.

Equation (12) must still be converted to one with $G(n)$ on the right hand side. Substituting the bound on $R(n)$ given by (9) in Eq. (8) yields

$$G(n) = \left(1 + \frac{O(1)A(n)}{\log n}\right) \frac{cn^{\phi}}{\phi},$$

so

$$\begin{aligned} \frac{[G(n)]^{\phi-1}}{\log G(n)} &= \frac{(1 + (O(1)A(n)/\log n))^{\phi-1}}{\log[(1 + (O(1)A(n)/\log n))]} \frac{(cn^{\phi}/\phi)^{\phi-1}}{\log(cn^{\phi}/\phi)} \\ &= \left(1 + \frac{O(1)A(n)}{\log n}\right) \frac{n}{c\phi \log n}. \end{aligned}$$

Putting this back in Eq. (12) yields

$$|E(G(n))| < \left(1 + \frac{O(1)A(n)}{\log n}\right) A(n) \frac{[G(n)]^{\phi-1}}{\log G(n)}.$$

What we have shown is that

$$A(G(n)) < \left(1 + \frac{O(1)A(n)}{\log n}\right) A(n).$$

Since $A(n) = o(\log n)$, it follows that one can choose an $n$ sufficiently large such that

$$\left(1 + \frac{O(1)A(n)}{\log n}\right) < 1.05.$$

Since $N_k > 3^{(1.1)^{k-1}}$, it follows that for a sufficiently large $n$

$$A(N_{k+1}) < \left(1 + \left(\frac{1.05}{1.1}\right)^k\right) A(N_k),$$

and the boundedness of $A(N_k)$ follows from the fact that

$$\prod_{k=1}^{\infty} (1 + x^k)$$

converges for $|x| < 1$. ∎

*Remark.* Note that if all errors terms are very small then from the third line of (10) the main term becomes

$$E(G(n) - r) \approx -\frac{1}{2} + \frac{r}{cn^{\phi - 1}},$$

so that the error approximates a linear function varying from $1/2$ to $-1/2$ as the argument goes from $G(n-1)+1$ to $G(n)$. This seems to explain Knuth's computation.

## 4. Computing $F(n)$

It turns out that one can compute large values of $F(n)$ efficiently by using the structure of Golomb's sequence. It is seen that $G(n)$ can be computed using (7), so that $F(G(n)) = n$ allows one to evaluate $F$ at larger arguments. There is a similar formula for $G(G(n))$:

$$G(G(n)) = \sum_{k=1}^{n} kF(k). \tag{13}$$

This implies that one can compute $F(G(G(n))) = G(n)$ almost as easily as $F(G(n)) = n$, but with an increase by a factor of $\phi$ in the number of digits of the argument to $F$.

To prove (13), note that the numbers $r$ in the interval $[G(G(k-1))+1, G(G(k))]$ are exactly those with

$$G(k-1)+1 \leqslant F(r) \leqslant G(k),$$

so that, as is seen from the proof of Eq. (7), $F(r)$ takes on the values

$$G(k-1)+1, G(k-1)+2, ..., G(k-1)+F(k).$$

On the other hand the numbers in this interval are also those with

$$F(F(r))=k.$$

So each distinct value $F(r)$ appears $F(F(r))=k$ times and there are $F(k)$ different values, so the formula follows.

This can be generalized by defining $G_1(n)=G(n)$, $G_m(n)=G(G_{m-1}(n))$, and then computing the number of elements in each interval $[G_m(k-1)+1, G_m(k)]$ by forming the following tree:

1. The terminal leaves are the integers in the interval $[G_m(k-1)+1, G_m(k)]$.

2. The parent of a leaf $r$ is equal to $F(r)$. This implies that two nodes $r_1, r_2$ are brothers if and only if $F(r_1)=F(r_2)$.

3. The tree has $m+2$ levels. This implies that this is a rooted tree with the root node having the value $F(k)$.

The problem is to compute how many terminal nodes the tree has. Instead of analyzing this directly it is more convenient to look at a generalization given by the (family) tree.

A royal family has the following tradition (see the Fig. 1):

1. Each king decides how many children he will have.

2. The number of children your successor has is equal to the number of children you have, if your successor is your brother, otherwise he has one more child than you do.

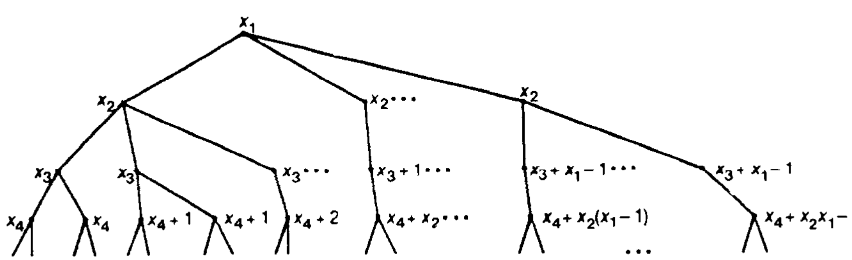


FIG. 1. The family tree.

The question is: How many descendents are there after $n$ generations? (The answer will be in terms of the number of children of king 1, king 2, ..., king $n$.)

Recall the rule for royal succession (we consider questions of succession to be for one generation at a time only): Your next younger brother is your successor. If you don't have a younger brother, the oldest child of your parent's successor is your successor.

*Remarks.* 1. We consider the case here when the first king has no brothers, i.e., the family tree is a rooted tree.

2. The $(n+1)$st king is the oldest child of the $n$th king.

It is fairly clear that the above recursive rules simplify to the following

PROPOSITION 2. *The number of children you have is the number of children of your king (i.e., of your generation) plus the order of succession of your grandparent.*

Translating Proposition 2 into a numerical formula gives:

PROPOSITION 3. *Let $x_i$ be the number of children of king $i$, we write $T_n(x_1, ..., x_n)$ for the number of children in generation $(n+1)$. We have*

$$T_1(x_1) = x_1,$$

$$T_n(x_1, ..., x_n) = \sum_{k=0}^{x_1 - 1} T_{n-1}(x_2, ..., x_i + T_{i-2}(k, x_2, x_3, ..., x_{i-2}), ..., x_n$$
$$+ T_{n-2}(k, x_2, x_3, ..., x_{n-2})).$$

*For example,*

$$T_4(x_1, x_2, x_3, x_4) = \sum_{k=0}^{x_1 - 1} T_3(x_2, x_3 + k, x_4 + kx_2).$$

*Proof.* One counts $T_n(x_1, ..., x_n)$ by counting recursively over the $x_1$ subtrees of the root node. The leftmost subtree clearly contains $T_{n-1}(x_2, x_3, ..., x_n)$ at the lowest level. In general, the $k$th subtree will contribute $T_{n-1}(y_2, y_3, ..., y_n)$, where the $y_i$'s are, by Proposition 2, equal to the number of descendents at generation $i-2$ of the tree where only the first $k-1$ children at generation 2 are taken into account. This number is $T_{i-2}(k-1, x_2, x_3, ..., x_{i-2})$. ∎

One can compute the $T_i$ explicitly by using the closed form for $\sum_{k<x} k^m$, $m$ an integer. For $n = 1, 2, 3, 4$ this is

COROLLARY 1.

$$T_1(x_1)=x_1$$

$$T_2(x_1, x_2)=x_1 x_2$$

$$T_3(x_1, x_2, x_3)=x_1 x_2 x_3+\frac{x_1(x_1-1)x_2}{2}$$

$$T_4(x_1, x_2, x_3, x_4)=x_1 x_2 x_3 x_4+\frac{x_1 x_2 x_3(x_2-1)}{2}+\left[x_2 x_4+x_3 x_2^2+\frac{x_2(x_2-1)}{2}\right]\frac{x_1(x_1-1)}{2}+x_2^2\frac{x_1(x_1-1)(2x_1-1)}{6}.$$

From the previous discussion it is seen that $G_m$ can be computed by

PROPOSITION 4.

$$G_m(n)=\sum_{k=1}^{n} T_m(F(k), k, G(k-1)+1, G_2(k-1)+1, ..., G_{m-2}(k-1)+1).$$

COROLLARY 2. *Let* $G(0)=0$, *then*

$$G_1(n)=\sum_{k=1}^{n} F(k),$$

$$G_2(n)=\sum_{k=1}^{n} kF(k),$$

$$G_3(n)=n\frac{G(n)[G(n)+1]}{2}-\sum_{k=1}^{n-1}\frac{G(k)[G(k)+1]}{2},$$

$$G_4(n)=\sum_{k=1}^{n}\left\{F(k)[G(k-1)+1]\left(kG_2(k-1)+\frac{k(k+1)}{2}\right)+\frac{[F(k)-1]F(k)}{2}\left(kG_2(k-1)+\frac{k(k+1)}{2}+k^2[G(k-1)+1]\right)+\frac{k^2[F(k)-1]F(k)[2F(k)-1]}{6}\right\}.$$

I have used these formulas to compute large values of $F(n)$. For example, by computing $G_4(100,000)$ and $G_3(100,000)$ one gets that

$$F(11132623751311908126247331173 09095)=31895648580638856178 3.$$

## 5. Numerical Results

It may be important in inductive arguments relating to Conjecture 2 to know that $E(k) > 0$ for $N < k < G(N)$, we do this in this section.

| $n$ | $F(n)$ | $H(n)$ | $E(n)$ |
|---|---|---|---|
| 719845 | 5008 | 5009.5527367956... | −1.5527367956... |
| 2302936 | 10277 | 10278.5680684822... | −1.5680684822... |
| 10790705 | 26700 | 26698.6934303285... | 1.3065696714... |
| 32875989 | 53160 | 53151.1639375215... | 8.8360624784... |
| 91374964 | 99996 | 99974.2086559435... | 21.7913440564... |
| 131178997 | 125037 | 125009.2738536491... | 27.7261463508... |
| 201636503 | 163090 | 163054.2699123597... | 35.7300876402... |
| 178053041255 | 10790705 | 10790662.1257056331... | 4.287429436... |
| 194195117253 | 11385212 | 11385211.9923031852... | .0076968147... |
| 194213527120 | 11385879 | 11385879.0416504430... | −.0416504430... |
| 919986484788 | 29772902 | 29774861.0721068723... | −1959.0721068723... |

Note that

$$G(26, 700) = 10790705$$

$$G(G(26, 700)) = 178053041255.$$

By above it is clear that if $E(k) > 0$, $N < kN'$ when $k = G(t)$, then $E(k) > 0$ for all $N < k < N'$. I have checked all the $E(G(k))$'s in this region so it follows $E(k) > 0$ for $10790705 < k < 178053401255$.

## 6. Problems

In this section I will give a number of problems related to Golomb's sequence. I will mark unsolved ones with a *.

**Problem 1** [1, p. 246]. *Show that*

$$F(n + F(F(n))) = F(n) + 1$$

*and use this to deduce the asymptotic formula for* $F(n)$.

**Problem 2.** *Show that*

$$\begin{aligned} \sum_{n=1}^{\infty} x^{F(n)} = \sum_{n=1}^{\infty} F(n) x^n &= \frac{x}{1-x} \sum_{n=1}^{\infty} x^{G(n)} \\ \sum_{n=1}^{\infty} G(n) x^n &= \frac{1}{1-x} \sum_{n=1}^{\infty} F(n)\, x^n. \end{aligned} \tag{14}$$

PROBLEM 3*. *Use the Littlewood–Hardy method to prove Conjecture* 1. *Bruno Salvy has noted that since* $G(n) \gg n^{\phi}$, $\Sigma x^{G(n)}$ *is a lacunary power series and so has a natural boundary at the unit circle.*

PROBLEM 4*. *Analyze the behavior of*

$$\sum_{n=1}^{\infty} \frac{F(n)}{n^s} = c\zeta(s+1-\phi) + \sum_{n=1}^{\infty} \frac{E(n)}{n^s}.$$

PROBLEM 5. *Prove the finite version of* (14)

$$\sum_{k=1}^{n} x^{F(k)} = \sum_{k=1}^{F(n)} F(k)x^k - [G(F(n))-n]\,x^{F(n)}$$

$$= \frac{x}{1-x} \sum_{k=1}^{F_2(n)} x^{G(k)} + \frac{x}{x-1} F_2(n)x^{F(n)} - [G(F(n))-n]\,x^{F(n)}.$$

PROBLEM 6*. *Show that for any fixed integer* $q > 0$

$$\sum_{k=1}^{n} e^{2\pi i E(k)/q} = o(n).$$

(*This would imply that* $\limsup |E(n)| = \infty$.)

PROBLEM 7*. *Let k be fixed: How is the sequence* $\{G_n(k)\}_{n=1}^{\infty}$ *related to* $\{L_{L_n+1}\}_{n=1}^{\infty}$, *where* $L_n = \phi^n + (-1)^n \phi^{-n}$ *are the Lucas numbers?*

PROBLEM 8. (a) *Let* $H_1(n) = cn^{\phi-1}$, $H_{k+1}(n) = H(H_k(n))$. *Show that*

$$\frac{n}{\phi} = \prod_{k=2}^{\infty} \frac{H_k(n)}{\phi}. \tag{15}$$

(b)* *What does* (15) *say about the "family tree" of Section* 4?

## REFERENCES


1. C. BENDER AND S. ORZAG, "Advanced Mathematical Methods for Scientists and Engineers," McGraw-Hill, New York, 1978.
2. S. GOLOMB, Problem 5407, *Amer. Math. Monthly* **73** (1966), 674.
3. D. MARCUS AND N. J. FINE (also J. H. van Lint and S. Golomb), Solution to problem 5407, *Amer. Math. Monthly* **74** (1967), 740–743.
4. R. L. GRAHAM, D. E. KNUTH, AND O. PATASHNIK, "Concrete Mathematics," Addison-Wesley, Reading, MA, 1988.
5. R. K. GUY, "Unsolved Problems in Number Theory," Springer-Verlag, New York/Berlin, 1981.
6. N. J. A. SLOANE, "A Handbook of Integer Sequences," Academic Press, Orlando, FL, 1973.
7. E. C. TITCHMARSH, "The Theory of the Riemann Zeta-Function," 2nd ed., revised by D. R. Heath-Brown, Oxford Univ. Press, London/New York, 1986.