126. *On the Sum of Digits of Prime Numbers*

By Iekata SHIOKAWA

Department of the Foundations of Mathematical Sciences,
Tokyo University of Education, Tokyo

(Comm. by Kôsaku YOSIDA, M. J. A., Oct. 12, 1974)

Let $r>1$ be a fixed integer. Then any positive integer n can be expressed in the form

$$(1)\qquad n=\sum_{i=1}^{k} a_i r^{k-i}=a_1a_2\cdots a_k,$$

where each a_i is one of $0,1,\cdots,r-1$ and

$$(2)\qquad k=k(n)=\left[\frac{\log n}{\log r}\right]+1,$$

where $[z]$ is the integral part of z. We put

$$\alpha(n)=\sum_{i=1}^{k} a_i.$$

I. Katai [2] proved, *assuming the validity of density hypothesis for the Riemann zeta function, that*

$$(3)\qquad \sum_{p\le k}\alpha(p)=\frac{r-1}{2}\frac{x}{\log r}+O\left(\frac{x}{(\log\log x)^{\frac{1}{8}}}\right)$$

hold, where in the sum p runs through the prime numbers.

In this paper we shall prove without any unsolved hypothesis the result (3) of Katai, even with an improved remainder term. Our method is to appeal to a simple combinatorial argument, and the deepest result on which we shall depend is the well-known prime number theorem in a rather weak form.

In what follows all the O-constants depend possibly on the given scale r.

Theorem. *We have*

$$\sum_{p\le x}\alpha(p)=\frac{r-1}{2}\frac{x}{\log r}+O\left(x\left(\frac{\log\log x}{\log x}\right)^{\frac{1}{2}}\right),$$

where in the sum p runs through the primes.

Proof. Let b be any fixed positive integer not greater than $r-1$. For any positive integer n, we denote by $F(b,n)$ the number of b's appearing in the r-adic representation (1) of n and set

$$D(b,n)=\left|F(b,n)-\frac{k(n)}{r}\right|.$$

Thus we have

$$(4)\qquad \sum_{p\le x}F(b,p)=\frac{1}{r}\sum_{p\le x}k(p)+O\left(\sum_{p\le x}D(b,p)\right).$$

It follows from the definition (2) of $k(n)$ and the well-known result

$$\sum_{p\le x}\log p=x+O\left(\frac{x}{\log x}\right)$$

that

$$\sum_{p\le x}k(p)=\frac{1}{\log r}\sum_{p\le x}\log p+O(\pi(x))=\frac{x}{\log r}+O\left(\frac{x}{\log x}\right).$$

Let ε be any positive number less than $\frac{1}{2}$. Then we have

$$\begin{aligned}\sum_{p\le x}D(b,p)&\le\sum_{p\le x}k(p)^{\frac{1}{2}+\varepsilon}+\sum_{\substack{p\le x\\ D(b,p)>k(p)^{\frac{1}{2}+\varepsilon}}}D(b,p)\\&=O\left(\sum_{p\le x}(\log p)^{\frac{1}{2}+\varepsilon}\right)+O\left(\sum_{\substack{n\le x\\ D(b,n)>k(n)^{\frac{1}{2}+\varepsilon}}}D(b,n)\right)\\&=O\left(\frac{x}{(\log x)^{\frac{1}{2}-\varepsilon}}\right)+O\left(\sum_{\substack{n\le x\\ D(b,n)>k(n)^{\frac{1}{2}+\varepsilon}}}k(n)\right),\end{aligned}\tag{6}$$

since $D(b,n)<k(n)$ for all $n\ge 1$.

In order to estimate the last sum, we show that there is a positive integer k_0 independent of ε such that the inequality

$$\sum_{\substack{n<r^k\\ D_k(b,n)>k^{\frac{1}{2}+\varepsilon}}}1<kr^k\exp\left(-\frac{1}{32}k^{2\varepsilon}\right)\tag{7}$$

holds for all $k\ge k_0$, where $D_k(b,n)=\left|F(b,n)-\frac{k}{r}\right|$. Note that in the right-hand side of this equality k is not necessarily identical with $k(n)$ and that if $k=k(n)$ or equivalently $r^{k-1}\le n<r^k$ then $D_k(b,n)=D(b,n)$.

Let j be an integer with $|j|>2$. Then we have

$$\binom{mr}{m+j}(r-1)^{mr-m-j}<r^{mr}\exp\left(-\frac{1}{4mr}j^2\right)\tag{8}$$

for all $m\ge 1$. (For the proof see [3; Lemmas 8.5 and 8.6].) If $k=mr$ we have from (8)

$$\begin{aligned}\sum_{\substack{n<r^{mr}\\ D_{mr}(b,n)>\frac{1}{2}(mr)^{\frac{1}{2}+\varepsilon}}}1&=\sum_{|l-m|>\frac{1}{2}(mr)^{\frac{1}{2}+\varepsilon}}\binom{mr}{l}(r-1)^{mr-l}\\&<r^{mr}\sum_{|j|>\frac{1}{2}(mr)^{\frac{1}{2}+\varepsilon}}\exp\left(-\frac{1}{4mr}j^2\right)\\&<r^{mr}mr\exp\left(-\frac{1}{16}(mr)^{2\varepsilon}\right)\end{aligned}\tag{9}$$

provided that $mr\ge 4$. Next, let $k=mr+q, 1\le q\le r-1$, and let $n=\sum_{i=1}^{k}a_ir^{k-i}=a_1a_2\cdots a_k$ be a positive integer less than r^k developed in the scale of r, where $a_i=0, 1\le i\le l$, if $n<r^{k-l}$. We set $n_0=\sum_{i=1}^{q}a_ir^{k-i}=a_1a_2\cdots a_q0\cdots 0$ and $n_1=\sum_{i=q+1}^{k}a_ir^{k-i}=a_{q+1}\cdots a_k$, so that $n=n_0+n_1$. Then we readily have

(10) $$D_k(b,n)<|F(b,n)-F(b,n_1)|+|F(b,n_1)-m|+\left|m-\frac{k}{r}\right|$$
$$<q+D_{mr}(b,n_1)+1.$$
Take a fixed integer k_0 such that

(11) $$k^{\frac{1}{2}}-r>\frac{1}{2}(mr)^{\frac{1}{2}+\varepsilon}\geq 2$$

for all $k\geq k_0$. Note that such k_0 can be chosen uniformly in ε. Hence, it follows from (10) and (11) that if $k\geq k_0$ then the inequality
$$D_k(b,n)>k^{\frac{1}{2}+\varepsilon}$$
implies
$$D_{mr}(b,n_1)>\frac{1}{2}(mr)^{\frac{1}{2}+\varepsilon}.$$
From this fact together with (9) we find

(12) $$\sum_{\substack{n<r^k\\ D_k(b,n)>k^{\frac{1}{2}+\varepsilon}}} 1<r^q r^{mr} mr\exp\left(-\frac{1}{16}(mr)^{2\varepsilon}\right)<r^k k\exp\left(-\frac{1}{32}k^{2\varepsilon}\right)$$

for all $k\geq k_0$, since we have assumed $0<\varepsilon<\frac{1}{2}$. Inequality (7) follows from (9) and (12).

By (7) we obtain (setting $k(x)=k([x])$)
$$\sum_{\substack{n\leq x\\ D(b,n)>k(n)^{\frac{1}{2}+\varepsilon}}} k(n)\leq\sum_{j=1}^{k(x)} j\sum_{\substack{r^{j-1}\leq n<r^j\\ D(b,n)>j^{\frac{1}{2}+\varepsilon}}} 1<\sum_{j=1}^{k_0-1} jr^j+\sum_{j=k_0}^{k(x)} j^2r^j\exp\left(-\frac{1}{32}j^{2\varepsilon}\right)$$
$$=O(1)+\sum_{k_0\leq j<k(x)/2}+\sum_{k(x)/2\leq j\leq k(x)}$$
$$=O(1)+O(x^{\frac{1}{2}}(\log x)^3)+O\left(x(\log x)^3\exp\left(-\frac{1}{64}\left(\frac{\log x}{\log r}\right)^{2\varepsilon}\right)\right)$$

(13) $$=O\left(x(\log x)^3\exp\left(-\frac{1}{64}\left(\frac{\log x}{\log r}\right)^{2\varepsilon}\right)\right),$$

where the O-constant is uniform in ε.

We now take a constant $B=B(r)$ large enough and then choose $\varepsilon=\varepsilon(x,r)$, $0<\varepsilon<\frac{1}{2}$, in such a way that

(14) $$(\log x)^{2\varepsilon}=B\log\log x.$$

This implies in particular

(15) $$(\log x)^3\exp\left(-\frac{1}{64}\left(\frac{\log x}{\log r}\right)^{2\varepsilon}\right)=O\left(\frac{1}{\log x}\right),$$

and we obtain from (6), (13), (14) and (15)

(16) $$\sum_{p\leq x} D(b,p)=O\left(x\left(\frac{\log\log x}{\log x}\right)^{\frac{1}{2}}\right).$$

Hence, it follows from (4), (5) and (16) that
$$\sum_{p\leq x} F(b,p)=\frac{1}{r}\,\frac{x}{\log r}+O\left(x\left(\frac{\log\log x}{\log x}\right)^{\frac{1}{2}}\right).$$
We have, therefore,

$$\sum_{p\le x}\alpha(p)=\sum_{b=1}^{r-1} b\sum_{p\le x}F(b,p)=\frac{r-1}{2}\,\frac{x}{\log r}+O\Big(x\Big(\frac{\log\log x}{\log x}\Big)^{\frac{1}{2}}\Big).$$

The proof of our theorem is now complete.

Remark. Copeland-Erdös [1] proved that *any increasing sequence of positive integers such that for every $\theta<1$ the number of m_j's up to x exceeds x^{θ} provided x is sufficiently large, is normal, in the sence of E. Borel, in any scale.* This theorem provides the only known proof of the sequence of prime numbers. And, the normality leads to the estimate

$$\sum_{p\le x}F(b,p)=\frac{1}{r}\,\frac{x}{\log r}+o(x).$$

Our proof may be regarded as a refinement of that due to Copeland-Erdös. The error term in the theorem would be replaced by $O(x^{\frac{1}{2}+\varepsilon})$, if there were some '*randomness*' in the distribution of the prime numbers.

The inequality (7) is a slight variant of Lemmas 8.7 and 8.8 in Niven's monograph [3].

Finally the present author should like to express his heartiest thanks to Prof. S. Uchiyama for his valuable advice.

References

[1] A. H. Copeland and P. Erdös: Note on normal numbers. Bull. Amer. Math. Soc., **52**, 857–860 (1946).

[2] I. Katai: On the sum of digits of prime numbers. Ann. Univ. Sci. Budapest Rolando Eötvös nom. Sect. Math., **10**, 89–93 (1967).

[3] I. Niven: Irrational Numbers. The Carus Math. Monogr. No. 11. Math. Assoc. Amer., Washington, D. C. (1956).