# A NOTE ON THE Sopfr($n$) FUNCTION.

Ruslan Sharipov

Abstract. The Sopfr($n$) function is defined as the sum of prime factors of $n$ each of which is taken with its multiplicity. This function is studied numerically. The analogy between Sopfr($n$) and the primes distribution function is drawn and some conjectures for prime numbers formulated in terms of the Sopfr($n$) function are suggested.

## 1. Introduction.

The Sopfr($n$) function is defined as the sum of prime factors of its positive integer argument $n$ (see [1]). For $n = 1$ this function is defined to be equal to zero: Sopfr$(1) = 0$. If $n$ is prime, then Sopfr$(n) = n$. If $n$ is a product of prime numbers

$$n = p_1^{k_1} \cdot \ldots \cdot p_s^{k_s}, \tag{1.1}$$

then Sopfr($n$) is calculated as the sum

$$\mathrm{Sopfr}(n) = k_1\, p_1 + \ldots + k_s\, p_s. \tag{1.2}$$

Note that the prime factors $p_1, \ldots, p_s$ in the sum (1.2) are taken with their multiplicities $k_1, \ldots, k_s$ in the expansion (1.1). Therefore the function Sopfr($n$) is similar to the logarithm. One can easily prove the following identity for it:

$$\mathrm{Sopfr}(n_1 \cdot n_2) = \mathrm{Sopfr}(n_1) + \mathrm{Sopfr}(n_2). \tag{1.3}$$

The Sopfr($n$) function is used in defining Ruth-Aaron pairs named after two famous baseball players George Herman Ruth Jr. and Henry Louis Aaron (see [2]). In mathematics a Ruth-Aaron pair is a pair of consecutive numbers $n$ and $n + 1$ whose sums of prime factors are equal to each other:

$$\mathrm{Sopfr}(n) = \mathrm{Sopfr}(n + 1). \tag{1.4}$$

The numbers 714 and 715 constitute the most famous Ruth-Aaron pair.

Let $x$ be an integer number and let $p$, $q$, $r$, and $s$ be four numbers expressed through $x$ by the following four polynomials:

$$\begin{aligned} p &= 8\,x + 5, & q &= 48\,x^2 + 24\,x - 1,\\ r &= 2\,x + 1, & s &= 48\,x^2 + 30\,x - 1. \end{aligned} \tag{1.5}$$

2000 *Mathematics Subject Classification.* 11N60, 11N64, 11-04.

Using the formulas (1.5), one easily derives that

$$p\,q+1=2^2\,r\,s, \qquad\qquad p+q=2\cdot 2+r+s. \tag{1.6}$$

Due to (1.6) and (1.3), if $p$, $q$, $r$, $s$ all are prime numbers, then the numbers $n=p\,q$ and $n+1=4\,r\,s$ constitute a Ruth-Aaron pair, i. e. they satisfy the equality (1.4). Schinzel's H-conjecture (see [3], [4], and [5]) implies that there are infinitely many integer numbers $x$ such that the numbers $p$, $q$, $r$, and $s$ given by the polynomials (1.5) all are prime.

In this paper we treat $\operatorname{Sopfr}(n)$ as an analog of the primes distribution function $\pi(n)$. The value $\pi(n)$ of this function is defined as the number of positive primes less than or equal to $n$. Gauss and Legendre (see [6]) in 1792–1808 conjectured the following asymptotic behavior of the function $\pi(n)$:

$$\pi(n)\sim\frac{n}{\ln(n)}\quad\text{as}\quad n\to\infty. \tag{1.7}$$

In 1849 and in 1852 P. L. Chebyshev proved two propositions very close to (1.7). The proposition (1.7) itself was proved in 1896 by Hadamard [7] and Valée Poussin [8]. See [9] for the modern explanation of their proof.

The main goal of this paper is to study the function $\operatorname{Sopfr}(n)$ numerically and formulate some conjectures similar to (1.7) for this function.

## 2. The averaged $\operatorname{Sopfr}(n)$ function.

The $\operatorname{Sopfr}(n)$ function is quite irregular. Looking at its graph (see [1]), one can find that the values of $\operatorname{Sopfr}(n)$ resemble random numbers. In order to make them more regular we average them over intervals between two consecutive squares:

$$A(n)=\sum_{i=n^2+1}^{(n+1)^2}\frac{\operatorname{Sopfr}(i)}{(n+1)^2-n^2}. \tag{2.1}$$

The function $A(n)$ in (2.1) is the averaged $\operatorname{Sopfr}(n)$ function. We study its values in two intervals $1\leqslant n\leqslant 998$ and $1000\leqslant n\leqslant 3161$. The graph of the function (2.1) in the first interval $1\leqslant n\leqslant 998$ is shown in Fig. 2.1. It is presented by a sequence of points whose coordinates are rendered in logarithmic scale, i. e. $A_n=(x_n,y_n)$, where $x_n=\ln(A(n))$ and $y_n=\ln(n)$.

Looking at Fig. 2.1 below, one can see that the points $A_n$ with $n\geqslant 122\approx e^{4.8}$ are approximated by a straight line. We write the equation of this straight line as

$$x=\alpha\,y+\beta. \tag{2.2}$$

In order to calculate the parameters $\alpha$ and $\beta$ in (2.2) we use the root mean squares method. For this purpose we use the following quadratic deviation function:

$$F(\alpha,\beta)=\sum_{n=122}^{998}(x_n-\alpha\,y_n-\beta)^2. \tag{2.3}$$

The quadratic function (2.3) has exactly one minimum point $\alpha=\alpha_{\text{min}}$, $\beta=\beta_{\text{min}}$.

This minimum point is determined by the following linear equations:

$$\frac{\partial F(\alpha,\beta)}{\partial \alpha} = 0, \qquad \frac{\partial F(\alpha,\beta)}{\partial \beta} = 0. \tag{2.4}$$

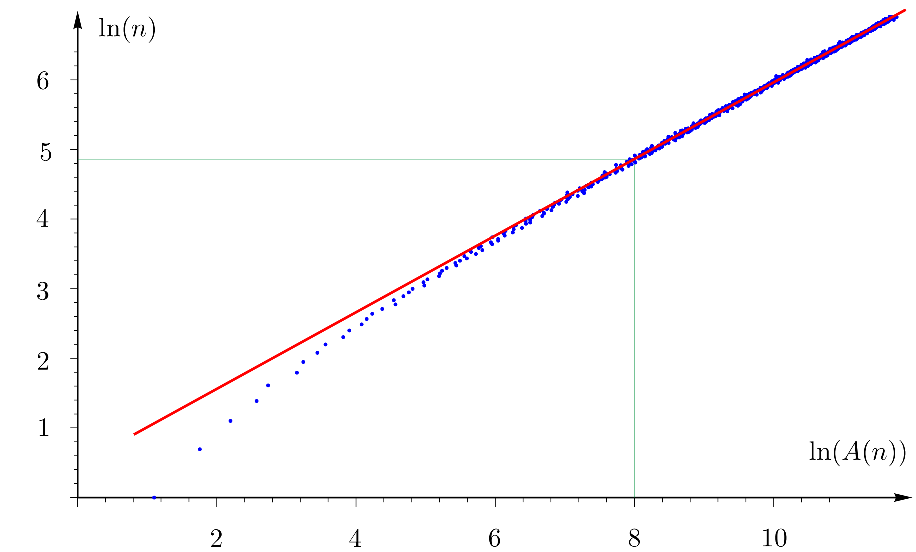


*Fig 2.1*

The function (2.3) and the equations (2.4) are computed numerically. Solving them,

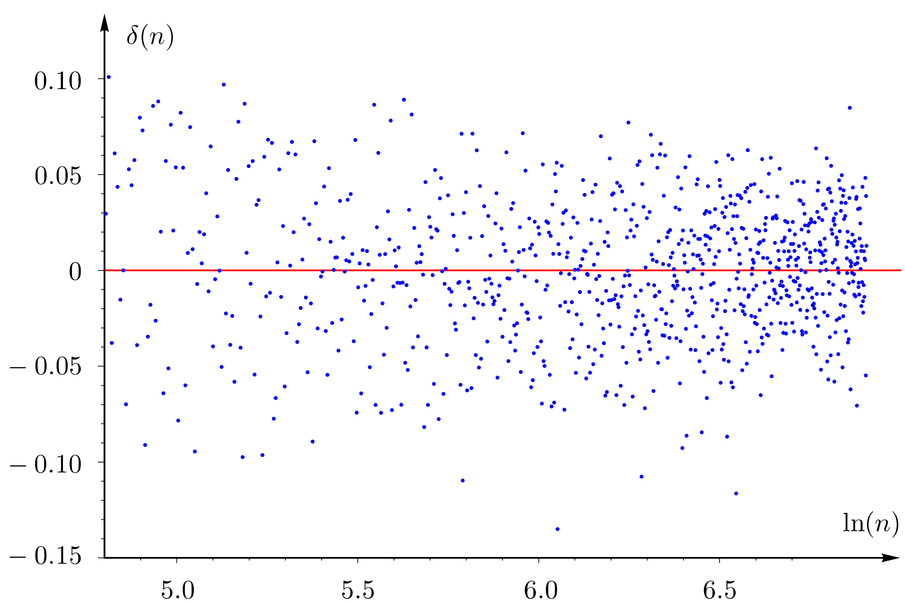


*Fig 2.2*

we find the numeric values of $\alpha$ and $\beta$ at the minimum of the function $F(\alpha,\beta)$:

$$\alpha \approx 1.820, \qquad\qquad \beta \approx -0.847. \tag{2.5}$$

Having calculated the constants (2.5), now we draw the graph of the deviation function $\delta(n) = \ln(A(n)) - \alpha\,\ln(n) - \beta$ in logarithmic scale. The graph of the function $\delta(n)$ in Fig. 2.2 is presented by a series of points $A_n = (x_n, y_n)$, where $x_n = \ln(n)$ and $y_n = \delta(n)$. Looking at this graph, we derive the following inequality for the deviation function $\delta(n)$:

$$-\delta_1 < \delta(n) < \delta_1, \ \text{ where } \ \delta_1 = 0.15 \ \text{ and } \ 122 \leqslant n \leqslant 998. \tag{2.6}$$

The next interval is $1000 \leqslant n \leqslant 3161$. The graph of the function (2.1) in this interval is shown in Fig. 2.3. Again it is presented by a sequence of points whose coordinates are rendered in logarithmic scale, i. e. $A_n = (x_n, y_n)$, where $x_n = \ln(A(n))$ and $y_n = \ln(n)$. The graph in Fig. 2.3 is also approximated by a

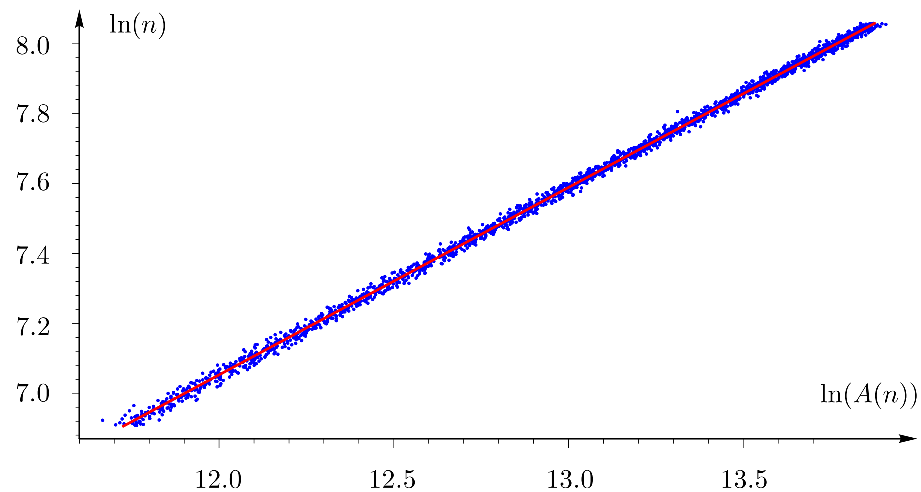


*Fig 2.3*

straight line. This straight line is given by the equation (2.2). The coefficients $\alpha$ and $\beta$ in this case are calculated by solving the equations (2.4) for the following quadratic deviation function, which is similar to (2.3):

$$F(\alpha,\beta) = \sum_{n=1000}^{3161} (x_n - \alpha\,y_n - \beta)^2. \tag{2.7}$$

The minimum of the function (2.7) corresponds to the following values of $\alpha$ and $\beta$:

$$\alpha \approx 1.860, \qquad\qquad \beta \approx -1.115. \tag{2.8}$$

The sharpness of the approximation of $A(n)$ by the straight line in Fig. 2.3 is expressed through the deviation function $\delta(n) = \ln(A(n)) - \alpha\,\ln(n) - \beta$, where $\alpha$

and $\beta$ are given by the formulas (2.8):

$$-\delta_2 < \delta(n) < \delta_1, \ \text{ where } \ \delta_2 = 0.1 \ \text{ and } \ 1000 \leqslant n \leqslant 3161. \tag{2.9}$$

The inequalities (2.9) are similar to the above inequalities (2.6). They are derived by drawing the graph of the function $\delta(n)$. This graph is shown in Fig. 2.4 below.

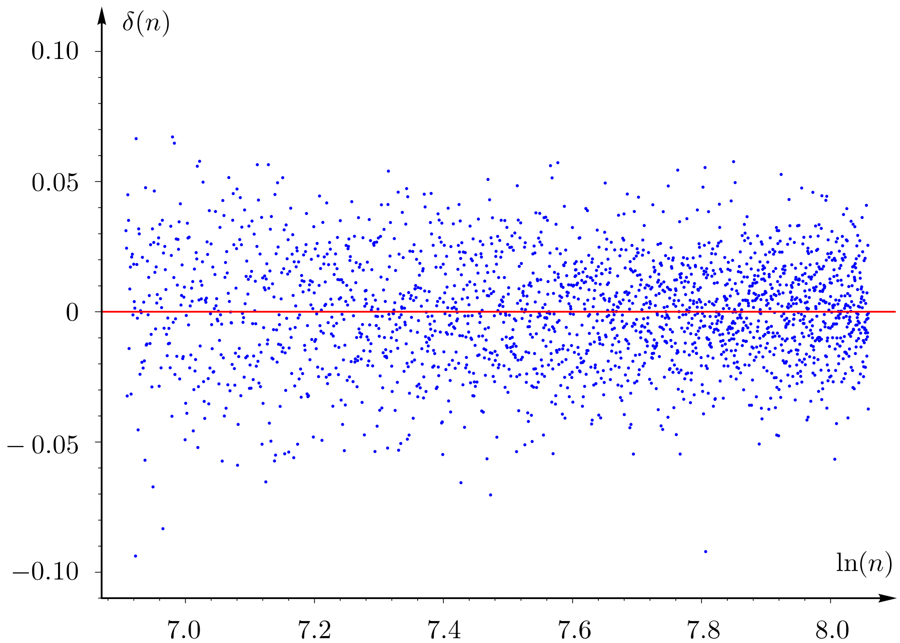


*Fig 2.4*

## 3. Approximation conjectures.

Note that the parameter $\alpha$ in (2.8) is greater than $\alpha$ in (2.5). This mean that the slope of the straight line approximating the graph of the function $A(n)$ slightly grows as $n \to \infty$. To take into account this growth we replace the linear approximation in (2.2) by a nonlinear one. We choose the following formula for it

$$x = \alpha\, y + \beta + \gamma\, \ln(y) + \lambda\, e^{-y} + \mu\, e^{-2\,y}. \tag{3.1}$$

The choice of (3.1) means that $A(n)$ is approximated by the formula

$$A(n) \approx B\, n^{\alpha}\, (\ln n)^{\gamma}\, \exp\!\left(\frac{\lambda}{n} + \frac{\mu}{n^2}\right), \ \text{ where } \ B = e^{\beta}. \tag{3.2}$$

In order to find the optimal values of the parameters $\alpha$, $\beta$, $\gamma$, $\lambda$, and $\mu$ for the approximation (3.2) we apply the root mean squares method. Instead of (2.3) and (2.7) in this case we use the following deviation function:

$$F = \sum_{n=4}^{3161} (x_n - \alpha\, y_n - \beta - \gamma\, \ln y_n - \lambda\, e^{-y_n} - \mu\, e^{-2\,y_n})^2. \tag{3.3}$$

Remember that $x_n = \ln(A(n)$ and $y_n = \ln(n)$ in (3.3). The optimal values of $\alpha$, $\beta$, $\gamma$, $\lambda$, and $\mu$ are determined by solving the equations

$$\frac{\partial F}{\partial \alpha} = 0, \qquad \frac{\partial F}{\partial \beta} = 0, \qquad \frac{\partial F}{\partial \gamma} = 0, \qquad \frac{\partial F}{\partial \lambda} = 0, \qquad \frac{\partial F}{\partial \mu} = 0. \tag{3.4}$$

The equations (3.4) are similar to the equations (2.4). Here is their solution:

$$\alpha \approx 2.001, \quad \beta \approx -0.047, \quad \gamma \approx -1.056, \quad \lambda \approx 1.187, \quad \mu \approx -2.240. \tag{3.5}$$

The exponential factor with $\lambda$ and $\mu$ in (3.2) is a decreasing function of $n$. For this reason we consider the function

$$B(n) = \frac{A(n)}{n^{\alpha}\,(\ln n)^{\gamma}}$$

and draw its graph. Like the graph of $A(n)$, it is presented as a sequence of points:

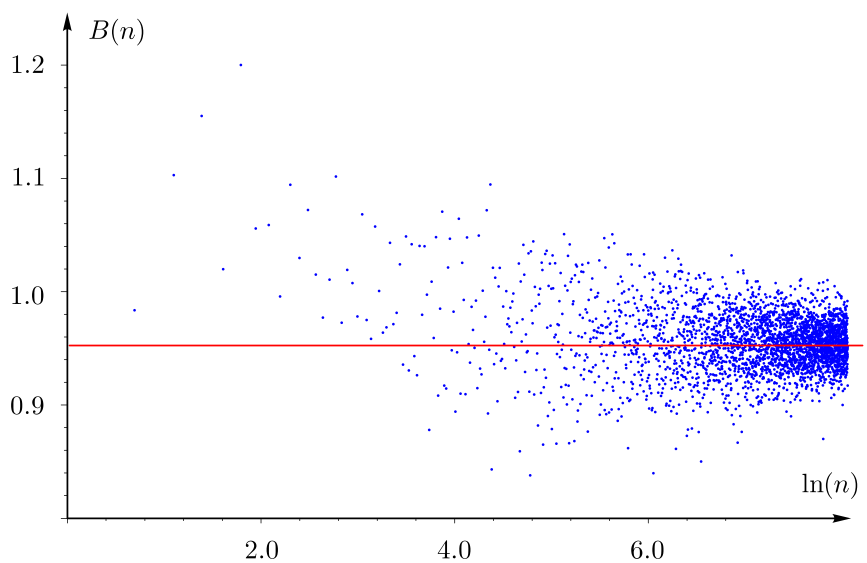


*Fig 3.1*

Looking at the graph in Fig. 3.1, we can formulate the following conjecture.

**Conjecture 3.1 (weak** Sopfr$(n)$ **conjecture).** *There are four constants $\alpha$, $\gamma$, $B_1$ and $B_2$ such that the averaged* Sopfr*-function $A(n)$ in (2.1) obey the inequalities*

$$B_1\, n^{\alpha}\,(\ln n)^{\gamma} \leqslant A(n) \leqslant B_2\, n^{\alpha}\,(\ln n)^{\gamma} \quad \text{for all} \quad n > 1. \tag{3.6}$$

The graph points in Fig. 3.1 condense to a band as $n \to \infty$. Its width is restricted by the constants $B_1$ and $B_2$ in (3.6). The width of this band can vanish at infinity. For this option we can formulate the following conjecture.

**Conjecture 3.2 (strong** Sopfr($n$) **conjecture).** *There are three constants* $\alpha$, $\gamma$, *and* $B$ *such that* $B > 0$ *and the following condition is fulfilled:*

$$A(n) \sim B\, n^{\alpha}\, (\ln n)^{\gamma} \quad \text{as} \quad n \to \infty. \tag{3.7}$$

Note that the constants $\alpha$ and $\gamma$ in (3.5) are are very close to integer numbers. Therefore we can formulate another conjecture.

**Conjecture 3.3.** *The constants* $\alpha$ *and* $\gamma$ *either in* (3.6) *or in* (3.7) *are explicit numbers* $\alpha = 2$ *and* $\gamma = -1$.

The averaged Sopfr($n$) function (2.1) is similar to the primes distribution function. The above formulas (3.6) and (3.7) are similar to the formula (1.7).

## References

1. Weisstein E. W., *Sum of Prime Factors*, MathWorld, Wolfram web resource http://mathworld.wolfram.com/SumofPrimeFactors.html.
2. Nelson C., Penney D.E., Pomerance C., *714 and 715*, Journal of Recreational Mathematics **7** (1974), no. 2, 87–89.
3. Shinzel A., Sierpiński C., *Sur certaines hypothèses concernant les nombres premiers*, Acta Arithm. **4** (1958), 185–208.
4. Shinzel A., Sierpiński C., *Erratum*, Acta Arithm. **5** (1959), 259.
5. Weisstein E. W., *Schinzel's Hypothesis*, MathWorld, Wolfram web resource http://mathworld.wolfram.com/SchinzelsHypothesis.html.
6. Legendre A. M., *Essai sur la théorie des nombres,* Seconde Édition, Chez Courcier, Imprimeur-Libraire pour les Mathématiques, quai des Augustins, $n^{\circ}$ 57, Paris, 1808.
7. Hadamard J., *Sur la distribution des zéros de la fonction* $\zeta(s)$ *et ses conséquences Arithmétiques*, Bull. Soc. Math. France **24** (1896), 199–220.
8. De la Vallee Poussin C.-J., *Recherches analytiques sur la theorie des nombres.* Premiere partie: *la fonction* $\zeta(s)$ *de Riemann et les nombres premiers en general*, Ann. Soc. Sci. Bruxelles 20 (1896), 183–256.
9. Tenenbaum G., France M. M., *The prime numbers and their distribution*, American Math. Soc., USA, 2000.

Bashkir State University, 32 Zaki Validi street, 450074 Ufa, Russia
*E-mail address*: r-sharipov@mail.ru