#A57 INTEGERS 12 (2012)

ON THE DIOPHANTINE EQUATION $\prod x_i = \sum x_i$

Andreas Weingartner
Department of Mathematics, Southern Utah University, Cedar City, Utah
weingartner@suu.edu

Received: 2/7/12, Accepted: 10/2/12, Published: 10/19/12

Abstract

For $n \geq 2$, let $f(n)$ denote the number of positive integer solutions to the equation $\prod_{i=1}^n x_i = \sum_{i=1}^n x_i$ with $x_1 \geq x_2 \geq \ldots \geq x_n$. We establish an asymptotic formula for the average order of $f(n)$.

1. Introduction

For $n \geq 2$, the equation (see [3, Problem D24])

$$\prod_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i \qquad (x_1 \geq x_2 \geq \ldots \geq x_n \geq 1) \tag{1}$$

has the solution $x_1 = n$, $x_2 = 2$, $x_3 = x_4 = \ldots = x_n = 1$. This is the only solution in positive integers when $n = 2, 3, 4, 6, 24, 114, 174$ and 444. It seems likely that these are the only values of n for which the solution is unique. Misiurewicz [4] verified this for $n \leq 10^3$, Brown [1] for $n \leq 5 \cdot 10^4$ and the author for $n \leq 10^{11}$. This problem seems to have first been asked by Trost [7] in connection with the problem of solving $\prod_{i=1}^n x_i = \sum_{i=1}^n x_i = 1$ in rationals.

For $n \geq 2$, let $f(n)$ be the number of solutions to (1), and let $f_k(n)$ denote the number of solutions to

$$\prod_{i=1}^{k} x_i - \sum_{i=1}^{k} (x_i - 1) = n, \qquad (x_1 \geq x_2 \geq \ldots \geq x_k \geq 1). \tag{2}$$

Any solution $(x_1, \ldots, x_k)$ of (2) with $k \leq n$ implies the solution of (1) given by $(x_1, \ldots, x_k, 1, \ldots, 1)$. Thus $f(n) \geq f_k(n)$ for $n \geq k$.

When $k = 2$, equation (2) simplifies to $(x_1 - 1)(x_2 - 1) = n - 1$, and hence the solution $x_1 = n$, $x_2 = 2$ is the only solution with $k = 2$ if and only if $n - 1$ is prime. That is $f_2(n) = 1$ if and only if $n - 1$ is prime. We clearly have $f_2(n) = \lfloor (d(n-1) + 1)/2 \rfloor$, where $d(n-1)$ is the number of divisors of $n - 1$, so that $\liminf f_2(n) = 1$ and $\limsup f_2(n) = \infty$, and therefore $\limsup f(n) = \infty$.

Computer experiments suggests that $f_4(n) = 1$ for only a finite number of n. Indeed, the 16 values of $n \leq 10^{11}$, for which there is no second solution to (2) with $k = 4$, are $n = 2, 3, 4, 6, 24, 42, 114, 174, 192, 252, 420, 444, 594, 6324, 27744$.

When $k = 3$, the numerical evidence is not as clear, since initially there are many n for which $f_3(n) = 1$, but it seems that even in this case there may only be finitely many such n.

Question 1. Is $\liminf_{n\to\infty} f(n) > 1$?

Question 2. Is $\liminf_{n\to\infty} f_k(n) > 1$ for $k \geq 3$?

We conjecture that both questions have a positive answer. Furthermore, it seems likely that each one of these limits inferior is ∞. An attempt to use the circle method to show that $\lim_{n\to\infty} f_3(n) = \infty$ was not successful, because we were not able to show that the contribution from the minor arcs was less than that from the major arcs.

It is not difficult to verify that for $n > 6$, $f(n) = 1$ implies $n \equiv 0 \pmod{30}$ or $n \equiv 24 \pmod{30}$: If $n - 1$ is composite, say $n - 1 = xy$, then $(x + 1, y + 1, 1, \ldots, 1)$ is a second solution to (1); if $2n - 1$ is composite, say $2n - 1 = xy$, then $(\frac{x+1}{2}, \frac{y+1}{2}, 2, 1, \ldots, 1)$ is a second solution; finally, if $n = 15x - 3$ for some $x \geq 1$, then $(x, 2, 2, 2, 2, 1, \ldots, 1)$ is a second solution.

Schinzel conjectured (see [2, p. 238]) that there is a k such that

$$\prod_{i=1}^{k} x_i - \sum_{i=1}^{k} x_i = n, \qquad (x_1 \geq x_2 \geq \ldots \geq x_k \geq 2) \tag{3}$$

has a solution for every sufficiently large n. This would imply $\liminf_{n\to\infty} f_k(n) > 1$. Viola [8] used the large sieve to show that $E_k(x)$, the number of $n \leq x$ such that (3) has no solution, satisfies

$$E_k(x) \ll x \exp\left\{-C(k)(\log x)^{1-\frac{1}{k-1}}\right\}.$$

To answer the above questions in the affirmative, we would want to show that $E_k(x) \ll 1$, which seems completely out of reach.

The average number of solutions is easier to estimate. To establish the following result, we make use of an asymptotic formula for the average of $d_k(n)$, the general divisor function.

Theorem 3. *Let $k \geq 2$ be fixed. For $x \geq 2$ we have*

$$\frac{1}{x} \sum_{2\leq n\leq x} f_k(n) = \frac{(\log x)^{k-1}}{k!(k-1)!}\left(1 + O\left(\frac{1}{\log x}\right)\right).$$

Throughout this paper, the implied constant in the error term may depend on k. We will derive the following result from an estimate for the average number of solutions in the Factorisatio Numerorum problem.

Theorem 4. *Let $m \geq 2$ be fixed. There are constants $c_1, \ldots, c_{m-1}$, such that, for $x \geq 2$,*

$$\frac{1}{x} \sum_{2 \leq n \leq x} f(n) = \frac{\exp(2\sqrt{\log x})}{2\sqrt{\pi}(\log x)^{3/4}} \left(1 + \sum_{j=1}^{m-1} \frac{c_j}{(\log x)^{j/2}} + O\left(\frac{1}{(\log x)^{m/2}} \right) \right).$$

2. Proof of Theorem 3

Let $d_k(n)$ denote the number of ordered k-tuples of positive integers whose product is n. That is,

$$d_k(n) := \left| \left\{ (x_1, \ldots, x_k) \in \mathbb{N}^k : \prod_{i=1}^{k} x_i = n \right\} \right|.$$

The study of the average value of $d_k(n)$ is a classic problem in number theory. For the following estimate see for example [5, Section 2.1, Exercise 18].

Lemma 5. *Let $k \geq 2$ be fixed. For $x \geq 2$ we have*

$$D_k(x) := \sum_{1 \leq n \leq x} d_k(n) = x P_k(\log x) + O\left(x^{1-1/k} (\log x)^{k-2} \right),$$

where P_k is a polynomial with degree $k-1$ and leading coefficient $1/(k-1)!$.

For $n \geq 2$, let

$$g_k(n) := \left| \left\{ (x_1, \ldots, x_k) \in \mathbb{N}^k : \prod_{i=1}^{k} x_i - \sum_{i=1}^{k} (x_i - 1) = n \right\} \right|,$$

and

$$G_k(x) := \sum_{2 \leq n \leq x} g_k(n) = \left| \left\{ (x_1, \ldots, x_k) \in \mathbb{N}^k : 2 \leq \prod_{i=1}^{k} x_i - \sum_{i=1}^{k} (x_i - 1) \leq x \right\} \right|.$$

Lemma 6. *Let $(x_1, \ldots, x_k) \in \mathbb{N}^k$. We have*

(i) $\prod_{i=1}^{k} x_i - \sum_{i=1}^{k} (x_i - 1) \geq 1$.

(ii) $\prod_{i=1}^{k} x_i - \sum_{i=1}^{k} (x_i - 1) = 1$ *if and only if at most one of the factors x_i exceeds 1.*

(iii) *If $2 \leq \prod_{i=1}^{k} x_i - \sum_{i=1}^{k} (x_i - 1) \leq x$ then $\prod_{i=1}^{k} x_i \leq 2x$.*

Proof. (i) The inequality is clearly satisfied when $k = 1$. We proceed by induction on k:

$$\sum_{i=1}^{k}(x_i - 1) = \sum_{i=1}^{k-1}(x_i - 1) + (x_k - 1) \leq \prod_{i=1}^{k-1} x_i - 1 + (x_k - 1)$$
$$= \prod_{i=1}^{k} x_i - 1 + \left(\prod_{i=1}^{k-1} x_i - 1\right)(1 - x_k) \leq \prod_{i=1}^{k} x_i - 1.$$

(ii) Let $s = \#\{i \leq k : x_i \geq 2\}$. If $s = 0$ or $s = 1$ we have $\prod_{i=1}^{k} x_i - \sum_{i=1}^{k}(x_i - 1) = 1$. Let $s \geq 2$ and assume $x_1 \geq x_2 \geq \ldots \geq x_k$. Then

$$\prod_{i=1}^{k} x_i - \sum_{i=1}^{k}(x_i - 1) = \prod_{i=1}^{s} x_i - \sum_{i=1}^{s}(x_i - 1)$$
$$\geq x_1 2^{s-1} - s(x_1 - 1) = x_1(2^{s-1} - s) + s \geq s \geq 2.$$

(iii) With s as in part (ii) we first show that

$$\prod_{i=1}^{s} x_i \geq 2\sum_{i=1}^{s}(x_i - 1), \quad (s \geq 2).$$

For $s = 2$, this is equivalent to $(x_1 - 2)(x_2 - 2) \geq 0$. We proceed by induction on s with the assumption that $x_1 \geq x_2 \geq \ldots \geq x_s$:

$$\prod_{i=1}^{s} x_i = x_s \prod_{i=1}^{s-1} x_i \geq 2x_s \sum_{i=1}^{s-1}(x_i - 1) = 2(x_s - 1)\sum_{i=1}^{s-1}(x_i - 1) + 2\sum_{i=1}^{s-1}(x_i - 1)$$
$$\geq 2(x_s - 1) + 2\sum_{i=1}^{s-1}(x_i - 1) = 2\sum_{i=1}^{s}(x_i - 1).$$

Part (ii) shows that the assumptions of (iii) imply $s \geq 2$. Thus

$$\prod_{i=1}^{k} x_i \leq x + \sum_{i=1}^{k}(x_i - 1) \leq x + \frac{1}{2}\prod_{i=1}^{k} x_i,$$

which gives the desired conclusion. □

Proposition 7. *Let $k \geq 2$ be fixed. For $x \geq 2$ we have $G_k(x) = D_k(x) + O(x)$.*

Proof. We have

$$G_k(x) = \left|\left\{(x_1, \ldots, x_k) \in \mathbb{N}^k : 2 \leq \prod_{i=1}^{k} x_i - \sum_{i=1}^{k}(x_i - 1) \leq x\right\}\right|$$
$$\geq \left|\left\{(x_1, \ldots, x_k) \in \mathbb{N}^k : 2 \leq \prod_{i=1}^{k} x_i - \sum_{i=1}^{k}(x_i - 1) \text{ and } \prod_{i=1}^{k} x_i \leq x\right\}\right|.$$

By Lemma 6, $\prod_{i=1}^k x_i - \sum_{i=1}^k (x_i - 1) < 2$ if and only if at most one of the x_i is greater than 1. There are at most kx such k-tuples with $\prod_{i=1}^k x_i \le x$. Thus

$$G_k(x) \ge -kx + \left| \left\{ (x_1, \ldots, x_k) \in \mathbb{N}^k : \prod_{i=1}^k x_i \le x \right\} \right| = -kx + D_k(x).$$

For the upper bound we write $G_k(x) \le D_k(x + 2k\sqrt{x}) + A$, where

$$\begin{aligned} A := & \Bigg| \Bigg\{ (x_1, \ldots, x_k) \in \mathbb{N}^k : \\ & \quad 2 \le \prod_{i=1}^k x_i - \sum_{i=1}^k (x_i - 1) \le x, \prod_{i=1}^k x_i > x, \ \sum_{i=1}^k (x_i - 1) > 2k\sqrt{x} \ \Bigg\} \Bigg| \\ & \le k \left| \left\{ (x_1, \ldots, x_k) \in \mathbb{N}^k : 2 \le \prod_{i=1}^k x_i - \sum_{i=1}^k (x_i - 1) \le x, \prod_{i=1}^k x_i > x, \ x_k > 2\sqrt{x} \ \right\} \right| \\ & \le k \sum_{2\sqrt{x} < x_k \le x} \left| \left\{ (x_1, \ldots, x_{k-1}) \in \mathbb{N}^{k-1} : \frac{x}{x_k} < \prod_{i=1}^{k-1} x_i \le \frac{x + \sum_{i=1}^k (x_i - 1)}{x_k} \right\} \right|. \end{aligned}$$

Note that

$$\frac{\sum_{i=1}^k (x_i - 1)}{x_k} < 1 + \frac{\sum_{i=1}^{k-1} (x_i - 1)}{x_k} \le 1 + \frac{\prod_{i=1}^{k-1} x_i}{x_k} \le 1 + \frac{2x}{x_k^2} \le \frac{3}{2},$$

by Lemma 6. Thus

$$\begin{aligned} A &\le k \sum_{2\sqrt{x} < x_k \le x} \left(d_{k-1} \left(\left\lfloor \frac{x}{x_k} + 1 \right\rfloor \right) + d_{k-1} \left(\left\lfloor \frac{x}{x_k} + 2 \right\rfloor \right) \right) \\ &\ll k \sum_{2\sqrt{x} < x_k \le x} \left(\frac{x}{x_k} \right)^{1/2} \ll kx. \end{aligned}$$

□

Proposition 8. *Let $k \ge 2$ be fixed and let P_k be as in Lemma 5. For $x \ge 2$ we have $G_k(x) = xP_k(\log x) + O(x)$.*

Proof. The result follows from Lemma 5 and Proposition 7. □

Proof of Theorem 3. From Proposition 8 we have

$$x \, \frac{(\log x)^{k-1}}{(k-1)!} \left(1 + O\left(\frac{1}{\log x} \right) \right) = \sum_{2 \le n \le x} g_k(n) = C + D,$$

where

$$\begin{aligned}
C &:= \left|\left\{(x_1,\ldots,x_k)\in\mathbb{N}^k : 2 \le \prod_{i=1}^{k} x_i - \sum_{i=1}^{k}(x_i-1) \le x,\ \exists i\neq j \text{ with } x_i = x_j\right\}\right| \\
&\le \binom{k}{2}\left|\left\{(x_1,\ldots,x_k)\in\mathbb{N}^k : 2 \le \prod_{i=1}^{k} x_i - \sum_{i=1}^{k}(x_i-1) \le x,\ x_1 = x_2\right\}\right| \\
&= \binom{k}{2}\left|\left\{(x_2,\ldots,x_k)\in\mathbb{N}^{k-1} : \right.\right. \\
&\qquad \left.\left. 2-(x_2-1)^2 \le x_2^2\prod_{i=3}^{k} x_i - (x_2^2-1) - \sum_{i=3}^{k}(x_i-1) \le x-(x_2-1)^2\right\}\right| \\
&\le E + F,
\end{aligned}$$

with

$$\begin{aligned}
E &:= \binom{k}{2}\left|\left\{(x_2,\ldots,x_k)\in\mathbb{N}^{k-1} : 2 \le x_2^2\prod_{i=3}^{k} x_i - (x_2^2-1) - \sum_{i=3}^{k}(x_i-1) \le x\right\}\right| \\
&= O\left(x(\log x)^{k-2}\right)
\end{aligned}$$

by Proposition 8, and

$$\begin{aligned}
F &:= \binom{k}{2}\left|\left\{(x_2,\ldots,x_k)\in\mathbb{N}^{k-1} : \right.\right. \\
&\qquad \left.\left. x_2 > 1,\ 1 = x_2^2\prod_{i=3}^{k} x_i - (x_2^2-1) - \sum_{i=3}^{k}(x_i-1) \le x-(x_2-1)^2\right\}\right| \\
&\le \binom{k}{2}\sqrt{x}.
\end{aligned}$$

Note that when estimating F we must have $x_3 = \ldots = x_k = 1$. Thus $C = O\left(x(\log x)^{k-2}\right)$ and

$$\begin{aligned}
D &:= \left|\left\{(x_1,\ldots,x_k)\in\mathbb{N}^k : 2 \le \prod_{i=1}^{k} x_i - \sum_{i=1}^{k}(x_i-1) \le x,\ x_i \neq x_j \text{ for all } i\neq j\right\}\right| \\
&= x\,\frac{(\log x)^{k-1}}{(k-1)!}\left(1+O\left(\frac{1}{\log x}\right)\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
G &:= \left|\left\{(x_1,\ldots,x_k)\in\mathbb{N}^k : 2 \le \prod_{i=1}^{k} x_i - \sum_{i=1}^{k}(x_i-1) \le x,\ x_1 > x_2 > \ldots > x_k\right\}\right| \\
&= x\,\frac{(\log x)^{k-1}}{k!(k-1)!}\left(1+O\left(\frac{1}{\log x}\right)\right),
\end{aligned}$$

which concludes the proof since $\sum_{2\leq n\leq x} f_k(n) = G + O(C)$. □

3. Proof of Theorem 4

Let $h(n)$ be the number of factorizations of the positive integer n into factors greater than 1, without regard to order. For example, $h(28) = 4$ since $28 = 2\cdot 14 = 4\cdot 7 = 2\cdot 2\cdot 7$. The following result is due to Oppenheim [6].

Lemma 9. *Let $m \geq 1$ be fixed. There are constants $c_1, c_2, \ldots, c_{m-1}$, such that, for $x \geq 2$,*

$$\frac{1}{x}\sum_{1\leq n\leq x} h(n) = \frac{\exp(2\sqrt{\log x})}{2\sqrt{\pi}(\log x)^{3/4}}\left(1 + \sum_{j=1}^{m-1}\frac{c_j}{(\log x)^{j/2}} + O\left(\frac{1}{(\log x)^{m/2}}\right)\right).$$

Proof of Theorem 4. We have

$$\begin{aligned}
&\sum_{2\leq n\leq x} f(n)\\
&\quad= \left|\left\{(x_1,\ldots,x_k)\in\mathbb{N}^k : k\geq 2,\ \prod_{i=1}^{k} x_i - \sum_{i=1}^{k}(x_i-1)\leq x,\ x_1\geq\ldots\geq x_k\geq 2\right\}\right|\\
&\quad\geq \left|\left\{(x_1,\ldots,x_k)\in\mathbb{N}^k : k\geq 1,\ \prod_{i=1}^{k} x_i\leq x,\ x_1\geq\ldots\geq x_k\geq 2\right\}\right|\\
&\qquad\qquad - \left|\left\{(x_1,\ldots,x_k)\in\mathbb{N}^k : k=1,\ x_1\leq x,\quad x_1\geq 2\right\}\right|\\
&\quad\geq -x + \sum_{1\leq n\leq x} h(n)\\
&\quad= \frac{x\ \exp(2\sqrt{\log x})}{2\sqrt{\pi}(\log x)^{3/4}}\left(1 + \sum_{j=1}^{m-1}\frac{c_j}{(\log x)^{j/2}} + O\left(\frac{1}{(\log x)^{m/2}}\right)\right),
\end{aligned}$$

by Lemma 9. For the upper bound we write

$$\begin{aligned}
&\sum_{2\leq n\leq x} f(n)\\
&\quad= \left|\left\{(x_1,\ldots,x_k)\in\mathbb{N}^k : k\geq 2,\ \prod_{i=1}^{k} x_i - \sum_{i=1}^{k}(x_i-1)\leq x,\ x_1\geq\ldots\geq x_k\geq 2\right\}\right|
\end{aligned}$$

$$
\begin{aligned}
&\le \left|\left\{(x_1,\dots,x_k)\in\mathbb{N}^k : k\ge 1,\ \prod_{i=1}^k x_i \le x+\sqrt{x},\quad x_1\ge x_2\ge\ldots\ge x_k\ge 2\right\}\right| \\
&\qquad + \left|\left\{(x_1,\dots,x_k)\in\mathbb{N}^k : k\ge 2,\ \prod_{i=1}^k x_i - \sum_{i=1}^k (x_i-1)\le x,\right.\right. \\
&\qquad\qquad\qquad \left.\left.\sum_{i=1}^k (x_i-1) > \sqrt{x},\quad x_1\ge x_2\ge\ldots\ge x_k\ge 2\right\}\right| \\
&=: A+B,
\end{aligned}
$$

say. Let $H(x) := \sum_{1\le n\le x} h(n)$. We have

$$
A = H(x+\sqrt{x}) = \frac{x\ \exp(2\sqrt{\log x})}{2\sqrt{\pi}(\log x)^{3/4}}\left(1+\sum_{j=1}^{m-1}\frac{c_j}{(\log x)^{j/2}} + O\left(\frac{1}{(\log x)^{m/2}}\right)\right),
$$

by Lemma 9. To estimate B, note that $\prod_{i=1}^k x_i \le 2x$ by Lemma 6, hence $x_1 < x$ as $k\ge 2$, $x_i\ge 2$. Moreover, $x_1 > \sqrt{x}/k$ with $k\le \log_2 2x$. Thus

$$
\begin{aligned}
B &\le \left|\left\{(x_1,\dots,x_k)\in\mathbb{N}^k : \prod_{i=1}^k x_i \le 2x,\ x_1 > \frac{\sqrt{x}}{\log_2 2x},\ x_1\ge x_2\ge\ldots\ge x_k\ge 2\right\}\right| \\
&\le \sum_{\sqrt{x}/\log_2 2x < x_1\le x} H\left(\frac{2x}{x_1}\right) \\
&\ll \sum_{2x^{1/3}<x_1\le x} \frac{x\ \exp\left(2\sqrt{\log(x^{2/3})}\right)}{x_1} \ll x(\log x)\exp\left(2\sqrt{\log(x^{2/3})}\right),
\end{aligned}
$$

which can be absorbed into the error term. □

References

[1] M. L. Brown, On the Diophantine Equation $\sum X_i = \prod X_i$, *Math. Comp.* 42 (1984), no. 165, 239–240.

[2] P. Erdős, Some unsolved problems, *Publ. Math. Inst. Hung. Acad. Sci.* 6 (1961), 221–254.

[3] R. Guy, *Unsolved Problems in Number Theory*, Springer, 2004.

[4] M. Misiurewicz, Ungelöste Probleme, *Elem. Math.* 21 (1966), 90.

[5] H. L. Montgomery and R. C. Vaughan, *Multiplicative Number Theory I. Classical Theory*, Cambridge University Press, 2007.

[6] A. Oppenheim, On an arithmetic function II, *J. London Math. Soc.* 2 (1927), 123–130.

[7] E. Trost, Ungelöste Probleme, Nr. 14, *Elem. Math.* 11 (1956), 134–135.

[8] C. Viola, On the diophantine equations $\prod_0^k x_i - \sum_0^k x_i = n$ and $\sum_0^k \frac{1}{x_i} = \frac{a}{n}$, *Acta Arith.*, 22 (1972/3) 339–352.