

# On the random variable $w_a$ and its connection with the Dirichlet divisor problem

Dmitry Pyatin / dmpyatin@petrsu.ru

03.07.2020

**Annotation:** In this paper, we study a discrete random variable  $w_a$ :

$$X = \left\{ x_i \mid \frac{a-i}{a} \right\}, \mathbb{P}(w_a = x_i) = \frac{1}{a}, i = 1, \dots, a, a \in \mathbb{N}^*$$

Its characteristics are found. It is shown that the variance of the sum of random variables:

$$\begin{aligned} \text{Var} \left[ \sum_{a=1}^n w_a \right] &= \sum_{i=1}^a \sum_{j=1}^b \text{Cov}(w_a, w_b), \\ \text{Cov}(w_a, w_b) &= \frac{\gcd(a, b)^2 - 1}{12ab} \end{aligned} \quad (1)$$

The variance estimate is given:

$$\text{Var} \left[ \sum_{a=1}^n w_a \right] = O(n \ln n - n + O(n) - H_n^2) \quad (2)$$

It is shown that when considering:

$$\left\lfloor \frac{x}{a} \right\rfloor = \frac{x}{a} - \mathbb{E}[w_a]$$

we can get the formula for counting  $D(n)$  – the number of lattice points under the hyperbola  $\frac{n}{xy}$ ,  $1 \leq x \leq n$ ,  $1 \leq y \leq n$ :

$$D(n) = n \ln n + (2\gamma - 1)n + H_{\sqrt{n}} + O(1)$$

satisfied with a standard deviation of the order  $O\left(\sqrt{\sqrt{n} \ln \sqrt{n} - \sqrt{n} + O(\sqrt{n}) - H_n^2}\right)$  with the hidden constant in main term  $C = \frac{1}{2\pi^2}$ . therefore, we can say that this estimate is  $O(n^{\frac{1}{4}+\epsilon})$  since:

$$\lim_{n \rightarrow \infty} \left( \frac{x^{\frac{1}{4}+\epsilon}}{\sqrt{\sqrt{n} \ln \sqrt{n} + \sqrt{n} - H_n^2}} \right) = 0 \quad (3)$$

for some very small  $\epsilon$ .

## 1. Random variable $w_a$

We assign to each number  $a, a \in \mathbb{N}^*$  a discrete random variable  $w_a$  with a set of values::

$$X = \left\{ x_i \mid \frac{a-i}{a} \right\}, i = 1, \dots, a \quad (4)$$

and probability:

$$\mathbb{P}(w_a = x_i) = \frac{1}{a} \quad (5)$$

The expected value of this random variable is:

$$\mathbb{E}[w_a] = \sum_{i=1}^a \left( \frac{1}{a} \cdot \frac{(a-i)}{a} \right) = \frac{a-1}{2a} \quad (6)$$

Variance is:

$$\text{Var}[w_a] = \sum_{i=1}^a \left( \frac{1}{a} \cdot \left( \frac{(a-i)}{a} - \frac{a-1}{2a} \right)^2 \right) = \frac{a^2-1}{12a^2} \quad (7)$$

Standard deviation:

$$\sigma[w_a] = \sqrt{\text{Var}[w_a]} = \frac{1}{2\sqrt{3}} \sqrt{\frac{a^2-1}{a^2}} \quad (8)$$

## 2. Random variable $w(n)$

Consider a random variable:

$$w(n) = \sum_{i=1}^n w_i \quad (9)$$

Its expected value is:

$$\mathbb{E}[w(n)] = \sum_{i=1..n} \mathbb{E}[w_i] = \sum_{i=1..n} \left( \frac{i-1}{2i} \right) = \frac{1}{2}(n - H_n) \quad (10)$$

Variance is:

$$\text{Var}[w(n)] = \sum_{a=1}^n \sum_{b=1}^n \text{Cov}(w_a, w_b) \quad (11)$$

where is the covariance:

$$\text{Cov}(w_a, w_b) = \quad (12)$$

$$= \frac{1}{\text{lcm}(a, b)} \sum_{i=1}^{\frac{a}{\text{gcd}(a, b)}} \sum_{j=1}^{\frac{b}{\text{gcd}(a, b)}} \sum_{k=1}^{\text{gcd}(a, b)} \left( \left( \frac{a - ((i-1)\text{gcd}(a, b) + k)}{a} - \frac{a-1}{2a} \right) \left( \frac{b - ((j-1)\text{gcd}(a, b) + k)}{b} - \frac{b-1}{2b} \right) \right) = \quad (13)$$

$$= \frac{\text{gcd}(a, b)^2 - 1}{12 \text{lcm}(a, b) \text{gcd}(a, b)} = \frac{\text{gcd}(a, b)^2 - 1}{12ab} \quad (14)$$

The formula (12) describes block structured covariance matrix  $A(d_1, d_2)$ :

$$A\left(\frac{a}{\gcd(a, b)}, \frac{b}{\gcd(a, b)}\right) = \begin{bmatrix} G_{11} & G_{12} & \dots \\ \vdots & \ddots & \\ G_{\frac{a}{\gcd(a, b)}, 1} & & G_{\frac{a}{\gcd(a, b)}, \frac{b}{\gcd(a, b)}} \end{bmatrix} \quad (15)$$

where is  $G(d_1, d_2)$  diagonal matrix:

$$G(\gcd(a, b), \gcd(a, b)) = \begin{bmatrix} \frac{1}{\text{lcm}(a, b)} & 0 & \dots \\ \vdots & \ddots & \\ 0 & & \frac{1}{\text{lcm}(a, b)} \end{bmatrix} \quad (16)$$

You can check the formula (12) using the link [2].

Let us estimate the variance:

$$\text{Var}[w(n)] = \sum_{a=1}^n \sum_{b=1}^n \frac{\gcd(a, b)^2 - 1}{12ab} = \quad (17)$$

$$= \sum_{a=1}^n \sum_{b=1}^n \frac{\gcd(a, b)^2}{12ab} - \sum_{a=1}^n \sum_{b=1}^n \frac{1}{12ab} = \frac{1}{12} \left( \sum_{a=1}^n \sum_{b=1}^n \frac{\gcd(a, b)^2}{ab} - H_n^2 \right) \quad (18)$$

We are interested in the principal term:

$$\sum_{a=1}^n \sum_{b=1}^n \frac{\gcd(a, b)^2}{ab} \quad (19)$$

We rewrite it:

$$\sum_{a=1}^n \sum_{b=1}^n \frac{\gcd(a, b)^2}{ab} = 2 \sum_{a=1}^n \sum_{b=1}^{a-1} \frac{\gcd(a, b)^2}{ab} + n \quad (20)$$

*I could not find any simplification for sums with two gcd factors. Therefore, here we give a rough estimate using only one gcd factor:*

$$\frac{\gcd(a, b)^2}{ab} \leq \frac{\gcd(a, b)}{a}, \quad \forall b < a \quad (21)$$

and we will continue to consider the sum of:

$$\sum_{a=1}^n \sum_{b=1}^{a-1} \frac{\gcd(a, b)}{a} = \sum_{a=1}^n \sum_{b=1}^a \frac{\gcd(a, b)}{a} - n \quad (22)$$

It is known (see [1]) that:

$$\begin{aligned} P(a) &= \sum_{b=1}^a \gcd(a, b) = \sum_{d|a} d \phi\left(\left\lfloor \frac{a}{d} \right\rfloor\right) \\ A(a) &= \frac{P(a)}{a} = \sum_{b=1}^a \frac{\gcd(a, b)}{a} = \sum_{d|a} \frac{\phi(d)}{d} \end{aligned} \quad (23)$$

It is also known (see [1]) that:

$$\sum_{a=1}^n A(a) = \frac{1}{2\zeta(2)} n \ln n + O(n) \quad (24)$$

From the reverse substitution (24), (22), (20) и (17) the variance estimate follows:

$$\sigma \left[ \sum_{i=1}^n w_i \right] = \sqrt{\text{Var} \left[ \sum_{i=1}^n w_i \right]} = O \left( \sqrt{\frac{1}{12} \left( \frac{1}{\zeta(2)} n \ln n - H_n^2 - n + O(n) \right)} \right) \quad (25)$$

Note that the hidden constant in main term is very small:

$$C = \frac{1}{12\zeta(2)} = \frac{1}{2\pi^2}. \quad (26)$$

### 3. Formula for $D(n)$

We apply the properties of the random variable  $w_a$  described above. Represent formula for finding  $D(n)$  – the number of lattice points  $(x, y)$  under hyperbola  $\frac{n}{xy}$ , таких, что  $1 \leq x \leq n$ ,  $1 \leq y \leq n$ .

Let:

$$\left\lfloor \frac{n}{x} \right\rfloor = \frac{n}{x} - E[w_x] \quad (27)$$

and let:

$$\Delta = \sum_{x=1.. \sqrt{n}} E[w_x] = \sum_{x=1.. \sqrt{n}} \frac{x-1}{2x} = \frac{1}{2} (\sqrt{n} - H_{\sqrt{n}}) \quad (28)$$

also let:

$$\sum_{x=1.. \sqrt{n}} \left( \frac{n}{x} - \left\lfloor \frac{n}{x} \right\rfloor \right) = nH_{\sqrt{n}} - \Delta \quad (29)$$

Then, applying combinatorial reasoning, we get:

$$D(n) = 2 \left( nH_{\sqrt{n}} - \Delta - \frac{\sqrt{n}^2}{2} \right) = 2 (nH_{\sqrt{n}} - \Delta) - n \quad (30)$$

откуда:

$$\begin{aligned} D(n) &= 2nH_{\sqrt{n}} - 2\Delta - n = \\ &= 2n \left( \ln \sqrt{n} + \gamma + \frac{1}{2\sqrt{n}} + O\left(\frac{1}{n}\right) \right) - 2 \left( \frac{1}{2} (\sqrt{n} - H_{\sqrt{n}}) \right) - n = \\ &= n \ln n + 2\gamma n - n + \sqrt{n} - \sqrt{n} + H_{\sqrt{n}} + O(1) = \\ &= n \ln n + (2\gamma - 1)n + H_{\sqrt{n}} + O(1) \end{aligned} \quad (31)$$

or in more precise form:

$$D(n) = (2n + 1)H_{\sqrt{n}} - n - \sqrt{n} \quad (32)$$

According (25), equality (30), (31), (32) satisfied with a standard deviation of the order  $O\left(\sqrt{\sqrt{n} \ln \sqrt{n} - \sqrt{n} + O(\sqrt{n}) - H_n^2}\right)$  with small hidden constant (26).

Note that (30) uses  $\sqrt{n}^2$ , not  $\lfloor \sqrt{n} \rfloor^2$  since we use harmonic summation from 1 to  $\sqrt{n}$  inclusive.

A similar result can be obtained using these arguments in the Dirichlet hyperbola method. We denote additionally:

$$\Delta_* = E[w_{\sqrt{n}}] = \frac{\sqrt{n} - 1}{2\sqrt{n}} \quad (33)$$

then:

$$D(n) = S_1(n) + S_2(n) = \sum_{x \leq \sqrt{n}} \sum_{xy \leq n} 1 + \sum_{\sqrt{n} < x \leq n} \sum_{xy \leq n} 1 \quad (34)$$

where the first term is:

$$\begin{aligned} S_1(n) &= \sum_{x \leq \sqrt{n}} \sum_{xy \leq n} 1 = \sum_{x \leq \sqrt{n}} \left\lfloor \frac{n}{x} \right\rfloor = \sum_{x \leq \sqrt{n}} \left( \frac{n}{x} - E[w_x] \right) = \\ &= \sum_{x \leq \sqrt{n}} \frac{n}{x} - \sum_{x \leq \sqrt{n}} E[w_x] = nH_{\sqrt{n}} - \Delta \end{aligned} \quad (35)$$

the second term is:

$$\begin{aligned} S_2(n) &= \sum_{\sqrt{n} < x \leq n} \sum_{xy \leq n} 1 = \sum_{y \leq \sqrt{n}} \sum_{xy \leq n} 1 - \sum_{y \leq \sqrt{n}} \sum_{x \leq \sqrt{n}} 1 = \sum_{y \leq \sqrt{n}} \left\lfloor \frac{n}{y} \right\rfloor - \left( \lfloor \sqrt{n} \rfloor^2 + 2\sqrt{n}\Delta_* \right) = \\ &= \sum_{y \leq \sqrt{n}} \left( \frac{n}{y} - E[w_y] \right) - \left( \lfloor \sqrt{n} \rfloor^2 + 2\sqrt{n}\Delta_* \right) = \\ &= \sum_{y \leq \sqrt{n}} \frac{n}{y} - \sum_{y \leq \sqrt{n}} E[w_y] - \left( \lfloor \sqrt{n} \rfloor^2 + 2\sqrt{n}\Delta_* \right) = \\ &= nH_{\sqrt{n}} - \Delta - \left( \lfloor \sqrt{n} \rfloor^2 + 2\sqrt{n}\Delta_* \right) \end{aligned} \quad (36)$$

summing it all together we get:

$$D(n) = 2(nH_{\sqrt{n}} - \Delta) - \left( \lfloor \sqrt{n} \rfloor^2 + 2\sqrt{n}\Delta_* \right) \quad (37)$$

consider  $\lfloor \sqrt{n} \rfloor^2$ , because:

$$\lfloor \sqrt{n} \rfloor^2 = (\sqrt{n} - \Delta_*)^2 = n - 2\sqrt{n}\Delta_* + \Delta_*^2 = n - \sqrt{n} + O\left(\frac{1}{n}\right) + 1 \quad (38)$$

then:

$$\lfloor \sqrt{n} \rfloor^2 + 2\sqrt{n}\Delta_* = n - \sqrt{n} + 1 + O\left(\frac{1}{n}\right) + \sqrt{n} - 1 = n + O\left(\frac{1}{n}\right) \quad (39)$$

so we get:

$$\begin{aligned} D(n) &= 2(nH_{\sqrt{n}} - \Delta) - n - O\left(\frac{1}{n}\right) = \\ &= n \ln n + (2\gamma - 1)n + H_{\sqrt{n}} + O(1) \end{aligned} \quad (40)$$

as already shown in (31). According to(25), the expression (40) satisfied with a standard deviation of the order  $O\left(\sqrt{\sqrt{n} \ln \sqrt{n} - \sqrt{n} + O(\sqrt{n}) - H_n^2}\right)$  with small hidden constant (26).

## References

- [1] Laszlo Toth. A Survey of Gcd-Sum Functions. Journal of Integer Sequences, Vol. 13 2010.  
URL: <https://cs.uwaterloo.ca/journals/JIS/VOL13/Toth/toth10.pdf>
- [2] Covariance formula 9-10-11.  
URL: <https://www.wolframalpha.com/input/?i=sum%5Bsum%5Bsum%5B%28%28a+-+%28%28i-1%29+gcd%28a%2Cb%29+%2B+k%29%29%2Fa+-+%28a-1%29%2F%282a%29%29%28%28b+-+%28%28j-1%29+gcd%28a%2Cb%29+%2B+k%29%29%2Fb+-+%28b-1%29%2F%282b%29%29%2Ck%3D1..gcd%28a%2Cb%29%5D%2Cj%3D1..b%2Fgcd%28a%2Cb%29%5D%2Ci%3D1..a%2Fgcd%28a%2Cb%29%5D>