Исследование случайной величины w_a и ее связи с проблемой делителей Дирихле

Дмитрий Пятин / dmpyatin@petrsu.ru

03.07.2020

Аннотация: В данной работе исследуется дискретная случайная величина w_a:

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

Найдены ее характеристики. Показано, что дисперсия суммы случайных величин:

$$\begin{gathered} \operatorname{Var}\Big[\sum_{a=1}^{n} w_a\Big] = \sum_{i=1}^{a}\sum_{j=1}^{b} \operatorname{Cov}(w_a, w_b), \\ \operatorname{Cov}(w_a, w_b) = \frac{\gcd(a,b)^2 - 1}{12ab} \end{gathered} \tag{1}$$

Дана оценка дисперсии:

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

Показано, что при рассмотрении:

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

мы можем получить формулу для подсчета $D(n)$ числа целочисленных точек под гиперболой $\frac{n}{xy}$, $1 \le x \le n$, $1 \le y \le n$:

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

со среднеквадратическим отклонением порядка $O\left(\sqrt{\sqrt{n}\ln\sqrt{n} - \sqrt{n} + O(\sqrt{n}) - H_n^2}\right)$ со скрытой константой $C = \frac{1}{2\pi^2}$ в главном подкоренном члене. Как следствие, мы можем говорить о оценке: $O(n^{\frac{1}{4}+\epsilon})$:

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

для некоторого очень малого ϵ.

1. Случайная величина w_a

Поставим в соответствие каждому числу $a, a \in \mathbb{N}^*$ дискретную случайную величину w_a со множеством значений:

$$X = \left\{ x_i \;\middle|\; \frac{a-i}{a} \right\}, i = 1, .., a \tag{4}$$

и вероятностью:

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

Математическое ожидание данной случайной величины равняется:

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

Дисперсия:

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

Среднеквадратическое отклонение:

$$\sigma[w_a] = \sqrt{\mathrm{Var}[w_a]} = \frac{1}{2\sqrt{3}} \sqrt{\frac{a^2-1}{a^2}} \tag{8}$$

2. Случайная величина $w(n)$

Рассмотрим случайную величину $w(n) = \sum_{i=1}^{n} w_i$. Ее матожидание:

$$\mathrm{E}\left[w(n)\right] = \sum_{i=1..n} \mathrm{E}[w_i] = \sum_{i=1..n} \left(\frac{i-1}{2i} \right) = \frac{1}{2}(n - H_n) \tag{9}$$

Дисперсия:

$$\mathrm{Var}\left[w(n)\right] = \sum_{a=1}^{n} \sum_{b=1}^{n} \mathrm{Cov}(w_a, w_b) \tag{10}$$

где ковариация:

$$\mathrm{Cov}(w_a, w_b) = \tag{11}$$

$$= \frac{1}{\mathrm{lcm}(a,b)} \sum_{i=1}^{\frac{a}{\gcd(a,b)}} \sum_{j=1}^{\frac{b}{\gcd(a,b)}} \sum_{i=1}^{\gcd(a,b)} \left(\left(\frac{a - ((i-1)\gcd(a,b) + k)}{a} - \frac{a-1}{2a} \right) \left(\frac{b - ((j-1)\gcd(a,b) + k)}{b} - \frac{b-1}{2b} \right) \right) = \tag{12}$$

$$= \frac{\gcd(a,b)^2 - 1}{12\,\mathrm{lcm}(a,b)\gcd(a,b)} = \frac{\gcd(a,b)^2 - 1}{12ab} \tag{13}$$

Формула (11) описывает ковариационную матрицу $A(d_1, d_2)$ блочной структуры:

$$A\left(\frac{a}{\gcd(a,b)}, \frac{b}{\gcd(a,b)}\right) = \begin{bmatrix} G_{11} & G_{12} & \cdots \\ \vdots & \ddots & \\ G_{\frac{a}{\gcd(a,b)},1} & & G_{\frac{a}{\gcd(a,b)},\frac{b}{\gcd(a,b)}} \end{bmatrix} \tag{14}$$

где $G(d_1, d_2)$ диагональная матрица:

$$G(\gcd(a,b), \gcd(a,b)) = \begin{bmatrix} \frac{1}{\operatorname{lcm}(a,b)} & 0 & \cdots \\ \vdots & \ddots & \\ 0 & & \frac{1}{\operatorname{lcm}(a,b)} \end{bmatrix} \tag{15}$$

Перепроверить формулу (11) можно в Wolfram Alpha по ссылке [2].

Дадим оценку дисперсии:

$$\operatorname{Var}[w(n)] = \sum_{a=1}^{n}\sum_{b=1}^{n} \frac{\gcd(a,b)^2 - 1}{12ab} = \tag{16}$$

$$= \sum_{a=1}^{n}\sum_{b=1}^{n} \frac{\gcd(a,b)^2}{12ab} - \sum_{a=1}^{n}\sum_{b=1}^{n} \frac{1}{12ab} = \frac{1}{12}\left(\sum_{a=1}^{n}\sum_{b=1}^{n} \frac{\gcd(a,b)^2}{ab} - H_n^2\right) \tag{17}$$

Нас интересует главный член:

$$\sum_{a=1}^{n}\sum_{b=1}^{n} \frac{\gcd(a,b)^2}{ab} \tag{18}$$

Перепишем его:

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

У меня не получилось найти какое-либо упрощение для сумм с двумя gcd множителями. Поэтому тут дадим грубую оценку, используя только один gcd множитель:

$$\frac{\gcd(a,b)^2}{ab} \le \frac{\gcd(a,b)}{a},\ \forall b < a \tag{20}$$

и будем далее рассматривать сумму:

$$\sum_{a=1}^{n}\sum_{b=1}^{a-1} \frac{\gcd(a,b)}{a} = \sum_{a=1}^{n}\sum_{b=1}^{a} \frac{\gcd(a,b)}{a} - n \tag{21}$$

Известно (см. [1]), что:

$$\begin{aligned} P(a) &= \sum_{b=1}^{a} \gcd(a,b) = \sum_{d|a} d\,\phi\left(\left\lfloor \frac{a}{d} \right\rfloor\right) \\ A(a) &= \frac{P(a)}{a} = \sum_{b=1}^{a} \frac{\gcd(a,b)}{a} = \sum_{d|a} \frac{\phi(d)}{d} \end{aligned} \tag{22}$$

Также известно (см. [1]), что:

$$\sum_{a=1}^{n} A(a) = \frac{1}{2\zeta(2)} n \ln n + O(n) \tag{23}$$

Из обратной подстановки (23), (21), (19) и (16) вытекает оценка дисперсии:

$$\sigma\left[\sum_{i=1}^{n} w_i\right] = \sqrt{\mathrm{Var}\left[\sum_{i=1}^{n} w_i\right]} = O\left(\sqrt{\frac{1}{12}\left(\frac{1}{\zeta(2)} n \ln n - H_n^2 - n + O(n)\right)}\right) \tag{24}$$

Отметим, что скрытая константа в главном члене равна:

$$C = \frac{1}{12\zeta(2)} = \frac{1}{2\pi^2}. \tag{25}$$

3. Формула для $D(n)$

Применим описанные выше свойства случайной величины w_a. Представим формулу для нахождения $D(n)$ – числа целочисленных точек (x, y) под гиперболой $\frac{n}{xy}$, таких, что $1 \le x \le n$, $1 \le y \le n$.

Пусть:

$$\left\lfloor \frac{n}{x} \right\rfloor = \frac{n}{x} - \mathrm{E}[w_x] \tag{26}$$

и пусть:

$$\Delta = \sum_{x=1..\sqrt{n}} \mathrm{E}[w_x] = \sum_{x=1..\sqrt{n}} \frac{x-1}{2x} = \frac{1}{2}\left(\sqrt{n} - H_{\sqrt{n}}\right) \tag{27}$$

также, пусть:

$$\sum_{x=1..\sqrt{n}} \left(\frac{n}{x} - \left\lfloor \frac{n}{x} \right\rfloor\right) = nH_{\sqrt{n}} - \Delta \tag{28}$$

Тогда, применив комбинаторные рассуждения, получим:

$$D(n) = 2\left(nH_{\sqrt{n}} - \Delta - \frac{\sqrt{n}^2}{2}\right) = 2\left(nH_{\sqrt{n}} - \Delta\right) - n \tag{29}$$

откуда:

$$\begin{aligned} D(n) =& 2nH_{\sqrt{n}} - 2\Delta - n = \\ =& 2n\left(\ln\sqrt{n} + \gamma + \frac{1}{2\sqrt{n}} + O\left(\frac{1}{n}\right)\right) - 2\left(\frac{1}{2}\left(\sqrt{n} - H_{\sqrt{n}}\right)\right) - n = \\ =& n\ln n + 2\gamma n - n + \sqrt{n} - \sqrt{n} + H_{\sqrt{n}} + O(1) = \\ =& n\ln n + (2\gamma - 1)n + H_{\sqrt{n}} + O(1) \end{aligned} \tag{30}$$

или в более точном виде:

$$D(n) = (2n+1)H_{\sqrt{n}} - n - \sqrt{n} \tag{31}$$

Согласно (24), равенства (29), (30), (31) выполняются выполняется со среднеквадратическим отклонением порядка $O\left(\sqrt{\sqrt{n}\ln\sqrt{n}+\sqrt{n}-H_n^2}\right)$ с малой скрытой константой (25).

Важно заметить, что в (29) используется $\sqrt{n}^2$, а не $\lfloor\sqrt{n}\rfloor^2$ т.к. мы используем суммирование гармоник от 1 до $\sqrt{n}$ *включительно*.

Аналогичный результат можно получить, используя данные рассуждения в методе гиперболы Дирихле. Обозначим вспомогательно:

$$\Delta_* = \mathrm{E}[w_{\sqrt{n}}] = \frac{\sqrt{n}-1}{2\sqrt{n}} \tag{32}$$

тогда:

$$D(n) = S_1(n) + S_2(n) = \sum_{x\le\sqrt{n}}\sum_{xy\le n} 1 + \sum_{\sqrt{n}<x\le n}\sum_{xy\le n} 1 \tag{33}$$

где первая сумма равняется:

$$\begin{aligned} S_1(n) = \sum_{x\le\sqrt{n}}\sum_{xy\le n} 1 = \sum_{x\le\sqrt{n}}\left\lfloor\frac{n}{x}\right\rfloor = \sum_{x\le\sqrt{n}}\left(\frac{n}{x} - \mathrm{E}[w_x]\right) = \\ = \sum_{x\le\sqrt{n}}\frac{n}{x} - \sum_{x\le\sqrt{n}}\mathrm{E}[w_x] = nH_{\sqrt{n}} - \Delta \end{aligned} \tag{34}$$

вторая сумма равняется:

$$\begin{aligned} S_2(n) &= \sum_{\sqrt{n}<x\le n}\sum_{xy\le n} 1 = \sum_{y\le\sqrt{n}}\sum_{xy\le n} 1 - \sum_{y\le\sqrt{n}}\sum_{x\le\sqrt{n}} 1 = \sum_{y\le\sqrt{n}}\left\lfloor\frac{n}{y}\right\rfloor - \left(\lfloor\sqrt{n}\rfloor^2 + 2\sqrt{n}\Delta_*\right) = \\ &= \sum_{y\le\sqrt{n}}\left(\frac{n}{y} - \mathrm{E}[w_y]\right) - \left(\lfloor\sqrt{n}\rfloor^2 + 2\sqrt{n}\Delta_*\right) = \\ &= \sum_{y\le\sqrt{n}}\frac{n}{y} - \sum_{y\le\sqrt{n}}\mathrm{E}[w_y] - \left(\lfloor\sqrt{n}\rfloor^2 + 2\sqrt{n}\Delta_*\right) = \\ &= nH_{\sqrt{n}} - \Delta - \left(\lfloor\sqrt{n}\rfloor^2 + 2\sqrt{n}\Delta_*\right) \end{aligned} \tag{35}$$

суммируя все вместе получаем:

$$D(n) = 2(nH_{\sqrt{n}} - \Delta) - \left(\lfloor\sqrt{n}\rfloor^2 + 2\sqrt{n}\Delta_*\right) \tag{36}$$

рассмотрим $\lfloor\sqrt{n}\rfloor^2$, т.к.:

$$\lfloor\sqrt{n}\rfloor^2 = (\sqrt{n} - \Delta_*)^2 = n - 2\sqrt{n}\Delta_* + \Delta_*^2 = n - \sqrt{n} + O\left(\frac{1}{n}\right) + 1 \tag{37}$$

то:

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

таким образом, мы получаем:

$$\begin{aligned} D(n) =& 2(nH_{\sqrt{n}} - \Delta) - n - O\left(\frac{1}{n}\right) = \\ =& n\ln n + (2\gamma - 1)n + H_{\sqrt{n}} + O(1) \end{aligned} \tag{39}$$

как уже было показано в (30). Согласно (24), выражение (39) выполняется со среднеквадратическим отклонением порядка $O\left(\sqrt{\sqrt{n}\ln\sqrt{n}+\sqrt{n}-H_n^2}\right)$ с малой скрытой константой (25).

Список литературы

[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] Формула для ковариации 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