Миронкин В.О.

Московский институт электроники и математики Национальный исследовательский университет Высшая школа экономики

ВЕРОЯТНОСТЬ КОЛЛИЗИИ ДВУХ СЛУЧАЙНЫХ ТОЧЕК ДЛЯ СТЕПЕНИ СЛУЧАЙНОГО ОТОБРАЖЕНИЯ

В работе получено точное выражение для вероятности коллизии двух случайно выбранных точек из некоторого конечного множества для произвольной степени случайного равновероятного отображения.

Введение

В настоящей работе вычисляется вероятность события, заключающегося в том, что две случайно выбранные точки некоторого конечного множества $S=\{1,...,N\}$, $N>1$, образуют коллизию при k-кратной итерации случайного отображении:

$$f^k : S \to S,\ k \in \mathbb{N}.$$

В терминах случайного отображения f рассматриваемое событие означает, что две случайные точки являются началом различных путей длины k в графе отображения f, сходящихся в одной точке через k шагов.

Рассмотрим конечное множество $S=\{1,...,N\}$, $N>1$, и множество $\Im$ всех отображений $f : S \to S$. Зададим на $\Im$ равномерную вероятностную меру и рассмотрим $k-$ю степень случайного отображения f.

Будем использовать представление случайных отображений в терминах случайных графов, хорошо известное по книгам [1-4]. Граф отображения f обозначим через G.

Пусть $x_0, y_0 \in S$ произвольные различные вершины в графе G. Вычислим вероятность события, заключающегося в том, что образы вершин x_0, y_0 совпадут при отображении f^k, $k \in \mathbb{N}$.

Здесь и далее для любых $i_0, i_1 \in \mathbb{N}$, $i_0 > i_1$, положим $\sum_{j=i_0}^{i_1}(...) = 0$.

Теорема. *Пусть* $f \in \Im$. *Тогда произвольные вершины* $x_0 \neq y_0 \in S$ *образуют коллизию при отображении* f^k *с вероятностью*

$$\mathbf{P}\left\{f^k(x_0) = f^k(y_0); x_0 \neq y_0\right\} = \sum_{l=1}^{N-1} \sum_{t_1=1}^{\min(k,N-l)} \frac{2(c_2-c_1+1)}{N^2} \prod_{i=2}^{t_1+l-1}\left(1-\tfrac{i}{N}\right) +$$

$$+\tfrac{1}{N^2}\sum_{l=1}^{N-2} \sum_{t_1=1}^{\min(k,N-l)} \sum_{p_1=\max(0,c_3)}^{c_4} \sum_{\gamma=0}^{l-1} \sum_{p_2=\max(0,c_5)}^{c_6} \prod_{i=2}^{2t_1-l(p_2-p_1-1)-\gamma-1}\left(1-\tfrac{i}{N}\right),$$

где $c_1 = c_1(l,t_1,\alpha) = \left\lceil \frac{k-\alpha-t_1+1}{l} \right\rceil$, $c_2 = c_2(l,\alpha) = \left[\frac{k-\alpha-1}{l}\right]$,

$c_3 = c_3(l,t_1) = \left\lceil \frac{k-t_1-l+1}{l} \right\rceil$, $c_4 = c_4(l,t_1) = \left[\frac{k-t_1}{l}\right]$,

$c_5 = c_5(l,t_1,p_1,\gamma) = \left\lceil \frac{2t_1+l(p_1+1)-\gamma-N}{l} \right\rceil$, $c_6 = c_6(l,t_1,p_1,\gamma) = \left[\frac{t_1+lp_1-\gamma-1}{l}\right]$.

Доказательство. Представим искомую вероятность в следующем виде:

$$\mathbf{P}\left\{f^k(x_0) = f^k(y_0); x_0 \neq y_0\right\} = \mathbf{P}\left\{f^k(x_0) = f^k(y_0); x_0 \neq y_0 \in 1\,\textit{компоненте}\, G\right\} +$$

$$\mathbf{P}\left\{f^k(x_0) = f^k(y_0); x_0 \neq y_0 \in \textit{разным компонентам}\, G\right\}.$$

Заметим, что если вершины x_0, y_0 лежат в разных компонентах графа G, то они не могут образовывать коллизию при отображении f^k, $k \in \mathbb{N}$, поэтому второе слагаемое в этом выражении равно нулю.

Выпишем первое слагаемое, используя формулу полной вероятности. Для этого рассмотрим возможные варианты расположения вершин x_0 и y_0:

1. Обе вершины x_0 и y_0 лежат на цикле некоторой компоненты графа G.

2. Одна из вершин x_0 или y_0 лежит на подходе к циклу, а другая на цикле.

3. Обе вершины x_0 и y_0 лежат на подходах к циклу.

Тогда по формуле полной вероятности имеем:

$$\mathbf{P}\left\{f^k(x_0) = f^k(y_0); x_0 \neq y_0\right\} =$$

$$= \mathbf{P}\left\{ f^k(x_0) = f^k(y_0); x_0 \neq y_0; x_0, y_0 \text{ на цикле} \right\} +$$

$$+2\mathbf{P}\left\{ f^k(x_0) = f^k(y_0); x_0 \neq y_0; x_0 \text{ на подходе}, y_0 \text{ на цикле} \right\} +$$

$$+\mathbf{P}\left\{ f^k(x_0) = f^k(y_0); x_0 \neq y_0; x_0, y_0 \text{ на подходе} \right\}.$$

Вычислим каждое слагаемое в отдельности.

В первом случае события, стоящие под знаком вероятности несовместны:

$$\left\{ f^k(x_0) = f^k(y_0) \right\} \cap \left\{ x_0 \neq y_0; x_0, y_0 \text{ на цикле} \right\} = \varnothing$$

Поэтому

$$\mathbf{P}\left\{ f^k(x_0) = f^k(y_0); x_0 \neq y_0; x_0, y_0 \text{ на цикле} \right\} = 0.$$

Второй случай. Пусть для определенности вершина x_0 лежит на подходе к циклу, а y_0 - на цикле. Тогда для выполнения события $\left\{ f^k(x_0) = f^k(y_0) \right\}$ вершина $f^k(x_0)$ должна лежать на соответствующем цикле.

Обозначим через $l \in \overline{1, N-1}$ длину цикла рассматриваемой компоненты графа G (вершина x_0 должна лежать на подходе, поэтому $l < N$).

Обозначим через t_1 расстояние от вершины x_0 до цикла. Так как вершина x_0 должна лежать на подходе, длина которого не превосходит k, то $t_1 \in \overline{1, \min(k, N-l)}$ (в противном случае не произойдет совпадения образов при отображении f^k).

Выделим на цикле отрезок длины $\alpha \in \overline{0, \min(l-1, k-t_1)}$, начинающийся в циклической вершине $f^{t_1}(x_0)$ и такой, что $f^{t_1+\alpha}(x_0) = f^k(x_0)$.

Для введенных величин имеет место следующее соотношение:

$$k = t_1 + lp_1 + \alpha, \text{ где } p_1 \geq 0,\ \alpha \in \overline{0, \min(l-1, k-t_1)}. \qquad (1)$$

При фиксированных значениях k, t_1, l из соотношения (1) значение α определяется однозначно.

Определим число возможных вариантов расположения вершины y_0 на соответствующем цикле, для которых выполняется событие $\left\{ f^k(x_0) = f^k(y_0) \right\}$.

Обозначим через $t_2 \in \overline{0, \min(l-1, k)}$ расстояние от вершины y_0 на соответствующем цикле до вершины $f^k(x_0)$. Тогда справедливо равенство:

$$k = t_2 + lp_2, \text{ где } p_2 \geq 0, t_2 \in \overline{0, \min(l-1, k)}. \qquad (2)$$

Число вариантов расположения вершины y_0 совпадает с числом возможных наборов значений параметра p_2 и t_2, для которых справедливо (2), а так как при заданном k соотношение (2) однозначно задает p_2 и t_2, то и расположение вершины y_0 определяется однозначно. Тогда с учетом соотношений (1) и (2) получим

$$\mathbf{P}\left\{f^k(x_0) = f^k(y_0); x_0 \neq y_0; x_0 \textit{ на подходе}, y_0 \textit{ на цикле}\right\} =$$

$$= \sum_{l=1}^{N-1} \sum_{t_1=1}^{\min(k, N-l)} \sum_{\substack{x_1, \ldots, x_{t_1+l-2} \\ \textit{различные}, x_i \notin \{x_0, y_0\}}} \mathbf{P}\left\{\begin{array}{c} f(x_0) = x_1, .., f(x_{t_1-1}) = x_{t_1}, ..., \\ f(x_{t_1}) = x_{t_1+1}, ..., y_0, ..., f(x_{t_1+l-2}) = x_{t_1} \end{array}\right\} =$$

$$= \sum_{l=1}^{N-1} \sum_{t_1=1}^{\min(k, N-l)} \sum_{\substack{x_1, \ldots, x_{t_1+l-2} \\ \textit{различные}, x_i \notin \{x_0, y_0\}}} \frac{1}{N^{t_1+l}} = \sum_{l=1}^{N-1} \sum_{t_1=1}^{\min(k, N-l)} \frac{1}{N^{t_1+l}} \prod_{i=2}^{t_1+l-1} (N-i) =$$

$$= \frac{1}{N^2} \sum_{l=1}^{N-1} \sum_{t_1=1}^{\min(k, N-l)} \prod_{i=2}^{t_1+l-1} \left(1 - \frac{i}{N}\right).$$

Рассмотрим третий случай. Здесь возможны следующие варианты расположения вершин x_0 и y_0:

1. Вершины $t_2 \in \overline{1, \min(k, N-l)}$, y_0 лежат на одном подходе к циклу.
2. Вершины x_0, y_0 лежат на разных подходах к циклу.

Пусть $t_2 \in \overline{1, \min(k, N-l)}$ и y_0 лежат на одном подходе к циклу. Для определенности будем считать, что y_0 лежит ближе к циклу, чем x_0. В силу равноправия вершин случай обратного расположения вершин рассматривается аналогично с точностью до замены x_0 на y_0.

Рассмотрим цикл длины $l \in \overline{1, N-2}$, не проходящий через вершины x_0, y_0 ($l \leq N-2$ так как вершины x_0, y_0 различны и лежат на подходе).

Как и выше обозначим через t_1 расстояние от вершины x_0 до цикла, а через t_2 расстояние от вершины y_0 до цикла. Так как в наших предположениях $t_1 > t_2$, то $t_1 \in \overline{2, \min(k, N-l)}$.

Выделим аналогичным образом на цикле отрезок длины $\alpha \in \overline{0, \min(l-1, k-t_1)}$, начинающийся в циклической вершине $f^{t_1}(x_0)$ и такой, что $f^{t_1+\alpha}(x_0) = f^k(x_0)$.

Для введенных величин имеет место равенство

$$k = t_1 + lp_1 + \alpha, \text{ где } p_1 \geq 0,\ \alpha \in \overline{0, \min(l-1, k-t_1)}.$$

Откуда при фиксированных значениях k, t_1, l значение α определяется однозначно.

Определим далее число возможных вариантов расположения вершины y_0 на подходе, соответствующем вершине x_0, для которых выполняется событие $\{f^k(x_0) = f^k(y_0)\}$. Для величины t_2 справедливо:

$$k = t_2 + lp_2 + \alpha, \text{ где } p_2 \geq 0, t_2 \in \overline{1, t_1 - 1}.$$

Откуда следует, что при заданных значениях k, l, t_1 значение t_2 определяется следующим образом

$$t_2 = k - lp_2 - \alpha, \tag{3}$$

$$p_2 \in \mathbb{N}_0 \text{ и } \left\lceil \tfrac{k-\alpha-t_1+1}{l} \right\rceil \leq p_2 \leq \left[\tfrac{k-\alpha-1}{l} \right],$$

где значения параметра p_2 определяются из неравенства

$$1 \leq k - lp_2 - \alpha \leq t_1 - 1.$$

Число вариантов расположения вершины y_0 совпадает с числом возможных значений величины p_2, удовлетворяющей (3). Обозначим через $c_1 = c_1(l, t_1, \alpha) = \left\lceil \tfrac{k-\alpha-t_1+1}{l} \right\rceil$ и $c_2 = c_2(l, \alpha) = \left[\tfrac{k-\alpha-1}{l} \right]$. Тогда с учетом замечания о равноправии вершин, получим

$$\mathbf{P}\{f^k(x_0) = f^k(y_0); x_0 \neq y_0; x_0, y_0 \textit{ на одном подходе}\} =$$

$$= 2\sum_{l=1}^{N-2}\sum_{t_1=2}^{\min(k,N-l)}\sum_{p_2=c_1}^{c_2}\sum_{\substack{x_1,\dots,x_{t_1+l-2}\\ \text{различные}, x_i\notin\{x_0,y_0\}}} \mathbf{P}\left\{\begin{array}{c} f(x_0)=x_1,\dots,y_0,\dots,f\left(x_{t_1-2}\right)=x_{t_1-1},\dots,\\ f\left(x_{t_1-1}\right)=x_{t_1},\dots,f\left(x_{t_1+l-2}\right)=x_{t_1-1}\end{array}\right\} =$$

$$= 2\sum_{l=1}^{N-2}\sum_{t_1=2}^{\min(k,N-l)}(c_2-c_1)\sum_{\substack{x_1,\dots,x_{t_1+l-2}\\ \text{различные}, x_i\notin\{x_0,y_0\}}} \frac{1}{N^{t_1+l}} = \sum_{l=1}^{N-2}\sum_{t_1=2}^{\min(k,N-l)}\frac{2(c_2-c_1)}{N^{t_1+l}}\prod_{i=2}^{t_1+l-1}(N-i) =$$

$$= \sum_{l=1}^{N-2}\sum_{t_1=2}^{\min(k,N-l)}\frac{2(c_2-c_1)}{N^2}\prod_{i=2}^{t_1+l-1}\left(1-\frac{i}{N}\right).$$

Пусть теперь вершины x_0, y_0 лежат на разных подходах к циклу.

Рассмотрим цикл длины $l\in\overline{1,N-2}$, не проходящий через вершины x_0, y_0. Пусть $t_1\in\overline{1,\min(k,N-l)}$ - расстояние от вершины x_0 и $\alpha\in\overline{0,\min(l-1,k-t_1)}$ - длина отрезка цикла, начинающегося в вершине $f^{t_1}(x_0)$ и такого, что $f^{t_1+\alpha}(x_0)=f^k(x_0)$.

Для указанных величин справедливо равенство

$$k=t_1+lp_1+\alpha,\ \text{где},\ \alpha\in\overline{0,\min(l-1,k-t_1)}, \qquad (4)$$

$$\max\left(0,\left\lceil\frac{k-t_1-l+1}{l}\right\rceil\right)\le p_1\le\left\lfloor\frac{k-t_1}{l}\right\rfloor,$$

где интервал значений для p_1 определяется из следующего соотношения:

$$0\le k-t_1-lp_1\le\min(l-1,k-t_1),\ \text{где}\ p_1\in\mathbb{N}_0.$$

Обозначим через $c_3=c_3(l,t_1)=\left\lceil\frac{k-t_1-l+1}{l}\right\rceil$ и $c_4=c_4(l,t_1)=\left\lfloor\frac{k-t_1}{l}\right\rfloor$.

Пусть далее t_2 - расстояние от вершины y_0 до цикла. И пусть $\gamma\in\overline{0,l-1}$ - длина отрезка рассматриваемого цикла от циклической вершины $z\equiv f^{t_2}(y_0)$ до циклической вершины $f^{t_1}(x_0)$ (отрезок выбирается по направлению роста степени отображения f). Тогда

$$k=t_2+lp_2+\gamma+\alpha,\ p_2\ge 0.$$

Или, что равносильно с учетом (4),

$$k=t_2+lp_2+\gamma+k-t_1-lp_1,\ p_2\ge 0\ \Leftrightarrow$$

$$\Leftrightarrow\ t_2 = t_1 - l\left(p_2 - p_1\right) - \gamma\ ,\ p_2 \geq 0\,.$$

При заданных значениях l, t_1, p_1, γ величина t_2 должна удовлетворять условию:

$$1 \leq t_2 \leq \min\left(N - l - t_1, k - \alpha - \gamma\right).$$

Откуда определяются границы значений величины $p_2 \in \mathbb{N}_0$:

$$1 \leq t_1 - l\left(p_2 - p_1\right) - \gamma \leq \min\left(N - l - t_1, k - \alpha - \gamma\right) \Leftrightarrow$$

$$\Leftrightarrow\ 1 \leq t_1 - l\left(p_2 - p_1\right) - \gamma \leq \min\left(N - l - t_1, t_1 + lp_1 - \gamma\right) \Leftrightarrow$$

$$\Leftrightarrow\ \max\left(\tfrac{2t_1 + l(p_1+1) - \gamma - N}{l}, 0\right) 1 \leq p_2 \leq \tfrac{t_1 + lp_1 - \gamma - 1}{l} \Leftrightarrow$$

$$\Leftrightarrow \max\left(\left\lceil \tfrac{2t_1 + l(p_1+1) - \gamma - N}{l} \right\rceil, 0\right) 1 \leq p_2 \leq \left\lceil \tfrac{t_1 + lp_1 - \gamma - 1}{l} \right\rceil.$$

Тогда, обозначив через $c_5 = c_5\left(l, t_1, p_1, \gamma\right) = \left\lceil \tfrac{2t_1 + l(p_1+1) - \gamma - N}{l} \right\rceil$ и $c_6 = c_6\left(l, t_1, p_1, \gamma\right) = \left\lceil \tfrac{t_1 + lp_1 - \gamma - 1}{l} \right\rceil$, получим:

$$\mathbf{P}\left\{ f^k\left(x_0\right) = f^k\left(y_0\right); x_0 \neq y_0; x_0, y_0 \textit{ на разных подходах} \right\} =$$

$$= \sum_{l=1}^{N-2} \sum_{t_1=1}^{\min(k, N-l)} \sum_{p_1=\max(0,c_3)}^{c_4} \sum_{\gamma=0}^{l-1} \sum_{p_2=\max(0,c_5)}^{c_6} \sum_{\substack{x_1,\ldots,x_{t_1+l-1},\\ y_1,\ldots,y_{t_1-l(p_2-p_1)-\gamma-1}\\ \textit{разл.}, x_i, y_i \notin \{x_0, y_0\}}} \mathbf{P}\left\{ \begin{array}{c} f\left(x_0\right) = x_1, \ldots, f\left(x_{t_1-1}\right) = x_{t_1}, \ldots, \\ f\left(x_{t_1}\right) = x_{t_1+1}, \ldots, z, \ldots, f\left(x_{t_1+l-1}\right) = x_{t_1}, \\ f\left(y_0\right) = y_1, \ldots, f\left(y_{t_1-l(p_2-p_1)-\gamma-1}\right) = z \end{array} \right\} =$$

$$= \sum_{l=1}^{N-2} \sum_{t_1=1}^{\min(k, N-l)} \sum_{p_1=\max(0,c_3)}^{c_4} \sum_{\gamma=0}^{l-1} \sum_{p_2=\max(0,c_5)}^{c_6} \sum_{\substack{x_1,\ldots,x_{t_1+l-1},\\ y_1,\ldots,y_{t_1-l(p_2-p_1)-\gamma-1}\\ \textit{разл.}, x_i, y_i \notin \{x_0, y_0\}}} \frac{1}{N^{2t_1 - l(p_2 - p_1 - 1) - \gamma}} =$$

$$= \sum_{l=1}^{N-2} \sum_{t_1=1}^{\min(k, N-l)} \sum_{p_1=\max(0,c_3)}^{c_4} \sum_{\gamma=0}^{l-1} \sum_{p_2=\max(0,c_5)}^{c_6} \frac{1}{N^{2t_1 - l(p_2 - p_1 - 1) - \gamma}} \prod_{i=2}^{2t_1 - l(p_2 - p_1 - 1) - \gamma - 1} \left(N - i\right) =$$

$$= \tfrac{1}{N^2} \sum_{l=1}^{N-2} \sum_{t_1=1}^{\min(k, N-l)} \sum_{p_1=\max(0,c_3)}^{c_4} \sum_{\gamma=0}^{l-1} \sum_{p_2=\max(0,c_5)}^{c_6} \prod_{i=2}^{2t_1 - l(p_2 - p_1 - 1) - \gamma - 1} \left(1 - \tfrac{i}{N}\right).$$

Объединяя полученные результаты, получим

$$\mathbf{P}\left\{ f^k\left(x_0\right) = f^k\left(y_0\right); x_0 \neq y_0 \right\} = \tfrac{2}{N^2} \sum_{l=1}^{N-1} \sum_{t_1=1}^{\min(k, N-l)} \prod_{i=2}^{t_1 + l - 1} \left(1 - \tfrac{i}{N}\right) +$$

$$+\sum_{l=1}^{N-2}\sum_{t_1=2}^{\min(k,N-l)}\frac{2(c_2-c_1)}{N^2}\prod_{i=2}^{t_1+l-1}\left(1-\tfrac{i}{N}\right)+$$

$$+\tfrac{1}{N^2}\sum_{l=1}^{N-2}\sum_{t_1=1}^{\min(k,N-l)}\sum_{p_1=\max(0,c_3)}^{c_4}\sum_{\gamma=0}^{l-1}\sum_{p_2=\max(0,c_5)}^{c_6}\prod_{i=2}^{2t_1-l(p_2-p_1-1)-\gamma-1}\left(1-\tfrac{i}{N}\right)=$$

$$=\sum_{l=1}^{N-1}\sum_{t_1=2}^{\min(k,N-l)}\frac{2(c_2-c_1+1)}{N^2}\prod_{i=2}^{t_1+l-1}\left(1-\tfrac{i}{N}\right)+$$

$$+\tfrac{1}{N^2}\sum_{l=1}^{N-2}\sum_{t_1=1}^{\min(k,N-l)}\sum_{p_1=\max(0,c_3)}^{c_4}\sum_{\gamma=0}^{l-1}\sum_{p_2=\max(0,c_5)}^{c_6}\prod_{i=2}^{2t_1-l(p_2-p_1-1)-\gamma-1}\left(1-\tfrac{i}{N}\right).\ \square$$

Библиографический список

1. *Колчин В.Ф.* Случайные отображения. – М., Наука, 1984.
2. *Михайлов В.Г.* Труды по дискретной математике. – М.:ФИЗМАЛИТ, 2002, том 5, 167-172.
3. *Степанов В.Е.* О распределении числа вершин в слоях случайного дерева. - Теория вероятностей и её применения, 1969, 14, № 1, 64-77.
4. *Flajolet P., Odlyzko A.* Random mapping statistics, Advances in Cryptology (J-J. Quisquater and J. Vandewalle, eds.), Lecture Notes in Computer Science, vol. 434, Springer Verlag, 1990, Proceedings of EUROCRYPT’89, Houtalen, Belgium, April 1989, pp. 329-354.