Доказательство теоремы о связности случайного графа

1. Неравенства Маркова и Чебышева

Неравенства Маркова и Чебышева будут сформулированы в случае, когда случайные величины принимают лишь конечное множество значений. Действительно, поскольку пространство случайных графов является конечным, любая функция, определенная на этом пространстве, будет принимать лишь конечное множество значений.

Утверждение (Неравенство Маркова) *Пусть ξ — случайная величина на конечном вероятностном пространстве, которая на любом элементарном исходе ω из этого пространства принимает неотрицательные значения $\xi(\omega) \geq 0$. Также пусть зафиксировано число $a > 0$. Тогда:*

$$\mathrm{P}(\xi \geq a) \leq \frac{\mathrm{M}\xi}{a},$$

где $\mathrm{M}\xi$ — математическое ожидание случайной величины ξ.

Следствием из неравенства Маркова является неравенство Чебышева.

Утверждение (Неравенство Чебышева) *Пусть ξ — случайная величина на конечном вероятностном пространстве, $a > 0$, тогда:*

$$\mathrm{P}\left(|\xi - \mathrm{M}\xi| \geq a\right) \leq \frac{\mathrm{D}\xi}{a^2},$$

где $\mathrm{D}\xi$ — дисперсия случайной величины ξ.

Доказательство этих неравенств приводится в курсе теории вероятностей.

С точки зрения задач о случайных графах, неравенства Маркова и Чебышева в некотором смысле дополняют друг друга. Неравенство Маркова позволяет доказать, что какие-то объекты почти наверняка не существуют, а неравенство Чебышева, наоборот, используется для того, чтобы доказать, что с высокой вероятностью случайный объект найдется.

Теорема о связности случайного графа

Теорема *Пусть $p = p(n) = \frac{c \ln n}{n}$, $c = \mathrm{const}$, $c > 0$.*

1. *Если $c > 1$, то а.п.н. случайный граф связен.*
2. *Если $c < 1$, то а.п.н. случайный граф не связен.*

Сначала будет доказан второй пункт теоремы.

2. Доказательство теоремы в случае $c < 1$

Доказательство Пусть вспомогательная случайная величина $\xi(G)$ — количество изолированных вершин в графе G. Достаточно показать, что существует хотя бы одна изолированная вершина, то есть $\mathrm{P}(\xi(G) \geq 1) \to 1$ при $n \to \infty$. В этом случае граф заведомо будет несвязным. Для этого можно воспользоваться неравенством Чебышева (при $a = \mathrm{M}\xi$):

$$\mathrm{P}(\xi \geq 1) = 1 - \mathrm{P}\left(\xi \leq 0\right) = 1 - \mathrm{P}\left(-\xi \geq 0\right) = 1 - \mathrm{P}\left(\mathrm{M}\xi - \xi \geq \mathrm{M}\xi\right) \geq 1 - \mathrm{P}\left(|\mathrm{M}\xi - \xi| \geq \mathrm{M}\xi\right) \geq 1 - \frac{\mathrm{D}\xi}{(\mathrm{M}\xi)^2}.$$

Случайную величину ξ можно представить в виде суммы индикаторных событий $\xi_i(G)$:

$$\xi = \sum_{i=1}^{n} \xi_i, \qquad \xi_i(G) = \begin{cases} 1, & \text{если вершина } i \text{ изолирована в } G \\ 0, & \text{иначе.} \end{cases}$$

Для математического ожидания индикаторной величины:

$$\mathrm{M}\xi_i = \mathrm{P}(\xi_i(G)=1) = (1-p)^{n-1},$$

так как вершина i будет изолированной только в случае, когда отсутствуют все ребра, соединяющие ее с каждой из оставшихся, и отсутствие ребер — взаимнонезависимые события. Таким образом, используя свойство линейности математического ожидания:

$$\mathrm{M}\xi = \sum_{i=1}^{n} \mathrm{M}\xi_i = n(1-p)^{n-1} = n\cdot e^{(n-1)\ln(1-p)} = n\cdot e^{-np(1+o(1))} = n\cdot e^{-c\ln n(1+o(1))} = \frac{n}{n^{c(1+o(1))}} \to \infty.$$

Поскольку для индикаторных случайных величин $\xi_i^2 = \xi_i$:

$$\mathrm{M}\xi^2 = \mathrm{M}(\xi_1+...+\xi_n)^2 = \mathrm{M}\left(\xi_1^2+...+\xi_n^2\right) + \sum_{i\neq j}\mathrm{M}\xi_i\xi_j = \mathrm{M}\left(\xi_1+...+\xi_n\right) + \sum_{i\neq j}\mathrm{M}\xi_i\xi_j = \mathrm{M}\xi + \sum_{i\neq j}\mathrm{M}\xi_i\xi_j.$$

Чтобы $\xi_i(G)\xi_j(G)=1$, необходимо, чтобы отсутствовали ребра, соединяющие i и j вершины с оставшимися, и ребро между i и j вершиной (то есть $2n-3$ ребра):

$$\mathrm{M}\xi_i\xi_j = \mathrm{P}(\xi_i(G)\xi_j(G)=1) = (1-p)^{2n-3}, \qquad \sum_{i\neq j}\mathrm{M}\xi_i\xi_j = n(n-1)(1-p)^{2n-3}.$$

Так как дисперсия $\mathrm{D}\xi = \mathrm{M}\xi^2 - (\mathrm{M}\xi)^2$, то:

$$\frac{\mathrm{D}\xi}{(\mathrm{M}\xi)^2} = \frac{n(1-p)^{n-1}+n(n-1)(1-p)^{2n-3}}{n^2(1-p)^{2n-2}} - 1 = o(1) - 1 + (1+o(1))(1-p)^{-1} = o(1) \to 0.$$

Таким образом, $\mathrm{P}(\xi\geq 1) \geq 1 - \frac{\mathrm{D}\xi}{(\mathrm{M}\xi)^2} \to 1$ при $n\to\infty$. Утверждение доказано.

3. Доказательство теоремы в случае $c > 1$

Замечание Для случайной величины $\xi(G)$, равной количеству изолированных вершин в случайном графе, в случае $c>1$ неравенство Маркова будет иметь вид:

$$\mathrm{M}\xi = \frac{n}{n^{c(1+o(1))}} \to 0, \qquad n\to\infty,$$

то есть в графе почти наверное не будет изолированных вершин.

Однако, чтобы доказать теорему, необходимо оценить количество не только изолированных вершин, но и компонент связности на двух или более вершинах.

Доказательство (случай $c>1$**)** Пусть введена случайная величина $\xi(G)$, равная числу нетривиальных компонент связности:

$$\xi(G) = \begin{cases} 0, \text{если граф связен} \\ k, \text{если число компонент связности равно } k. \end{cases}$$

Вероятность того, что случайный граф G связен, равна $\mathrm{P}(\xi=0)$. Согласно неравенству Маркова:

$$\mathrm{P}(\xi=0) = 1-\mathrm{P}(\xi\geq 1) \geq 1-\mathrm{M}\xi.$$

Так как случайная величина ξ может быть представлена в виде $\xi = \sum_{i=1}^{n-1}\xi_i$, где $\xi_i(G)$ — случайная величина, равная числу компонент связности на i вершинах. Тогда, используя свойство линейности математического ожидания:

$$\mathrm{M}\xi = \mathrm{M}\left(\sum_{i=1}^{n-1}\xi_i\right) = \sum_{i=1}^{n-1}\mathrm{M}\xi_i,$$

Каждая из случайных величин ξ_i, в свою очередь, может быть представлена в виде суммы по индикаторным случайным величинам. Пусть $A_1,...,A_{C_n^i}$ — всевозможные i-элементные подмножества множества вершин,

$$\xi_i = \sum_{j=1}^{C_n^i}\xi_{i,j}, \qquad \xi_{i,j}(G) = \begin{cases} 1, \text{если множество } A_j \text{ образует связную компоненту в } G \\ 0, \text{иначе.} \end{cases}$$

Еще раз используя линейность математического ожидания, а также свойство индикаторных величин:

$$\mathrm{M}\xi_i = \sum_{j=1}^{C_n^i} \mathrm{M}\xi_{i,j}, \qquad \mathrm{M}\xi_{i,j} = \mathrm{P}(A_j \text{ образует компоненту связности в случ. графе}).$$

Событие, состоящее в том, что A_j образует компоненту связности в случайном графе:

$$\{A_j \text{ обр. компоненту связности в } G\} = \{\text{Подграф } G|_{A_j} \text{ связен}\} \bigcap \{\forall v \in A_j \forall w \in V \;\; A_j \quad (v,w) \notin E\}.$$

Следовательно его вероятность можно оценить как:

$$\mathrm{P}(A_j \text{ обр. компоненту связности в } G) \leq \mathrm{P}\left(\{\forall v \in A_j \forall w \in V \;\; A_j \quad (v,w) \notin E\}\right) = (1-p)^{i(n-i)}.$$

Таким образом, можно получить оценку сверху для математического ожидания ξ:

$$\mathrm{M}\xi_i = \sum_{j=1}^{C_n^i} \mathrm{M}\xi_{i,j} \leq \sum_{j=1}^{C_n^i} (1-p)^{i(n-i)} = C_n^i (1-p)^{i(n-i)}, \quad \Longrightarrow \quad \mathrm{M}\xi \leq \sum_{i=1}^{n-1} C_n^i (1-p)^{i(n-i)}.$$

Пусть $a_i(n) = C_n^i(1-p)^{i(n-i)}$. Так как $a_i = a_{n-i}$, для доказательства теоремы **достаточно показать**, что:

$$\sum_{i=1}^{n/2} a_i = \sum_{i=1}^{n/\ln\ln n} a_i(n) + \sum_{i=n/\ln\ln n}^{n/2} a_i(n) \to 0, \qquad n \to \infty.$$

Так как в случае $i < \frac{n}{\ln\ln n}$ при $n \to \infty$:

$$\frac{a_{i+1}(n)}{a_i(n)} = \frac{C_n^{i+1}(1-p)^{(i+1)(n-i-1)}}{C_n^i(1-p)^{i(n-i)}} = \frac{n-i}{i+1}(1-p)^{-i+n-i-1} \leq n(1-p)^{-i+n-i-1} = n(1-p)^{n(1+o(n))} := q_n \to 0,$$

то первую сумму можно оценить следующим образом:

$$\sum_{i=1}^{n/\ln\ln n} a_i(n) = a_1(n)\left(1 + \frac{a_2(n)}{a_1(n)} + \frac{a_3(n)}{a_2(n)}\frac{a_2(n)}{a_1(n)} + \frac{a_4(n)}{a_3(n)}\frac{a_3(n)}{a_2(n)}\frac{a_2(n)}{a_1(n)} + \ldots\right) \leq a_1(n)(1+q(n)+q^2(n)+\ldots) = \frac{a_1(n)}{1-q(n)}.$$

Таким образом, для первой суммы:

$$\sum_{i=1}^{n/\ln\ln n} a_i(n) \leq \frac{a_1(n)}{1-q(n)} \to 0, \qquad n \to \infty.$$

Чтобы оценить вторую сумму, можно грубо оценить каждое слагаемое:

$$C_n^i(1-p)^{i(n-i)} \leq 2^n e^{i(n-i)\ln(1-p)} \leq 2^n e^{-pi(n-i)} \leq 2^n e^{-p\frac{n}{\ln\ln n}\frac{n}{2}} = 2^n e^{-\frac{c\ln n}{n}\frac{n}{\ln\ln n}\frac{n}{2}} = e^{n\ln 2 - \frac{c}{2}\frac{n\ln n}{\ln\ln n}}.$$

Тогда, поскольку $\ln n$ растет значительно быстрее $\ln\ln n$, а n — быстрее $\ln n$, для второй суммы:

$$\sum_{i=n/\ln\ln n}^{n/2} C_n^i(1-p)^{i(n-1)} \leq n e^{n\ln 2 - \frac{c}{2}\frac{n\ln n}{\ln\ln n}} = e^{\ln n + n\ln 2 - \frac{c}{2}\frac{n\ln n}{\ln\ln n}} \to 0, \qquad n \to \infty.$$

Теорема доказана.