# Математические основы Computer Science Часть 2: Вероятностный метод. Лекция 8. Метод второго момента

Дмитрий Ицыксон

ПОМИ РАН

6 декабря 2009

# Содержание лекции

1. Дисперсия
2. Неравенство Чебышева
3. Порог для 4-клики
4. Сепаратор

## Литература


1. Н. Алон, Дж. Спенсер. Вероятностный метод.
2. S. Jukna. Extremal combinatorics.

# Дисперсия

Определение. Дисперсией случайной величины $\xi$ называется величина $\mathsf{D}\,\xi = \mathsf{E}[(\xi - \mathsf{E}\,\xi)^2]$.
$\mathsf{D}\,\xi = \mathsf{E}[\xi - \mathsf{E}[\xi]]^2 = \mathsf{E}[\xi^2 - 2\xi\,\mathsf{E}[\xi] + (\mathsf{E}[\xi])^2] = \mathsf{E}[\xi^2] - (\mathsf{E}[\xi])^2 \geq 0$.
Лемма. (Неравенство Чебышева). $\mathsf{D}[\xi] = \sigma^2$, тогда

$$\Pr[|\xi - \mathsf{E}[\xi]| \geq k\sigma] \leq \frac{1}{k^2}.$$

Доказательство. $\eta = (\xi - \mathsf{E}[\xi])^2$, $\mathsf{E}[\eta] = \sigma^2$.
$\Pr[|\xi - \mathsf{E}\,\xi| \geq k\sigma\} = \Pr\{\eta \geq k^2\sigma^2] \overset{\text{(нер-во Маркова)}}{\leq} \frac{1}{k^2}$.

# Дисперсия суммы

- $Cov[\xi_1, \xi_2] = \mathsf{E}[\xi_1 \xi_2] - \mathsf{E}[\xi_1]\,\mathsf{E}[\xi_2]$
- $\mathsf{D}[\sum_{i=1}^n \xi_1] = \sum_{i=1}^n D[\xi_i] + 2\sum_{i<j} Cov[\xi_i, \xi_j]$
  - $\mathsf{D}[\sum_i \xi_i] = \mathsf{E}[(\sum_i \xi_i)^2] - (\mathsf{E}[\sum_i \xi_i])^2$
  - $\mathsf{E}[(\sum_i \xi_i)^2] = \mathsf{E}[\sum_i \xi_i^2 + 2\sum_{i<j} \xi_i \xi_j] = \sum_i \mathsf{E}[\xi_i^2] + 2\sum_{i<j} \mathsf{E}[\xi_i \xi_j]$
  - $(\mathsf{E}[\sum_i \xi_i])^2 = (\sum_i \mathsf{E}[\xi_i])^2 = \sum_i (\mathsf{E}[\xi_i])^2 + 2\sum_{i<j} \mathsf{E}[\xi_i]\,\mathsf{E}[\xi_j]$
  - $\mathsf{D}[\sum_i \xi_i] = \sum_i \mathsf{E}[\xi_i^2] - \sum_i (\mathsf{E}[\xi_i])^2 + 2\sum_{i<j} \mathsf{E}[\xi_i \xi_j] - 2\sum_{i<j} \mathsf{E}[\xi_i]\,\mathsf{E}[\xi_j] = \sum_i D[\xi_i] + 2\sum_{i<j} Cov[\xi_i, \xi_j]$
- Если $\xi_1, \xi_2, \ldots, \xi_n$ попарно независимы, то $\mathsf{D}[\xi_1 + \xi_2 + \cdots + \xi_n] = \mathsf{D}[\xi_1] + \mathsf{D}[\xi_2] + \cdots + \mathsf{D}[\xi_n]$.

# Порог существования 4-клики

- $G(n, p)$ — случайный граф на $n$ вершинах, каждое ребро добавляется независимо с вероятностью $p$.
- Задача: при каких $p$ в графе $G(n, p)$ с большой вероятностью есть клика на 4 вершинах?
- Теорема. Если $p = o(n^{-2/3})$, то в графе $G(n, p)$ есть 4-клика с вероятностью $o(1)$. Если $p = \omega(n^{-2/3})$, то в графе $G(n, p)$ есть 4-клика с вероятностью $1 - o(1)$.
  Доказательство.
  - $S \subseteq V$, $|S| = 4$, $A_S$: $S$ образует клику.
  - $X_S = \begin{cases} 1, \text{если } S \text{ образует клику } G \\ 0, \text{иначе} \end{cases}$
  - $E[X_S] = \Pr[X_S = 1] = p^6$
  - $X = \sum X_S$
  - $\Pr[X \geq 1] \leq E[X] = C_n^4 p^6 \leq \frac{n^4 p^6}{24} = o(1)$, если $p = o(n^{-2/3})$

## Порог существования 4-клики

- $\Pr[X = 0] \leq \frac{\mathrm{D}[X]}{\mathrm{E}[X]^2}$
- $\mathrm{D}[X] = \sum_S \mathrm{D}[X_S] + \sum_{S \neq T} Cov(X_S, X_T)$
- $\mathrm{D}[X_S] = \mathrm{E}[X_S^2] - \mathrm{E}[X_S]^2 \leq \mathrm{E}[X_S^2] = \mathrm{E}[X_S] = p^6$
- $\sum_S \mathrm{D}[X_S] = O(n^4 p^6)$
- Если $A_S$ и $A_T$ независимы, то $Cov(X_S, X_T) = 0$.
- Если $A_S$ и $A_T$ зависимы, если $|S \cap T| = 2$ или $|S \cap T| = 3$.
- Если $|S \cap T| = 2$, таких пар $O(n^6)$,
  $Cov[X_S, X_T] \leq \mathrm{E}[X_S X_T] = p^{11}$
- Если $|S \cap T| = 3$, таких пар $O(n^5)$,
  $Cov[X_S, X_T] \leq \mathrm{E}[X_S X_T] = p^9$
- $\mathrm{D}[X] = O(n^4 p^6 + n^6 p^{11} + n^5 p^9) = o(n^8 p^{12}) = o((\mathrm{E}[X])^2)$

## Сепараторы

Теорема. $X = \{1, 2, \dots, n\}$, $\mathcal{F} \subseteq 2^X$. 1) $\forall A \in \mathcal{F}, |A| \le r$; 2) $d_x = |\{A \in \mathcal{F} \mid x \in A\}|$; 3) $d = \frac{\sum_{x \in X} d_x}{n}$;
Тогда существует пара множеств $S, T \subseteq X$, что $S \cap T = \emptyset$ и для любого $A \in \mathcal{F}$ выполняется либо $A \cap S = \emptyset$, либо $A \cap T = \emptyset$, кроме того $|S|, |T| \ge (1 - \delta)2^{-d}n$, где $\delta = \sqrt{\frac{dr2^{d+1}}{n}}$.

Доказательство.

- Покрасим все множества из $\mathcal{F}$ случайным образом в синий и красные цвета.
- $S = \{x \in X \mid \forall A \in \mathcal{F}, x \in A \implies A \text{ синее}\}$
- $T = \{x \in X \mid \forall A \in \mathcal{F}, x \in A \implies A \text{ красное}\}$
- $S \cap T = \emptyset$, $\forall A \in \mathcal{F}$ либо $A \cap S = \emptyset$, либо $A \cap T = \emptyset$.
- Цель: показать, что с положительной вероятностью $|S|, |T| \ge (1 - \delta)2^{-d}n$.
- $Z_x = \begin{cases} 1, x \in S \\ 0, \text{иначе} \end{cases}$, $\mathrm{E}[Z_x] = \Pr[Z_x = 1] = 2^{-d_x}$
- $Z = \sum_x Z_x$, $Z = |S|$

# Сепараторы

- $Z_x = \begin{cases} 1, x \in S \\ 0, \text{иначе} \end{cases}$, $\mathsf{E}[Z_x] = \Pr[Z_x = 1] = 2^{-d_x}$
- $Z = \sum_x Z_x$, $Z = |S|$
- $\mathsf{E}[Z] = \sum_x 2^{-d_x} \geq n 2^{-\sum_x d_x / n} = n 2^{-d}$
- $\mathsf{D}[Z] = \sum_x \mathsf{D}[Z_x] + \sum_{x \neq y} Cov[Z_x, Z_y]$
- Если нет $A \in \mathcal{F}$, что $x, y \in A$, то $Z_x$ и $Z_y$ независимы, поэтому $Cov(Z_x, Z_y) = 0$.
- Для $x \in X$ существует не более $(r-1)d_x$ элементов $y$, что $x, y$ лежат в одном $A \in \mathcal{F}$
- $Cov(Z_x, Z_y) \leq \mathsf{E}[Z_x Z_y] \leq \mathsf{E}[Z_x] \leq 2^{-d_x}$
- $\sum_{y \neq x} Cov(Z_x, Z_y) \leq (r-1) \sum_x d_x 2^{-d_x} \leq \frac{r-1}{n} (\sum_x 2^{-d_x})(\sum_x d_x) = d(r-1)E[Z]$
- $\mathsf{D}[Z] \leq (d(r-1)+1)\,\mathsf{E}[Z] \leq dr\,\mathsf{E}[Z]$
- $\Pr[Z < (1-\delta)\,\mathsf{E}[Z]] < \frac{\mathsf{D}[Z]}{\delta^2 (\mathsf{E}[Z])^2} \leq \frac{dr}{\delta^2\,\mathsf{E}[Z]} = \frac{drn}{dr 2^{d+1}\,\mathsf{E}[Z]} = \frac{n}{2^{d+1}\,\mathsf{E}[Z]} \leq \frac{n}{2^{d+1} 2^{-d} n} = \frac{1}{2}$.