

# с/к “Эффективные алгоритмы”

## Лекция 19: Дерандомизация

А. Куликов

Computer Science клуб при ПОМИ  
<http://logic.pdmi.ras.ru/~infclub/>

## План лекции

- 1 Метод условных вероятностей
  - Общая идея
  - Общий метод
  - Нахождение двудольного подграфа
  - Максимальная выполнимость
- 2 Метод малых пространств событий
  - Независимость по  $k$
  - Нахождение двудольного подграфа
- 3 Свободные от сумм множества

## Методы

### Методы

- Естественным способом **дерандомизации** вероятностного алгоритма  $A$  является нахождение эффективного метода поиска в соответствующем пространстве событий  $\Omega$  точки  $w$ , хорошей относительно входа  $x$ .
- Точка  $w$  называется хорошей относительно  $x$ , если  $A(x, w) = f(x)$ .
- Проблема состоит в том, что пространство событий, как правило, имеет экспоненциальный размер.
- Два метода:
  - 1 **Метод условных вероятностей**. Начинает с тривиального пространства  $\Omega$ , но ищет в нём хитро.
  - 2 **Метод малых пространств событий**. Пытается построить вероятностное пространство  $\Omega'$  меньшего (полиномиального) размера, после чего перебирает все его точки.

## План лекции

- 1 Метод условных вероятностей
  - Общая идея
  - Общий метод
  - Нахождение двудольного подграфа
  - Максимальная выполнимость
- 2 Метод малых пространств событий
  - Независимость по  $k$
  - Нахождение двудольного подграфа
- 3 Свободные от сумм множества

## План лекции

- 1 Метод условных вероятностей
  - Общая идея
  - Общий метод
  - Нахождение двудольного подграфа
  - Максимальная выполнимость
- 2 Метод малых пространств событий
  - Независимость по  $k$
  - Нахождение двудольного подграфа
- 3 Свободные от сумм множества

## Общая идея

### Общая идея

- Основная цель метода — преобразовывать вероятностные доказательства **существования** комбинаторных объектов в эффективные детерминированные алгоритмы **построения** таких объектов.
- На каждом шаге текущее пространство событий делится на две равные по размеру части, для каждой из которых вычисляется **условная вероятность** нахождения в ней хорошей точки.
- После этого поиск продолжается в половине, для которой подсчитанная вероятность оказалась выше.
- Поиск завершается через  $\log_2 |\Omega|$  шагов, когда в пространстве окажется всего одна точка (которая должна быть хорошей).
- Метод работает, если такие условные вероятности могут быть эффективно подсчитаны (или хотя бы приближены).

## Пример: задача максимальной выполнимости

### Пример: задача максимальной выполнимости

- Рассмотрим КНФ формулу, содержащую только 3-клозы.
- Если присвоить каждой переменной значение случайным образом, то в среднем выполнится  $7/8$  всех клозов.
- Такой метод, однако, гарантирует лишь **существование** такого набора, но ничего не говорит о том, как его **искать**.

## Обезьяна и бананы

### Обезьяна и бананы

- Построим бинарное дерево с  $2^n$  листьями, соответствующими всем возможным наборам.
- Корень дерева будет снизу, листья — где-то в небе.
- Левая ветка каждого узла на уровне  $i$  будет соответствовать случаю  $x_i = 0$ , правая —  $x_i = 1$ .
- В каждый лист повесим чёрный ящик, в который положим столько бананов, сколько клозов выполняет соответствующий набор.
- После этого пригласим обезьяну достать какой-нибудь из ящиков.

### Обезьяна и бананы

- Повторив эксперимент много раз для разных формул с разными количествами клозов и переменных, с удивлением обнаружим, что обезьяна **всегда** достаёт ящик, количество бананов в котором не менее  $7/8$  от общего количества клозов.
- Обезьяна никогда не смотрит на формулу (хотя ей разрешается это делать).
- Спросив её, в чём же её секрет (у неё определённно он есть, это не могло быть просто удачей), получим такой ответ:

*“Элементарно! В каждом узле я делаю одно и то же: я смотрю, в каком поддереве больше бананов (ветка, поддерживающая это поддерево, наклонена ниже) и именно в него и иду.”*

- 1 Метод условных вероятностей
  - Общая идея
  - Общий метод
  - Нахождение двудольного подграфа
  - Максимальная выполнимость
- 2 Метод малых пространств событий
  - Независимость по  $k$
  - Нахождение двудольного подграфа
- 3 Свободные от сумм множества

## Общий метод

### Общий метод

- Рассмотрим пространство событий  $\Omega = \{0, 1\}^n$ .
- Пусть  $A_1, \dots, A_m$  — набор событий, а  $X_1, \dots, X_m$  являются их характеристическими переменными.
- Рассмотрим также случайную переменную  $X = X_1 + \dots + X_m$ .
- $E(X) = \sum_{i=1}^m \text{Prob}(A_i)$ .
- Допустим, что у нас есть доказательство того факта, что  $E(X) \geq k$ .
- Значит, есть точка  $(\epsilon_1, \dots, \epsilon_n)$  в пространстве событий, в которой выполнены хотя бы  $k$  событий.
- Наша задача — найти такую точку **детерминированно**.

## Общий метод (продолжение)

### Общий метод (продолжение)

- Введём  $n$  случайных переменных  $Y_1, \dots, Y_n$ , где каждая из  $Y_i$  принимает значение  $0$  и  $1$  независимо и равномерно.
- Мы находим биты  $\epsilon_1, \dots, \epsilon_n$ .
- Допустим,  $\epsilon_1, \dots, \epsilon_{j-1}$  уже выбраны.
- Выбираем значение для  $\epsilon_j$  в зависимости от значения условного мат. ожидания

$$E(X | \epsilon_1, \dots, \epsilon_j) = \sum_{i=1}^m \text{Prob}(A_i | \epsilon_1, \dots, \epsilon_j),$$

где под “ $\epsilon_1, \dots, \epsilon_j$ ” понимается событие “ $Y_1 = \epsilon_1, \dots, Y_j = \epsilon_j$ ”.

## Общий метод (продолжение)

### Общий метод (продолжение)

- Ясно, что

$$\text{Prob}(A_i | \epsilon_1, \dots, \epsilon_j) = \frac{\text{Prob}(A_i | \epsilon_1, \dots, \epsilon_j, 0) + \text{Prob}(A_i | \epsilon_1, \dots, \epsilon_j, 1)}{2}$$

- Таким образом,

$$\mathbf{E}(X | \epsilon_1, \dots, \epsilon_j) = \frac{\mathbf{E}(X | \epsilon_1, \dots, \epsilon_j, 0) + \mathbf{E}(X | \epsilon_1, \dots, \epsilon_j, 1)}{2},$$

откуда заключаем, что

$$\mathbf{E}(X | \epsilon_1, \dots, \epsilon_j) \leq \max\{\mathbf{E}(X | \epsilon_1, \dots, \epsilon_j, 0), \mathbf{E}(X | \epsilon_1, \dots, \epsilon_j, 1)\}.$$

## Общий метод (завершение)

### Общий метод (завершение)

- Таким образом, если выбирать значения  $\epsilon_{j+1}$  так, чтобы максимизировать значение  $\mathbf{E}(X | \epsilon_1, \dots, \epsilon_{j+1})$ , то его значение не будет уменьшаться.
- Поскольку в начале оно было не менее  $k$ , то будет оно не менее  $k$  и в конце.
- Но в конце у каждого  $\epsilon_i$  уже есть значение, а значит,  $\mathbf{E}(X | \epsilon_1, \dots, \epsilon_n)$  будет равно количеству событий, выполненных в данной точке.

### Замечание

Метод работает, если  $n$  не слишком велико (это, как правило, выполнено) и, что более важно, если условные вероятности  $\text{Prob}(A_i | \epsilon_1, \dots, \epsilon_j)$  могут быть вычислены эффективно.

## План лекции

- 1 Метод условных вероятностей
  - Общая идея
  - Общий метод
  - Нахождение двудольного подграфа
  - Максимальная выполнимость
- 2 Метод малых пространств событий
  - Независимость по  $k$
  - Нахождение двудольного подграфа
- 3 Свободные от сумм множества

## Существование двудольного подграфа

### Теорема

В любом графе с  $m$  рёбрами существует двудольный подграф с не менее чем  $m/2$  рёбрами.

### Доказательство

- Рассмотрим граф  $G = (V, E)$ , где  $V = \{1, \dots, n\}$ .
- Рассмотрим также случайное подмножество его вершин  $U$ , включая в это подмножество каждую вершину независимо с вероятностью  $1/2$ .
- Ребро  $e = \{i, j\}$  назовём пересекающим, если ровно один из его концов лежит в  $U$ .
- Через  $X$  обозначим количество пересекающих рёбер.
- Тогда  $X = \sum_{e \in E} X_e$ , где  $X_e$  — индикатор того, что  $e$  является пересекающим.
- Ясно, что  $\mathbf{E}(X_e) = 1/2$ .

## Доказательство (завершение)

### Доказательство

- Таким образом,

$$\mathbf{E}(X) = \sum_{e \in E} \mathbf{E}(X_e) = \frac{|E|}{2}.$$

- Значит,  $X \geq |E|/2$  для некоторого подмножества  $U$ .  $\square$

### Замечание

Данная теорема показывает, что существует вероятностный алгоритм нахождения двудольного подграфа с мат. ожиданием числа рёбер не менее  $|E|/2$ . Мы покажем, как превратить этот алгоритм в детерминированный.

## Дерандомизация

### Дерандомизация

- Заведём  $n$  случайных переменных  $Y_1, \dots, Y_n$ , где  $Y_i = 1 \Leftrightarrow i \in U$ .
- Выбираем  $\epsilon_1 \in \{0, 1\}$  так, чтобы  $\mathbf{E}(X \mid Y_1 = \epsilon_1) \geq \mathbf{E}(X \mid Y_1 = \epsilon_1 \oplus 1)$  и присваиваем  $Y_1$  значение  $\epsilon_1$ .
- Повторим этот процесс для всех  $Y_j$ .
- В конце получим  $\mathbf{E}(X \mid Y_1 = \epsilon_1, \dots, Y_n = \epsilon_n) \geq |E|/2$ .
- При этом  $X$  уже не является случайной переменной (поскольку  $U$  уже задано), поэтому  $X \geq |E|/2$ .
- Ясно, что необходимые условные вероятности просто вычисляются.

## План лекции

### 1 Метод условных вероятностей

- Общая идея
- Общий метод
- Нахождение двудольного подграфа
- Максимальная выполнимость

### 2 Метод малых пространств событий

- Независимость по  $k$
- Нахождение двудольного подграфа

### 3 Свободные от сумм множества

## Существование

### Теорема

Для любой формулы  $F = C_1 \wedge \dots \wedge C_m$  существует набор, выполняющий хотя бы  $(1 - 2^{-l})$  от общего числа кловов, где  $l$  — длина минимального клова формулы  $F$ .

### Доказательство

- Присвоим каждой переменной значение 0 или 1 с вероятностью  $1/2$ . Полученный набор обозначим через  $w$ .
- Пусть  $l_j = |C_j|$ ,  $Z_j$  — индикатор того, что клов  $C_j$  выполнен.
- $\mathbf{E}(Z_j) = \mathbf{Prob}(C_j(w) = 1) = 1 - \mathbf{Prob}(C_j(w) = 0) = 1 - 2^{-l_j}$ .
- Пусть  $Z$  — количество выполненных кловов, тогда

$$\mathbf{E}(Z) = \sum_{j=1}^m (1 - 2^{-l_j}) \geq m(1 - 2^{-l}).$$

$\square$

## Дерандомизация

### Дерандомизация

- Покажем, как найти соответствующий набор  $x_1 = w_1, \dots, x_n = w_n$ .
- Пусть  $w_1, \dots, w_i$  уже выбраны. Введём обозначения:  
 $e_i = \mathbf{E}(Z \mid x_1 = w_1, \dots, x_i = w_i)$ ,  
 $e_{i0} = \mathbf{E}(Z \mid x_1 = w_1, \dots, x_i = w_i, x_{i+1} = 0)$ ,  
 $e_{i1} = \mathbf{E}(Z \mid x_1 = w_1, \dots, x_i = w_i, x_{i+1} = 1)$ .
- Тогда  $e_0 = \mathbf{E}(Z)$  и  $e_i = (e_{i0} + e_{i1})/2$ .
- На каждом шаге присваиваем  $x_{i+1}$  такое значение  $w_{i+1}$ , для которого  $e_{i w_{i+1}} \geq e_i$ .
- В данном случае нет необходимости точного вычисления условных вероятностей  $e_{i0}$  и  $e_{i1}$ .
- Мы присваиваем  $x_{i+1}$  значение 1, если  $e_{i1} - e_i \geq 0$ .

## Дерандомизация (завершение)

### Дерандомизация (завершение)

- Присвоение значения переменной  $x_{i+1}$  влияет, конечно, только на клозы, содержащие эту переменную.
- Каждый клоз  $C_j$ , содержащий литерал  $x_{i+1}$ , будет выполнен (при присвоении  $x_{i+1} = 1$ ) и увеличит  $e_i$  на  $1 - (1 - 2^{-l_j}) = 2^{-l_j}$ .
- Клоз  $C_j$ , содержащий литерал  $\bar{x}_{i+1}$ , станет на один литерал короче, что уменьшит его вероятность быть выполненным с  $(1 - 2^{-l_j})$  до  $(1 - 2^{-(l_j-1)})$  (то есть опять на  $2^{-l_j}$ ).
- Таким образом,

$$e_{i1} - e_i = \sum_{C_j \in \mathcal{C}: C_j \ni x_{i+1}} 2^{-l_j} - \sum_{C_j \in \mathcal{C}: C_j \ni \bar{x}_{i+1}} 2^{-l_j},$$

где  $\mathcal{C}$  — множество оставшихся к данному моменту клозов,  $l_j$  — длина клоза  $C_j$ .

## План лекции

- 1 Метод условных вероятностей
  - Общая идея
  - Общий метод
  - Нахождение двудольного подграфа
  - Максимальная выполнимость
- 2 Метод малых пространств событий
  - Независимость по  $k$
  - Нахождение двудольного подграфа
- 3 Свободные от сумм множества

## План лекции

- 1 Метод условных вероятностей
  - Общая идея
  - Общий метод
  - Нахождение двудольного подграфа
  - Максимальная выполнимость
- 2 Метод малых пространств событий
  - Независимость по  $k$
  - Нахождение двудольного подграфа
- 3 Свободные от сумм множества

## Введение

### Введение

- Рассмотрим некоторую функцию  $f$ , определенную на  $\{0, 1\}^n$ .
- Вероятностный алгоритм  $A$  для вычисления  $f$  имеет некоторое вероятностное пространство  $(\Omega, P)$ , где  $\Omega$  — пространство событий, а  $P$  — вероятностная мера.
- Считаем, что все точки  $\Omega$  равновероятны.
- Называем точку  $w \in \Omega$  хорошей для входа  $x$ , если  $A(x, w) = f(x)$ .
- Говорим, что алгоритм  $A$  вычисляет  $f$ , если  $\text{Prob}(A(x, w) = f(x)) \geq 1/2$  для всех  $x$ .
- Проблема в том, что  $\Omega$ , как правило, имеет экспоненциальный размер.

## Существование меньшего пространства событий

### Теорема

Существует множество  $S \subseteq \Omega$  размера не более  $n$ , такое что для любого входа  $x \in \{0, 1\}^n$  найдётся хотя бы одна хорошая точка в  $S$ .

### Доказательство

- Рассмотрим 0/1-матрицу  $M = (m_{x,w})$  размера  $2^n \times |\Omega|$ , такую что  $m_{x,w} = 1 \Leftrightarrow w$  является хорошей точкой для  $x$ .
- В каждой строчке этой матрицы стоит хотя бы половина 1.
- Значит, есть и столбец, в котором есть хотя бы половина 1.
- Положим соответствующий элемент в  $S$ , выкинем соответствующие строчки, повторим.  $\square$

## Независимость по $k$

### Определение

Случайные переменные  $X_1, \dots, X_n$ , принимающие значения в конечном множестве  $S$ , называются **независимыми по  $k$**  ( $k$ -wise independent), если для любой последовательности  $(s_1, \dots, s_k)$  из  $k$  значений ( $s_j \in S$ ) выполнено равенство

$$\text{Prob}(X_{i_1} = s_{i_1}, \dots, X_{i_k} = s_{i_k}) = \prod_{j=1}^k \text{Prob}(X_{i_j} = s_{i_j}).$$

## План лекции

- 1 Метод условных вероятностей
  - Общая идея
  - Общий метод
  - Нахождение двудольного подграфа
  - Максимальная выполнимость
- 2 Метод малых пространств событий
  - Независимость по  $k$
  - Нахождение двудольного подграфа
- 3 Свободные от сумм множества

## Существование двудольного графа

### Существование двудольного графа

- Мы доказали, что в любом графе  $G = (V, E)$  есть бинарный подграф с хотя бы  $|E|/2$  рёбрами.
- Другими словами, найдётся подмножество вершин  $U \subseteq V$ , такое что количество рёбер между  $U$  и  $V \setminus U$  не менее  $|E|/2$ .
- Мы использовали случайные переменные для нахождения такого  $U$ .
- Именно, для каждой вершины  $i \in V$  мы бросали монетку для принятия решения, добавлять  $i$  в  $U$  или нет.
- Это требовало  $n$  подбрасываний монеток, поэтому соответствующее пространство событий имело огромный размер —  $2^n$ .

## Где использовалась независимость?

### Где использовалась независимость?

- Независимость использовалась только для того, чтобы заключить, что для любых двух вершин  $i \neq j$  события  $i \in U$  и  $j \in U$  независимы.
- Используя это, мы показывали, что

$$\text{Prob}(i \in U, j \notin U) = \text{Prob}(i \in U) \cdot \text{Prob}(j \notin U) = 1/4.$$

- В этом доказательстве, таким образом, достаточно независимости по 2 индикаторов  $X_i$  событий  $i \in U$ .
- Это замечание позволяет нам значительно уменьшить размер соответствующего пространства событий.

## Малое пространство событий

### Малое пространство событий

- Будем рассматривать множество  $U$  как последовательность раскрасок  $X_1, \dots, X_n: V \rightarrow \{0, 1\}$ , где  $X_i = 1 \Leftrightarrow i \in U$ .
- Наша задача состоит в том, чтобы построить как можно меньшее пространство событий для этих переменных так, чтобы они были независимыми по 2.
- Предположим, что  $n = 2^d$ . Сопоставим вершинам графа элементы поля  $\mathbb{F}_n$ .
- Выберем случайно и независимо два элемента  $a$  и  $b$  этого поля и сопоставим каждому элементу  $i$  случайную переменную  $Z_i = a \cdot i + b$ .

### Лемма

Случайные переменные являются независимыми по 2.

## Доказательство

### Доказательство

$$\begin{aligned} \text{Prob}(Z_i = x, Z_j = y) &= \text{Prob}(ai + b = x, aj + b = y) \\ &= \text{Prob}\left(a = \frac{x-y}{i-j}, b = \frac{yi-xj}{i-j}\right) = \frac{1}{n^2} \\ &= \text{Prob}(Z_i = x) \cdot \text{Prob}(Z_j = y). \end{aligned}$$

□

## Построение малого пространства событий

### Построение малого пространства событий

- Присвоим теперь переменной  $X_i$  значение первого бита двоичной записи  $Z_i$ .
- По лемме переменные  $X_i$  также независимы по 2.
- Красим теперь каждую вершину в цвет  $X_i$ .
- Каждая такая раскраска задаётся парой  $(a, b)$  элементов поля  $\mathbb{F}_n$ .
- Пространство событий, таким образом, имеет размер всего  $n^2$ . Значит, для нахождения требуемой раскраски можно произвести полный перебор элементов пространства.

## План лекции

- 1 Метод условных вероятностей
  - Общая идея
  - Общий метод
  - Нахождение двудольного подграфа
  - Максимальная выполнимость
- 2 Метод малых пространств событий
  - Независимость по  $k$
  - Нахождение двудольного подграфа
- 3 Свободные от сумм множества

## Свободные от сумм множества

### Идея

Мы уже увидели два способа дерандомизации вероятностных алгоритмов. Приведём теперь пример вероятностного доказательства, в котором скрыт детерминированный алгоритм.

### Определение

Подмножество  $B$  некоторой аддитивной группы называется **свободным от сумм** (sum-free), если  $x + y \notin B$  для любых  $x, y \in B$ .

## Теорема

### Определение

- $\mathbb{Z}_n = \{0, 1, 2, \dots, n-1\}$  — множество всех вычетов (остатков) по модулю  $n$ .
- $\mathbb{Z}_n^*$  — множество остатков, взаимно простых с  $n$  (для простого  $n$   $\mathbb{Z}_n^* = \mathbb{Z}_n \setminus \{0\}$ ).

### Теорема

Пусть  $p = 3k + 2 \in \mathbb{P}$ ,  $w(x)$  — неотрицательная функция, определённая на  $\mathbb{Z}_p^*$ . Допустим, что  $W = \sum_{x \in \mathbb{Z}_p^*} w(x) > 0$ . Тогда найдётся свободное от сумм подмножество  $E$  группы  $\mathbb{Z}_p^*$ , для которого

$$\sum_{x \in E} w(x) > W/3.$$

## Доказательство

### Доказательство

- Пусть  $S = \{k + 1, \dots, 2k + 1\}$ .
- Заметим, что  $|S| > (p - 1)/3$  и  $S$  является свободным от сумм подмножеством группы  $\mathbb{Z}_p$ .
- Пусть  $t$  выбрано случайно из  $\mathbb{Z}_p^*$ .
- Пусть  $f(t) = \sum w(x)$ , где сумма берётся по всем  $x$ , для которых  $x \cdot t \in S$  ( $x \cdot t$  вычисляется в  $\mathbb{Z}_p$ ).
- Поскольку  $\mathbb{Z}_p^*$  является мультипликативной группой,  $E(f(t)) = W \cdot (|S|/(p - 1)) > W/3$ .
- Рассмотрим  $t \in \mathbb{Z}_p^*$ , для которого выполнено такое неравенство, и определим  $E$  как  $t^{-1}S$ .  $\square$

## Идеи алгоритма

### Идеи алгоритма

- Итак, нам дан набор чисел  $\{a_1, \dots, a_N\}$ , длина записи которого равна  $l = \sum_{i=1}^N \log_2 |a_i|$ .
- Наша задача состоит в нахождении свободного от сумм подмножества размера хотя бы  $N/3$  за полиномиальное от  $l$  время.
- Считаем, что  $l$  достаточно велико.
- Ясно, что у любого  $a_i$  есть не более  $l$  различных простых делителей.
- Известно, что для любых двух взаимно простых чисел  $b$  и  $c$  количество простых чисел  $p \leq x$  вида  $p = bk + c$  растёт как  $x/(\varphi(x) \cdot \ln x)$ , где  $\varphi(b) = |\{y \in \mathbb{Z}_b: \gcd(y, b) = 1\}|$  — функция Эйлера.
- Значит, найдётся простое число  $p \leq 3/\log_2 l$  вида  $p = 3k + 2$ , которое не делит ни одно из  $a_i$ .

## Идеи алгоритма (продолжение)

### Идеи алгоритма (продолжение)

- Пусть  $w(x) = |\{t \in A: t \equiv x \pmod{p}\}|$ .
- Поскольку  $p$  не делит ни одно  $a_i$ , то  $W = N$ .
- Значит, найдётся свободное от сумм множество  $E \subseteq \mathbb{Z}_p^*$ , для которого множество  $B = \{t \in A: t \bmod p \in E\}$  содержит хотя бы  $N/3$  элементов.
- Более того, данное множество  $B$  также является свободным от сумм, поскольку из  $x + y = z$  для некоторых  $x, y, z \in B$  следовало бы, что  $x + y \equiv z \pmod{p}$ , и  $E$  не было бы свободным от сумм.

## Алгоритм

### Алгоритм

- Посчитать все простые числа до  $3/\log_2 l$ .
- Найти среди них число  $p = 3k + 2$ , которое не делит ни одно число из набора  $A$ .
- Вычислить значения  $w(x)$  для всех  $x \in \mathbb{Z}_p^*$ .
- Полным перебором найти такое  $t \in \mathbb{Z}_p^*$ , для которого  $f(t) > N/3$  и построить множество  $E = t^{-1}S$ .
- Построить множество  $B = \{t \in A: t \bmod p \in E\}$ .

## Что мы узнали за сегодня?

### Что мы узнали за сегодня?

- Метод условных вероятностей на каждом шаге уменьшает пространство событий вдвое и в конце концов находит хорошую точку.
- Метод малых пространств строит меньшее пространство событий и находит в нём хорошую точку полным перебором.
- Иногда детерминированный алгоритм можно извлечь непосредственно из вероятностного доказательства.