Вычислительно трудные задачи и дерандомизация
Лекция 5: Псевдо-случайный генератор Нисана-Вигдерсона

Дмитрий Ицыксон

ПОМИ РАН

22 марта 2009

Дерандомизация

- Вероятностная машина Тьюринга:
  - Специальная лента со случайными битами
- **BPP** — это множество языков L, для которых $\exists$ полиномиальная по времени вероятностная машина Тьюринга M:

$$\forall x \Pr[M(x) = L(x)] \geq \frac{3}{4}$$

- Задача дерандомизации:
  - Избавиться от случайных битов
  - Не сильно проиграть по времени
- $\mathbf{P} \subseteq \mathbf{BPP} \subseteq \mathbf{EXP}$, $\mathbf{P} \subsetneq \mathbf{EXP}$
  - Если $\mathbf{P} = \mathbf{BPP}$, то дерандомизация есть
  - Если $\mathbf{BPP} = \mathbf{EXP}$, то дерандомизации нет

Когда случайные числа по делу?

Проверка равенства нулю многочлена

- Многочлен задан формулой в неканоническом виде.
- $(p^2 + qr - 21r)^3(p^{20}q - r^3p + s) + \ldots$
- Задача: проверить, является ли этот многочлен тождественным нулем.
- Раскрыть скобки и привести подобные: экспоненциальное время!
- Вероятностное решение:
  1. Оценить степень многочлена $d \leq 2^m$, где m — число умножений
  2. Найти $p > 10d$ — простое число.
  3. Взять случайный вектор из $\mathbb{Z}_p^n$ и подставить его в многочлен. Если получится 0, сказать, что многочлен нулевой, если не 0, то многочлен ненулевой.
- (Упражнение) Доказать корректность этого алгоритма.

Псевдослучайный генератор

Псевдослучайное распределение

Распределение R на $\{0,1\}^m$ называется (S,ϵ)-псевдослучайным, если для каждой схемы C размера $\leq S$:

$$|\Pr[C(R)=1]-\Pr[C(U_m)=1]|<\epsilon$$

Псевдослучайный генератор

$S(\ell)$ — правильная неубывающая функция. Функция $G:\{0,1\}^*\to\{0,1\}^*$ называется $S(\ell)$-псевдослучайным генератором, если

- G вычислима за время $2^{O(n)}$
- $\forall z\in\{0,1\}^\ell$, $|G(z)|=S(\ell)$
- $\forall\ell$ распределение $G(U_\ell)$ является $(S(\ell)^3,\frac{1}{10})$-псевдослучайным.

Чем помогает псевдослучайный генератор?

Лемма. Если существует $2^{\epsilon\ell}$-псевдослучайный генератор ($\epsilon > 0$), то $\mathbf{BPP} = \mathbf{P}$.

Доказательство.

- $L \in \mathbf{BPP}$, решается за n^c на вероятностной машине A.
- $\Pr_{r \leftarrow U(\{0,1\}^m)}[A(x, r) = L(x)] \geq \frac{3}{4}$.
- $m \leq n^c$, $2^{\epsilon\ell} = n^c \implies \ell = \frac{c}{\epsilon} \log n$
- Алгоритм B: для всех строк $z \in \{0,1\}^\ell$ (их $n^{\frac{c}{\epsilon}}$) запустим $A(x, G(z))$ и выдадим наиболее частый ответ
- Время работы B не превосходит $n^c n^{\frac{c}{\epsilon}} n^{O(\frac{c}{\epsilon})} = poly(n)$
- Пусть $\exists x : \Pr[A(x, G(z)) = L(x)] < \frac{3}{4} - 0.1$
- Взломщик для генератора: $r \mapsto A(x, r)$, вычислим схемой размера $(n^c)^2 < S(\ell)^3 = (n^c)^3$.

Схемная сложность функций

$f : \{0,1\}^n \to \{0,1\}$.

Сложность в наихудшем случае

$$H_{wrs}(f) = \max\{S | \forall \text{ схемы } C \text{ размера } \leq S \\ \exists x \in \{0,1\}^n : C(x) \neq f(x)\}$$

Сложность в среденем случае

$$H_{avg}(f) = \max\{S | \forall \text{ схемы } C \text{ размера } \leq S \\ \Pr_{x \leftarrow U(\{0,1\}^n)}[C(x) = f(x)] < \frac{1}{2} + \frac{1}{S}\}$$

- $H_{avg}(f) \leq H_{wrs}(f) \leq cn2^n$
- (Упражнение) Покажите, что для случайной f выполняется $H_{avg}(f) \geq 2^{n/10}$

Генератор из сложных функций

Теорема. (Нисан-Вигдерсон) Если существует $f : \{0,1\}^* \to \{0,1\}$, вычислимая за время $2^{O(n)}$, что $H_{avg}(f) \geq 2^{\epsilon n}$, тогда существует $2^{\epsilon' n}$-псевдослучайный генератор.

Следствие. При тех же предположениях $\mathbf{P} = \mathbf{BPP}$.

- Позже мы покажем, как по функции f, вычислимой за $2^{O(n)}$ с $H_{wrs}(f) = 2^{\Omega(n)}$ построить функцию f', вычислимую за $2^{O(n)}$ с $H_{avg}(f) = 2^{\Omega(n)}$.
- Но это будет не сегодня...

Непредсказуемость $\Longrightarrow$ псевдослучайность

Теорема. (Яо, 1982) Пусть Y — это распределение на $\{0,1\}^m$. Существует $S > 10m$, $\epsilon > 0$, что для каждой схемы C размера не больше $2S$ и всех $1 \leq i \leq m$ выполняется $\Pr_{r \leftarrow Y}[C(r_1, r_2, \ldots, r_{i-1}) = r_i] \leq \frac{1}{2} + \frac{\epsilon}{2m}$. Тогда Y является (S, ϵ)-псевдослучайным.

Доказательство. Гибридный метод. $r \leftarrow Y, u \leftarrow U(\{0,1\}^m)$

- $X_0 = (u_1, u_2, \ldots, u_m)$
- $X_1 = (r_1, u_2, \ldots, u_m)$
- $X_2 = (r_1, r_2, \ldots, u_m)$
- $\ldots$
- $X_m = (r_1, r_2, \ldots, r_m)$

- $p_0 = \Pr[C(X_0) = 1]$
- $p_1 = \Pr[C(X_1) = 1]$
- $p_2 = \Pr[C(X_2) = 1]$
- $\ldots$
- $p_m = \Pr[C(X_m) = 1]$

Продолжение доказательства

- Пусть $p_m - p_0 > \epsilon$
- $\exists i$, что $p_i - p_{i-1} > \frac{\epsilon}{m}$
- Предсказатель $D(r_1, r_2, \ldots, r_{i-1})$ — схема, которая использует случайные биты:
- $u_i, u_{i+1} \ldots, u_m, \rho$ — случайные биты
- $a := C(r_1, r_2, \ldots, r_{i-1}, u_i, u_{i+1} \ldots, u_m)$
- $D(r_1, r_2, \ldots, r_{i-1}) = \begin{cases} u_i, & a = 1 \\ \rho, & a = 0 \end{cases}$.

Продолжение доказательства

	$a = 0$	$a = 1$
$r_i = u_i$	α	β
$r_i \neq u_i$	γ	δ

- $\alpha + \beta = \gamma + \delta = \frac{1}{2}$
- $p_{i-1} = \beta + \delta$
- $p_i = \Pr[C(r_1, r_2, \ldots, r_i, u_{i+1}, \ldots, u_m) = 1] =$
  $\Pr[C(r_1, \ldots, r_{i-1}, u_i, \ldots, u_m) = 1 | u_i = r_i] =$
  $2\Pr[C(r_1, \ldots, r_{i-1}, u_i, \ldots, u_m) = 1] = 2\beta$
- $\Pr[D(r_1, r_2, \ldots, r_{i-1}) = r_i] = \frac{\alpha+\gamma}{2} + \beta = \frac{1}{2} + \frac{\beta-\delta}{2} =$
  $\frac{1}{2} + \frac{p_i - p_{i-1}}{2} \geq \frac{1}{2} + \frac{\epsilon}{2m}$
- Осталось выбрать “лучшую” последовательность случайных битов и зашить их в схему D.

Игрушечный $(\ell+1)$-генератор

- Пусть f вычислима за время $2^{O(n)}$ и $H_{avg}(f) \geq n^4$.
- Тогда $G(z) = z \circ f(z)$ есть $(\ell+1)$-генератор.
- Первые ℓ битов не предсказать по предыдущим, так как они случайные.
- Пусть $(\ell+1)$-й бит можно предсказать:
  $\Pr_{z \leftarrow U(\{0,1\}^\ell)}[C(z) = f(z)] \geq \frac{1}{2} + \frac{\epsilon}{2(\ell+1)} = \frac{1}{2} + \frac{1}{20(\ell+1)} > \frac{1}{2} + \frac{1}{\ell^4}$
- Противоречие со сложностью f.

Игрушечный $(\ell + 2)$-генератор

- Пусть f вычислима за время $2^{O(n)}$ и $H_{avg}(f) \geq n^4$.
- $G(z) = z_1 \dots z_{\ell/2} \circ f(z_1 \dots z_{\ell/2}) \circ z_{\ell/2+1} \dots z_\ell \circ f(z_{\ell/2+1} \dots z_\ell)$ есть $(\ell + 2)$-генератор.
- Первые $\ell + 1$ битов не предсказать аналогично предыдущему
- Пусть $(\ell + 2)$-й бит можно предсказать: $\Pr_{z,z' \leftarrow U(\{0,1\}^{\ell/2})}[C(z \circ f(z) \circ z') = f(z')] \geq \frac{1}{2} + \frac{1}{20(\ell+2)}$
- Из принципа усреднения можно выбрать такое z, что $\Pr_{z' \leftarrow U(\{0,1\}^{\ell/2})}[C(z \circ f(z) \circ z') = f(z')] \geq \frac{1}{2} + \frac{1}{20(\ell+2)}$
- Пусть схема D получается из C подстановкой z. $\Pr_{z' \leftarrow U(\{0,1\}^{\ell/2})}[D(z') = f(z')] \geq \frac{1}{2} + \frac{1}{20(\ell+2)}$
- Противоречие со сложностью f.
- Аналогично $G(z_1, \dots, z_\ell) = z^{(1)} \circ f(z^{(1)}) \dots z^{(k)} \circ f(z^{(k)})$

Генератор Нисана-Вигдерсона

Определение. (Комбинаторный дизайн) Семейство $\mathcal{I} = \{I_1, I_2, \ldots, I_m\}$ подмножеств $\{1, 2, \ldots, \ell\}$ называется (ℓ, d, n)-дизайном, если $\forall j, |I_j| = n$ и $\forall j \neq k, |I_j \cap I_k| \leq d$.

Определение. (Генератор Нисана-Вигдерсона) Пусть $\mathcal{I} = \{I_1, I_2, \ldots, I_m\}$ — дизайн. $f : \{0,1\}^* \to \{0,1\}$. $NW^f_{\mathcal{I}} : \{0,1\}^\ell \to \{0,1\}^m$:

$$NW^f_{\mathcal{I}}(z) = f(z_{I_1}) \circ f(z_{I_2}) \circ \cdots \circ f(z_{I_m})$$

Лемма. (Конструкция дизайна) За время $2^{O(\ell)}$ можно построить (ℓ, d, n)-дизайн, где $n > d, \ell > 10n^2/d$. В этом дизайне будет $2^{d/10}$ множеств.

Основная лемма

Лемма. Пусть $\mathcal{I}$ — это (ℓ, d, n)-дизайн, в котором $|\mathcal{I}| = 2^{d/10}$, $H_{avg}(f) > 2^{2d}$, тогда распределение $NW^f_{\mathcal{I}}(U_\ell)$ является $(H_{avg}(f)/10, 1/10)$-псевдослучайным.

Следствие. Если f вычислима за время $2^{O(n)}$, $H_{avg}(f) > 2^{\epsilon n}$, то выберем $d = \epsilon n/2$, $\ell = 100n$, тогда $NW^f_{\mathcal{I}}$ — это $2^{\epsilon\ell/1000}$-генератор.

Доказательство. $S = H_{avg}(f)$. (Доказываем, что $NW^f_{\mathcal{I}}(U_\ell)$ является $(S/10, 1/10)$-псевдослучайным.) Достаточно проверить непредсказуемость. Пусть найдется схема C размера $S/2$ и число i, что

$$\Pr_{z \leftarrow U_\ell}[C(f(z_{I_1}) \dots f(z_{I_{i-1}})) = f(z_{I_i})] \geq \frac{1}{2} + \frac{1}{20 \cdot 2^{d/10}}$$

Продолжение доказательства

- $\Pr_{z \leftarrow U_\ell}[C(f(z_{I_1}) \dots f(z_{I_{i-1}})) = f(z_{I_i})] \geq \frac{1}{2} + \frac{1}{20 \cdot 2^{d/10}}$
- Пусть z_1 — это часть z, которая соответствует I_1, а z_2 — все остальное.
- $\Pr_{z_1 \leftarrow U_n, z_2 \leftarrow U_{\ell-n}}[C(f_1(z_1, z_2) \dots f_{i-1}(z_1, z_2)) = f(z_1)] \geq \frac{1}{2} + \frac{1}{20 \cdot 2^{d/10}}$
- По принципу усреднения $\exists \tilde{z}_2$, что
- $\Pr_{z_1 \leftarrow U_n}[C(f_1(z_1, \tilde{z}_2) \dots f_{i-1}(z_1, \tilde{z}_2)) = f(z_1)] \geq \frac{1}{2} + \frac{1}{20 \cdot 2^{d/10}}$
- Так как $|I_i \cap I_j| \leq d$, то $z_1 \mapsto f_j(z_1, \tilde{z}_2)$ можно представить схемой размера $d2^d$. Итого $C(f_1(z_1, \tilde{z}_2))$ можно представить схемой B размера $2^{d/10} d2^d + S/2 < S$: $\Pr_{z_1 \leftarrow U_n}[B(z_1) = f(z_1)] > \frac{1}{2} + \frac{1}{20 \cdot 2^{d/10}} > \frac{1}{2} + \frac{1}{S}$.
- Противоречие со сложностью f.