Вычислительно трудные задачи и дерандомизация
Лекция 10: Экстракторы

Дмитрий Ицыксон

ПОМИ РАН

3 мая 2009

План

1. Слабые источники случайных битов
2. Экстракторы
3. Существование экстракторов
4. Конструкция экстрактора, основанная на случайном блуждании по экспандеру

Постановка задачи

- Пусть алгоритму требуется k случайных битов.
- Достаточно иметь равномерное распределение на $\{0,1\}^k$.
- Или почти равномерное...
- Пусть D_1, D_2 — распределения на $\{0,1\}^n$. Статистическое расстояние: $\delta(D_1, D_2) = \sum_{x \in \{0,1\}^n} |D_1(x) - D_2(x)|$.
- Достаточно иметь распределение на расстоянии $\frac{1}{10}$ от равномерного

Слабые источники случайных битов

- Пусть D распределение на $\{0,1\}^n$.
- Минимальная энтропия
  $H_\infty(D) = \max\{k \in \mathbb{R} \mid \forall z\ D(z) \leq 2^{-k}\}$.
- $0 \leq H_\infty(D) \leq n$; если $H_\infty(D) = n$, то X — равномерное.
- $H_\infty(D)$ — это число “настоящих” случайных битов, которые содержит распределение D.
- Распределение D на $\{0,1\}^n$ — это (n,k)-источник, если, $H_\infty(D) \geq k \iff \forall z\ D(z) \leq 2^{-k}$.

Примеры распределений

1. (Зафиксированные биты) $S \subseteq [n]$, $|S| = k$. Распределение равномерное на битах из S и фиксированная строчка на битах из $[n] \setminus S$. $H_\infty = k$.
2. (Линейное подпространство) Равномерное распределение на линейном подпространстве $\mathbb{Z}_2^n$ размерности k. $H_\infty = k$.
3. (Равномерное на множестве) $S \subseteq \{0,1\}^n$, $|S| = 2^k$, распределение равномерное на S. $H_\infty = k$.
4. (Несимметричная монета) n раз независимо бросается монета, которая выдает 1 с вероятностью $\delta < \frac{1}{2}$ и 0 с вероятностью $1 - \delta$. $H_\infty = \log \frac{1}{1-\delta} n$

Определение экстрактора

Определение. Функция $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ называется (k, ϵ)-экстрактором, если для любого (n, k)-источника D, распределение $\mathrm{Ext}(D, U_d)$ находится на расстоянии $\leq \epsilon$ от U_m.

- Зачем дополнительные настоящие случайные биты?
- $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}^m$
- Пусть $|\mathrm{Ext}^{-1}(1*)| \geq 2^{n-1}$, D — равномерное распределение на на $\mathrm{Ext}^{-1}(1*)$, $H_\infty(D) = n - 1$. $\mathrm{Ext}(D)$ находится на расстоянии хотя бы $\frac{1}{2}$ от равномерного.
- Если $d = O(\log n)$, то все строки длины d можно перебрать.

Существование экстрактора

Теорема. $\forall k, n, \epsilon$ существует (k, ϵ)-экстрактор $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^k$, где $d = \log n + 2\log(1/\epsilon) + O(1)$.

Доказательство.

- (n, k)-источник D плоский, если D — это равномерное распределение на множестве размера 2^k.
- Любой (n, k)-источник — это выпуклая комбинация плоских. Достаточно доказывать только для плоских (n, k)-источников.
- Выберем $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^k$ случайным образом.
- Пусть D — это плоский (n, k)-источник с носителем S, $|S| = 2^k$.
- $f : \{0,1\}^k \to \{0,1\}$ — некоторая функция.
- $\mathrm{E}[f(\mathrm{Ext}(D, U_d))] = \frac{1}{2^k \times 2^d} \sum_{x \in S, y \in \{0,1\}^d} f(Ext(x, y))$.

Существование экстрактора

- Пусть D — это плоский (n, k)-источник с носителем S, $|S| = 2^k$.
- $f : \{0,1\}^k \to \{0,1\}$ — некоторая функция.
- $\mathrm{E}[f(\mathrm{Ext}(D, U_d))] = \frac{1}{2^k \times 2^d} \sum_{x \in S, y \in \{0,1\}^d} f(Ext(x, y))$.
- Вычислить $\mathrm{E}[f(\mathrm{Ext}(D, U_d))]$ — это тоже самое, что $2^k \times 2^d$ раз запустить f на случайной строке из $\{0,1\}^k$.
- Из оценок Чернова $\mathrm{E}[f(\mathrm{Ext}(D, U_d))]$ отличается от $\mathrm{E}[f(U_k)]$ на $\frac{\epsilon}{2}$ с вероятностью меньше, чем $2^{-2^{k+d}\epsilon^2/4}$
- При $d > \log n + 2\log(1/\epsilon) + 3$ эта вероятность меньше, чем $2^{-(2n)2^k}$. Число плоских источников не больше, чем $(2^n)^{2^k}$, чило функций $\{0,1\}^k \to \{0,1\}$ ровно 2^{2^k}.
- Найдется такой Ext, что для всех f и D $|\mathrm{E}[f(\mathrm{Ext}(D, U_d))] - \mathrm{E}[f(U_k)]| < \frac{\epsilon}{2}$

Экстрактор из случайного блуждания по экспандеру

Лемма. $\forall \epsilon > 0, n \geq k$ можно построить (k, ϵ)-экстрактор $\text{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^n$, где $t = O(n - k + \log \frac{1}{\epsilon})$.

Доказательство.

- Пусть A — нормированная матрица смежности $(2^n, d, \frac{1}{2})$-алгебраического расширителя. D — это (n, k)-источник.
- t битов описывают случайное блуждание.
- $A^m D$ — распределение после m шагов блуждания.
- $\|A^m D - \frac{1}{2^n}\mathbf{1}\|_1 \leq 2^{n/2}\|A^m(D - \frac{1}{2^n}\mathbf{1})\|_2 \leq 2^{n/2}\frac{1}{2^m}\|D - \frac{1}{2^n}\mathbf{1}\|_2$
- $\|D - \frac{1}{2^n}\mathbf{1}\|^2 = \sum_x (D(x) - \frac{1}{2^n})^2 = \sum_x (D(x))^2 + \frac{1}{2^n} - 2\sum_x D(x)\frac{1}{2^n} \leq \sum_x (D(x))^2 \leq 2^k \cdot 2^{-2k} = 2^{-k}$
- $\|A^k D - \frac{1}{2^n}\mathbf{1}\|_1 \leq \frac{2^{(n-k)/2}}{2^m}$
- Выберем $m = n - k + \log \frac{1}{\epsilon}$