Экстракторы случайности

Всеволод Нагибин, Даниил Мусатов

Московский Физико-Технический Институт

20 мая, 2025

Введение

Обычно вероятностным алгоритмам требуются независимые равномерно распределенные случайные биты. Однако реальные источники случайности, обычно основанные на физических процессах, редко обладают такими свойствами.
Экстрактор случайности — это функция, преобразующая «не очень равномерную» случайную величину в близкую к равномерной.
В 1950-х задача экстракции случайности была впервые поставлена фон Нейманом, для специфического случая смещенных бросков монеты. Также изучались и другие классы слабых источников случайности, позднее была предложена нотация мин-энтропии [CG85].
Помимо множественных связей с другими псевдослучайными объектами, экстракторы имеют приложения в криптографии [Tre01], теории кодирования [Vad07], теории графов [Li23].

Определения

Мин-энтропия

Мин-энтропия случайной величины X это

$$H_\infty(X) = \sup\{k \in \mathbb{R} | \forall x \mathbb{P}[X = x] \leq 2^{-k}\}$$

Если X это случайная величина, принимающая значения в $\{0,1\}^n$, с $H_\infty(x) \geq k$, то она называется (n,k)-*источником*.

Статистическое расстояние

Пусть X и Y это два распределения на множестве Ω. *Статистическое расстояние* между X и Y это

$$\Delta(X,Y) = \max_{S \subset \Omega} |\mathbb{P}[X \in S] - \mathbb{P}[Y \in S]| = \frac{1}{2} \sum_{\omega \in \Omega} |\mathbb{P}[X = \omega] - \mathbb{P}[Y = \omega]|.$$

Распределения X и Y называются ε-*близкими*, если $\Delta(X,Y) \leq \varepsilon$.

Определения

Можно доказать, что детерминированной функцией даже для класса $(n, n-1)$-источников невозможно получить распределение, близкое к равномерному. Поэтому строятся экстракторы, использующие дополнительный сид — небольшое число независимых равномерных случайных бит, а также экстракторы для нескольких независимых в совокупности источников.

Экстрактор

Функция $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ это (k, ε)-*экстрактор*, если для любого (n, k)-источника X распределение $\mathrm{Ext}(X, U_d)$ ε-близко к U_m.

Экстрактор для нескольких источников

Функция $\mathrm{Ext} : (\{0,1\}^n)^C \to \{0,1\}^m$ это (k, ε)-*экстрактор для C источников*, если для любых независимых в совокупности (n, k)-источников $X_1, ..., X_C$ распределение $\mathrm{Ext}(X_1, ..., X_k)$ ε-близко к U_m.

Экстракторы с одним источником

Вероятностным методом можно получить неявную конструкцию экстрактора с хорошими параметрами.

Теорема

Для любых $0 \leq k \leq n \in \mathbb{N}, \varepsilon > 0$ существует (k, ε)-экстрактор $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ с параметрами $d = \log_2(n-k) + 2\log_2 \frac{1}{\varepsilon} + O(1)$ и $m = k + d - 2\log_2 \frac{1}{\varepsilon} - O(1)$.

Однако для приложений нужны явные (эффективно вычислимые) конструкции. В [GUV09] изложена конструкция, достигающая этих параметров с точностью до константы.

Теорема

Для любых констант $\alpha, \varepsilon > 0$, для любых $k \leq n \in \mathbb{N}$ существует явная конструкция (k, ε)-экстрактора $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ с параметрами $d = O(\log n)$ и $m = (1-\alpha) \cdot k$.

Экстракторы с двумя источниками

Вероятностным методом также можно получить конструкцию экстрактора для двух источников с хорошими параметрами.

Теорема

Для любых $0 \leq k \leq n \in \mathbb{N}, \varepsilon > 0$ существует (k, ε)-экстрактор для двух источников $\mathrm{Ext} : (\{0,1\}^n)^2 \to \{0,1\}^m$ с длиной выхода $m = k - \log_2 n - 2\log_2 \frac{1}{\varepsilon} - O(1)$.

Одна из лучших на сегодняшний день явных конструкций представлена в [Li23].

Теорема

Для любого $\varepsilon > 0$ существует константа $c > 1$ и явный (k, ε)-экстрактор для двух источников $\mathrm{Ext} : (\{0,1\}^n)^2 \to \{0,1\}$ для $k \geq c \log_2 n$.

Экстракторы с двумя источниками

Сравнение некоторых конструкций экстракторов с двумя источниками.

Мин-энтропия k	Длина выхода m	Ошибка ε	Ссылка	Год
$0.51n$	$\Theta(n)$	$2^{-\Omega(n)}$	[CG85]	1985
$0.51n, O(\log n)$	$\Theta(k)$	$2^{-\Omega(k)}$	[Raz05]	2005
$0.499n$	$\Theta(n)$	$2^{-\Omega(n)}$	[BOU05]	2005
$\mathrm{polylog}\ n$	1	$2^{-\Omega(1)}$	[CZ16]	2016
$\mathrm{polylog}\ n$	$k^{\Omega(1)}$	$2^{-\Omega(1)}$	[Li16]	2016
$O(\log n \log\log n)$	1	$O(1)$	[Li17]	2017
$O(\frac{\log n \log\log n}{\log\log\log n})$	1	$O(1)$	[Li20]	2019
$O(\log n)$	1	$O(1)$	[Li23]	2023
$O(\log n)$	$\Theta(k)$	$2^{-\Omega(k)}$	Неявная конструкция	

Конструкция

Устойчивый экстрактор

Функция $\mathrm{nmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ это *t-устойчивый* (k, ε)*-экстрактор*, если для любого (n, k)-источника X, независимой от X случайной величины Y, равномерно распределенной на $\{0,1\}^d$ и функций $f_1, ..., f_t : \{0,1\}^d \to \{0,1\}^d$ без неподвижных точек,

$$(\mathrm{nmExt}(X, Y), \mathrm{nmExt}(X, f_1(Y)), ..., \mathrm{nmExt}(X, f_t(Y))) \approx_\varepsilon (U_m, \mathrm{nmExt}(X, f_1(Y)), ..., \mathrm{nmExt}(X, f_t(Y)))$$

Неформально, t-устойчивый экстрактор это такой экстрактор, что его значение на сиде Y нельзя узнать даже если знать значения при t других значениях сида.

Конструкция

Несколько последних конструкций экстракторов для двух источников получены на основе конструкций t-устойчивых экстракторов и следующей теоремы.

Теорема

Предположим, что существует $f(t, \varepsilon)$ такая что для $n, t \in \mathbb{N}, \varepsilon > 0$ существует явная конструкция сильного t-устойчивого (k', ε)-экстрактора с длиной сида $k' = d = f(t, \varepsilon)$. Тогда для любой константы $\varepsilon > 0$ существуют $t' = t'(\varepsilon), c' = c'(\varepsilon)$ и явная конструкция (k, ε)-экстрактора $\mathrm{Ext} : (\{0,1\}^n)^2 \to \{0,1\}$ для двух источников, где $k = f(t', \frac{1}{n^{c'}}) + O(1)$.

Итоги

Для явных конструкций экстракторов с одним источником остается открытым вопрос об исключении каждой из констант по отдельности ([Vad12]).

Открытая проблема 1

Предоставить явную конструкцию $(k, 0.01)$-экстрактора $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ с длиной сида $d = O(\log n)$ и длиной выхода $m = k + d - O(1)$.

Открытая проблема 2

Предоставить явную конструкцию $(k, 0.01)$-экстрактора $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ с длиной сида $d = \log n + O(1)$ и длиной выхода $m = \Omega(k)$.

Итоги

На сегодняшний день неизвестна явная конструкция экстрактора для двух источников с мин-энтропией $k = O(\log n)$ и выходом неконстантной длины. Также, существующие конструкции весьма сложны: для них строятся не только экстракторы с различными параметрами, но и другие псевдослучайные объекты — декорреляторы, конденсаторы, семплеры, устойчивые экстракторы. Поэтому кроме улучшения параметров явных конструкций экстракторов с двумя источниками стоит задача упрощения существующих конструкций.

Список литературы I

J. BOURGAIN.

More on the sum-product phenomenon in prime fields and its applications.
International Journal of Number Theory, 01(01):1–32, 2005.

Benny Chor and Oded Goldreich.

Unbiased bits from sources of weak randomness and probabilistic communication complexity.
In *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*, pages 429–442, 1985.

Eshan Chattopadhyay and David Zuckerman.

Explicit two-source extractors and resilient functions.
In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, page 670–683, New York, NY, USA, 2016. Association for Computing Machinery.

Venkatesan Guruswami, Christopher Umans, and Salil Vadhan.

Unbalanced expanders and randomness extractors from parvaresh–vardy codes.
J. ACM, 56(4), July 2009.

Xin Li.

Improved two-source extractors, and affine extractors for polylogarithmic entropy.
In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 168–177, 2016.

Xin Li.

Improved non-malleable extractors, non-malleable codes and independent source extractors.
In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, page 1144–1156, New York, NY, USA, 2017. Association for Computing Machinery.

Список литературы II

Xin Li.

Non-malleable extractors and non-malleable codes: partially optimal constructions.
In *Proceedings of the 34th Computational Complexity Conference*, CCC '19, Dagstuhl, DEU, 2020. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Xin Li.

Two Source Extractors for Asymptotically Optimal Entropy, and (Many) More .
In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1271–1281, Los Alamitos, CA, USA, November 2023. IEEE Computer Society.

Ran Raz.

Extractors with weak random seeds.
In *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '05, page 11–20, New York, NY, USA, 2005. Association for Computing Machinery.

L. Trevisan.

Extractors and pseudorandom generators.
J. ACM, 48(4):860–879, July 2001.

S. Vadhan.

The unified theory of pseudorandomness: guest column.
SIGACT News, 38(3):39–54, September 2007.

Salil P. Vadhan.

Pseudorandomness.
Foundations and Trends® in Theoretical Computer Science, 7(1–3):1–336, 2012.

Приложения

Непосредственно из конструкций экстракторов для двух источников строятся конструкции графов Рамсея.

Теорема

Существует константа $c > 1$ такая, что для любого N существует строго явная конструкция K-графа Рамсея (то есть графа без клики размера K и без антиклики размера K) на N вершинах с $K = \log_2^c N$.