

maximal load: $O(\log n / \log \log n)$ — aim

Space required: $O(\log^2 n / \log \log n)$

Construction

each function described in $O(\log n / \log \log n)$

evaluated in $O(\log n / \log \log n)$

IDEA: concatenate output of $O(\log \log n)$ functions which are gradually more independent. Each function f is described using d functions h_1, \dots, h_d

$$f(x) = \underbrace{h_1(x)}_{\substack{\uparrow \\ \text{binary string}}} \circ h_2(x) \circ \dots \circ \underbrace{h_d(x)}_{\substack{\uparrow \\ \text{concatenate}}}$$

Gradually more independent

$$\begin{array}{c} h_1: O(1)\text{-wise indep.} \\ h_2: O(h_1)\text{-wise indep.} \\ \dots \\ h_d: O(\log n / \log \log n)\text{-wise indep.} \end{array}$$

↓
k wise
k increase

Note: ① output length decreases at the same time

$$h_1: \sum \log n \Rightarrow h_d: O(\log \log n)$$

② each of h_1, \dots, h_d could be described/computed in $O(\log n)$ bits/time

Definitions

$$[n] \Rightarrow \{1, \dots, n\}$$

$U_n \Rightarrow$ uniform distribution over set $\{0, 1\}^n$

$x \in X \Rightarrow$ sample x from X for a r.v. X .

$x \in S \Rightarrow$ sample x uniformly from finite set S .

$$SD(X, Y) \Rightarrow \text{statistical distance between two r.v. over finite domain } \Sigma \\ = \frac{1}{2} \sum_{\omega \in \Sigma} |\Pr[X=\omega] - \Pr[Y=\omega]|$$

$x \circ y \Rightarrow$ concatenate x, y bit string

unit cost RAM model

\Rightarrow elements are taken from a universe of size n and each element can be stored using $c = O(\log n)$ bits.

k -wise f -dependent

For a family of $f: [n] \rightarrow [v]$, we say k -wise f -dependent iff

$$SD(X, Y) \leq \delta$$

where $X = \text{distribution}(f(x_1), f(x_2), \dots, f(x_k))$ for any distinct $x_1, \dots, x_k \in [n]$
 $Y = \text{uniform distribution over } [v]^k$

Describe Space: $O(k \max\{\log n, \log v\})$ bits

Evaluation Time: $O(k)$

ϵ -biased Distribution [N/N93]

r.v.s X_1, X_2, \dots, X_n over $\{0, 1\}$ is ϵ -biased if for any $S \neq \emptyset \subseteq [n]$,

$$|\Pr[\bigoplus_{i \in S} X_i = 1] - \Pr[\bigoplus_{i \in S} X_i = 0]| \leq \epsilon$$

\uparrow
XOR operation

[Ahl'92, Sec. 5] constructs an ϵ -biased distribution over $\{0, 1\}^n$ where each point could be specified using $O(\log(n/\epsilon))$ bits where each bit could be calculated using $O(\log(n/\epsilon))$ times.

In RAM, t bits could be calculated in time $O(\log(n/\epsilon)t)$

may need proof

min-entropy

for a r.v. X , its min-entropy is

$$H_{\infty}(X) = -\log(\max_x \Pr[X=x])$$

negative log of max probability of $X=x$.

k-source

k-source is a r.v. X with its min-entropy $H_{\infty}(X) > k$.

(T, k)-block source

r.v. $X = (X_1, \dots, X_T)$, for any $i \in [T]$

$$H_{\infty}(X_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \geq k$$

(T, k, ε)-block source

r.v. $X = (X_1, \dots, X_T)$, for any $i \in [T]$

$$\Pr[(H_{\infty}(X_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \geq k)] \geq 1 - \epsilon$$

$$(x_1, \dots, x_{i-1}) \leftarrow (X_1, \dots, X_{i-1})$$

Pr of X is (T, k) -block source $\geq 1 - \epsilon$

7 lemmas

Corollary 2.1

Fact

for any k an ϵ -biased distribution is also k -wise δ -dependent

$$\text{for } \delta = \epsilon 2^{k/2}$$

For any $u, v, v=2^i$, there exists a family of k -wise δ -dependent function $f:[u] \rightarrow [v]$ described in $\underline{O(\log u + k \log v + \log(1/\delta))}$ bits and calculated in $\underline{O(\log u + k \log v + \log(1/\delta))}$ in RAM.

check AGW92 for proof.

Corollary 2.2

Let $X_1, \dots, X_n \in \{0,1\}$ be $2k$ -wise δ -dependent r.v. for $k \in \mathbb{N}, 0 \leq \delta < 1$,

let $X = \sum X_i$ and $\mu = E[X]$. Then, for any $t > 0$,

$$\Pr[|X - \mu| > t] \leq 2 \left(\frac{2^{nk}}{t^2} \right)^k + \delta \left(\frac{n}{t} \right)^{2k}.$$

How X change from μ . proceed on paper using Markov Ineq.

Lemma 2.2 from [BR94]

may not be used by our construction.

Lemma 2.4 [GW97] ← may need or proof (not sure from 4.1/4.2 main results)

r.v. $X_1 \in \{0,1\}^{n_1}, X_2 \in \{0,1\}^{n_2}, H_\infty(X_1, X_2) \geq h_1 + h_2 - \Delta$

① $H_\infty(X_1) \geq h_1 - \Delta$

② for any $\epsilon > 0$,

$$\Pr_{x_1 \in X_1} [H_\infty(X_2 | X_1 = x_1) < h_2 - \Delta - \log(1/\epsilon)] < \epsilon$$

Corollary 2.5

Any $X = (X_1, \dots, X_T)$ over $(\{0,1\}^n)^T, H_\infty X \geq Tn - \Delta$

is a $(T, n-d \cdot \log(\gamma\epsilon), \epsilon)$ -block source for any $\epsilon > 0$.

↑ def of (T, k, ϵ) -block source
and lemma 2.4.

Theorem 3.1

any constant $c > 0$, integer n , $u = \text{poly}(n)$, there exists a family F of $f: [u] \rightarrow [n]$ that:

- ① describe using $O(\log n \log \log n)$ bits
- ② $f(x)$ can be computed in $O(\log n \log \log n)$ time
- ③ $\exists \gamma > 0$ such that for any $S \subseteq [n]$

$$\Pr_{f \in F} \left[\max_{i \in [u]} |f^{-1}(i) \cap S| \leq \frac{\gamma \log n}{\log \log n} \right] > 1 - \frac{1}{n^c}$$

↑
maximum load = $O\left(\frac{\gamma \log n}{\log \log n}\right)$ with high probability.

Construction

$$\text{assume } n = 2^k$$

let $d = O(\log \log n)$, and for every $i \in [d]$, let H_i be a family of k_i -wise f -dependent functions $[u] \rightarrow \{0, 1\}^{L_i}$, where:

$$① n_0 = n \quad n_v := \frac{n}{2^{L_v}} \text{ for every } v \in [d]$$

$$n_1 = \frac{n}{2^{L_1}} \quad n_2 = \frac{n_1}{2^{L_2}} = \frac{n}{2^{L_1+L_2}} \dots n_d = \frac{n}{2^{\sum_{v=1}^{d-1} L_v}}$$

$$② L_i = \left\lfloor \frac{\log n_{i-1}}{4} \right\rfloor \text{ for every } i \in [d-1], \quad d = \log n - \sum_{v=1}^{d-1} L_v$$

$$L_i = \left\lfloor \frac{\log n_{i-1}}{4} \right\rfloor = \left\lfloor \frac{\log n - \sum_{j=1}^{i-1} L_j}{4} \right\rfloor \text{ for } i \in [d-1]$$

$$d = \log n - \sum_{i=1}^{d-1} L_i$$

$$③ k_i L_i = O(\log n) \text{ for every } i \in [d-1], \quad k_d = O(\log n / \log \log n)$$

$$k_i L_i = k_i \left\lfloor \frac{\log n - \sum_{j=1}^{i-1} L_j}{4} \right\rfloor$$

$$④ f = \text{poly}(\frac{1}{n})$$

By corollary 2.1, we have a family H_i :

need detail.

h_i is represented using $O(\log n + k_i l_i + \log \frac{1}{\delta}) = \log n$ space/time

we defined our function family F as

$$f(x) = h_1(x) \circ h_2(x) \circ \dots \circ h_d(x)$$

for $h_i \in H_i$

We visualize the construction as a reversed tree with d layers.

- has last layer using h_1 \hookrightarrow Layer 0 has 1 bin with all balls.
- has last layer using h_2 \hookrightarrow Layer 1 has 2^{L_1} intermediate bins.
-
- has last layer using h_d \hookrightarrow Layer d has n final bins

Lemma 3.2

For any $i = \{0, \dots, d-1\}$, $\alpha = \Delta(1/\log \log n)$, $0 < \alpha_i < 1$ and set $S_i \subseteq [n]$ of size at most $(1+\alpha_i)n_i$,

$$\Pr_{h_{i+1} \in H_{i+1}} \left[\max_{y \in [0, 1]^{L_{i+1}}} |h_{i+1}^{-1}(y) \cap S_i| \leq (1+\alpha)(1+\alpha_i)n_{i+1} \right] \geq 1 - \frac{1}{n^{c+1}}$$

$$\text{maximum load of layer } i+1 \leq (1 + \Delta(1/\log \log n))(1 + \alpha_i)n_{i+1}$$

with high prob.

proof for lemma 3.2

proof for lemma 3.1

Application

- ① storing elements using linear probing
- ② maintaining k-wise independence function using an construction without affect the space/time requirement

read
some
evidence
or proof.