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 $h_1, \dots, h_d$

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

$h_1(x)$: binary string; $\circ$: concatenate

Gradually more independent

$h_1 : O(1)$-wise indep.

$h_2 : O(h_1)$-wise indep.

...

$h_d : O(\log n / \log\log n)$-wise indep.

(k-wise, k increase ↓)

Note: ① output length decreases at the same time

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

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.

$SD(X, Y) \Rightarrow$ statistical distance between two r.v. over finite domain Ω

$$= \frac{1}{2} \sum_{\omega \in \Omega} |Pr[X=\omega] - Pr[Y=\omega]|$$

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

unit cost RAM model

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

k-wise δ-dependent

For a family of $f : [n] \to [v]$, it is k-___ δ-___ iff

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

where X = distribution $(f(x_1), f(x_2), \ldots, f(x_k))$ for any distinct $x_1, \ldots x_k \in [n]$

Y = uniform distribution over $[v]^k$

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

Evaluation Time: $O(k)$

ε-biased Distribution [NN93]

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

$$|Pr(\oplus_{i \in S} X_i = 1) - Pr(\oplus_{i \in S} X_i = 0)| \leq \varepsilon$$

($\oplus$: XOR operation)

[AGHP92, Sec. 5] constructs an ε-biased distribution over $\{0,1\}^n$ where each point could be specified using $O(\log(n/\varepsilon))$ bits where each bit could be calculated using $O(\log(n/\varepsilon))$ times.

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

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 $H_\infty(X) \geq k$.

(T, k)-block source

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

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

(T, k, ε)-block source

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

$$\Pr_{(x_1,\ldots,x_{i-1}) \leftarrow (X_1,\ldots,X_{i-1})} [H_\infty(X_i | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) \geq k] \geq 1 - \varepsilon$$

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

Theorem

Corollary 2.1

Fact

for any k an ε-biased distribution is also k-wise δ-dependent for $\delta = \varepsilon 2^{k/2}$

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

check AGHP92 for proof.

Corollary 2.2

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

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

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

How X diverge from μ.

proved in paper using Markov Ineq.

Lemma 2.2 from [BR94]

Lemma 2.4 [GW97]

may not be used by our construction.

← may need a proof (but same from 4.1/4.2 main rescale)

r.v. $X_1 \in [0,1]^{n_1}$, $X_2 \in [0,1]^{n_2}$, $H_\infty(X_1 X_2) \ge n_1 + n_2 - \Delta$

① $H_\infty(X_1) \ge n_1 - \Delta$

② for any $\varepsilon > 0$,

$$\Pr_{x_1 \leftarrow X_1}\left[H_\infty(X_2 | X_1 = x_1) < n_2 - \Delta - \log(1/\varepsilon)\right] < \varepsilon$$

Corollary 2.5

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

is a $(T, n-o-\log(1/\varepsilon), \varepsilon)$-block source for any $\varepsilon > 0$.

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