The complexity of distributions

Emanuele Viola

Northeastern University
Local functions

• $f : \{0,1\}^n \rightarrow \{0,1\}$ d-local : output depends on d input bits

• **Fact**: $\text{Parity}(x) = 1 \iff \sum x_i = 1 \mod 2$
  is not n-1 local

• **Proof**: Flip any input bit $\Rightarrow$ output flips $\diamondsuit$
Local generation of (Y, parity(Y))

- **Theorem** [Babai; Boppana Lagarias '87]
  There is \( f : \{0,1\}^n \rightarrow \{0,1\}^{n+1} \), each bit is 2-local
  Distribution \( f(X) \equiv (Y, \text{parity}(Y)) \) \( (X, Y \in \{0,1\}^n \) uniform)
• Complexity theory of distributions (as opposed to functions)

How hard is it to generate distribution D given random bits?

E.g., $D = (Y, \text{parity}(Y))$, $D = W_k := \text{uniform } n\text{-bit with } k 1's$
Our results

- **Theorem:** \( f : \{0,1\}^n \rightarrow \{0,1\}^n \), \( \varepsilon \log(n) \) – local.
  \[
  \text{Distance}(f(X), W_{n/2} = \text{uniform set of size } n/2) > 1 - n^{-\Omega(1)}
  \]

- Tight up to \( \Omega() : f(x) = x \)

- **Corollary:**
  Data structure lower bound for storing \( S \subseteq [n] \), \(|S| = n/2\)
Results for AC₀

- Model: small constant-depth circuits (AC₀)

- Challenge: ∃ explicit boolean f : cannot generate ( Y, f(Y) ) ?

- Theorem [Matias Vishkin, Hagerup, Czumaj Kanarek Lorys Kutyłowski, V.]
  Can generate ( Y, majority(Y) ) (exp. small error)

- Theorem [Lovett V.] Cannot generate error-correcting code
• Thank you
Rest of this talk

- Connection with succinct data structures
- Lower bound for generating $W_{n/2} = \text{uniform } n\text{-bit with } n/2 \text{ 1's}$
- Other results and conclusion
Succinct data structures for sets

• Store $S \subseteq \{1, 2, \ldots, n\}$ of size $|S| = k$

In $u$ bits $b_1, \ldots, b_u \in \{0,1\}$

• Want:
  Small space $u$  (optimal $= \lceil \log_2 \binom{n}{k} \rceil$)
  Answer “$i \in S$?” by probing few bits (optimal $= 1$)

• In combinatorics: Nešetřil Pultr, …, Körner Monti
Previous results

• Store $S \subseteq \{1, 2, \ldots, n\}$, $|S| = k$, in bits, answer "$i \in S?$"

• [Minsky Papert '69, Buhrman Miltersen Radhakrishnan Venkatesh; Pagh; ...; Pătraşcu; V. '09]

• Surprising upper bounds
  space = optimal + $o(n)$, probe $O(\log n)$

• No lower bounds for $k = n / 2^a$
Claim: If store $S \subseteq \{1, 2, \ldots, n\}$, $|S| = k$ in $u = \text{optimal} + r$ bits answer “$i \in S$?” by (non-adaptively) probing $d$ bits. Then $\exists f : \{0,1\}^u \rightarrow \{0,1\}^n$, $d$-local
Distance($f(X)$, $W_k =$ uniform set of size $k) < 1 - 2^{-r}$

$\left( \text{distance}(A, B) := \max_T \left| \text{Pr}[A \in T] - \text{Pr}[B \in T] \right| \right)$

Proof: $f_i := \text{“}i \in S\text{”}$

$f(X) = W_k$ with probability $(\text{n choose k}) / 2^u = 2^{-r}$
Rest of this talk

- Connection with succinct data structures

- Lower bound for generating $W_{n/2} = \text{uniform n-bit with n/2 1's}$

- Other results and conclusion
Our result

- **Theorem**: Let $f : \{0,1\}^n \rightarrow \{0,1\}^n : (d=O(1))$-local.
  
  There is $T \subseteq \{0,1\}^n : \left| \Pr[f(x) \in T] - \Pr[W_{n/2} \in T] \right| > 1 - n^{-\omega(1)}$

- **Warm-up scenarios**:

  - $f(x) = 000111$ \textbf{Low-entropy} \quad $T := \{ 000111 \}$
    
    $\left| \Pr[ f(x) \in T] - \Pr[W_{n/2} \in T] \right| = \left| 1 - |T| / (\text{n choose n/2}) \right|$

  - $f(x) = x$ \textbf{“Anti-concentration”} \quad $T := \{ z : \sum_i z_i = n/2 \}$
    
    $\left| \Pr[ f(x) \in T] - \Pr[W_{n/2} \in T] \right| = \left| 1/\sqrt{n} - 1 \right|$
Proof

- Partition input bits \( X = (X_1, X_2, \ldots, X_s, H) \)

- Fix \( H \). Output block \( B_i \) depends only on bit \( X_i \)

- Many \( B_i \) constant ( \( B_i(0,H) = B_i(1,H) \) ) \( \Rightarrow \) low-entropy

- Many \( B_i \) depend on \( X_i \) ( \( B_i(0,H) \neq B_i(1,H) \) )
  
  Intuitively, anti-concentration: output bits can't sum to \( n/2 \)
If many \( B_i(0,H) \), \( B_i(1,H) \) have different sum of bits, use Anti-concentration Lemma \([\text{Littlewood Offord}]\)

For \( a_1, a_2, \ldots, a_s \neq 0 \), any \( c \),
\[
\Pr_{X \in \{0,1\}^s} \left[ \sum_i a_i X_i = c \right] < \frac{1}{\sqrt{n}}
\]

- **Problem**: \( B_i(0,H) = 100 \), \( B_i(1,H) = 010 \)
  high entropy but no anti-concentration

- **Fix**: want many blocks 000, so high entropy \( \Rightarrow \) different sum
Rest of this talk

- Connection with succinct data structures

- Lower bound for generating $W_{n/2} = \text{uniform } n\text{-bit with } n/2 \text{ 1's}$

- Other results and conclusion
Conclusion

- Complexity of distributions = uncharted territory

- Lower bound for generating $W_k$ locally

- $\Rightarrow$ lower bound for succinct data structures for storing sets of size $n / 2^a$
• $\sum \land \notin \cup \subset \subseteq \in \Rightarrow \forall \geq \Omega \alpha \beta \epsilon \gamma \delta \rightarrow$

• $\neq \approx$
More connections

- More uses of generating $W_k := \text{uniform } n\text{-bit string with } k \text{ 1's}$
- McEliece cryptosystem
- Switching networks, …
Previous results

- Store $S \subseteq \{1, 2, \ldots, n\}$, $|S| = k$, in bits, answer “$i \in S$?”

- [Minsky Papert '69] Average-case study

- [Buhrman Miltersen Radhakrishnan Venkatesh; Pagh '00] Space $O(\text{optimal})$, probe $O(1)$ when $k = \Theta(n)$

  Lower bounds for $k < n^{1-\varepsilon}$

- [..., Pagh, Pătraşcu] space = optimal + $o(n)$, probe $O(\log n)$

- [V. '09] lower bounds for $k = \Omega(n)$, except $k = n / 2^a$