On the complexity of distributions

Emanuele Viola

Northeastern University

Laci Babai's 60th birthday conference
March 2010
Local functions

- \( f : \{0,1\}^n \rightarrow \{0,1\} \) \textbf{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 ♦

Local generation of ( Y, parity(Y) )

- **Theorem [Babai '87]**
  
  There is \( f : \{0,1\}^n \rightarrow \{0,1\}^{n+1} \), each bit is 2-local
  
  Distribution \( f(X) \equiv ( Y, \text{parity}(Y) ) \quad (X, Y \in \{0,1\}^n \text{ uniform}) \)
Message

• Complexity theory of distributions (as opposed to functions)

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

E.g., $D = (Y, \text{parity}(Y))$, $D = W_k := \text{uniform n-bit with k 1's}$
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 $= \lceil \lg_2 (n \text{ choose } k) \rceil$)
  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 $k = n / 2^a$
General connection

• **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 $\exists f : \{0,1\}^u \to \{0,1\}^n$, $d$-local

Distance( $f(X)$, $W_k$ = uniform set of size $k$) $< 1 - 2^{-r}$

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

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

$f(X) = W_k$ with probability $(n \text{ choose } k) / 2^u = 2^{-r}$ ♦
Our result

- **Theorem:** \( f : \{0,1\}^{\text{optimal} + n^{o(1)}} \rightarrow \{0,1\}^n \), \((d < \varepsilon \log(n))\)-local.

  \[ \text{Distance}(f(X), W_k = \text{uniform set of size } k = \Theta(n)) > 1 - n^{-\Omega(1)} \]

- Tight up to \( \Omega() \) if \( k = n/2 \): \( f(x) = x \), \((n \text{ choose } n/2) = O(2^{n/\sqrt{n}}) \)

- **Corollary:** To store \( S \subseteq \{1, 2, \ldots, n\}, |S| = k = n / 2^a \) answer “\( i \in S? \)” probing \( d < \varepsilon \log(n) \) bits:

  Need space > optimal + \( \Omega(\log n) \)
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
Our result

- **Theorem:** Let $f : \{0,1\}^n \rightarrow \{0,1\}^n : (d=O(1))-\text{local}$. 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:**
  - $f(x) = 000111$ \text{ Low-entropy } 
    \[ T := \{ 000111 \} \]
    \[ \left| \Pr[ f(x) \in T] – \Pr[W_{n/2} \in T] \right| = \left| 1 - \frac{|T|}{\binom{n}{n/2}} \right| \]
  - $f(x) = x$ \text{ “Anti-concentration” } 
    \[ T := \{ z : \sum_i z_i = n/2 \} \]
    \[ \left| \Pr[ f(x) \in T] – \Pr[W_{n/2} \in T] \right| = \left| \frac{1}{\sqrt{n}} – 1 \right| \]
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 [Littlewood Offord]

\[
\text{Anti-concentration Lemma [Littlewood Offord]}
\]

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
Other directions and results

• More general model: small bounded-depth circuits ($\text{AC}^0$)

\[ \text{Input } x \]

\[
\begin{array}{c}
\neg \\
\land \\
\land \\
\land \\
\land \\
\land \\
\neg \\
\lor \\
\lor \\
\lor \\
\lor \\
\lor \\
\lor \\
\land \\
\land \\
\end{array}
\]

constant depth

\[
V = \text{or} \\
\land = \text{and} \\
\neg = \text{not}
\]

• Challenge: $\exists$ explicit boolean $f : \text{cannot generate } (Y, f(Y))$?

• Theorem [Matias Vishkin, Hagerup, Czumaj Kanarek Lorys Kutylowski, V.]
  Can generate $(Y, \text{majority}(Y))$ (exp. small error)

• Theorem [Lovett V.] Cannot generate error-correcting code
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 $n / 2^a$
• \( \sum \neg \cup \notin \cap \emptyset \subseteq \subseteq \sqrt{\uparrow} \Rightarrow \uparrow \leftarrow \leftrightarrow \nabla \geq \leq \forall \exists \Omega \alpha \beta \varepsilon \gamma \delta \to \)
• \( \neq \approx \)
More connections

- More uses of generating $W_k := \text{uniform n-bit string with k 1's}$
- 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 $k = n / 2^a$