Deterministic extraction from independent Santha-Vazirani sources

Grigory Yaroslavtsev

The Pennsylvania State University
http://www.cse.psu.edu/~gyy5026

December 2, 2010

Plan

Motivation and defintions

Definition (α-unpredictable bit source)

For $\alpha \in [0,1]$, random variable $X = (X_1, \ldots, X_n)$ taking values in $\{0,1\}^n$ is an α-unpredictable bit source if for every $i \in [n]$, and every $x_1, \ldots, x_{i-1} \in \{0,1\}^{i-1}$, we have

$$\alpha \leq \frac{Pr[X_i = 0 | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}]}{Pr[X_i = 1 | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}]} \leq 1/\alpha$$

Motivation and defintions

Definition (α-unpredictable bit source)

For $\alpha \in [0,1]$, random variable $X = (X_1, \ldots, X_n)$ taking values in $\{0,1\}^n$ is an α-unpredictable bit source if for every $i \in [n]$, and every $x_1, \ldots, x_{i-1} \in \{0,1\}^{i-1}$, we have

$$\alpha \leq \frac{Pr[X_i = 0 | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}]}{Pr[X_i = 1 | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}]} \leq 1/\alpha$$

Motivation:

- Model for correlations in sequences of random bits [SV86].
- Applications to two-party differential privacy [MMP+10].

Basic properties of Santha-Vazirani sources

Properties

- *No string has probability mass greater than* $1/(1+\alpha)^n$
- *Min-entropy (*$\min_x \log_2(1/Pr[X=x])$*), is at least* βn*, where* $\beta = \log_2(1+\alpha) \geq \alpha$.

Basic properties of Santha-Vazirani sources

Properties

- *No string has probability mass greater than* $1/(1+\alpha)^n$
- *Min-entropy* ($\min_x \log_2(1/Pr[X=x])$)*, is at least* βn*, where* $\beta = \log_2(1+\alpha) \geq \alpha$.

Definition (Total variation and δ-closeness)

For random variables X *and* X' *taking values in* Ω*, we say that* X *and* X' *are* δ*-close if the total variation between their distributions is at most* δ*, i.e.,*

$$||X - X'||_{TV} := \frac{1}{2} \sum_{x \in \Omega} |Pr[X=x] - Pr[X'=x]| \leq \delta$$

Santha-Vazirani sources

Definition (Quasi-randomness)

If S is source of length n, with probability distribution $P_n(x)$. S is quasi-random if for every $t > 0$, and for large n: $||P_n(x) - U_n||_{TV} < 1/n^{\epsilon}$

Santha-Vazirani sources

Definition (Quasi-randomness)

If S is source of length n, with probability distribution $P_n(x)$. S is quasi-random if for every $t > 0$, and for large n: $||P_n(x) - U_n||_{TV} < 1/n^{\epsilon}$

- Vazirani [Vaz87]: Inner product modulo 2 — almost uniform bit.
- Vazirani [Vaz87]: Inner product modulo 2 of blocks of length $k = \omega(\alpha) \log n$ — quasi-random sequence.
- Vazirani [Vaz87]: Extracting n quasi-random bits from sources of length $O(\alpha^2 n)$.
- Not possible for one source (no function can be more than α-unpredictable ([SV86]
- Generalization [MMP+10]: Inner product modulo m extracts an almost-uniform element of $\mathbb{Z}_m$, if n is at least roughly m^2. Even conditioned on bits of one of the sources.

Deterministic extraction from Santha-Vazirani sources

Theorem (Randomness extraction)

There is a constant c such that if:

1. *X is an α-unpredictable bit source on $\{0,1\}^n$,*
2. *Y is a source on $\{0,1\}^n$ with min-entropy at least βn,*
3. *Y is independent from X,*
4. *$Z = \langle X, Y \rangle \bmod m$ for some $m \in \mathbb{N}$,*

then for every $\delta \in [0,1]$, such that

$$n \geq c \cdot \frac{m^2}{\alpha\beta} \cdot \log\left(\frac{m}{\beta}\right) \cdot \log\left(\frac{m}{\delta}\right),$$

the random variable (Y, Z) is δ-close to (Y, U) where U is uniform on $\mathbb{Z}_m$ and independent of Y.

Proof technique

- Using Cauchy-Schwarz inequality to go from L_1 to L_2 norm.

Proof technique

- Using Cauchy-Schwarz inequality to go from L_1 to L_2 norm.
- Using complex DFT to calculate L_2 norm $||p_Z - p_U||_2$.

$$||Z - U||_{SD} = \frac{1}{2}||p_Z - p_U||_1 \leq \frac{\sqrt{m}}{2}||p_Z - p_U||_2$$

Proof technique

- Using Cauchy-Schwarz inequality to go from L_1 to L_2 norm.
- Using complex DFT to calculate L_2 norm $||p_Z - p_U||_2$.

$$||Z - U||_{SD} = \frac{1}{2}||p_Z - p_U||_1 \leq \frac{\sqrt{m}}{2}||p_Z - p_U||_2$$

- Estimating Fourier coefficients $\hat{p}_Z$ separately for all $y \in Y$.

Proof technique

- Using Cauchy-Schwarz inequality to go from L_1 to L_2 norm.
- Using complex DFT to calculate L_2 norm $||p_Z - p_U||_2$.

$$||Z - U||_{SD} = \frac{1}{2}||p_Z - p_U||_1 \leq \frac{\sqrt{m}}{2}||p_Z - p_U||_2$$

- Estimating Fourier coefficients $\hat{p}_Z$ separately for all $y \in Y$.
- To get bound on number of y's with large Fourier coefficients, estimate $2t$'th moment.

Project goal

- Understand randomness extraction technique and its limitations.
- Understand proof technique for randomness extraction $\mathbb{Z}_m$.
- Give motivation for using this technique.

Project goal

- Understand randomness extraction technique and its limitations.
- Understand proof technique for randomness extraction $\mathbb{Z}_m$.
- Give motivation for using this technique.
- Try other approaches for extracting random element of $\mathbb{Z}_m$:
  - Bits of Z are not independent, we cannot directly use results like [DF90].
  - Using inner products modulo 2 in blocks as in [Vaz87] doesn't give desired independence.
  - Relaxing mixing requirement, because original motivation in [MMP+10] allows this.

References

- [DF90] Persi Diaconis, James Allen Fill, "Strong stationary times via new form of duality".
- [MMP+10] Andrew McGregor, Ilya Mironov, Toniann Pitassi, Omer Reingold, Kunal Talwar, Salil Vadhan, "The Limits of Two-Party Differential Privacy" (http://research.microsoft.com/pubs/137029/2dplimits.pdf)
- [Vaz87] U. V. Vazirni, "Strong communication complexity or generating quasirandom sequences from two communicating semirandom sources"
- [SV86] Miklos Santha, Umesh V. Vazirani, "Generating quasi-random sequences from semi-random sources"
- Ryan O'Donnell, "Analysis of Boolean functions" (http://www.cs.cmu.edu/~odonnell/boolean-analysis/).