Simple lower-bounds for small-bias spaces

Preetum Nakkiran

Jun 03, 2016

I was reading about PRGs recently, and I think a lemma mentioned last time (used for Johnson-Lindenstrauss lower-bounds) can give simple lower-bounds for ε-biased spaces.

Notice:

- 2^n mutually orthogonal vectors requires dimension at least 2^n, but 2^n "almost orthogonal" vectors with pairwise inner-products $|\langle v_i, v_j \rangle| \leq \varepsilon$ exists in dimension $O(n/\varepsilon^2)$, by Johnson-Lindenstrauss.
- Sampling n iid uniform bits requires a sample space of size 2^n, but n ε-biased bits can be sampled from a space of size $O(n/\varepsilon^2)$.

First, let's look at k-wise independent sample spaces, and see how the lower-bounds might be extended to the almost k-wise independent case.

Note: To skip the background, just see Lemma 2, and its application in Claim 4.

1 Preliminaries

What "size of the sample space" means is: For some sample space S, and ± 1 random variables X_i, we will generate bits $x_1, \dots x_n$ as an instance of the r.vs X_i. That is, by drawing a sample $s \in S$, and setting $x_i = X_i(s)$. We would like to have $|S| \ll 2^n$, so we can sample from it using less than n bits.

Also, any random variable X over S can be considered as a vector $\widetilde{X} \in \mathbb{R}^{|S|}$, with coordinates $\widetilde{X}[s] := \sqrt{\Pr[s]} X(s)$. This is convenient because $\langle \widetilde{X}, \widetilde{Y} \rangle = \mathbb{E}[XY]$.

2 Exact k-wise independence

A distribution D on n bits is *k-wise independent* if any subset of k bits are iid uniformly distributed. Equivalently, the distribution $D : \{\pm 1\}^n \to \mathbb{R}_{\geq 0}$ is k-wise independent iff the Fourier coefficients $\hat{D}(S) = 0$ for all $S \neq 0, |S| \leq k$.

n such k-wise independent bits can be generated from a seed of length $O(k \log n)$ bits, using say Reed-Solomon codes. That is, the size of the sample space is $n^{O(k)}$. For k-wise independent bits, size is optimal, as the below claim shows (adapted from Umesh Vazirani's lecture notes [Vaz99]).

Claim 1. *Let D be a k-wise independent distribution on $\{\pm 1\}$ random variables $x_1, \dots, x_n$, over a sample space S. Then, $|S| = \Omega_k(n^{k/2})$.*

Proof. For subset $T \subseteq [n]$, let $\chi_T(x) = \prod_{i \in T} x_i$ be the corresponding Fourier character. Consider these characters as vectors in $\mathbb{R}^{|S|}$ as described above, with

$$\langle \chi_A, \chi_B \rangle = \mathop{\mathbb{E}}_{x \sim D}[\chi_A(x) \chi_B(x)]$$

Let J be the family of all subsets of size $\le k/2$. Note that, for $A, B \in J$, the characters χ_A, χ_B are orthogonal:

$$\begin{aligned}
\langle \chi_A, \chi_B \rangle &= \mathop{\mathbb{E}}_{x \sim D}[\chi_A(x)\chi_B(x)] \\
&= \mathop{\mathbb{E}}_{x \sim D}[(\prod_{i \in A \cap B} x_i^2)(\prod_{i \in A \Delta B} x_i)] \\
&= \mathop{\mathbb{E}}_{x \sim D}[\chi_{A \Delta B}(x)] && \text{(since } x_i^2 = 1\text{)} \\
&= 0 && \text{(since } |A \Delta B| \le k \text{, and } D \text{ is } k\text{-wise independent)}
\end{aligned}$$

Here $A \Delta B$ denotes symmetric difference, and the last equality is because $\chi_{A\Delta B}$ depends on $\le k$ variables, so the expectation over D is the same as over iid uniform bits.

Thus, the characters $\{\chi_A\}_{A \in J}$ form a set of $|J|$ mutually-orthogonal vectors in $\mathbb{R}^{|S|}$. So we must have $|S| \ge |J| = \Omega_k(n^{k/2})$. ■

The key observation was relating independence of random variables to linear independence (orthogonality). Similarly, we could try to relate ε-almost k-wise independent random variables to almost-orthogonal vectors.

3 Main Lemma

This result is Theorem 9.3 from Alon's paper [Alo03]. The proof is very clean, and Section 9 can be read independently. [1]

Lemma 2. *Let $\{v_i\}_{i \in [N]}$ be a collection of N unit vectors in $\mathbb{R}^d$, such that $|\langle v_i, v_j \rangle| \le \varepsilon$ for all $i \ne j$. Then, for $\frac{1}{\sqrt{N}} \le \varepsilon \le 1/2$,*

$$d \ge \Omega\left(\frac{\log N}{\varepsilon^2 \log(1/\varepsilon)}\right)$$

This lower-bound on the dimension of "almost-orthogonal" vectors translates to a nearly-tight lower-bound on Johnson-Lindenstrauss embedding dimension, and will also help us below.

4 Small bias spaces

A distribution D on n bits is *ε-biased w.r.t linear tests* (or just "ε-biased") if all $\mathbb{F}_2$-linear tests are at most ε-biased. That is, for $x \in \{\pm 1\}^n$, the following holds for all subsets $S \subseteq [n]$:

$$\left| \mathop{\mathbb{E}}_{x \sim D}[\chi_S(x)] \right| = \left| \Pr_{x \sim D}[\chi_S(x) = 1] - \Pr_{x \sim D}[\chi_S(x) = -1] \right| \le \varepsilon$$

Similarly, a distribution is *ε-biased w.r.t. linear tests of size k* (or "k-wise ε-biased) if the above holds for all subsets S of size $\le k$.

There exists an ε-biased space on n bits of size $O(n/\varepsilon^2)$: a set of $O(n/\varepsilon^2)$ random n-bit strings will be ε-biased w.h.p. Further, explicit constructions exist that are nearly optimal: the such first construction was in [NN93], and was nicely simplified by [AGHP92] (both papers are very readable).

[1] Theorem 9.3 is stated in terms of lower bounding the rank of a matrix $B \in \mathbb{R}^{N \times N}$ where $B_{i,i} = 1$ and $|B_{i,j}| \le \varepsilon$. The form stated here follows by defining $B_{i,j} := \langle v_i, v_j \rangle$.

These can be used to sample n bits that are k-wise ε-biased, from a space of size almost $O(k\log(n)/\varepsilon^2)$; much better than the size $\Omega(n^k)$ required for perfect k-wise independence. For example[2], see [AGHP92] or the lecture notes [Vaz99].

4.1 Lower Bounds

The best lower bound on size of an ε-biased space on n bits seems to be $\Omega(\frac{n}{\varepsilon^2\log(1/\varepsilon)})$, which is almost tight. The proofs of this in the literature (to my knowledge) work by exploiting a nice connection to error-correcting codes: Say we have a sample space S under the uniform measure. Consider the characters $\chi_T(x)$ as vectors $\widetilde{\chi}_T \in \{\pm 1\}^{|S|}$ defined by $\widetilde{\chi}_T[s] = \chi_T(x(s))$, similar to what we did in Section 2. The set of 2^n vectors $\{\widetilde{\chi}_T\}_{T\subseteq[n]}$ defines the codewords of a linear code of length $|S|$ and dimension n. Further, the hamming-weight of each codeword (number of -1s in each codeword, in our context), is within $n(\frac{1}{2}\pm\varepsilon)$, since each parity χ_T is at most ε-biased. Thus this code has relative distance at least $\frac{1}{2}-\varepsilon$, and we can use sphere-packing-type bounds from coding-theory to lower-bound the codeword length $|S|$ required to achieve such a distance. Apparently the "McEliece-Rodemich-Rumsey-Welch bound" works in this case; a more detailed discussion is in [AGHP92, Section 7].

We can also recover this same lower bound using Lemma 2 in a straightforward way.

Claim 3. *Let D be an ε-biased distribution on n bits $x_1, \dots, x_n$, over a sample space S. Then,*

$$|S| = \Omega\left(\frac{n}{\varepsilon^2\log(1/\varepsilon)}\right)$$

Proof. Following the proof of Claim 1, consider the Fourier characters $\chi_T(x)$ as vectors $\widetilde{\chi}_T \in \mathbb{R}^{|S|}$, with $\widetilde{\chi}_T[s] = \sqrt{\Pr[s]}\chi_T(x(s))$. Then, for all distinct subsets $A, B \subseteq [n]$, we have

$$\langle \widetilde{\chi}_A, \widetilde{\chi}_B \rangle = \mathop{\mathbb{E}}_{x\sim D}[\chi_A(x)\chi_B(x)] = \mathop{\mathbb{E}}_{x\sim D}[\chi_{A\Delta B}(x)]$$

Since D is ε-biased, $|\mathbb{E}_{x\sim D}[\chi_{A\Delta B}(x)]| \le \varepsilon$ for all $A \neq B$. Thus, applying Lemma 2 to the collection of $N = 2^n$ unit vectors $\{\widetilde{\chi}_T\}_{T\subseteq[n]}$ gives the lower bound $|S| = \Omega\left(\frac{n}{\varepsilon^2\log(1/\varepsilon)}\right)$. ■

This also nicely generalizes the proof of Claim 1, to give an almost-tight lower bound on spaces that are ε-biased w.r.t linear tests of size k.

Claim 4. *Let D be a distribution on n bits that is ε-biased w.r.t. linear tests of size k. Then, the size of the sample space is*

$$|S| = \Omega\left(\frac{k\log(n/k)}{\varepsilon^2\log(1/\varepsilon)}\right)$$

Proof. As before, consider the Fourier characters $\chi_T(x)$ as vectors $\widetilde{\chi}_T \in \mathbb{R}^{|S|}$, with $\widetilde{\chi}_T[s] = \sqrt{\Pr[s]}\chi_T(x(s))$. Let J be the family of all subsets $T \subseteq [n]$ of size $\le k/2$. Then, for all distinct subsets $A, B \in J$, we have

$$|\langle \widetilde{\chi}_A, \widetilde{\chi}_B \rangle| = \left|\mathop{\mathbb{E}}_{x\sim D}[\chi_{A\Delta B}(x)]\right| \le \varepsilon$$

since $|A\Delta B| \le k$, and D is ε-biased w.r.t such linear tests. Applying Lemma 2 to the collection of $|J|$ unit vectors $\{\widetilde{\chi}_T\}_{T\in J}$ gives $|S| = \Omega(\frac{k\log(n/k)}{\varepsilon^2\log(1/\varepsilon)})$. ■

[2] This can be done by composing an (n, k') ECC with dual-distance k and an ε-biased distribution on $k' = k\log n$ bits. Basically, use a linear construction for generating n exactly k-wise independent bits from k' iid uniform bits, but use an ε-biased distribution on k' bits as the seed instead.

Note: I couldn't find the lower bound given by Claim 4 in the literature, so please let me know if you find a bug or reference.

Also, these bounds do not directly imply nearly tight lower bounds for ε-almost k-wise independent *distributions (that is, distributions s.t. their marginals on all sets of k variables are* ε*-close to the uniform distribution, in* ℓ_∞ *or* ℓ_1 *norm). Essentially because of the loss in moving between closeness in Fourier domain and closeness in distributions.* [3]

References

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992. URL: `http://www.tau.ac.il/~nogaa/PDFS/aghp4.pdf`.

[Alo03] Noga Alon. Problems and results in extremal combinatorics, part i. *Discrete Math*, 273:31–53, 2003. URL: `http://www.tau.ac.il/~nogaa/PDFS/extremal1.pdf`.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM journal on computing*, 22(4):838–856, 1993. URL: `http://www.wisdom.weizmann.ac.il/~naor/PAPERS/bias.pdf`.

[Vaz99] Umesh Vazirani. k-wise independence and epsilon-biased k-wise indepedence. 1999. URL: `https://people.eecs.berkeley.edu/~vazirani/s99cs294/notes/lec4.pdf`.

[3] Eg, ε-biased $\implies$ ε-close in ℓ_∞, but ε-close in ℓ_∞ can be up to $2^{k-1}\varepsilon$-biased. And $2^{-k/2}\varepsilon$-biased $\implies$ ε-close in ℓ_1, but not the other direction.