Note on Direct Product Testing with Nearly Identical Sets

Irit Dinur

May 5, 2016

We show a simple reduction from direct product testing with large intersection size \((1 - \delta)n\) to direct product testing with linear intersection size \(\theta(n)\). The linear intersection regime was analyzed in [1] by the author and Steurer. Moshkovitz [3] is interested in the large-intersection regime because of its possible connection to unique games questions, see [2].

Let \(f : U^n \to \mathbb{R}^n\). We will consider the direct product test in which the function \(f\) is queried in two locations \(x, x'\) and the values \(f(x)\) and \(f(x')\) are compared on a set of indices ("the intersection") where \(x_i = x'_i\).

Let \(B_\alpha\) be the distribution over triples \(x, T, x'\) parameterized by \(0 < \alpha < 1\) as follows.

1. Select \(T \subset [n]\) from the binomial distribution \(B(n, \alpha)\), i.e. for each \(i\) independently put \(i\) in \(T\) with probability \(\alpha\).
2. Select \(x \in U^n\) uniformly, select \(x' \in U^n\) uniformly conditioned on \(x'_T = x_T\).

Let

\[
\text{test}(\alpha) = \mathbb{P}_{x, T, x' \sim B_\alpha} [f(x)_T = f(x')_T].
\]

The main point in this note is that

**Proposition 1.** For every \(0 < \delta < 1\), \(\text{test}((1 - \delta)^2) \geq (\text{test}(1 - \delta))^2\).

**Corollary 2.** Fix \(0 < \delta, \beta < 1\). If \(\text{test}(1 - \delta) > \beta^n\) then \(\text{test}(\alpha) > \beta^n\), where \(\frac{1}{2^\delta} \leq \alpha \leq \frac{1}{e}\).

**Proof.** Let \(p \geq 0\) be an integer such that that \(2^{-p} < \delta \leq 2^{-p+1}\). By repeating the inequality \(p\) times we can deduce that

\[
\text{test}(\alpha) \geq (\text{test}(1 - \delta))^{2^p} \geq (\text{test}(1 - \delta))^{2^p/\delta}
\]

where \(\alpha = (1 - \delta)^{2^p}\) is a constant which is between \(1/e^2\) and \(1/e\). \(\square\)

*Department of Computer Science and Applied Mathematics, Weizmann Institute.
This almost suffices for applying the local-structure lemma of the author and Steurer [1, Lemma 1.2]. We must make one small tweak since the test distribution in [1] is not $B_α$, but rather $B_{α,k}$, defined by selecting a random subset $T \subset [n]$ of size exactly $k$ and then two random strings $x, x' \in U^n$ such that $x_T = (x')_T$. Denote

$$test(k) = \mathbb{P}_{(x,T,x') \sim B_{α,k}} [f(x)_T = f(x')_T].$$

Clearly $test(α) = \sum_{k=0}^{n} \binom{n}{k}α^k(1 - α)^{n-k}test(k)$ and using standard tail bounds we can deduce that if $test(α) > β^n$ then $test(k) > β^n - \exp(-n)$ for some $k \approx αn$ where $\exp(-n)$ is an error term that comes from a tail inequality. Thus,

**Corollary 3.** Fix $0 < δ, β < 1$. If $test(1 - δ) > β^n$ then there is some $n/10 < k < n/2$ such that $test(k) > β^n - \exp(-n)$.

The main theorem in [3] follows directly from this corollary together with the direct product testing result [1, Lemma 1.2] for linear intersection size, (a more friendly version appears as Lemma 1.1 in [3]). This is meaningful even for $δ = 1/n$, i.e. for the largest possible intersection, of $n - 1$ elements (in expectation).

**Proof.** (of Proposition 1) For an event $A$ denote by $1(A)$ the corresponding indicator variable.

$$\begin{align*}
(test(1 - δ))^2 &= (\mathbb{E}_{x, T, x'|x} 1(f(x)_T = f(x')_T))^2 \\
&\leq \mathbb{E}_{x, T, x'|x} 1(f(x)_T = f(x')_T))^2 \\
&= \mathbb{E}_{x, T_1, x_1|x} 1(f(x)_T_1 = f(x)_T_1) \mathbb{E}_{T_2, x_2|x} 1(f(x)_T_2 = f(x)_T_2)) \\
&= \mathbb{E}_{x, x_1, T_1, x_2, T_2|x} 1(f(x)_T_1 = f(x)_T_1) \cdot 1(f(x)_T_2 = f(x)_T_2) \\
&\leq \mathbb{E}_{x, T_1, x_1, T_2, x_2} 1(f(x)_T_1 = f(x)_T_1 \land f(x)_T_2 = f(x)_T_2) \\
&= test((1 - δ)^2)
\end{align*}$$

where the first inequality is Jensen’s inequality, and the last equality is because the triple $x_1, T_1 \cap T_2, x_2$ is distributed exactly according to $B_{(1-δ)^2}$. $\square$

**References**

