Contents

1 The basic idea

This is a feasibility result regarding ds-IPPs for distributions, where 'ds' stands for doubly-sublinear. We shall show a property of distributions for which the prover's sample complexity is close to the complexity of testing, whereas the sample complexity of the verifier is much lower. Actaully, we shall show two such related properties.

For $d \ll n$, we consider the set of degree d polynomials over a field of size n, denoted $\mathcal{F}$. For each such polynomial p, we let X_p be a distribution that is uniform over $S_p \stackrel{\text{def}}{=} \{(e, p(e)) : e \in \mathcal{F}\}$, and let $\mathcal{D}$ be the set of all such distributions. We also consider $\mathcal{D}'$ that consist of all distributions with a support that is a subset of S_p for some degree d polynomial p.

We first note that testing $\mathcal{D}$ (resp., $\mathcal{D}'$) requires more than d samples. This is the case because when using at most d samples one cannot distinguish the following two random d-long sequences.

1. A sequence of the form $((i_1, f(i_1)), ..., (i_d, f(i_d)))$, where f is a random function from $\mathcal{F}$ to $\mathcal{F}$ and $i_1, ..., i_d$ are uniformly and independently selected in $\mathcal{F}$.

2. A sequence of the form $((i_1, p(i_1)), ..., (i_d, p(i_d)))$, where p is a random degree d polynomial over $\mathcal{F}$ and $i_1, ..., i_d$ are uniformly and independently selected in $\mathcal{F}$.

The point is that any d values of a random degree d polynomial are uniformly and independent distributed in $\mathcal{F}$, just as the case with a random function.

On the other hand, assuming that $d = \omega(\sqrt{n})$, an ds-IPP for $\mathcal{D}$ proceeds as follows. The honest prover uses $O(d/\epsilon^2)$ samples of X_p in order to reconstruct the polynomial p, which it sends to the verifier. The verifier takes a sample of size $O(\sqrt{n}/\epsilon^2)$ and invokes the uniformity tester on the input distribution, while using a 1-1 correspondance between $[n]$ and S_p. Needless to say, the verifier rejects if it sees any sample that is not in S_p. Note that the hypothesis $d = \omega(\sqrt{n})$ implies that the sample complexity of the verifier is lower than the sample complexity of a tester.

An ds-IPP for $\mathcal{D}'$ can also be obtained. In this case the sample (of size $O(d/\epsilon^2)$) may not determine a unique degree d polynomial, but the honest prover may just select an arbitrary degree d polynomial p that is consistent with this sample, and send p to the verifier. The verifier accepts if and only if a random sample of size $O(1/\epsilon^2)$ that it takes yields values that are all in S_p. Note that in this case, regardless of the relation between d and n, the sample complexity of the verifier is lower than the sample complexity of a tester.

Summary. Both $\mathcal{D}$ and $\mathcal{D}'$ require more than d samples for testing, but they both have IPPs in which the honest prover uses $O(d/\epsilon^2)$ samples. The verifier's sample complexity is $O(\sqrt{n}/\epsilon^2)$ in case of $\mathcal{D}$, and $O(1/\epsilon^2)$ in case of $\mathcal{D}'$. Recall that each distribution in $\mathcal{D}$ is uniform over a set of size n, whereas each distribution in $\mathcal{D}'$ has support size at most n.

2 Ramifications

Beyond being artifical, the foregoing examples have two deficiencies, which we fix next.

1. The complexity of testing $\mathcal{D}$ (resp., $\mathcal{D}'$) equals the one of learning distributions in $\mathcal{D}$ (resp., $\mathcal{D}'$), whereas the interesting cases in property testing are those in which testing is significanly more efficient than learning.
2. For constant $\epsilon > 0$, the communication complexity of our IPP equals the sample complexity of testing; hence, in this case, the running-time of the verifier is not lower than its the running-time of the tester.

Needless to say, the resulting properties are even more artifical. Still this does not harm their demostrtative feature.

2.1 Making testing significantly easier than learning

For the class $\mathcal{D}$, this can be done by augmenting the class of distributions so that the complexity of learning it becomes higher than the complexity of testing, which is maintained. For example, we may replace S_p by $S_{p,g} = \{(e, p(e), g(e)) : e \in \mathcal{F}\}$ such that $g : \mathcal{F} \to \{0,1\}$ is an arbitrary Boolean function. Recall that the complexity of learning an arbitrary Boolean function over $\mathcal{F} = [n]$ is $\Omega(n)$, but testing the class of all Boolean functions is trivial.

The verifier in the ds-IPP proceeds as the basic one, except that it also checks that samples that are equal on the first element (i.e., e) are also equal on the third element (i.e., $g(e)$). Given that the verifier takes $O(\sqrt{n}/\epsilon^2)$ samples anyhow, this allows checking that the distribution of the third element fits *some* function $g : [n] \to \{0,1\}$.

For the class $\mathcal{D}'$ (and actually al;o for $\mathcal{D}$), we can augment the distribution on pairs by an arbitrary distribution on an additioanl bit (equiv., the foregoing function g can be replaced by an arbitrary random process $G : [n] \to \{0,1\}$). In this case, we do not augment the basic verifier.

2.2 Significantly reducing the communication complexity of our IPP

Here a more significant modification is required. Rather than using a single univariate polynomial of degree $d = \omega(\sqrt{n})$ over a field of size n, we shall use $t = n/|\mathcal{F}|$ univariate polynomials of degree $d = O(\log n)$ over $\mathcal{F}$ such that $|\mathcal{F}| \gg d$ (e.g., $|\mathcal{F}| \geq 10d$). Specifically, for $\overline{p} = (p_1, ..., p_t)$, where each p_i is a degree d polynomial over $\mathcal{F}$, we consider the set

$$S_{\overline{p}} \stackrel{\text{def}}{=} \{(i, e, p_i(e)) : i \in [t] \ \& \ e \in \mathcal{F}\}$$

and let $\mathcal{D}$ (resp., $\mathcal{D}'$) be the set of all distributions that are uniform over some $S_{\overline{p}}$ (resp., have a support that is a subset of some $S_{\overline{p}}$). Note that $|S_{\overline{p}}| = t \cdot |\mathcal{F}| = n$, and that we can pick $\mathcal{F}$ such that $|\mathcal{F}| = O(\log n)$.

We first observe that that $\mathcal{D}$ (resp., $\mathcal{D}'$) cannot be tested using $o(t \cdot d)$ samples. This holds because using $m = o(t \cdot d)$ samples does not allow to distinguish the following two distributions over m-long sequences.

1. A sequence of the form $((i_1, e_1, f(e_1, i_1)), ..., (i_m, e_m, f(i_m, e_m)))$, where f is a random function from $[t] \times \mathcal{F}$ to $\mathcal{F}$ and $(i_1, e_1), ..., (i_m, e_m)$ are uniformly and independently selected in $[t] \times \mathcal{F}$.

2. A sequence of the form $((i_1, e_1, p_{i_1}(e_1)), ..., (i_m, e_m, p_{i_m}(e_m)))$, where the p_i's are random degree d polynomial over $\mathcal{F}$ and $(i_1, e_1), ..., (i_m, e_m)$ are uniformly and independently selected in $[t] \times \mathcal{F}$.

The point is that m samples of the form (i, e, v) are highly unlikely to include more than d triples with the same first element (i.e., same i), and otherwise the samples that correspond to each $i \in [t]$ are uniformly and independently distributed in $\mathcal{F} \times \mathcal{F}$.

On the other hand, assuming $t = \omega(\sqrt{n})$, an ds-IPP for $\mathcal{D}$ proceeds as follows. The honest prover uses $O(t \cdot d/\epsilon^2)$ samples of $X_{\overline{p}}$ in order to reconstruct $\overline{p} = (p_1, ..., p_t)$, but (in this case it) does not send $\overline{p}$ to the verifier. The verifier uses $O(\sqrt{n}/\epsilon^2)$ samples in order to check whether the first two elements of the input distribution X are uniformly distributed in $[t] \times \mathcal{F}$. In addition, for each sample (i, e, v), the verifier sends i to the prover, which replies with a degree d polynomial $\widetilde{p}_i$, and the verifier checks that $\widetilde{p}_i$ (is a degree d polynomial that) satisfies $\widetilde{p}_i(e) = v$.

An ds-IPP for $\mathcal{D}'$ can also be obtained. In this case a sample (of size $O(t \cdot d/\epsilon^2)$) may not determine a unique sequence of t polynomials (of degree d), but the honest prover may just select an arbitrary sequence $\overline{p} = (p_1, ..., p_t)$ that is consistent with this sample. The verifier takes $O(1/\epsilon^2)$ samples, send the corresponding first elements to the prover, which again replies with the corresponding degree d polynomials, and the verifier accepts if and only if the polynomials sent by the prover fit the corresponding other elements. That is, for each sample (i, e, v), the verifier sends i to the prover, which replies with p_i (and the verifier checks that $p_i(e) = v$ (and that p_i is a polynomial of degree d)).

Summary. Both $\mathcal{D}$ and $\mathcal{D}'$ require $\Omega(t \cdot d)$ samples for testing and $\Omega(n)$ samples for learning, but they both have IPPs in which the honest prover uses $O(t \cdot d/\epsilon^2)$ samples. The verifier's sample complexity is $O(\sqrt{n}/\epsilon^2)$ in case of $\mathcal{D}$, and $O(1/\epsilon^2)$ in case of $\mathcal{D}'$. The communication complexity is $O(d)$ times larger than the sample complexity of the verifier. Recall that each distribution in $\mathcal{D}$ is uniform over a set of size $t \cdot |\mathcal{F}| = n$, and that we may use $|\mathcal{F}| = O(d) = O(\log n)$.

Combining both augmentations. Using the suggestion made in the last paragraph of Section 2.1, we can augment the forgoing $\mathcal{D}$ (resp., $\mathcal{D}'$) by replacing $S_{\overline{p}}$ with $S_{\overline{p},G}$ such that

$$S_{\overline{p},G} \stackrel{\text{def}}{=} \{(i, e, p_i(e), G(i, e)) : i \in [t] \ \& \ e \in \mathcal{F}\}$$

where $G : [n] \times \mathcal{F} \to \{0, 1\}$ is an arbitrary random process.