Uniformity Testing for High-Dimensional Distributions: Subcube Conditioning, Random Restrictions, and Mean Testing

Clément Canonne (IBM Research)

December 17, 2019

Joint work with Xi Chen, Gautam Kamath, Amit Levi, and Erik Waingarten

Outline

Introduction

Property Testing

Sublinear-time,

Property Testing

Sublinear-time, approximate,

Property Testing

Sublinear-time, approximate, randomized

Property Testing

Sublinear-time, approximate, randomized decision algorithms that make local queries to their input.

Property Testing

Sublinear-time, approximate, randomized decision algorithms that make local queries to their input.

- Big dataset: *too* big

Property Testing

Sublinear-time, approximate, randomized decision algorithms that make local queries to their input.

- Big dataset: *too* big
- Expensive access: *pricey* data

Property Testing

Sublinear-time, approximate, randomized decision algorithms that make local queries to their input.

- Big dataset: *too* big
- Expensive access: *pricey* data
- "Model selection": *many* options

Property Testing

Sublinear-time, approximate, randomized decision algorithms that make local queries to their input.

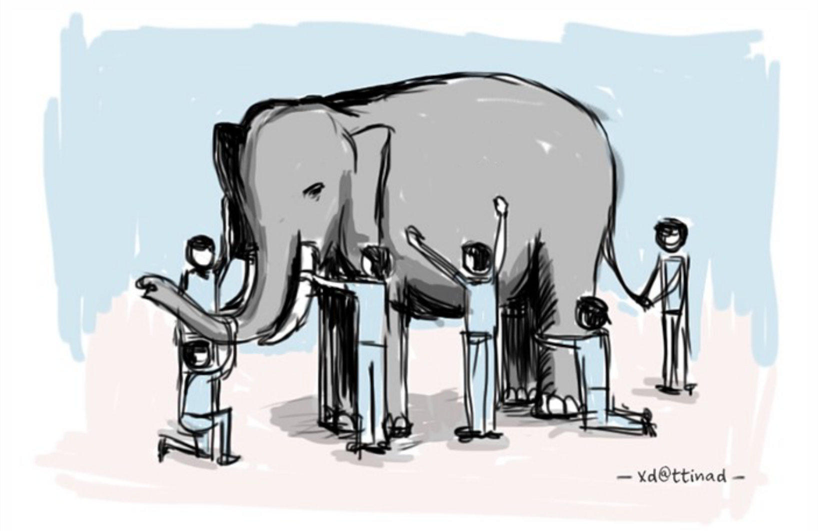
- Big dataset: *too* big
- Expensive access: *pricey* data
- “Model selection”: *many* options
- “Good enough:” *a priori* knowledge

Property Testing

Sublinear-time, approximate, randomized decision algorithms that make local queries to their input.

- Big dataset: *too* big
- Expensive access: *pricey* data
- "Model selection": *many* options
- "Good enough:" *a priori* knowledge

Need to infer information – *one bit* – from the data: quickly, or with very few lookups.



"Is it far from a kangaroo?"

Property Testing

Introduced by [RS96, GGR96] – has been a *very* active area since.

- Known space (e.g., $\{0,1\}^N$)
- Property $\mathcal{P} \subseteq \{0,1\}^N$
- Oracle access to unknown $x \in \{0,1\}^N$
- Proximity parameter $\varepsilon \in (0,1]$

Must decide

$$x \in \mathcal{P} \quad \text{vs.} \quad \text{dist}(x, \mathcal{P}) > \varepsilon$$

(has the property, or is ε-*far* from it)

Distribution Testing

Now, our "big object" is a probability distribution over a (finite) domain.

Distribution Testing

Now, our "big object" is a probability distribution over a (finite) domain.

- type of queries: *independent samples**

Distribution Testing

Now, our "big object" is a probability distribution over a (finite) domain.

- type of queries: *independent samples**
- type of distance: *total variation*

Distribution Testing

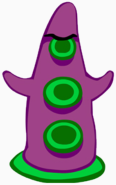
Now, our “big object” is a probability distribution over a (finite) domain.

- type of queries: *independent samples**
- type of distance: *total variation*
- type of object: *distributions*

Distribution Testing

Now, our "big object" is a probability distribution over a (finite) domain.

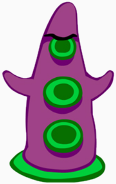
- type of queries: *independent samples**
- type of distance: *total variation*
- type of object: *distributions*



Distribution Testing

Now, our "big object" is a probability distribution over a (finite) domain.

- type of queries: *independent samples**
- type of distance: *total variation*
- type of object: *distributions*

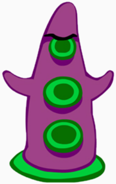


*Disclaimer: not always, as we will see.

Distribution Testing

Now, our “big object” is a probability distribution over a (finite) domain.

- type of queries: *independent samples**
- type of distance: *total variation*
- type of object: *distributions*



*Disclaimer: not always, as we will see.

Our Problem

Uniformity testing

We focus on arguably the simplest and most fundamental property: uniformity.

Given samples from $\mathbf{p}$: is $\mathbf{p} = \mathbf{u}$, or $\mathrm{TV}(\mathbf{p}, \mathbf{u}) > \varepsilon$?

Uniformity testing

We focus on arguably the simplest and most fundamental property: uniformity.

Given samples from $\mathbf{p}$: is $\mathbf{p} = \mathbf{u}$, or $\mathrm{TV}(\mathbf{p}, \mathbf{u}) > \varepsilon$?

Oh, and we would like to do that for high-dimensional distributions.

Uniformity testing: Good News

Its is well-known ([Pan08, VV14], and then [DGPP16, DGPP18] and more) that testing uniformity over a domain of size N takes $\Theta(\sqrt{N}/\varepsilon^2)$ samples.

Uniformity testing: Bad News

In the high-dimensional setting (we think of $\{-1,1\}^n$ with $n \gg 1$) that means $\Theta(2^{n/2}/\varepsilon^2)$ samples, exponential in the dimension.

Uniformity testing: Good News

In the high-dimensional setting with structure* testing uniformity over $\{-1,1\}^n$ takes $\Theta(\sqrt{n}/\varepsilon^2)$ samples [CDKS17].

* when we assume product distributions.

Uniformity testing: Bad News

We do not want to make any structural assumption. $\mathbf{p}$ is, a priori, arbitrary.

Uniformity testing: Bad News

We do not want to make any structural assumption. $\mathbf{p}$ is, a priori, arbitrary.

So what to do?

Subcube Conditioning

Variant of conditional sampling [CRS15, CFGM16] suggested in [CRS15] and first studied in [BC18]: can specify assignments of any of the n bits, and get a sample from $\mathbf{p}$ conditioned on those bits being fixed.

Subcube Conditioning

Variant of conditional sampling [CRS15, CFGM16] suggested in [CRS15] and first studied in [BC18]: can specify assignments of any of the n bits, and get a sample from $\mathbf{p}$ conditioned on those bits being fixed.

Very well suited to this high-dimensional setting.

Testing Result

[BC18] showed that subcube conditional queries allow uniformity testing with $\tilde{O}(n^3/\varepsilon^3)$ samples (no longer exponential!). Surprinsingly, we show it is sublinear:

Theorem (Main theorem)

Testing uniformity with subcube conditional queries has sample complexity $\tilde{O}(\sqrt{n}/\varepsilon^2)$.

(immediate $\Omega(\sqrt{n}/\varepsilon^2)$ lower bound from the product case)

Ingredients

This relies on two main ingredients: a structural result analyzing random restrictions of a distribution; and a subroutine for a related testing task, mean testing.

Structural Result (I)

Definition (Projection)

Let $\mathbf{p}$ be any distribution over $\{-1,1\}^n$, and $S \subseteq [n]$. The projection $\mathbf{p}_S$ of $\mathbf{p}$ on S is the marginal distribution of $\mathbf{p}$ on $\{-1,1\}^{|S|}$.

Definition (Mean)

Let $\mathbf{p}$ be as above. $\mu(\mathbf{p}) \in \mathbb{R}^n$ is the mean vector of $\mathbf{p}$, $\mu(\mathbf{p}) = \mathbb{E}_{\mathbf{x}\sim\mathbf{p}}[\mathbf{x}]$.

Structural Result (II)

Definition (Restriction)

Let $\mathbf{p}$ be any distribution over $\{-1, 1\}^n$, and $\sigma \in [0, 1]$. A random restriction $\rho = (\mathbf{S}, \mathbf{x})$ is obtained by (i) sampling $\mathbf{S} \subseteq [n]$ by including each element i.i.d. w.p. σ; (ii) sampling $\mathbf{x} \sim \mathbf{p}$.

Conditioning $\mathbf{p}$ on $x_i = \mathbf{x}_i$ for all $i \in \mathbf{S}$ gives the distribution $\mathbf{p}_{|\rho}$.

Structural Result (III)

Theorem (Restriction theorem, Informal)

Let $\mathbf{p}$ *be any distribution over* $\{-1,1\}^n$. *Then, when* $\mathbf{p}$ *is "hit" by a random restriction* ρ *as above,*

$$\mathbb{E}_\rho\left[\|\mu(\mathbf{p}_{|\rho})\|_2\right] \geq \sigma \cdot \mathbb{E}_{\mathbf{S}}\left[\mathrm{TV}(\mathbf{p}_{\overline{\mathbf{S}}}, \mathbf{u})\right].$$

Structural Result (IV)

Theorem (Pisier's inequality [Pis86, NS02])

Let $f\colon \{-1,1\}^n \to \mathbb{R}$ *be s.t.* $\mathbb{E}_{\mathbf{x}}[f(\mathbf{x})] = 0$. *Then*

$$\mathbb{E}_{\mathbf{x}\sim\{-1,1\}^n}[|f(\mathbf{x})|] \lesssim \log n \cdot \mathbb{E}_{\mathbf{x},\mathbf{y}\sim\{-1,1\}^n}\left[\left|\sum_{i=1}^n \mathbf{y}_i \mathbf{x}_i L_i f(\mathbf{x})\right|\right].$$

Structural Result (IV)

Theorem (Pisier's inequality [Pis86, NS02])

Let $f\colon \{-1,1\}^n \to \mathbb{R}$ *be s.t.* $\mathbb{E}_{\mathbf{x}}[f(\mathbf{x})] = 0$*. Then*

$$\mathbb{E}_{\mathbf{x}\sim\{-1,1\}^n}[|f(\mathbf{x})|] \lesssim \log n \cdot \mathbb{E}_{\mathbf{x},\mathbf{y}\sim\{-1,1\}^n}\left[\left|\sum_{i=1}^{n} \mathbf{y}_i \mathbf{x}_i L_i f(\mathbf{x})\right|\right].$$

Theorem (Robust version)

Let $f\colon \{-1,1\}^n \to \mathbb{R}$ *be s.t.* $\mathbb{E}_{\mathbf{x}}[f(\mathbf{x})] = 0$ *and* $G = (\{-1,1\}^n, E)$ *be any orientation of the hypercube. Then,*

$$\mathbb{E}_{\mathbf{x}\sim\{-1,1\}^n}[|f(\mathbf{x})|] \lesssim \log n \cdot \mathbb{E}_{\mathbf{x},\mathbf{y}\sim\{-1,1\}^n}\left[\left|\sum_{\substack{i\in[n] \\ (\mathbf{x},\mathbf{x}^{(i)})\in E}} \mathbf{y}_i \mathbf{x}_i L_i f(\mathbf{x})\right|\right].$$

Mean Testing Result (I)

Consider the following question: from i.i.d. ("standard") samples from $\mathbf{p}$ on $\{-1, 1\}^n$, distinguish (i) $\mathbf{p} = \mathbf{u}$ and (ii) $\|\mu(\mathbf{p})\|_2 > \varepsilon$.

Mean Testing Result (I)

Consider the following question: from i.i.d. ("standard") samples from $\mathbf{p}$ on $\{-1,1\}^n$, distinguish (i) $\mathbf{p} = \mathbf{u}$ and (ii) $\|\mu(\mathbf{p})\|_2 > \varepsilon$.

Remarks

No harder than uniformity testing.

Mean Testing Result (I)

Consider the following question: from i.i.d. ("standard") samples from $\mathbf{p}$ on $\{-1,1\}^n$, distinguish (i) $\mathbf{p} = \mathbf{u}$ and (ii) $\|\mu(\mathbf{p})\|_2 > \varepsilon$.

Remarks

No harder than uniformity testing. Can ask the same for Gaussians: $\mathbf{p} = N(0_n, I_n)$ vs. $\mathbf{p} = N(\mu, \Sigma)$ with $\|\mu(\mathbf{p})\|_2 > \varepsilon$.

Mean Testing Result (II)

Theorem (Mean Testing theorem)

For $\varepsilon \in (0, 1]$, ℓ_2 mean testing has (standard) sample complexity $\Theta^*(\sqrt{n}/\varepsilon^2)$, for both Boolean and Gaussian cases.

Mean Testing Result (III)

Main idea
Use a nice unbiased estimator that works well in the product case:

$$Z = \left\langle \frac{1}{m}\sum_{j=1}^{m} X^{(2i)}, \frac{1}{m}\sum_{j=1}^{m} X^{(2i-1)} \right\rangle$$

Mean Testing Result (III)

Main idea
Use a nice unbiased estimator that works well in the product case:

$$Z = \left\langle \frac{1}{m} \sum_{j=1}^{m} X^{(2i)}, \frac{1}{m} \sum_{j=1}^{m} X^{(2i-1)} \right\rangle$$

$\mathbb{E}[Z] = \|\mu(\mathbf{p})\|_2^2$, and $\mathrm{Var}[Z] \approx \|\Sigma(\mathbf{p})\|_F^2$.

Mean Testing Result (III)

Main idea
Use a nice unbiased estimator that works well in the product case:

$$Z = \left\langle \frac{1}{m} \sum_{j=1}^{m} X^{(2i)}, \frac{1}{m} \sum_{j=1}^{m} X^{(2i-1)} \right\rangle$$

$\mathbb{E}[Z] = \|\mu(\mathbf{p})\|_2^2$, and $\mathrm{Var}[Z] \approx \|\Sigma(\mathbf{p})\|_F^2$.

If the test fails...
it means the variance is too big, and so $\|\Sigma(\mathbf{p})\|_F^2 \gg n$. So build $\mathbf{p}'$ on $\{-1, 1\}^{\binom{n}{2}}$ with $\mu(\mathbf{p}') = \Sigma(\mathbf{p})$ and test that one...

Mean Testing Result (III)

Main idea
Use a nice unbiased estimator that works well in the product case:

$$Z = \left\langle \frac{1}{m}\sum_{j=1}^{m} X^{(2i)}, \frac{1}{m}\sum_{j=1}^{m} X^{(2i-1)} \right\rangle$$

$\mathbb{E}[Z] = \|\mu(\mathbf{p})\|_2^2$, and $\mathrm{Var}[Z] \approx \|\Sigma(\mathbf{p})\|_F^2$.

If the test fails...
it means the variance is too big, and so $\|\Sigma(\mathbf{p})\|_F^2 \gg n$. So build $\mathbf{p}'$ on $\{-1,1\}^{\binom{n}{2}}$ with $\mu(\mathbf{p}') = \Sigma(\mathbf{p})$ and test that one... again, and again, and again.

Putting it together (I)

To get from the above ingredients to our main theorem, we rely on the following (simple) lemma:

Lemma (Recursion Lemma)

Let $\mathbf{p}$ *be a distribution on* $\{-1,1\}^n$*. For any* $\sigma \in [0,1]$,

$$\mathsf{TV}(\mathbf{p},\mathbf{u}) \leq \mathbb{E}_{\mathbf{S}}\left[\mathsf{TV}(\mathbf{p}_{\overline{\mathbf{S}}},\mathbf{u})\right] + \mathbb{E}_{\rho}\left[\mathsf{TV}(\mathbf{p}_{|\rho},\mathbf{u})\right].$$

Putting it together (I)

To get from the above ingredients to our main theorem, we rely on the following (simple) lemma:

Lemma (Recursion Lemma)

Let $\mathbf{p}$ *be a distribution on* $\{-1,1\}^n$. *For any* $\sigma \in [0,1]$,

$$\mathrm{TV}(\mathbf{p},\mathbf{u}) \leq \mathbb{E}_{\mathbf{S}}\left[\mathrm{TV}(\mathbf{p}_{\overline{\mathbf{S}}},\mathbf{u})\right] + \mathbb{E}_{\rho}\left[\mathrm{TV}(\mathbf{p}_{|\rho},\mathbf{u})\right].$$

Now we recurse...

Putting it together (II)

Start with **p**, far from uniform. Hit it with a random restriction: one of the two terms has to be at least $\varepsilon/2$.

- If it's the first, by our structural lemma the mean has to be large.* Apply our mean testing algorithm.
- If it's the second, then we have the same testing question on $n/2$ variables. Recurse.

Conclusion

Upshot

- In high dimensions, testing is expensive. Either you assume structure, or assume access.

Upshot

- In high dimensions, testing is expensive. Either you assume structure, or assume access.
- Surprinsingly, both lead to the same (huge) savings!

Upshot

- In high dimensions, testing is expensive. Either you assume structure, or assume access.
- Surprinsingly, both lead to the same (huge) savings!
- New notion of random restriction for distributions, analysis via isoperimetry. Further applications?

Upshot

- In high dimensions, testing is expensive. Either you assume structure, or assume access.
- Surprinsingly, both lead to the same (huge) savings!
- New notion of random restriction for distributions, analysis via isoperimetry. Further applications?
- Mean testing for Gaussians. Previously unknown (?*)

Thank you.

Questions?

Rishiraj Bhattacharyya and Sourav Chakraborty.
Property testing of joint distributions using conditional samples.
Transactions on Computation Theory, 10(4):16:1–16:20, 2018.

Clément L. Canonne, Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart.
Testing Bayesian networks.
In *Proceedings of the 30th Annual Conference on Learning Theory*, COLT '17, pages 370–448, 2017.

Sourav Chakraborty, Eldar Fischer, Yonatan Goldhirsh, and Arie Matsliah.
On the power of conditional samples in distribution testing.
SIAM Journal on Computing, 45(4):1261–1296, 2016.

Clément L. Canonne, Dana Ron, and Rocco A. Servedio.
Testing probability distributions using conditional samples.
SIAM Journal on Computing, 44(3):540–616, 2015.

Ilias Diakonikolas, Themis Gouleakis, John Peebles, and Eric Price.
Collision-based testers are optimal for uniformity and closeness.
arXiv preprint arXiv:1611.03579, 2016.

Ilias Diakonikolas, Themis Gouleakis, John Peebles, and Eric Price.
Sample-optimal identity testing with high probability.
In *Proceedings of the 45th International Colloquium on Automata, Languages, and Programming*, ICALP '18, pages 41:1–41:14, 2018.

Oded Goldreich, Shafi Goldwasser, and Dana Ron.
Property testing and its connection to learning and approximation.
In *Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '96, pages 339–348, Washington, DC, USA, 1996. IEEE Computer Society.

Assaf Naor and Gideon Schechtman.
Remarks on non-linear type and Pisier's inequality.
Journal fur die Reine und Angewandte Mathematik, 552:213–236, 2002.

Liam Paninski.
A coincidence-based test for uniformity given very sparsely sampled discrete data.
IEEE Transactions on Information Theory, 54(10):4750–4755, 2008.

Gilles Pisier.
Probabilistic methods in the geometry of Banach spaces.
In Giorgio Letta and Maurizio Pratelli, editors, *Probability and Analysis*, pages 167–241. Springer, 1986.

Ronitt Rubinfeld and Madhu Sudan.
Robust characterization of polynomials with applications to program testing.
SIAM Journal on Computing, 25(2):252–271, 1996.

Gregory Valiant and Paul Valiant.
An automatic inequality prover and instance optimal identity testing.
In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '14, pages 51–60, Washington, DC, USA, 2014. IEEE Computer Society.