Frontiers in Distribution Testing: A Sample of What to Expect

Too Early for Puns?

Clément Canonne

October 14, 2017

~~Columbia University~~ Stanford University

Background, Context, and Motivation

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.

Property Testing

Figure 1: Property Testing: Inside the yolk, or outside the egg.

Property Testing

Introduced by [RS96, GGR98] – 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

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

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

Many variants, subareas, with a plethora of results (see e.g. [Ron08, Ron10, Gol10, Gol17, BY17]).

Distribution Testing

Now, our "big object" is a probability distribution over a (discrete*) domain Ω (e.g., $\Omega = [n]$).

Distribution Testing

Now, our "big object" is a probability distribution over a (discrete*) domain Ω (e.g., $\Omega = [n]$).

- instead of queries: samples*

Distribution Testing

Now, our "big object" is a probability distribution over a (discrete*) domain Ω (e.g., $\Omega = [n]$).

- instead of queries: samples*
- instead of Hamming distance: total variation*

Distribution Testing

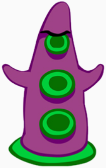
Now, our "big object" is a probability distribution over a (discrete*) domain Ω (e.g., $\Omega = [n]$).

- instead of queries: samples*
- instead of Hamming distance: total variation*
- instead of functions/graphs/strings: distributions

Distribution Testing

Now, our "big object" is a probability distribution over a (discrete*) domain Ω (e.g., $\Omega = [n]$).

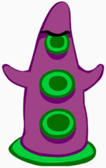
- instead of queries: samples*
- instead of Hamming distance: total variation*
- instead of functions/graphs/strings: distributions



Distribution Testing

Now, our "big object" is a probability distribution over a (discrete*) domain Ω (e.g., $\Omega = [n]$).

- instead of queries: samples*
- instead of Hamming distance: total variation*
- instead of functions/graphs/strings: distributions

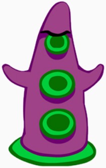


Focus on the sample complexity, with efficiency as ancillary goal.

Distribution Testing

Now, our "big object" is a probability distribution over a (discrete*) domain Ω (e.g., $\Omega = [n]$).

- instead of queries: samples*
- instead of Hamming distance: total variation*
- instead of functions/graphs/strings: distributions



Focus on the sample complexity, with efficiency as ancillary goal.

*usually.

Background

Over the past 15+ years, many results on many properties:

Caveat: The above is not *entirely* accurate, and only the (usually) dominant term is included. For instance, the sample complexity of equivalence is actually $\Theta(\max(n^{2/3}/\varepsilon^{4/3}, \sqrt{n}/\varepsilon^2))$; for monotonicity, the current best upper bound has an additional $1/\varepsilon^4$ term, while for PBDs the lower bound of $\Omega(n^{1/4}/\varepsilon^2)$ is almost matched by an $O(n^{1/4}/\varepsilon^2 + \log^2(1/\varepsilon)/\varepsilon^2)$ upper bound. Don't sue me.

Background

Over the past 15+ years, many results on many properties:

- Uniformity: $\Theta(\sqrt{n}/\varepsilon^2)$ [GR00, BFR+00, Pan08, DGPP16]

Caveat: The above is not *entirely* accurate, and only the (usually) dominant term is included. For instance, the sample complexity of equivalence is actually $\Theta(\max(n^{2/3}/\varepsilon^{4/3}, \sqrt{n}/\varepsilon^2))$; for monotonicity, the current best upper bound has an additional $1/\varepsilon^4$ term, while for PBDs the lower bound of $\Omega(n^{1/4}/\varepsilon^2)$ is almost matched by an $O(n^{1/4}/\varepsilon^2 + \log^2(1/\varepsilon)/\varepsilon^2)$ upper bound. Don't sue me.

Background

Over the past 15+ years, many results on many properties:

- Uniformity: $\Theta(\sqrt{n}/\varepsilon^2)$ [GR00, BFR+00, Pan08, DGPP16]
- Identity: $\Theta(\sqrt{n}/\varepsilon^2)$, $\Phi(p, \Theta(\varepsilon))$ [BFF+01, VV14, DKN15, BCG17]

Caveat: The above is not *entirely* accurate, and only the (usually) dominant term is included. For instance, the sample complexity of equivalence is actually $\Theta(\max(n^{2/3}/\varepsilon^{4/3}, \sqrt{n}/\varepsilon^2))$; for monotonicity, the current best upper bound has an additional $1/\varepsilon^4$ term, while for PBDs the lower bound of $\Omega(n^{1/4}/\varepsilon^2)$ is almost matched by an $O(n^{1/4}/\varepsilon^2 + \log^2(1/\varepsilon)/\varepsilon^2)$ upper bound. Don't sue me.

Background

Over the past 15+ years, many results on many properties:

- Uniformity: $\Theta(\sqrt{n}/\varepsilon^2)$ [GR00, BFR+00, Pan08, DGPP16]
- Identity: $\Theta(\sqrt{n}/\varepsilon^2)$, $\Phi(p, \Theta(\varepsilon))$ [BFF+01, VV14, DKN15, BCG17]
- Equivalence: $\Theta(n^{2/3}/\varepsilon^{4/3})$ [BFR+00, Val11, CDVV14, DK16]

Caveat: The above is not *entirely* accurate, and only the (usually) dominant term is included. For instance, the sample complexity of equivalence is actually $\Theta(\max(n^{2/3}/\varepsilon^{4/3}, \sqrt{n}/\varepsilon^2))$; for monotonicity, the current best upper bound has an additional $1/\varepsilon^4$ term, while for PBDs the lower bound of $\Omega(n^{1/4}/\varepsilon^2)$ is almost matched by an $O(n^{1/4}/\varepsilon^2 + \log^2(1/\varepsilon)/\varepsilon^2)$ upper bound. Don't sue me.

Background

Over the past 15+ years, many results on many properties:

- Uniformity: $\Theta(\sqrt{n}/\varepsilon^2)$ [GR00, BFR+00, Pan08, DGPP16]
- Identity: $\Theta(\sqrt{n}/\varepsilon^2)$, $\Phi(p, \Theta(\varepsilon))$ [BFF+01, VV14, DKN15, BCG17]
- Equivalence: $\Theta(n^{2/3}/\varepsilon^{4/3})$ [BFR+00, Val11, CDVV14, DK16]
- Independence: $\Theta(m^{2/3}n^{1/3}/\varepsilon^{4/3})$ [BFF+01, LRR13, DK16]

Caveat: The above is not *entirely* accurate, and only the (usually) dominant term is included. For instance, the sample complexity of equivalence is actually $\Theta(\max(n^{2/3}/\varepsilon^{4/3}, \sqrt{n}/\varepsilon^2))$; for monotonicity, the current best upper bound has an additional $1/\varepsilon^4$ term, while for PBDs the lower bound of $\Omega(n^{1/4}/\varepsilon^2)$ is almost matched by an $O(n^{1/4}/\varepsilon^2 + \log^2(1/\varepsilon)/\varepsilon^2)$ upper bound. Don't sue me.

Background

Over the past 15+ years, many results on many properties:

- Uniformity: $\Theta(\sqrt{n}/\varepsilon^2)$ [GR00, BFR+00, Pan08, DGPP16]
- Identity: $\Theta(\sqrt{n}/\varepsilon^2)$, $\Phi(p, \Theta(\varepsilon))$ [BFF+01, VV14, DKN15, BCG17]
- Equivalence: $\Theta(n^{2/3}/\varepsilon^{4/3})$ [BFR+00, Val11, CDVV14, DK16]
- Independence: $\Theta(m^{2/3}n^{1/3}/\varepsilon^{4/3})$ [BFF+01, LRR13, DK16]
- Monotonicity: $\Theta(\sqrt{n}/\varepsilon^2)$ [BKR04, BFRV11, ADK15]

Caveat: The above is not *entirely* accurate, and only the (usually) dominant term is included. For instance, the sample complexity of equivalence is actually $\Theta(\max(n^{2/3}/\varepsilon^{4/3}, \sqrt{n}/\varepsilon^2))$; for monotonicity, the current best upper bound has an additional $1/\varepsilon^4$ term, while for PBDs the lower bound of $\Omega(n^{1/4}/\varepsilon^2)$ is almost matched by an $O(n^{1/4}/\varepsilon^2 + \log^2(1/\varepsilon)/\varepsilon^2)$ upper bound. Don't sue me.

Background

Over the past 15+ years, many results on many properties:

- Uniformity: $\Theta(\sqrt{n}/\varepsilon^2)$ [GR00, BFR+00, Pan08, DGPP16]
- Identity: $\Theta(\sqrt{n}/\varepsilon^2)$, $\Phi(p, \Theta(\varepsilon))$ [BFF+01, VV14, DKN15, BCG17]
- Equivalence: $\Theta(n^{2/3}/\varepsilon^{4/3})$ [BFR+00, Val11, CDVV14, DK16]
- Independence: $\Theta(m^{2/3}n^{1/3}/\varepsilon^{4/3})$ [BFF+01, LRR13, DK16]
- Monotonicity: $\Theta(\sqrt{n}/\varepsilon^2)$ [BKR04, BFRV11, ADK15]
- Poisson Binomial Distributions: $\tilde{\Theta}(n^{1/4}/\varepsilon^2)$ [AD15, CDGR16, CDS17]

Caveat: The above is not *entirely* accurate, and only the (usually) dominant term is included. For instance, the sample complexity of equivalence is actually $\Theta(\max(n^{2/3}/\varepsilon^{4/3}, \sqrt{n}/\varepsilon^2))$; for monotonicity, the current best upper bound has an additional $1/\varepsilon^4$ term, while for PBDs the lower bound of $\Omega(n^{1/4}/\varepsilon^2)$ is almost matched by an $O(n^{1/4}/\varepsilon^2 + \log^2(1/\varepsilon)/\varepsilon^2)$ upper bound. Don't sue me.

Background

Over the past 15+ years, many results on many properties:

- Uniformity: $\Theta(\sqrt{n}/\varepsilon^2)$ [GR00, BFR+00, Pan08, DGPP16]
- Identity: $\Theta(\sqrt{n}/\varepsilon^2)$, $\Phi(p, \Theta(\varepsilon))$ [BFF+01, VV14, DKN15, BCG17]
- Equivalence: $\Theta(n^{2/3}/\varepsilon^{4/3})$ [BFR+00, Val11, CDVV14, DK16]
- Independence: $\Theta(m^{2/3}n^{1/3}/\varepsilon^{4/3})$ [BFF+01, LRR13, DK16]
- Monotonicity: $\Theta(\sqrt{n}/\varepsilon^2)$ [BKR04, BFRV11, ADK15]
- Poisson Binomial Distributions: $\tilde{\Theta}(n^{1/4}/\varepsilon^2)$ [AD15, CDGR16, CDS17]
- histograms, MHR, log-concavity, k-wise independence, SIIRV, PMD, clusterability, juntas... *and it goes on.* [Rub12, Can15]

Caveat: The above is not *entirely* accurate, and only the (usually) dominant term is included. For instance, the sample complexity of equivalence is actually $\Theta(\max(n^{2/3}/\varepsilon^{4/3}, \sqrt{n}/\varepsilon^2))$; for monotonicity, the current best upper bound has an additional $1/\varepsilon^4$ term, while for PBDs the lower bound of $\Omega(n^{1/4}/\varepsilon^2)$ is almost matched by an $O(n^{1/4}/\varepsilon^2 + \log^2(1/\varepsilon)/\varepsilon^2)$ upper bound. Don't sue me.

Background

Over the past 15+ years, many results on many properties:

- Uniformity: $\Theta(\sqrt{n}/\varepsilon^2)$ [GR00, BFR+00, Pan08, DGPP16]
- Identity: $\Theta(\sqrt{n}/\varepsilon^2)$, $\Phi(p, \Theta(\varepsilon))$ [BFF+01, VV14, DKN15, BCG17]
- Equivalence: $\Theta(n^{2/3}/\varepsilon^{4/3})$ [BFR+00, Val11, CDVV14, DK16]
- Independence: $\Theta(m^{2/3}n^{1/3}/\varepsilon^{4/3})$ [BFF+01, LRR13, DK16]
- Monotonicity: $\Theta(\sqrt{n}/\varepsilon^2)$ [BKR04, BFRV11, ADK15]
- Poisson Binomial Distributions: $\tilde{\Theta}(n^{1/4}/\varepsilon^2)$ [AD15, CDGR16, CDS17]
- histograms, MHR, log-concavity, k-wise independence, SIIRV, PMD, clusterability, juntas... *and it goes on.* [Rub12, Can15]

So much has been done;

Caveat: The above is not *entirely* accurate, and only the (usually) dominant term is included. For instance, the sample complexity of equivalence is actually $\Theta(\max(n^{2/3}/\varepsilon^{4/3}, \sqrt{n}/\varepsilon^2))$; for monotonicity, the current best upper bound has an additional $1/\varepsilon^4$ term, while for PBDs the lower bound of $\Omega(n^{1/4}/\varepsilon^2)$ is almost matched by an $O(n^{1/4}/\varepsilon^2 + \log^2(1/\varepsilon)/\varepsilon^2)$ upper bound. Don't sue me.

Background

Over the past 15+ years, many results on many properties:

- Uniformity: $\Theta(\sqrt{n}/\varepsilon^2)$ [GR00, BFR+00, Pan08, DGPP16]
- Identity: $\Theta(\sqrt{n}/\varepsilon^2)$, $\Phi(p, \Theta(\varepsilon))$ [BFF+01, VV14, DKN15, BCG17]
- Equivalence: $\Theta(n^{2/3}/\varepsilon^{4/3})$ [BFR+00, Val11, CDVV14, DK16]
- Independence: $\Theta(m^{2/3}n^{1/3}/\varepsilon^{4/3})$ [BFF+01, LRR13, DK16]
- Monotonicity: $\Theta(\sqrt{n}/\varepsilon^2)$ [BKR04, BFRV11, ADK15]
- Poisson Binomial Distributions: $\tilde{\Theta}(n^{1/4}/\varepsilon^2)$ [AD15, CDGR16, CDS17]
- histograms, MHR, log-concavity, k-wise independence, SIIRV, PMD, clusterability, juntas... *and it goes on.* [Rub12, Can15]

So much has been done; and yet so much remains...

Caveat: The above is not *entirely* accurate, and only the (usually) dominant term is included. For instance, the sample complexity of equivalence is actually $\Theta(\max(n^{2/3}/\varepsilon^{4/3}, \sqrt{n}/\varepsilon^2))$; for monotonicity, the current best upper bound has an additional $1/\varepsilon^4$ term, while for PBDs the lower bound of $\Omega(n^{1/4}/\varepsilon^2)$ is almost matched by an $O(n^{1/4}/\varepsilon^2 + \log^2(1/\varepsilon)/\varepsilon^2)$ upper bound. Don't sue me.

Background

Over the past 15+ years, many results on many properties:

- Uniformity: $\Theta(\sqrt{n}/\varepsilon^2)$ [GR00, BFR+00, Pan08, DGPP16]
- Identity: $\Theta(\sqrt{n}/\varepsilon^2)$, $\Phi(p, \Theta(\varepsilon))$ [BFF+01, VV14, DKN15, BCG17]
- Equivalence: $\Theta(n^{2/3}/\varepsilon^{4/3})$ [BFR+00, Val11, CDVV14, DK16]
- Independence: $\Theta(m^{2/3}n^{1/3}/\varepsilon^{4/3})$ [BFF+01, LRR13, DK16]
- Monotonicity: $\Theta(\sqrt{n}/\varepsilon^2)$ [BKR04, BFRV11, ADK15]
- Poisson Binomial Distributions: $\tilde{\Theta}(n^{1/4}/\varepsilon^2)$ [AD15, CDGR16, CDS17]
- histograms, MHR, log-concavity, k-wise independence, SIIRV, PMD, clusterability, juntas... *and it goes on.* [Rub12, Can15]

So much has been done; and yet so much remains...

Caveat: The above is not *entirely* accurate, and only the (usually) dominant term is included. For instance, the sample complexity of equivalence is actually $\Theta(\max(n^{2/3}/\varepsilon^{4/3}, \sqrt{n}/\varepsilon^2))$; for monotonicity, the current best upper bound has an additional $1/\varepsilon^4$ term, while for PBDs the lower bound of $\Omega(n^{1/4}/\varepsilon^2)$ is almost matched by an $O(n^{1/4}/\varepsilon^2 + \log^2(1/\varepsilon)/\varepsilon^2)$ upper bound. Don't sue me.

Many questions remain

Techniques

Most algorithms, results are somewhat ad hoc, and property-specific.

Many questions remain

Techniques

Most algorithms, results are somewhat ad hoc, and property-specific.

Hardness

Most properties are depressingly hard to test: $\Omega(\sqrt{n})$ samples are required.

Many questions remain

Techniques

Most algorithms, results are somewhat ad hoc, and property-specific.

Hardness

Most properties are depressingly hard to test: $\Omega(\sqrt{n})$ samples are required.

Tolerance and estimation

Testing is good; but what about tolerant testing and functional estimation?

Many questions remain

Techniques

Most algorithms, results are somewhat ad hoc, and property-specific.

Hardness

Most properties are depressingly hard to test: $\Omega(\sqrt{n})$ samples are required.

Tolerance and estimation

Testing is good; but what about tolerant testing and functional estimation?

Beyond?

Only a preliminary step! What if…

Some Notation

Glossary

- Probability distributions over discrete Ω (e.g. $[n] := \{1, \ldots, n\}$)

$$\Delta([n]) = \left\{ p \colon \Omega \to [0,1] \ : \ \sum_{i \in \Omega} p(i) = 1 \right\}$$

Glossary

- Probability distributions over discrete Ω (e.g. $[n] := \{1, \ldots, n\}$)

$$\Delta([n]) = \left\{ p \colon \Omega \to [0,1] \, : \, \sum_{i \in \Omega} p(i) = 1 \right\}$$

- Property (or class) of distributions over Ω:

$$\mathcal{P} \subseteq \Delta(\Omega)$$

Glossary

- Probability distributions over discrete Ω (e.g. $[n] := \{1, \ldots, n\}$)

$$\Delta([n]) = \left\{ p \colon \Omega \to [0,1] \,:\, \sum_{i \in \Omega} p(i) = 1 \right\}$$

- Property (or class) of distributions over Ω:

$$\mathcal{P} \subseteq \Delta(\Omega)$$

- Total variation distance (statistical distance, ℓ_1 distance):

$$\mathrm{d}_{\mathrm{TV}}(p, q) = \sup_{S \subseteq \Omega} (p(S) - q(S)) = \frac{1}{2} \sum_{x \in \Omega} |p(x) - q(x)| \in [0, 1]$$

Glossary

- Probability distributions over discrete Ω (e.g. $[n] := \{1, \ldots, n\}$)

$$\Delta([n]) = \left\{ p\colon \Omega \to [0,1] \;:\; \sum_{i \in \Omega} p(i) = 1 \right\}$$

- Property (or class) of distributions over Ω:

$$\mathcal{P} \subseteq \Delta(\Omega)$$

- Total variation distance (statistical distance, ℓ_1 distance):

$$\mathrm{d}_{\mathrm{TV}}(p,q) = \sup_{S \subseteq \Omega}(p(S) - q(S)) = \frac{1}{2}\sum_{x \in \Omega} |p(x) - q(x)| \in [0,1]$$

Domain size/parameter $n \in \mathbb{N}$ is big ("goes to ∞").

Glossary

- Probability distributions over discrete Ω (e.g. $[n] := \{1, \dots, n\}$)

$$\Delta([n]) = \left\{ p \colon \Omega \to [0,1] \;:\; \sum_{i \in \Omega} p(i) = 1 \right\}$$

- Property (or class) of distributions over Ω:

$$\mathcal{P} \subseteq \Delta(\Omega)$$

- Total variation distance (statistical distance, ℓ_1 distance):

$$\mathrm{d}_{\mathrm{TV}}(p,q) = \sup_{S \subseteq \Omega} (p(S) - q(S)) = \frac{1}{2} \sum_{x \in \Omega} |p(x) - q(x)| \in [0,1]$$

Domain size/parameter $n \in \mathbb{N}$ is big ("goes to ∞"). Proximity parameter $\varepsilon \in (0,1]$ is small.

Glossary

- Probability distributions over discrete Ω (e.g. $[n] := \{1, \dots, n\}$)

$$\Delta([n]) = \left\{ p\colon \Omega \to [0,1] \;:\; \sum_{i\in\Omega} p(i) = 1 \right\}$$

- Property (or class) of distributions over Ω:

$$\mathcal{P} \subseteq \Delta(\Omega)$$

- Total variation distance (statistical distance, ℓ_1 distance):

$$\mathrm{d}_{\mathrm{TV}}(p,q) = \sup_{S\subseteq\Omega}(p(S) - q(S)) = \frac{1}{2}\sum_{x\in\Omega}|p(x) - q(x)| \in [0,1]$$

Domain size/parameter $n \in \mathbb{N}$ is big ("goes to ∞"). Proximity parameter $\varepsilon \in (0,1]$ is small. Lowercase Greek letters are in $(0,1]$.

Glossary

- Probability distributions over discrete Ω (e.g. $[n] := \{1, \dots, n\}$)

$$\Delta([n]) = \left\{ p\colon \Omega \to [0,1] \;:\; \sum_{i\in\Omega} p(i) = 1 \right\}$$

- Property (or class) of distributions over Ω:

$$\mathcal{P} \subseteq \Delta(\Omega)$$

- Total variation distance (statistical distance, ℓ_1 distance):

$$\mathrm{d}_{\mathrm{TV}}(p,q) = \sup_{S\subseteq\Omega}(p(S) - q(S)) = \frac{1}{2}\sum_{x\in\Omega} |p(x) - q(x)| \in [0,1]$$

Domain size/parameter $n \in \mathbb{N}$ is big ("goes to ∞"). Proximity parameter $\varepsilon \in (0,1]$ is small. Lowercase Greek letters are in $(0,1]$. Asymptotics $\tilde{O}$, $\tilde{\Omega}$, $\tilde{\Theta}$ hide logarithmic factors.*

General Approaches, Unified Paradigms, and Many-Birded Stones

Testing By Learning

Trivial baseline in property testing: *"you can learn, so you can test."*

Testing By Learning

Trivial baseline in property testing: "*you can learn, so you can test.*"

(i) Learn p without assumptions using a learner for $\Delta(\Omega)$

Testing By Learning

Trivial baseline in property testing: *"you can learn, so you can test."*

(i) Learn p without assumptions using a learner for $\Delta(\Omega)$

(ii) Check if $d_{\mathrm{TV}}(\hat{p}, \mathcal{P}) \leq \frac{\varepsilon}{3}$ (Computational)

Testing By Learning

Trivial baseline in property testing: *"you can learn, so you can test."*

(i) Learn p without assumptions using a learner for $\Delta(\Omega)$

(ii) Check if $\mathrm{d_{TV}}(\hat{p}, \mathcal{P}) \leq \frac{\varepsilon}{3}$ (Computational)

Yes, but...

(i) has sample complexity $\Theta(n/\varepsilon^2)$.

Testing By Learning

"Folklore" baseline in property testing: *"if you can learn, you can test."*

Testing By Learning

"Folklore" baseline in property testing: *"if you can learn, you can test."*

(i) Learn p as if $p \in \mathcal{P}$ using a learner for $\mathcal{P}$

Testing By Learning

"Folklore" baseline in property testing: *"if you can learn, you can test."*

(i) Learn p as if $p \in \mathcal{P}$ using a learner for $\mathcal{P}$

(ii) Test $\mathrm{d}_{\mathrm{TV}}(\hat{p}, p) \leq \frac{\varepsilon}{3}$ vs. $\mathrm{d}_{\mathrm{TV}}(\hat{p}, p) \geq \frac{2\varepsilon}{3}$

Testing By Learning

"Folklore" baseline in property testing: *"if you can learn, you can test."*

(i) Learn p as if $p \in \mathcal{P}$ using a learner for $\mathcal{P}$

(ii) Test $\mathrm{d}_{\mathrm{TV}}(\hat{p}, p) \leq \frac{\varepsilon}{3}$ vs. $\mathrm{d}_{\mathrm{TV}}(\hat{p}, p) \geq \frac{2\varepsilon}{3}$

(iii) Check if $\mathrm{d}_{\mathrm{TV}}(\hat{p}, \mathcal{P}) \leq \frac{\varepsilon}{3}$ (Computational)

Testing By Learning

"Folklore" baseline in property testing: *"if you can learn, you can test."*

(i) Learn p as if $p \in \mathcal{P}$ using a learner for $\mathcal{P}$

(ii) Test $\mathrm{d}_{\mathrm{TV}}(\hat{p}, p) \leq \frac{\varepsilon}{3}$ vs. $\mathrm{d}_{\mathrm{TV}}(\hat{p}, p) \geq \frac{2\varepsilon}{3}$

(iii) Check if $\mathrm{d}_{\mathrm{TV}}(\hat{p}, \mathcal{P}) \leq \frac{\varepsilon}{3}$ (Computational)

The triangle inequality does the rest.

Testing By Learning?

"Folklore" baseline in property testing: *"if you can learn, you can test."*

(i) Learn p as if $p \in \mathcal{P}$ using a learner for $\mathcal{P}$

(ii) Test if $\mathrm{d}_{\mathrm{TV}}(\hat{p}, p) \leq \frac{\varepsilon}{3}$ vs. $\mathrm{d}_{\mathrm{TV}}(\hat{p}, p) \geq \frac{2\varepsilon}{3}$

(iii) Check if $\mathrm{d}_{\mathrm{TV}}(\hat{p}, \mathcal{P}) \leq \frac{\varepsilon}{3}$ (Computational)

Not quite.

(ii) fine for functions. But for distributions? Requires $\Omega(\frac{n}{\log n})$ samples [VV11a, JYW17]

Testing By Learning!

All is doomed, there is no hope, and every dream ends up shattered on this unforgiving Earth.

Testing By Learning!

All is doomed, there is no hope, and every dream ends up shattered on this unforgiving Earth. Although...

(i) Learn p as if $p \in \mathcal{P}$ using a learner for $\mathcal{P}$ in χ^2 distance

Testing By Learning!

All is doomed, there is no hope, and every dream ends up shattered on this unforgiving Earth. Although...

(i) Learn p as if $p \in \mathcal{P}$ using a learner for $\mathcal{P}$ in χ^2 distance

(ii) Test if $\chi^2(\hat{p} \,||\, p) \leq \varepsilon^2$ vs. $d_{\mathrm{TV}}(\hat{p}, p) \geq \frac{2\varepsilon}{3}$

Testing By Learning!

All is doomed, there is no hope, and every dream ends up shattered on this unforgiving Earth. Although...

(i) Learn p as if $p \in \mathcal{P}$ using a learner for $\mathcal{P}$ in χ^2 distance

(ii) Test if $\chi^2(\hat{p} \,||\, p) \leq \varepsilon^2$ vs. $\mathrm{d}_{\mathrm{TV}}(\hat{p}, p) \geq \frac{2\varepsilon}{3}$

(iii) Check if $\mathrm{d}_{\mathrm{TV}}(\hat{p}, \mathcal{P}) \leq \frac{\varepsilon}{3}$ (Computational)

Testing By Learning!

All is doomed, there is no hope, and every dream ends up shattered on this unforgiving Earth. Although…

(i) Learn p as if $p \in \mathcal{P}$ using a learner for $\mathcal{P}$ in χ^2 distance

(ii) Test if $\chi^2(\hat{p} \,||\, p) \leq \varepsilon^2$ vs. $\mathrm{d}_{\mathrm{TV}}(\hat{p}, p) \geq \frac{2\varepsilon}{3}$

(iii) Check if $\mathrm{d}_{\mathrm{TV}}(\hat{p}, \mathcal{P}) \leq \frac{\varepsilon}{3}$ (Computational)

Success.

Acharya, Daskalakis, and Kamath [ADK15]: now (i) is harder, but (ii) becomes cheap!

Testing By Learning!

All is not doomed, there is some hope, and not every dream ends up shattered on this unforgiving Earth.

Testing By Learning!

All is not doomed, there is some hope, and not every dream ends up shattered on this unforgiving Earth. And...

(i) Test that p satisfies a strong structural guarantee of $\mathcal{P}$: succinct approximation by histograms ("shape restrictions")

Testing By Learning!

All is not doomed, there is some hope, and not every dream ends up shattered on this unforgiving Earth. And...

(i) Test that p satisfies a strong structural guarantee of $\mathcal{P}$: succinct approximation by histograms ("shape restrictions")

(ii) Learn p efficiently (in a weird KL/ℓ_2 sense) using this structure

Testing By Learning!

All is not doomed, there is some hope, and not every dream ends up shattered on this unforgiving Earth. And...

(i) Test that p satisfies a strong structural guarantee of $\mathcal{P}$: succinct approximation by histograms ("shape restrictions")

(ii) Learn p efficiently (in a weird KL/ℓ_2 sense) using this structure

(iii) Check if $\mathrm{d}_{\mathrm{TV}}(\hat{p}, \mathcal{P}) \leq \frac{\varepsilon}{3}$ (Computational)

Testing By Learning!

All is not doomed, there is some hope, and not every dream ends up shattered on this unforgiving Earth. And...

(i) Test that p satisfies a strong structural guarantee of $\mathcal{P}$: succinct approximation by histograms ("shape restrictions")

(ii) Learn p efficiently (in a weird KL/ℓ_2 sense) using this structure

(iii) Check if $\mathrm{d}_{\mathrm{TV}}(\hat{p}, \mathcal{P}) \leq \frac{\varepsilon}{3}$ (Computational)

Success.

Canonne, Diakonikolas, Gouleakis, and Rubinfeld [CDGR16]: now $\mathrm{d}_{\mathrm{TV}}(\hat{p}, p) \leq O(\varepsilon)$ comes for free!

Testing By Learning!

All is hope, there is no doom, and every dream ends up bright and shiny on this wonderful Earth.

Testing By Learning!

All is hope, there is no doom, and every dream ends up bright and shiny on this wonderful Earth. And...

(i) Test that p satisfies a strong structural guarantee of $\mathcal{P}$: nice discrete Fourier transform (Fourier sparsity)

Testing By Learning!

All is hope, there is no doom, and every dream ends up bright and shiny on this wonderful Earth. And...

(i) Test that p satisfies a strong structural guarantee of $\mathcal{P}$: nice discrete Fourier transform (Fourier sparsity)

(ii) Learn p efficiently (in ℓ_2 sense) using this structure

Testing By Learning!

All is hope, there is no doom, and every dream ends up bright and shiny on this wonderful Earth. And...

(i) Test that p satisfies a strong structural guarantee of $\mathcal{P}$: nice discrete Fourier transform (Fourier sparsity)

(ii) Learn p efficiently (in ℓ_2 sense) using this structure

(iii) Check if $d_{\mathrm{TV}}(\hat{p}, \mathcal{P}) \leq \frac{\varepsilon}{3}$ (Computational)

Testing By Learning!

All is hope, there is no doom, and every dream ends up bright and shiny on this wonderful Earth. And…

(i) Test that p satisfies a strong structural guarantee of $\mathcal{P}$: nice discrete Fourier transform (Fourier sparsity)

(ii) Learn p efficiently (in ℓ_2 sense) using this structure

(iii) Check if $\mathrm{d}_{\mathrm{TV}}(\hat{p}, \mathcal{P}) \leq \frac{\varepsilon}{3}$ (Computational)

Success.

Canonne, Diakonikolas, and Stewart [CDS17]: "all your (Fourier) basis are belong to…"

Testing in TV via ℓ_2

Testing in ℓ_2 distance is well-understood [CDVV14]; testing in TV ℓ_1) is trickier.

Testing in TV via ℓ_2

Testing in ℓ_2 distance is well-understood [CDVV14]; testing in TV ℓ_1) is trickier. Can we reduce one to the other?

Testing in TV via ℓ_2

Testing in ℓ_2 distance is well-understood [CDVV14]; testing in TV ℓ_1) is trickier. Can we reduce one to the other?

(i) Map $p \in \Delta([n])$ to a "nicer, smoother" $p' \in \Delta([O(n)])$

Testing in TV via ℓ_2

Testing in ℓ_2 distance is well-understood [CDVV14]; testing in TV ℓ_1) is trickier. Can we reduce one to the other?

(i) Map $p \in \Delta([n])$ to a "nicer, smoother" $p' \in \Delta([O(n)])$

(ii) Test p' using an ℓ_2 tester

Testing in TV via ℓ_2

Testing in ℓ_2 distance is well-understood [CDVV14]; testing in TV ℓ_1) is trickier. Can we reduce one to the other?

(i) Map $p \in \Delta([n])$ to a "nicer, smoother" $p' \in \Delta([O(n)])$

(ii) Test p' using an ℓ_2 tester

(iii) That's all.

Testing in TV via ℓ_2

Testing in ℓ_2 distance is well-understood [CDVV14]; testing in TV ℓ_1) is trickier. Can we reduce one to the other?

(i) Map $p \in \Delta([n])$ to a "nicer, smoother" $p' \in \Delta([O(n)])$

(ii) Test p' using an ℓ_2 tester

(iii) That's all.

Success.

Diakonikolas and Kane [DK16]: "It works."

Tolerant Testing and Estimation

Theorem (Everything is $\frac{n}{\log n}$)

Pretty much every tolerant testing question or functional estimation (entropy, support size, ...) has sample complexity $\Theta_\varepsilon(\frac{n}{\log n})$.

Technically, and as Jiantao's talk will describe: a more accurate description is that *whatever estimation can be performed in $k \log k$ samples via the plug-in empirical estimator, the optimal scheme does with k*. "Enlarge your sample," if you will.

Tolerant Testing and Estimation

Theorem (Everything is $\frac{n}{\log n}$)

Pretty much every tolerant testing question or functional estimation (entropy, support size, ...) has sample complexity $\Theta_\varepsilon(\frac{n}{\log n})$.

Technically, and as Jiantao's talk will describe: a more accurate description is that *whatever estimation can be performed in $k \log k$ samples via the plug-in empirical estimator, the optimal scheme does with k*. "Enlarge your sample," if you will.

Tolerant Testing and Estimation

- Paul Valiant [Val11]: the canonical tester for symmetric properties (not quite, but near-optimal)

Tolerant Testing and Estimation

- Paul Valiant [Val11]: the canonical tester for symmetric properties (not quite, but near-optimal)
- Valiant–Valiant [VV11a]: learn the histogram with $O(\frac{n}{\varepsilon^2 \log n})$ samples, then plug in – and we're done

Tolerant Testing and Estimation

- Paul Valiant [Val11]: the canonical tester for symmetric properties (not quite, but near-optimal)
- Valiant–Valiant [VV11a]: learn the histogram with $O(\frac{n}{\varepsilon^2 \log n})$ samples, then plug in – and we're done
- Valiant–Valiant [VV11b]: actually, can even do it with a linear estimator

Tolerant Testing and Estimation

- Paul Valiant [Val11]: the canonical tester for symmetric properties (not quite, but near-optimal)
- Valiant–Valiant [VV11a]: learn the histogram with $O(\frac{n}{\varepsilon^2 \log n})$ samples, then plug in – and we're done
- Valiant–Valiant [VV11b]: actually, can even do it with a linear estimator
- Acharya, Das, Orlitsky, Suresh [ADOS17]: actually, the (Profile) Maximum Likelihood Estimator (PMLE) does it

Tolerant Testing and Estimation

- Paul Valiant [Val11]: the canonical tester for symmetric properties (not quite, but near-optimal)
- Valiant–Valiant [VV11a]: learn the histogram with $O(\frac{n}{\varepsilon^2 \log n})$ samples, then plug in – and we're done
- Valiant–Valiant [VV11b]: actually, can even do it with a linear estimator
- Acharya, Das, Orlitsky, Suresh [ADOS17]: actually, the (Profile) Maximum Likelihood Estimator (PMLE) does it
- Jiao et al. [JVHW15], Wu and Yang [WY16]: actually, best polynomial approximation is the tool for the job

Tolerant Testing and Estimation

- Paul Valiant [Val11]: the canonical tester for symmetric properties (not quite, but near-optimal)
- Valiant–Valiant [VV11a]: learn the histogram with $O(\frac{n}{\varepsilon^2 \log n})$ samples, then plug in – and we're done
- Valiant–Valiant [VV11b]: actually, can even do it with a linear estimator
- Acharya, Das, Orlitsky, Suresh [ADOS17]: actually, the (Profile) Maximum Likelihood Estimator (PMLE) does it
- Jiao et al. [JVHW15], Wu and Yang [WY16]: actually, best polynomial approximation is the tool for the job
- Han, Jiao, and Weissman [HJW17]: actually, moment-matching is *also* the tool for the job

General Approaches To Sadness, Too

Unified algorithms and techniques for upper bounds are nice, but what about this feeling of despair in the face of impossibility?

General Approaches To Sadness, Too

- Paul Valiant [Val11]: lower bounds for symmetric properties via moment-matching: "Wishful Thinking Theorem."

General Approaches To Sadness, Too

- Paul Valiant [Val11]: lower bounds for symmetric properties via moment-matching: “Wishful Thinking Theorem.”
- Valiant-Valiant [VV14]: blackbox statement for Le Cam’s two point method

General Approaches To Sadness, Too

- Paul Valiant [Val11]: lower bounds for symmetric properties via moment-matching: "Wishful Thinking Theorem."
- Valiant-Valiant [VV14]: blackbox statement for Le Cam's two point method
- Diakonikolas and Kane [DK16]: information-theoretic framework to proving lower bounds via mutual information.

General Approaches To Sadness, Too

- Paul Valiant [Val11]: lower bounds for symmetric properties via moment-matching: "Wishful Thinking Theorem."
- Valiant-Valiant [VV14]: blackbox statement for Le Cam's two point method
- Diakonikolas and Kane [DK16]: information-theoretic framework to proving lower bounds via mutual information.
- Canonne, Diakonikolas, Gouleakis, and Rubinfeld [CDGR16]: lower bounds by reductions from (distribution testing+agnostic learning): "if you can learn, you *can't* test."

General Approaches To Sadness, Too

- Paul Valiant [Val11]: lower bounds for symmetric properties via moment-matching: "Wishful Thinking Theorem."
- Valiant-Valiant [VV14]: blackbox statement for Le Cam's two point method
- Diakonikolas and Kane [DK16]: information-theoretic framework to proving lower bounds via mutual information.
- Canonne, Diakonikolas, Gouleakis, and Rubinfeld [CDGR16]: lower bounds by reductions from (distribution testing+agnostic learning): "if you can learn, you *can't* test."
- Blais, Canonne, and Gur [BCG17]: lower bounds by reductions from communication complexity: "Alice and Bob say I can't test."

General Approaches To Sadness, Too

- Paul Valiant [Val11]: lower bounds for symmetric properties via moment-matching: "Wishful Thinking Theorem."
- Valiant-Valiant [VV14]: blackbox statement for Le Cam's two point method
- Diakonikolas and Kane [DK16]: information-theoretic framework to proving lower bounds via mutual information.
- Canonne, Diakonikolas, Gouleakis, and Rubinfeld [CDGR16]: lower bounds by reductions from (distribution testing+agnostic learning): "if you can learn, you *can't* test."
- Blais, Canonne, and Gur [BCG17]: lower bounds by reductions from communication complexity: "Alice and Bob say I can't test."
- Valiant–Valiant, Jiao et al., Wu and Yang: lower bounds for tolerant testing via best polynomial approximation (dual of the u.b.'s).

For More and Better on This...

Ilias Diakonikolas (USC)

Optimal Distribution Testing via Reductions

Ilias Diakonikolas (USC)

Optimal Distribution Testing via Reductions

Jiantao Jiao (Stanford University)

Three Approaches towards Optimal Property Estimation and Testing

Ilias Diakonikolas (USC)

Optimal Distribution Testing via Reductions

Jiantao Jiao (Stanford University)

Three Approaches towards Optimal Property Estimation and Testing

Alon Orlitsky (UCSD)

A Unified Maximum Likelihood Approach for Estimating Symmetric Distribution Properties

Ilias Diakonikolas (USC)

Optimal Distribution Testing via Reductions

Jiantao Jiao (Stanford University)

Three Approaches towards Optimal Property Estimation and Testing

Alon Orlitsky (UCSD)

A Unified Maximum Likelihood Approach for Estimating Symmetric Distribution Properties

Gautam Kamath (MIT)

Testing with Alternative Distances

The Curse of Dimensionality, and How to Deal with It

Costis Daskalakis (MIT)

High-Dimensional Distribution Testing

Now, Make It Quantum.

Ryan O'Donnell (CMU)

Distribution testing in the 21½$^{\text{th}}$ century

“Correct Me If I’m Wrong”

Ronitt Rubinfeld (MIT and Tel Aviv University)

Sampling Correctors

Samples are fun, but... Testing with Merlin?

Tom Gur (UC Berkeley)

Proofs of Proximity for Distribution Testing

Thank you.

Jayadev Acharya and Constantinos Daskalakis.
Testing Poisson Binomial Distributions.
In *Proceedings of SODA*, pages 1829–1840, 2015.

Jayadev Acharya, Constantinos Daskalakis, and Gautam C. Kamath.
Optimal Testing for Properties of Distributions.
In C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 3577–3598. Curran Associates, Inc., 2015.

Jayadev Acharya, Hirakendu Das, Alon Orlitsky, and Ananda Theertha Suresh.
A unified maximum likelihood approach for optimal distribution property estimation.
In *Proceedings of ICML*, 2017.

Eric Blais, Clément L. Canonne, and Tom Gur.
Distribution testing lower bounds via reductions from communication complexity.

In *Computational Complexity Conference*, volume 79 of *LIPIcs*, pages 28:1–28:40. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017.

Tuğkan Batu, Eldar Fischer, Lance Fortnow, Ravi Kumar, Ronitt Rubinfeld, and Patrick White.
Testing random variables for independence and identity.
In *Proceedings of FOCS*, pages 442–451, 2001.

Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White.
Testing that distributions are close.
In *Proceedings of FOCS*, pages 189–197, 2000.

Arnab Bhattacharyya, Eldar Fischer, Ronitt Rubinfeld, and Paul Valiant.
Testing monotonicity of distributions over general partial orders.
In *Proceedings of ITCS*, pages 239–252, 2011.

Tuğkan Batu, Ravi Kumar, and Ronitt Rubinfeld.

Sublinear algorithms for testing monotone and unimodal distributions.
In *Proceedings of STOC*, pages 381–390, New York, NY, USA, 2004. ACM.

Arnab Bhattacharyya and Yuichi Yoshida.
***Property Testing*.**
Forthcoming, 2017.

Clément L. Canonne.
A Survey on Distribution Testing: your data is Big. But is it Blue?
Electronic Colloquium on Computational Complexity (ECCC), 22:63, April 2015.

Clément L. Canonne, Ilias Diakonikolas, Themis Gouleakis, and Ronitt Rubinfeld.
Testing Shape Restrictions of Discrete Distributions.
In *Proceedings of STACS*, 2016.
See also [CDGR17] (full version).

Clément L. Canonne, Ilias Diakonikolas, Themis Gouleakis, and Ronitt Rubinfeld.
Testing shape restrictions of discrete distributions.
Theory of Computing Systems, pages 1–59, 2017.

Clément L. Canonne, Ilias Diakonikolas, and Alistair Stewart.
Fourier-based testing for families of distributions.
Electronic Colloquium on Computational Complexity (ECCC), 24:75, 2017.

Siu-on Chan, Ilias Diakonikolas, Gregory Valiant, and Paul Valiant.
Optimal algorithms for testing closeness of discrete distributions.
In *Proceedings of SODA*, pages 1193–1203, 2014.

Ilias Diakonikolas, Themis Gouleakis, John Peebles, and Eric Price.
Collision-based testers are optimal for uniformity and closeness.
Electronic Colloquium on Computational Complexity (ECCC), 23:178, 2016.

Ilias Diakonikolas and Daniel M. Kane.
A new approach for testing properties of discrete distributions.
In *Proceedings of FOCS*. IEEE Computer Society, 2016.

Ilias Diakonikolas, Daniel M. Kane, and Vladimir Nikishkin.
Testing Identity of Structured Distributions.
In *Proceedings of SODA*, 2015.

Oded Goldreich, Shafi Goldwasser, and Dana Ron.
Property testing and its connection to learning and approximation.
Journal of the ACM, 45(4):653–750, July 1998.

Oded Goldreich, editor.
***Property Testing: Current Research and Surveys*.**
Springer, 2010.
LNCS 6390.

Oded Goldreich.
***Introduction to Property Testing*.**

Forthcoming, 2017.

Oded Goldreich and Dana Ron.
On testing expansion in bounded-degree graphs.
Technical Report TR00-020, Electronic Colloquium on Computational Complexity (ECCC), 2000.

Yanjun Han, Jiantao Jiao, and Tsachy Weissman.
Local moment matching: a unified methodology for optimal functional estimation and distribution estimation under Wasserstein distance, 2017.

Jiantao Jiao, Kartik Venkat, Yanjun Han, and Tsachy Weissman.
Minimax estimation of functionals of discrete distributions.
IEEE Transactions on Information Theory, 61(5):2835–2885, May 2015.

Jiantao Jiao, Han Yanjun, and Tsachy Weissman.
Minimax Estimation of the L_1 Distance.
ArXiv e-prints, May 2017.

Reut Levi, Dana Ron, and Ronitt Rubinfeld.
Testing properties of collections of distributions.
Theory of Computing, 9:295–347, 2013.

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

Dana Ron.
Property Testing: A Learning Theory Perspective.
Foundations and Trends in Machine Learning, 1(3):307–402, 2008.

Dana Ron.
Algorithmic and analysis techniques in property testing.
Foundations and Trends in Theoretical Computer Science, 5:73–205, 2010.

Ronitt Rubinfeld and Madhu Sudan.

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

Ronitt Rubinfeld.
Taming big probability distributions.
XRDS: Crossroads, The ACM Magazine for Students, 19(1):24, sep 2012.

Paul Valiant.
Testing symmetric properties of distributions.
SIAM Journal on Computing, 40(6):1927–1968, 2011.

Gregory Valiant and Paul Valiant.
A CLT and tight lower bounds for estimating entropy.
Electronic Colloquium on Computational Complexity (ECCC), 17:179, 2010.

Gregory Valiant and Paul Valiant.
Estimating the unseen: A sublinear-sample canonical estimator of distributions.

Electronic Colloquium on Computational Complexity (ECCC), 17:180, 2010.

Gregory Valiant and Paul Valiant.
Estimating the unseen: An $n/\log n$-sample estimator for entropy and support size, shown optimal via new clts.
In *Proceedings of STOC*, pages 685–694, 2011.

Gregory Valiant and Paul Valiant.
The power of linear estimators.
In *Proceedings of FOCS*, pages 403–412, October 2011.
See also [VV10a] and [VV10b].

Gregory Valiant and Paul Valiant.
An automatic inequality prover and instance optimal identity testing.
In *Proceedings of FOCS*, 2014.

Yihong Wu and Pengkun Yang.
Minimax rates of entropy estimation on large alphabets via best polynomial approximation.

IEEE Transactions on Information Theory, 62(6):3702–3720, June 2016.