Testing Monotone Continuous Distributions on High-dimensional Real Cubes

Michal Adamaszek

DIMAP, University of Warwick

Joint work with Artur Czumaj and Christian Sohler

Testing probability distributions

- Test if a probability distribution has a given property $\mathcal{P}$.
- Distribution is accessed by drawing random samples.

Testing probability distributions

- Test if a probability distribution has a given property $\mathcal{P}$.
- Distribution is accessed by drawing random samples.
- Goal: distinguish between
  - distributions with the property $\mathcal{P}$,
  - distributions which are far from $\mathcal{P}$

Testing probability distributions

- Test if a probability distribution has a given property $\mathcal{P}$.
- Distribution is accessed by drawing random samples.
- Goal: distinguish between
  - distributions with the property $\mathcal{P}$,
  - distributions which are far from $\mathcal{P}$

  minimizing the number of samples
  and with error probability $\leq 1/3$.

Testing probability distributions

- Test if a probability distribution has a given property $\mathcal{P}$.
- Distribution is accessed by drawing random samples.
- Goal: distinguish between
  - distributions with the property $\mathcal{P}$,
  - distributions which are far from $\mathcal{P}$

  minimizing the number of samples
  and with error probability $\leq 1/3$.
- Examples:
  - is the distribution uniform?
  - is it equal to a fixed distr.?
  - are two distributions identical?
  - are they independent?
  - estimate support size etc...

Classical/typical results

- Is a distribution on k points uniform

  $\tilde{O}(\sqrt{k})$ samples.
- Are two distributions on k points close in L_1-norm

  $\tilde{O}(k^{2/3})$ samples.
- Is a distribution on $\{0, 1, \ldots, k\}$ close to monotone

  $\tilde{O}(\sqrt{k})$ samples.
- Is a distribution on $[k] \times [k]$ a product of its marginals

  $\tilde{O}(k)$ samples.

Batu, Fischer, Fortnow, Kumar, Rubinfeld, Smith, White et al.

Infinite domains

Infinite domains

$\Omega = [0,1]^n$

Infinite domains

$\Omega = [0,1]^n$

- continuous distributions with density f so that

$$\mathbf{Pr}_f[A] = \int_A f \,\mathrm{d}\mu.$$

Infinite domains

$\Omega = [0,1]^n$

- continuous distributions with density f so that

$$\mathbf{Pr}_f[A] = \int_A f \mathrm{d}\mu.$$

- distributions with atoms

$$f + \sum p_i \delta_{x_i}$$

Non-testable properties

Non-testable properties

Is a distribution continuous, or purely discrete?

Non-testable properties

Is a distribution continuous, or purely discrete?

Is a continuous distribution uniform or is it ϵ-far from uniform in the L_1 metric?

Non-testable properties

Is a distribution continuous, or purely discrete?

Is a continuous distribution uniform or is it ϵ-far from uniform in the L_1 metric?

- "fatten" a discrete distribution on M random points,
- up to $\sim \sqrt{M}$ draws this looks like a random distribution,
- but is very L_1-far from uniform.

A testable property - discreteness on M points

For arbitrary Ω distinguish between

- $$f = \sum_{i=1}^{M} p_i \delta_{x_i}$$

  for some $x_1, \ldots, x_M$,

- $$\mathbf{Pr}_f[A] < 1 - \epsilon$$

  for any set $A \subset \Omega$ of size M.

A testable property - discreteness on M points

Tester for discreteness on M points:

- Take $2M/\epsilon$ random samples
- If there are $\leq M$ distinct values accept.
- If there are $> M$ distinct values reject.

A testable property - discreteness on M points

Tester for discreteness on M points:

- Take $2M/\epsilon$ random samples
- If there are $\leq M$ distinct values accept.
- If there are $> M$ distinct values reject.

Lower bound $\Omega(M^{1-o(1)})$ follows from bounds for estimating distribution support size (eg. Raskhodnikova et al'09, Valiant'08). Match these bounds?

Monotone distributions and uniformity

Find a class of distributions for which being uniform is testable.

Monotone distributions and uniformity

Find a class of distributions for which being uniform is testable.

- $\Omega = [0,1]^n$
- The density f is monotone if $f(x) \leq f(y)$ whenever $x_i \leq y_i$ for all i.

Monotone distributions and uniformity

Find a class of distributions for which being uniform is testable.

- $\Omega = [0,1]^n$
- The density f is monotone if $f(x) \leq f(y)$ whenever $x_i \leq y_i$ for all i.

Given a distribution with monotone density f,

- Is f the uniform distribution $\mathcal{U}$?
- Or is it ϵ-far from $\mathcal{U}$ in the L_1 metric

$$d(f,g) = \frac{1}{2}\int_{\Omega} |f-g|.$$

Discrete vs. continuous cubes

Rubinfeld, Servedio'05

Testing uniformity of monotone distributions on the boolean cube $\{0,1\}^n$ with L_1 distance

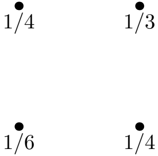
- Is possible with $O(n\log(n/\epsilon)/\epsilon^2)$ samples.
- Requires $\Omega(n/\log^2 n)$ samples.

Discrete vs. continuous cubes

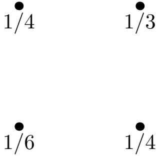
Rubinfeld, Servedio'05

Testing uniformity of monotone distributions on the boolean cube $\{0,1\}^n$ with L_1 distance

- Is possible with $O(n\log(n/\epsilon)/\epsilon^2)$ samples.
- Requires $\Omega(n/\log^2 n)$ samples.



1/4	1/3
1/6	1/4



1/4	1/3
1/6	1/4

lower bound $\longrightarrow$ lower bound
upper bound $\longleftarrow$ upper bound

Discrete vs. continuous cubes

Rubinfeld, Servedio'05

Testing uniformity of monotone distributions on the boolean cube $\{0,1\}^n$ with L_1 distance

- Is possible with $O(n\log(n/\epsilon)/\epsilon^2)$ samples.
- Requires $\Omega(n/\log^2 n)$ samples.

Discrete vs. continuous cubes

Rubinfeld, Servedio'05

Testing uniformity of monotone distributions on the boolean cube $\{0,1\}^n$ with L_1 distance

- Is possible with $O(n\log(n/\epsilon)/\epsilon^2)$ samples.
- Requires $\Omega(n/\log^2 n)$ samples.

Our result

Testing uniformity of monotone distributions on the real cube $[0,1]^n$ with L_1 distance

- Is possible with $O(n/\epsilon^2)$ samples.

Tester

Idea: estimate $\|x\|_1 = x_1 + x_2 + \ldots + x_n$.

Tester

Idea: estimate $\|x\|_1 = x_1 + x_2 + \ldots + x_n$.

- If $\mathcal{U}$ is the uniform distribution then

$$\mathbf{E}_{\mathcal{U}}[\|x\|_1] = \frac{n}{2}.$$

Tester

Idea: estimate $\|x\|_1 = x_1 + x_2 + \ldots + x_n$.

- If $\mathcal{U}$ is the uniform distribution then

$$\mathbf{E}_{\mathcal{U}}[\|x\|_1] = \frac{n}{2}.$$

Theorem

If f is a monotone distribution, ϵ-far from uniform then

$$\mathbf{E}_f[\|x\|_1] \geq \frac{n}{2} + \frac{\epsilon}{2}.$$

Tester

$\Omega = [0,1]^n$, f - unknown monotone distribution.

- Draw C samples $x_1, \ldots, x_C$,

$$\tilde{E} = \frac{1}{C} \sum \|x_i\|_1.$$

- If $\tilde{E} > \frac{n}{2} + \frac{\epsilon}{4}$ say ϵ far from uniform.
- If $\tilde{E} \leq \frac{n}{2} + \frac{\epsilon}{4}$ say uniform.

$C = 40n/\epsilon^2$ is good (use Feige's inequality).

A word on the proof

$$\mathbf{E}_f[\|x\|_1] \geq \frac{n}{2} + \frac{\epsilon}{2}.$$

A word on the proof

$$\mathbf{E}_f[\|x\|_1] \geq \frac{n}{2} + \frac{\epsilon}{2}.$$

or

$$\int_\Omega \|x\|_1 g(x)\mathrm{d}x \geq \frac{1}{4}\int_\Omega |g(x)|\mathrm{d}x$$

for $\int_\Omega g(x)\mathrm{d}x = 0$, $g : [0,1]^n \to \mathbf{R}$ - monotone.

A word on the proof

$$\mathbf{E}_f[\|x\|_1] \geq \frac{n}{2} + \frac{\epsilon}{2}.$$

or

$$\int_\Omega \|x\|_1 g(x)\mathrm{d}x \geq \frac{1}{4}\int_\Omega |g(x)|\mathrm{d}x$$

for $\int_\Omega g(x)\mathrm{d}x = 0$, $g : [0,1]^n \longrightarrow \mathbf{R}$ - monotone.

- $$\int_0^1 t \cdot g(t)\mathrm{d}t = \frac{1}{4}\int_{t,s} |g(t) - g(s)|\mathrm{d}s\mathrm{d}t - \frac{1}{2}\int_0^1 g(t)\mathrm{d}t$$

A word on the proof

$$\mathbf{E}_f[\|x\|_1] \geq \frac{n}{2} + \frac{\epsilon}{2}.$$

or

$$\int_\Omega \|x\|_1 g(x)\mathrm{d}x \geq \frac{1}{4}\int_\Omega |g(x)|\mathrm{d}x$$

for $\int_\Omega g(x)\mathrm{d}x = 0$, $g : [0,1]^n \to \mathbf{R}$ - monotone.

- $$\int_0^1 t \cdot g(t)\mathrm{d}t = \frac{1}{4}\int_{t,s} |g(t) - g(s)|\mathrm{d}s\mathrm{d}t - \frac{1}{2}\int_0^1 g(t)\mathrm{d}t$$
- If g is a function defined on the vertices of a boolean cube

$$\sum_{\textit{diagonals}} |g(u) - g(v)| \leq \sum_{\textit{edges}} |g(u) - g(v)|.$$

Conclusions

- We can test if a monotone distribution on $[0, 1]^n$ is uniform.

Conclusions

- We can test if a monotone distribution on $[0,1]^n$ is uniform. Same for monotone distributions on $\{0,1,\ldots,k\}^n$.

Conclusions

- We can test if a monotone distribution on $[0,1]^n$ is uniform. Same for monotone distributions on $\{0,1,\ldots,k\}^n$.
- Other testable classes of distributions?

Conclusions

- We can test if a monotone distribution on $[0,1]^n$ is uniform. Same for monotone distributions on $\{0,1,\ldots,k\}^n$.
- Other testable classes of distributions?
- Other closeness measures instead of L_1?
  Earth-mover distance? (Ba, Nguyen, Nguyen, Rubinfeld '09)