Testing Continuous Distributions

Artur Czumaj

DIMAP (Centre for Discrete Maths and its Applications) & Department of Computer Science
University of Warwick

Joint work with Michal Adamaszek & Christian Sohler
Testing probability distributions

• General question:
  – Test if a given probability distribution has a given property

  Distribution is available by accessing only samples drawn from the distribution

Examples:
- is given probability uniform?
- are two prob. distributions identical?
- are two prob. distributions independent?
Trend change analysis

Transactions of 20-30 yr olds

Transactions of 30-40 yr olds

trend change?

(slide due to R. Rubinfeld)
Outbreak of diseases

- Do two diseases follow similar patterns?
- Are they correlated with income level or zip code?
- Are they more prevalent near certain areas?
Is the lottery uniform?

• New Jersey Pick-k Lottery (k =3,4)
  – Pick k digits in order.
  – $10^k$ possible values.
• Data:
  – Pick 3 - 8522 results from 5/22/75 to 10/15/00
    • $\chi^2$-test gives 42% confidence
  – Pick 4 - 6544 results from 9/1/77 to 10/15/00.
    • fewer results than possible outcomes
    • $\chi^2$-test gives no confidence

(slide due to R. Rubinfeld)
Testing probability distributions

Lots of research in statistics
Some recent research in algorithms

• Typical result:
  – Given a probability distribution on n points, we can test if it’s uniform after seeing $\sim \sqrt{n}$ random samples

  \[ \text{Testing} = \text{distinguish between uniform distribution and distributions which are } \epsilon\text{-far from uniform} \]

  $\epsilon$-far from uniform:

  \[ \sum_{x \in \Omega} |\Pr[x] - \frac{1}{n}| \geq \epsilon \]

  \[ \text{error probab. } \leq \frac{1}{3} \]

  [Batu et al ‘01]
Testing probability distributions

Given a probability distribution on n points, we can test if it’s uniform after seeing \( \sim \sqrt{n} \) random samples [Batu et al ’01]

Testing = distinguish between uniform distribution and distributions which are \( \epsilon \)-far from uniform

\( \epsilon \)-far from uniform: \( \sum_{x \in \Omega} |\Pr[x] - \frac{1}{n}| \geq \epsilon \)

error probab. \( \leq 1/3 \)

Choose \( \Theta(n^{1/2}) \) random samples \( s_1, s_2, \ldots, s_T \)
Count the number of collisions: \( \text{coll} = \#\{i < j: s_i = s_j\} \)
Accept iff \( \text{coll} < T^2 \left(1 + \epsilon/4\right)/2n \)

Let \( p_i \) be the probability that the \( i \)th element is chosen, \( 1 \leq i \leq n \).
If distribution is uniform then \( p_i = \frac{1}{n} \) for every \( i \).
For given \( p_1, \ldots, p_n \), we have: \( E[\text{coll}] = \left(\frac{T}{2}\right) \cdot \sum_{i=1}^{n} p_i^2 \)
In particular, if the distribution is uniform then \( E[\text{coll}] = \left(\frac{T}{n}\right) \)

Assuming that \( \sum_{i=1}^{n} |p_i - \frac{1}{n}| \geq \epsilon \), when is \( \sum_{i=1}^{n} p_i^2 \) minimized?
When \( \epsilon n/2 \) elements have zero probability and the remaining are uniform:

\[
\sum_{i=1}^{n} p_i^2 \geq (1 - \epsilon/2)n \cdot \left(\frac{1}{n} \cdot \frac{2}{2 - \epsilon}\right)^2 \approx \frac{1}{n} \cdot (1 + \epsilon/2) .
\]

Hence, \( E[\text{coll}] \geq \left(\frac{T}{n}\right) \cdot (1 + \epsilon/2) \)

Therefore, if we choose \( T = \Theta(\sqrt{n}\log(1/\epsilon)/\epsilon^2) \) then we can distinguish between these two cases with probability \( \geq 0.99 \).
Testing probability distributions

• Typical result:
  – Given a probability distribution on n points, we can test if it’s uniform after seeing $\sim \sqrt{n}$ random samples

[Batu et al ‘01]

• Similar bounds for testing
  • if a distribution is monotone
  • if two distributions are independent
  • …
Testing probability distributions

Lots of research in statistics

Some recent research in algorithms

• Typical result:
  – Given a probability distribution on n points, we can test if it’s uniform after seeing $\sim \sqrt{n}$ random samples

  [Batu et al ‘01]

Many properties of distributions can be tested in time sublinear in the domain/support size
Testing probability distributions

Lots of research in statistics

Some recent research in algorithms

• Typical result:
  – Given a probability distribution on $n$ points, we can test if it’s uniform after seeing $\sim \sqrt{n}$ random samples

[Batu et al ‘01]

• What if distribution has infinite support?
• Continuous probability distributions?
Testing continuous probability distributions

• Typical result:
  – Given a probability distribution on n points, we can test if it’s uniform after seeing $\sim \sqrt{n}$ random samples
  – $\sim \sqrt{n}$ random samples are necessary

Distinguish between

• uniform probability on $n$ points
• uniform probability on $\frac{1}{2}n$ points.

Random sampling won’t repeat any single sample after $o(\sqrt{n})$ draws (“Birthday paradox”)
Testing continuous probability distributions

• Typical result:
  – Given a probability distribution on $n$ points, we can test if it’s uniform after seeing $\sim \sqrt{n}$ random samples
  – $\sim \sqrt{n}$ random samples are necessary

• Given a continuous probability distribution on $[0,1]$, can we test if it’s uniform?
  
  • Impossible
    • Follows from lower bound for discrete case with $n \to \infty$
More direct proof:

Suppose tester A distinguishes in at most t steps between uniform distribution and $\epsilon$-far from uniform

- $D_1$ – uniform distribution
- $D_2$ is $\frac{1}{2}$-far from uniform and is defined as follows:
  - Choose $t^3$ points from $[0,1]$ independently and uniformly at random
  - $D_2$ is defined uniformly on the chosen points ($D_2$ is discrete on $t^3$ points!)
  - In t steps, no point from support of $D_2$ will be chosen more than once

A cannot distinguish between $D_1$ and $D_2$
Testing continuous probability distributions

• What can be tested?

• First question:
  test if the distribution is indeed continuous
Testing continuous probability distributions

• Dual question:
  Test if a probability distribution is **discrete**

• Prob. distribution $D$ on $\Omega$ is discrete on $N$ points if there is a set $X \subseteq \Omega$, $|X| \leq N$, st. $\Pr_D[X]=1$

• $D$ is $\epsilon$-far from discrete on $N$ points if
  \[
  \forall X \subseteq \Omega, |X| \leq N \hspace{1cm} \Pr_D[X] \leq 1-\epsilon
  \]
Testing if distribution is discrete on $N$ points

- We repeatedly draw random points from $D$
- All what can we see:
  - Count frequency of each point
  - Count number of points drawn

For some $D$ (eg, uniform or close):
- we need $\Omega(\sqrt{N})$ to see first multiple occurrence

Gives a hope that can be solved in sublinear-time
Testing if distribution is discrete on $N$ points

Raskhodnikova et al ’07 (Valiant’08):

**Distinct Elements Problem:**
- $D$ discrete with each element with prob. $\geq \frac{1}{N}$
- Estimate the support size

$\Omega(N^{1-o(1)})$ queries are needed to distinguish instances with $\leq N/100$ and $\geq N/11$ support size

Key step: two distributions that have identical first $\log^{\Theta(1)} N$ moments
- their expected frequencies up to $\log^{\Theta(1)} N$ are identical
Testing if distribution is discrete on N points

Raskhodnikova et al ’07 (Valiant’08):

Distinct Elements Problem:

• D discrete with each element with prob. \( \geq \frac{1}{N} \)
• Estimate the support size

\( \Omega(N^{1-o(1)}) \) queries are needed to distinguish instances with \( \leq \frac{N}{100} \) and \( \geq \frac{N}{11} \) support size

Corollary:

Testing if a distribution is discrete on N points requires \( \Omega(N^{1-o(1)}) \) samples
Testing if distribution is discrete on N points

• We repeatedly draw random points from D
• All what can we see:
  – Count frequency of each point
  – Count number of points drawn
• Can we get O(N) time?
Testing if a distribution is discrete on N points:

- Draw a sample $S = (s_1, ..., s_t)$ with $t = \frac{cN}{\epsilon}$.
- If $S$ has more than $N$ distinct elements, then REJECT.
- Else ACCEPT.

If $D$ is discrete on $N$ points then we will accept $D$.

We only have to prove that:

- if $D$ is $\epsilon$-far from discrete on $N$ points, then we will reject with probability $> \frac{2}{3}$. 
• Testing if a distribution is discrete on \( N \) points:

\[
\text{Draw a sample } S = (s_1, \ldots, s_t) \text{ with } t = \frac{cN}{\epsilon}
\]

• If \( S \) has more than \( N \) distinct elements then \textbf{REJECT}
else \textbf{ACCEPT}

\( D \) is \( \epsilon \)-far from discrete on \( N \) points, then reject with prob \( >2/3 \)

\( D \) is \( \epsilon \)-far from discrete on \( N \) points \( \Rightarrow \)

\( \forall X \subseteq \Omega, \text{ if } |X| \leq N \text{ then } Pr_D[\Omega \setminus X] \geq \epsilon \)

• Assuming that we haven’t chosen \( n \) points yet, we choose a new point with probability at least \( \epsilon \)

After \( (1 + o(1))N/\epsilon \) samples, we choose \( N + 1 \) points with prob. \( \geq 0.99 \)
Testing if a distribution is discrete on N points:

- Draw a sample $S = (s_1, \ldots, s_t)$ with $t = \frac{cN}{\epsilon}$
- If $S$ has more than $N$ distinct elements then REJECT
  else ACCEPT

Can we do better (if we only count distinct elements)?

D: has 1 point with prob. $1 - 4\epsilon$
  2N points with prob. $2\epsilon/N$

D is $\epsilon$-far from discrete on N points

We need $\Omega(N/\epsilon)$ samples to see at least N points
Testing continuous probability distributions

• What can we test efficiently?
  – Complexity for discrete distributions should be “independent” on the support size

• Uniform distribution … under some conditions

• Rubinfeld & Servedio’05:
  – testing monotone distributions for uniformity
Rubinfeld & Servedio’05:

• Testing monotone distributions for uniformity

D: distribution on n-dimensional cube; D: \{0,1\}^n \rightarrow \mathbb{R}

x, y \in \{0,1\}^n, x \preceq y \iff \forall i: x_i \leq y_i

D is monotone if x \preceq y \Rightarrow \Pr[x] \leq \Pr[y]

Goal: test if a monotone distribution is uniform

Rubinfeld & Servedio’05:

Testing if a monotone distribution on n-dimensional binary cube is uniform:

• Can be done with \(O(n \log(1/\epsilon)/\epsilon^2)\) samples
• Requires \(\Omega(n/\log^2 n)\) samples
Rubinfeld & Servedio’05:
• Testing monotone distributions for uniformity

\( \mathbf{D} : \) distribution on n-dimensional cube; \( \mathbf{D} : \{0,1\}^n \rightarrow \mathbb{R} \)

\( x, y \in \{0,1\}^n, x \preceq y \iff \forall i: x_i \leq y_i \)

\( \mathbf{D} \) is monotone if \( x \preceq y \Rightarrow \Pr[x] \leq \Pr[y] \)

Goal: test if a monotone distribution is uniform

\( \mathbf{D} : \) distribution on n-dimensional cube; density function \( f : [0,1]^n \rightarrow \mathbb{R} \)

\( x, y \in [0,1]^n, x \preceq y \iff \forall i: x_i \leq y_i \)

\( \mathbf{D} \) is monotone if \( x \preceq y \Rightarrow f(x) \leq f(y) \)
Rubinfeld & Servedio’05:
Testing if a monotone distribution on n-dimensional binary cube is uniform:
• Can be done with $O(n \log(1/\epsilon)/\epsilon^2)$ samples
• Requires $\Omega(n/\log^2 n)$ samples
D is $\epsilon$-far from uniform if

$$\frac{1}{2} \int_{x \in \Omega} |f(x) - 1| dx \geq \epsilon$$

To test uniformity, we need to characterize monotone distributions that are $\epsilon$-far from uniform.

On the high level:
- we follow approach of Rubinfeld & Servedio’05;
- details are quite different
Testing monotone distributions for uniformity

D is $\epsilon$-far from uniform if

$$\frac{1}{2} \int_{x \in \Omega} |f(x) - 1| dx \geq \epsilon$$

**Key Technical Lemma:**

Let $g: [0,1]^n \rightarrow \mathbb{R}$ be a monotone function with $\int x \ g(x) \ dx = 0$ then

$$\int x \cdot g(x) dx \geq \frac{1}{4} \int |g(x)| dx$$

**Key Lemma:**

If $D$ is a monotone distribution on $[0,1]^n$ with density function $f$ and which is $\epsilon$-far from uniform then

$$E_f[\|x\|_1] = \int x \cdot f(x) dx \geq \frac{n}{2} + \frac{\epsilon}{2}$$
Key Lemma:
If D is a monotone distribution on $[0,1]^n$ with density function $f$ and which is $\epsilon$-far from uniform then

$$E_f[\|x\|_1] = \int_x \|x\|_1 \cdot f(x) dx \geq \frac{n}{2} + \frac{\epsilon}{2}$$

Uniform distribution:
If D is uniform on $[0,1]^n$ with density function $f$ then

$$E_f[\|x\|_1] = \int_x \|x\|_1 \cdot f(x) dx = \frac{n}{2}$$
Key Lemma:
If $D$ is a monotone distribution on $[0,1]^n$ with density function $f$ and which is $\epsilon$-far from uniform then

\[ E_f[\|x\|_1] = \int_x \|x\|_1 \cdot f(x) dx \geq \frac{n}{2} + \frac{\epsilon}{2} \]

\[ s = \frac{cn}{\epsilon^2} \]

Repeat 20 times
   
   Draw a sample $S=(x_1, \ldots, x_s)$ from $[0,1]^n$
   
   If $\sum_i \|x_i\|_1 \geq s \left( \frac{n}{2} + \frac{\epsilon}{4} \right)$ then REJECT and exit

ACCEPT
Testing monotone distributions for uniformity

Theorem:
The algorithm below tests if $D$ is uniform. Its complexity is $O(n/\epsilon^2)$.

Slightly better bound than the one by RS'05

$s = cn/\epsilon^2$
Repeat 20 times
   Draw a sample $S=(x_1,...,x_s)$ from $[0,1]^n$
   If $\sum_i ||x_i||_1 \geq s (n/2+\epsilon/4)$ then REJECT and exit
ACCEPT
Testing monotone distributions for uniformity

s = cn/\epsilon^2

Repeat 20 times

Draw a sample S=(x_1,...,x_s) from [0,1]^n

If \sum_i \|x_i\|_1 \geq s(n/2+\epsilon/4) then REJECT and exit

ACCEPT

Lemma 1: If D is uniform then
Pr[\sum_i \|x_i\|_1 \geq s(n/2+\epsilon/4)] \leq 0.01

Easy application of Chernoff bound

Lemma 2: If D is \epsilon-far from uniform then
Pr[\sum_i \|x_i\|_1 < s(n/2+\epsilon/4)] \leq 12/13

By Key Lemma + Feige lemma
Testing monotone distributions for uniformity

$s = cn/\epsilon^2$
Repeat 20 times

Draw a sample $S=(x_1,\ldots,x_s)$ from $[0,1]^n$

If $\sum_i ||x_i||_1 \geq s \left(n/2+\epsilon/4\right)$ then REJECT and exit

**Lemma 2**: If $D$ is $\epsilon$-far from uniform then
\[ \Pr[\sum_i ||x_i||_1 < s(n/2+\epsilon/4)] \leq 12/13 \]

**Proof:**

D is $\epsilon$-far from uniform $\Rightarrow$ $E[\sum_i ||x_i||_1] \geq s(n+\epsilon)/2$

**Markov inequality:**
\[ \Pr[\sum_i ||x_i||_1 < s(n/2+\epsilon/4)] \geq 1/O(\epsilon) \]

**Weak Feige’s lemma:** $Y_1, \ldots, Y_s$ independent r.v., $Y_i \geq 0$, $E[Y_i \leq 1] \Rightarrow$
\[ \Pr[\sum_i Y_i < s + 1/12] \geq 1/13 \]

Choose $Y_i = 2-2||x_i||_1/(n+\epsilon)$

Then, Feige’s lemma yields the desired claim
Testing monotone distributions for uniformity

\[ s = \frac{cn}{\epsilon^2} \]

Repeat 20 times

Draw a sample \( S = (x_1, \ldots, x_s) \) from \([0,1]^n\)

If \( \sum_i ||x_i||_1 \geq s \left( \frac{n}{2} + \frac{\epsilon}{4} \right) \) then REJECT and exit

ACCEPT

Lemma 2: If \( D \) is \( \epsilon \)-far from uniform then

\[ \Pr[\sum_i ||x_i||_1 < s(n/2+\epsilon/4)] \leq \frac{12}{13} \]

Proof:

D is \( \epsilon \)-far from uniform \( \Rightarrow \) \( E[\sum_i ||x_i||_1] \geq s(n+\epsilon)/2 \)

**Feige’s lemma:** \( Y_1, \ldots, Y_s \) independent r.v., \( Y_i \geq 0, E[Y_i \leq 1] \)

\[ \Pr[\sum_i Y_i < s + 1/12] \geq \frac{1}{13} \]

Choose \( Y_i = 2-2||x_i||_1/(n+\epsilon) \)

Then, Feige’s lemma yields the desired claim
Testing monotone distributions for uniformity

**Key Lemma:**
If $D$ is a monotone distribution on $[0,1]^n$ with density function $f$ and which is $\epsilon$-far from uniform then

$$E_f[\|x\|_1] = \int_x \|x\|_1 \cdot f(x) dx \geq \frac{n}{2} + \frac{\epsilon}{2}$$

$s = cn/\epsilon^2$

Repeat 20 times

Draw a sample $S=(x_1,\ldots,x_s)$ from $[0,1]^n$

If $\sum_i \|x_i\|_1 \geq s \left(\frac{n}{2} + \frac{\epsilon}{4}\right)$ then REJECT and exit

ACCEPT
Testing monotone distributions for uniformity

**Key Lemma:**
If D is a monotone distribution on $[0,1]^n$ with density function $f$ and which is $\varepsilon$-far from uniform then

$$E_f[\|x\|_1] = \int_x \|x\|_1 \cdot f(x) dx \geq \frac{n}{2} + \frac{\varepsilon}{2}$$

**Key Technical Lemma:**
Let $g:[0,1]^n \rightarrow \mathbb{R}$ be a monotone function with $\int_x g(x) dx = 0$ then

$$\int_x \|x\|_1 \cdot g(x) dx \geq \frac{1}{4} \int_x |g(x)| dx$$
Key Technical Lemma:

Let \( g: [0,1]^n \to \mathbb{R} \) be a monotone function with \( \int g(x) \, dx = 0 \) then

\[
\int_x ||x||_1 \cdot g(x) \, dx \geq \frac{1}{4} \int_x |g(x)| \, dx
\]

Why such a bound:

Tight for \( g(x) = \text{sgn}(x_1 - \frac{1}{2}) \)

\[
\int_{x: x_1 > \frac{1}{2}} ||x||_1 \cdot g(x) \, dx = \frac{1}{2} \int_{x: x_1 > \frac{1}{2}} (x_1 + \ldots + x_n) \, dx = \frac{1}{2} \left( \frac{3}{4} + \frac{1}{2} + \ldots + \frac{1}{2} \right) \, dx = \frac{n+1}{4}.
\]

Similarly,

\[
\int_{x: x_1 < \frac{1}{2}} ||x||_1 \cdot g(x) \, dx = \frac{1}{2} \left( \frac{1}{4} + \frac{1}{2} + \ldots + \frac{1}{2} \right) = \frac{n - 1}{8},
\]

and hence,

\[
\int_x ||x||_1 \cdot g(x) \, dx = \int_{x: x_1 > \frac{1}{2}} ||x||_1 \cdot g(x) \, dx - \int_{x: x_1 < \frac{1}{2}} ||x||_1 \cdot g(x) \, dx = \frac{1}{4} = \frac{1}{4} \cdot \int_x |g(x)| \, dx.
\]
Key Technical Lemma:

Let \( g: [0,1]^n \to \mathbb{R} \) be a monotone function with \( \int_x g(x) \, dx = 0 \) then

\[
\int_x \|x\|_1 \cdot g(x) \, dx \geq \frac{1}{4} \int_x |g(x)| \, dx
\]
Let $P = \{x : g(x) \geq 0\}$ and $N = \{x : g(x) < 0\}$. Consider:

$$
\int_{x \in N} \int_{y \in P} |g(x) - g(y)| \, dy \, dx .
$$

For $g(x) < 0 \leq g(y)$, we have $|g(x) - g(y)| = |g(x)| + |g(y)|$.

Moreover $\int_{x \in N} |g(x)| \, dx = \int_{y \in P} |g(y)| \, dy = \frac{1}{2} \int_{x} |g(x)| \, dx$.

Hence:

$$
\int_{x \in N} \int_{y \in P} (|g(x)| + |g(y)|) = \int_{y \in P} \int_{x \in N} |g(x)| + \int_{x \in N} \int_{y \in P} |g(y)|
$$

$$
= \frac{1}{2} \int_{y \in P} \int_{x} |g(x)| + \frac{1}{2} \int_{x \in N} \int_{y} |g(y)| = \frac{1}{2} \int_{y} \int_{x} |g(x)| = \frac{1}{2} \int_{x} |g(x)| .
$$

Since every pair $(x, y)$ can satisfy at most one of the conditions $(x, y) \in P \times N$ and $(x, y) \in N \times P$, we obtain:

$$
\int_{x \in N} \int_{y \in P} |g(x) - g(y)| \, dy \, dx \leq \frac{1}{2} \int \int_{x,y} |g(x) - g(y)| \, dy \, dx .
$$

Hence:

$$
\frac{1}{2} \int_{x} |g(x)| \, dx = \int_{x \in N} \int_{y \in P} |g(x) - g(y)| \, dx \, dy \leq \frac{1}{2} \int \int_{x,y} |g(x) - g(y)| \, dx \, dy .
$$
Testing monotone distributions for uniformity

Reductions via discrete cubes:

Let $D(\{0,1\}^n)$ be the set of all main diagonals of discrete cube $\{0,1\}^n$:

$$D(\{0,1\}^n) = \{(x, y) \in \{0,1\}^n \times \{0,1\}^n : x_i = 1 - y_i \text{ for every } i\}$$

Let $E_i(\{0,1\}^n)$ be the set of all edges in the $i$th direction:

$$E_i(\{0,1\}^n) = \{(x, y) \in \{0,1\}^n \times \{0,1\}^n : x_i = 1 - y_i \text{ and } x_j = y_j \text{ for every } j \neq i\}$$

Let $E(\{0,1\}^n) = \bigcup_i E_i(\{0,1\}^n)$.

For any function $g : \{0, 1\}^n \to \mathbb{R}$:

$$\sum_{(x,y) \in D(\{0,1\}^n)} |g(x) - g(y)| \leq \sum_{(x,y) \in E(\{0,1\}^n)} |g(x) - g(y)|.$$
For any $x \preceq y$, define a discrete cube $K_{x,y}$ by the affine transformation sending $x$ to $0^n$ and $y$ to $1^n$

Let $D(x,y)$, $E_i(x,y)$, and $E(x,y)$ denote the diagonals and edges of $K_{x,y}$

By the previous claim, for any function $g:[0,1]^n \to \mathbb{R}$ and any $x \preceq y$:

$$\sum_{(u,v) \in D(x,y)} |g(u) - g(v)| \leq \sum_{(u,v) \in E(x,y)} |g(u) - g(v)| \leq \sum_{i=1}^{n} \sum_{(u,v) \in E_i(x,y)} |g(u) - g(v)| .$$

For any monotone function $g:[0,1]^n \to \mathbb{R}$ and any $i = 1, \ldots, n$ we have

$$\int_{x \preceq y} \left( \sum_{(u,v) \in E_i(x,y)} |g(u) - g(v)| \right) \, dx \, dy = 2 \int_{x} (2x_i - 1)g(x) \, dx$$

Proof by induction on $n$. 
By considering all the possible relative placements of $x$ and $y$ within $[0, 1]^n$ and splitting the domain accordingly, one can see that

\[
\int \int_{x,y} |g(x) - g(y)| \, dy \, dx = \int \int_{x \prec y} \left( \sum_{(u,v) \in D(x,y)} |g(u) - g(v)| \right) \, dy \, dx.
\]
Testing monotone distributions for uniformity

Key Technical Lemma:
Let \( g: [0, 1]^n \to \mathbb{R} \) be a monotone function with \( \int_{\mathbb{R}} g(x) \, dx = 0 \) then

\[
\int_{\mathbb{R}} \|x\|_1 \cdot g(x) \, dx \geq \frac{1}{4} \int_{\mathbb{R}} |g(x)| \, dx
\]

Key inequalities in the proof:
\[
\frac{1}{4} \int_{\mathbb{R}} |g(x)| \, dx \leq \frac{1}{4} \int_{\mathbb{R}} \int_{\mathbb{R}} |g(x) - g(y)| \, dx \, dy
\]
\[
\leq \int_{\mathbb{R}} \int_{\mathbb{R}} \left( \sum_{(u, v) \in D(x, y)} |g(u) - g(v)| \right) \, dx \, dy
\]
\[
\leq \frac{1}{2} \sum_{i=1}^{n} \int_{\mathbb{R}} \int_{\mathbb{R}} \left( \sum_{(u, v) \in E_i(x, y)} |g(u) - g(v)| \right) \, dx \, dy
\]
\[
\leq \frac{1}{2} \sum_{i=1}^{n} \int_{\mathbb{R}} (2x_i - 1)g(x) \, dx
\]
\[
\leq \int_{\mathbb{R}} \|x\|_1 g(x) \, dx
\]
Rubinfeld & Servedio’05:
Testing if a monotone distribution on n-dimensional binary cube is uniform:
• Can be done with $O(n \log(1/\epsilon)/\epsilon^2)$ samples
• Requires $\Omega(n/\log^2 n)$ samples

Here:
Testing if a monotone distribution on n-dimensional continuous cube is uniform:
• Can be done with $O(n/\epsilon^2)$ samples
• (Requires $\Omega(n/\log^2 n)$ samples)
Can be easily extended to $\{0,1,\ldots,k\}^n$ cubes
Conclusions

• Testing continuous distributions or distribution on infinite/uncountable support is different from testing discrete distributions
  – Continuous distributions are harder

• Challenge: understand when it’s possible to test
  – Usually some additional conditions are to be imposed

• Tight(er) bounds?
Conclusions

• Continuous distributions are harder
• Is the $L_1$-norm the right one?
  – It doesn’t work well for continuous distributions
• Earth mover norm?
  – How much mass have to be moved and how far to obtain a given distribution
  – Testing uniformity on $[0,1]$ can be done in time $f(1/\epsilon)$