Robust List Decoding

Daniel M. Kane

Departments of CS/Math
University of California, San Diego
dakane@ucsd.edu

June 23rd, 2019
Outline

- Problem Setup
- Multifilter Idea
- Basic Algorithm
- Application: Mixtures of Spherical Gaussians
Robust Mean Estimation

- Gaussian $G = N(\mu, I) \subset \mathbb{R}^n$
Robust Mean Estimation

- Gaussian $G = N(\mu, I) \subset \mathbb{R}^n$
- $X = (1 - \epsilon)G + \epsilon E$ for small $\epsilon$
Robust Mean Estimation

- Gaussian $G = N(\mu, I) \subset \mathbb{R}^n$
- $X = (1 - \epsilon)G + \epsilon E$ for small $\epsilon$
- Given $m$ independent samples $x_i$ of $X$
Robust Mean Estimation

- Gaussian $G = N(\mu, I) \subset \mathbb{R}^n$
- $X = (1 - \epsilon)G + \epsilon E$ for small $\epsilon$
- Given $m$ independent samples $x_i$ of $X$
- Learn Approximation to $\mu$
Very Robust Mean Estimation

- Gaussian $G = N(\mu, I) \subset \mathbb{R}^n$
Very Robust Mean Estimation

- Gaussian $G = N(\mu, I) \subset \mathbb{R}^n$
- $X = \alpha G + (1 - \alpha)E$ for small $\alpha$
Very Robust Mean Estimation

- Gaussian $G = N(\mu, I) \subset \mathbb{R}^n$
- $X = \alpha G + (1 - \alpha)E$ for small $\alpha$
- Given $m$ independent samples $x_i$ of $X$
Very Robust Mean Estimation

- Gaussian $G = N(\mu, I) \subset \mathbb{R}^n$
- $X = \alpha G + (1 - \alpha)E$ for small $\alpha$
- Given $m$ independent samples $x_i$ of $X$
- Learn Approximation to $\mu$
Problem

What if \( X = \sum_i \alpha_i G_i \)? Which is the “real” \( G \)?
Problem

What if $X = \sum_i \alpha_i G_i$? Which is the “real” $G$?

List decoding: return several hypotheses $h_i$ with guarantee that at least one is close.
Problem

What if $X = \sum_i \alpha_i G_i$? Which is the “real” $G$?

List decoding: return several hypotheses $h_i$ with guarantee that at least one is close. Considered by [Charikar-Steinhardt-Valiant ’17].
Lower Bounds

- Suppose $X = N(0, I)$. 

Adding a bit to $X$, can hide $\alpha - \Omega(C)$ such Gaussians.
Lower Bounds

- Suppose $X = \mathcal{N}(0, I)$.
- Any $\alpha N(\mu, I)$ with $|\mu| \leq \sqrt{\log(1/\alpha)/C}$ nearly hides under $X$ (up to $\alpha^{\Omega(C)}$ error).
Lower Bounds

- Suppose $X = N(0, I)$.
- Any $\alpha N(\mu, I)$ with $|\mu| \leq \sqrt{\frac{\log(1/\alpha)}{C}}$ nearly hides under $X$ (up to $\alpha^{\Omega(C)}$ error).
- Adding a bit to $X$, can hide $\alpha^{-\Omega(C)}$ such Gaussians.
Lower Bounds

Proposition

There is no algorithm that returns $\text{poly}(1/\alpha)$ many hypothesis so that with at least $2/3$ probability, at least one is within $o(\sqrt{\log(1/\alpha)})$ of the true mean.

- Let $X$ be the slightly modified Gaussian.
- There are $\alpha^{-\Omega(C)}$ possibilities, no two within $\sqrt{\log(1/\alpha)}/C$.
- Algorithm cannot tell which possibility is correct, and must return a hypothesis for each.
Lower Bounds

Proposition

There is no algorithm that returns $\text{poly}(1/\alpha)$ many hypothesis so that with at least $2/3$ probability, at least one is within $o(\sqrt{\log(1/\alpha)})$ of the true mean.

- Let $X$ be the slightly modified Gaussian.
- There are $\alpha^{-\Omega(C)}$ possibilities, no two within $\sqrt{\log(1/\alpha)}/C$.
- Algorithm cannot tell which possibility is correct, and must return a hypothesis for each.

We will show $\tilde{O}(1/\sqrt{\alpha})$ error. Next talk: near optimal error.
Moderately Robust Algorithm

With few errors algorithm looks like:

1. Compute Covariance
2. If large eigenvalue produce filter and repeat
3. Return sample mean
Moderately Robust Algorithm

With few errors algorithm looks like:

1. Compute Covariance
2. If large eigenvalue produce filter and repeat
3. Return sample mean

Would like to do the same thing in the high noise case. It almost works.
Multifilters

If $\alpha < 1/2$, might not be able to tell where the real samples are.
Multifilters

If $\alpha < 1/2$, might not be able to tell where the real samples are.

Split into several overlapping sets of samples $S_i$
Multifilters

If $\alpha < 1/2$, might not be able to tell where the real samples are.

Split into several overlapping sets of samples $S_i$ so that:
- At least one $S_i$ has higher fraction of good samples than $S$
- $\sum |S_i|^2 \leq |S|^2$
Analysis

Split into cases

- **Case 1:** Almost all of the samples are in the same small interval.
- **Case 2:** There are clusters of samples far apart from each other.
Suppose that there is an interval $I$ containing all but an $\alpha/3$-fraction of samples.
Filter Case

Suppose that there is an interval $I$ containing all but an $\alpha/3$-fraction of samples.

- With high probability, true mean in $I$. 
Suppose that there is an interval $I$ containing all but an $\alpha/3$-fraction of samples.

- With high probability, true mean in $I$.
- All but a tiny fraction of good samples within $O(\sqrt{\log(1/\alpha)})$ of $I$. 
Filter Case

Suppose that there is an interval \( I \) containing all but an \( \alpha/3 \)-fraction of samples.

- With high probability, true mean in \( I \).
- All but a tiny fraction of good samples within \( O(\sqrt{\log(1/\alpha)}) \) of \( I \).
- Unless variance is \( O(|I|^2 + \log(1/\alpha)) \), at most an \( \alpha^2 \)-fraction of removed samples were good.
Multifilter Case

Suppose that there is an interval $I$ with at least an $\alpha/6$-fraction of samples on either side of it.
Multifilter Case

Suppose that there is an interval $I$ with at least an $\alpha/6$-fraction of samples on either side of it.

- Find some $x$, let $S_1 = \{\text{samples $\leq x + 10\sqrt{\log(1/\alpha)}$}\}$,
  $S_2 = \{\text{samples $\geq x - 10\sqrt{\log(1/\alpha)}$}\}$.

All but an $\alpha^2$-fraction of removed samples (on the correct side) are bad:

- If $\mu \geq x$, all but $\alpha^3$-fraction of good samples in $S_2$.
- If $\mu \leq x$, all but $\alpha^3$-fraction in $S_1$.
- Always throw away at least $\alpha/6$ samples.

Need:
$$|S_1|^2 + |S_2|^2 \leq |S|^2.$$
Multifilter Case

Suppose that there is an interval $I$ with at least an $\alpha/6$-fraction of samples on either side of it.

- Find some $x$, let $S_1 = \{\text{samples} \leq x + 10\sqrt{\log(1/\alpha)}\}$, $S_2 = \{\text{samples} \geq x - 10\sqrt{\log(1/\alpha)}\}$.

- All but an $\alpha^2$-fraction of removed samples (on the correct side) are bad:
  - If $\mu \geq x$, all but $\alpha^3$-fraction of good samples in $S_2$.
  - If $\mu \leq x$, all but $\alpha^3$-fraction in $S_1$.
  - Always throw away at least $\alpha/6$ samples.
Suppose that there is an interval $I$ with at least an $\alpha/6$-fraction of samples on either side of it.

- Find some $x$, let $S_1 = \{\text{samples} \leq x + 10\sqrt{\log(1/\alpha)}\}$, $S_2 = \{\text{samples} \geq x - 10\sqrt{\log(1/\alpha)}\}$.

- All but an $\alpha^2$-fraction of removed samples (on the correct side) are bad:
  - If $\mu \geq x$, all but $\alpha^3$-fraction of good samples in $S_2$.
  - If $\mu \leq x$, all but $\alpha^3$-fraction in $S_1$.
  - Always throw away at least $\alpha/6$ samples.

- **Need:** $|S_1|^2 + |S_2|^2 \leq |S|^2$. 
Let $f(x)$ be the fraction of samples less than $x$. 
Analysis

- Let $f(x)$ be the fraction of samples less than $x$.
- Need $x \in I$ so that $(1 - f(x))^2 + f(x + 20\sqrt{\log(1/\alpha)})^2 \leq 1$. 

Happens unless $f(x + 20\sqrt{\log(1/\alpha)}) \gg f(x)^{1/2}$.

Good unless $f(x + 20t\sqrt{\log(1/\alpha)}) \gg \alpha^{1/2}t$, only works for $t \ll \log \log(1/\alpha)$.

Can find such sets unless $|I| = O(\sqrt{\log(1/\alpha) \log \log(1/\alpha)})$. 
Let \( f(x) \) be the fraction of samples less than \( x \).

Need \( x \in I \) so that \((1 - f(x))^2 + f(x + 20\sqrt{\log(1/\alpha)})^2 \leq 1.\)

Happens unless \( f(x + 20\sqrt{\log(1/\alpha)}) \gg f(x)^{1/2}.\)
Let $f(x)$ be the fraction of samples less than $x$.

Need $x \in I$ so that $(1 - f(x))^2 + f(x + 20\sqrt{\log(1/\alpha)})^2 \leq 1$.

Happens unless $f(x + 20\sqrt{\log(1/\alpha)}) \gg f(x)^{1/2}$.

Good unless $f(x + 20t\sqrt{\log(1/\alpha)}) \gg \alpha^{1/2t}$, only works for $t \ll \log \log(1/\alpha)$. 
Let $f(x)$ be the fraction of samples less than $x$.

Need $x \in I$ so that $(1 - f(x))^2 + f(x + 20\sqrt{\log(1/\alpha)})^2 \leq 1$.

Happens unless $f(x + 20\sqrt{\log(1/\alpha)}) \gg f(x)^{1/2}$.

Good unless $f(x + 20t\sqrt{\log(1/\alpha)}) \gg \alpha^{1/2^t}$, only works for $t \ll \log \log(1/\alpha)$.

Can find such sets unless $|I| = O(\sqrt{\log(1/\alpha)} \log \log(1/\alpha))$. 
General Situation

Can create a filter or multifilter if either:

- No interval $I$ of length $O(\sqrt{\log(1/\alpha) \log \log(1/\alpha)})$ contains all but an $\alpha/3$-fraction of samples.
- An interval $I$ of length $O(\sqrt{\log(1/\alpha) \log \log(1/\alpha)})$ contains all but an $\alpha/3$-fraction of samples, and the variance is $\Omega(|I|^2)$.
General Situation

Can create a filter or multifilter if either:

- No interval $I$ of length $O(\sqrt{\log(1/\alpha) \log \log(1/\alpha)})$ contains all but an $\alpha/3$-fraction of samples.
- An interval $I$ of length $O(\sqrt{\log(1/\alpha) \log \log(1/\alpha)})$ contains all but an $\alpha/3$-fraction of samples, and the variance is $\Omega(|I|^2)$.

Proposition

*If the variance in some direction is more than a sufficient multiple of $\log(1/\alpha)$ (with a slight refinement of the argument) then we can find at most two sets of samples $S_i$ so that*

1. For some $i$, at most an $\alpha^2$-fraction of $S \setminus S_i$ is good samples.
2. $\sum_i |S_i|^2 \leq |S|^2$. 
Basic Multifilter Algorithm

1. Maintain several sets $S_i$ of samples
2. For each $i$, compute empirical covariance matrix $\hat{\Sigma}_i$
3. If some $\hat{\Sigma}_i$ has a large eigenvalue
   - Create multifilter
   - Apply to $S_i$
   - Replace $S_i$ by resulting sets in list
   - Go to step 2.
4. Return list of all $\mu S_i$
Analysis

At each step:

- At least one $S_i$ has an $\alpha$-fraction of good samples (in fact at least half of the total good samples)
- $\sum |S_i|^2 \leq |S|^2$
Analysis

At each step:

- At least one $S_i$ has an $\alpha$-fraction of good samples (in fact at least half of the total good samples)
- $\sum |S_i|^2 \leq |S|^2$

When return if:

- $S_i$ has $\alpha$-fraction of good samples AND
- $\hat{\Sigma}_i$ has no large eigenvalues
Analysis

At each step:

- At least one $S_i$ has an $\alpha$-fraction of good samples (in fact at least half of the total good samples)
- \[ \sum |S_i|^2 \leq |S|^2 \]

When return if:

- $S_i$ has $\alpha$-fraction of good samples AND
- $\hat{\Sigma}_i$ has no large eigenvalues

Then for all $|\mathbf{v}| = 1$,

\[ \log(1/\alpha) \gg \text{Var}(\mathbf{v} \cdot S_i) \geq \alpha[\mathbf{v} \cdot (\mu_{S_i} - \mu)]^2, \]

so

\[ |\mu_{S_i} - \mu| = O(\alpha^{-1/2} \sqrt{\log(1/\alpha)}). \]
Note: [CSV17] Show how to do list decoding in the stochastic optimization setting (an $\alpha$-fraction of functions come from good distribution). Additional work shows how to do this with filters.
Learning Mixtures of Spherical Gaussians

Application: Let $X = \frac{1}{k} \sum_{i=1}^{k} G_i$ with each $G_i \sim N(\mu_i, I)$. 
Learning Mixtures of Spherical Gaussians

Application: Let $X = \frac{1}{k} \sum_{i=1}^{k} G_i$ with each $G_i \sim N(\mu_i, I)$. Want to learn the $\mu_i$. 
[Regev-Vijayraghavan ’17] show information-theoretically impossible to learn the means unless have separation $\Omega(\sqrt{\log(k)})$. 

[Regev-Vijayraghavan ’17] show how to improve a rough approximation to $\mu_i$ to a precise one.

[Vempala-Wang ’02] Give algorithm with separation $\Omega(k^{1/4})$.

Question: How much separation is actually needed?
[Regev-Vijayraghavan ’17] show information-theoretically impossible to learn the means unless have separation \( \Omega(\sqrt{\log(k)}) \).

[Regev-Vijayraghavan ’17] show how to improve a rough approximation to \( \mu_i \) to a precise one.
History

- [Regev-Vijayraghavan ’17] show information-theoretically impossible to learn the means unless have separation $\Omega(\sqrt{\log(k)})$.
- [Regev-Vijayraghavan ’17] show how to improve a rough approximation to $\mu_i$ to a precise one.
- [Vempala-Wang ’02] Give algorithm with separation $\Omega(k^{1/4})$. 
History

- [Regev-Vijayraghavan ’17] show information-theoretically impossible to learn the means unless have separation $\Omega(\sqrt{\log(k)})$.
- [Regev-Vijayraghavan ’17] show how to improve a rough approximation to $\mu_i$ to a precise one.
- [Vempala-Wang ’02] Give algorithm with separation $\Omega(k^{1/4})$.

**Question:** How much separation is actually needed?
Run list decoding algorithm. Since $X$ is a noisy version of each $G_i$, our list contains approximations to all means with error $D$. 
Clustering

Round samples to nearest hypothesis. With high probability samples round to one of hypotheses within $O(D)$ of the mean.

Cluster used hypotheses.

Recover original Gaussians to estimate means.
Clustering

Round samples to nearest hypothesis. With high probability samples round to one of hypotheses within $O(D)$ of the mean.
Clustering

Round samples to nearest hypothesis. With high probability samples round to one of hypotheses within $O(D)$ of the mean. Cluster used hypotheses.
Clustering

Round samples to nearest hypothesis. With high probability samples round to one of hypotheses within $O(D)$ of the mean. Cluster used hypotheses. Recover original Gaussians to estimate means.
If we can do list decoding with $\alpha = 1/k$ and error $D$, we can learn equal mixtures of $k$ Gaussians with separation $\Omega(D)$. 
Results

If we can do list decoding with $\alpha = 1/k$ and error $D$, we can learn equal mixtures of $k$ Gaussians with separation $\Omega(D)$.

This Talk: We showed how to do this with $D \approx k^{1/2}$.
Next talk: We will show how to achieve $D = k^\epsilon$. 


