Average-Case Averages: Private Algorithms for Smooth Sensitivity and Mean Estimation

Thomas Steinke

IBM Research – Almaden

http://stein.ke/

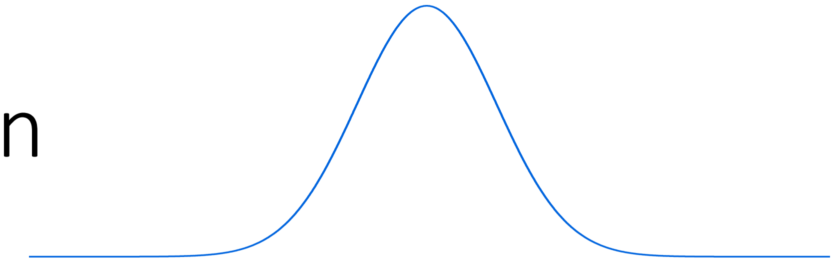
Based on joint work with Mark Bun https://arxiv.org/abs/1906.02830

Talk Outline

- Motivating example: (Gaussian) mean estimation
- Trimmed mean
- Smooth Sensitivity & Differential Privacy
- New Smooth Sensitivity-based algorithms
- Applying Smooth Sensitivity to Gaussian mean estimation
- Conclusion & further work

- **Theme of this work: Connecting robustness and privacy.**

(Gaussian) Mean Estimation



- Data: $X_1, X_2, \cdots, X_n$ i.i.d. samples from $N(\mu, \sigma^2)$.
- Goal: Learn μ.
- Know: $\mu \in [a, b]$ and σ (for simplicity).
- Constraint: Must satisfy ε-differential privacy or similar.

- Extremely fundamental task. Embarrassingly under-studied.
- Note: Distributional assumption on data for utility, but privacy must hold for any input.

Preview: Our Algorithm for Gaussian Mean

Theorem. Let $n \geq O(\log((b-a)/\sigma)\,/\varepsilon)$. Then there exists a ε-DP (or $\frac{1}{2}\varepsilon^2$-CDP) algorithm $M : \mathbb{R}^n \to \mathbb{R}$ such that, for all $\mu \in [a, b]$, we have

$$\mathbf{E}[(M(X)-\mu)^2] \leq \frac{\sigma^2}{n} + \frac{\sigma^2}{n^2} \cdot O\left(\frac{\log\left(\frac{b-a}{\sigma}\right)}{\varepsilon} + \frac{\log n}{\varepsilon^2}\right)$$

when $X \leftarrow N(\mu, \sigma^2)^n$.

- Matches previous work [Karwa-Vadhan18].
- Extends to unknown σ.
- **Extends to non-Gaussian data.**

Non-Privately: Empirical Mean

- Data: $X_1, X_2, \cdots, X_n$ i.i.d. samples from $N(\mu, \sigma^2)$
- Non-private estimator: $\bar{X} = \frac{1}{n}\sum_i^n X_i$
- Unbiased $\mathbf{E}[\bar{X}] = \mu$ and minimal variance $\mathbf{Var}[\bar{X}] = \frac{\sigma^2}{n}$

- Problem: Global sensitivity = ∞ so cannot just add noise to achieve DP.
- In contrast, for distribution with bounded support $[a, b]$ simple ε-DP algorithm: $M(x) = \bar{x} + \mathrm{Lap}\left(\frac{b-a}{\varepsilon n}\right)$.

Truncation [Karwa-Vadhan18]

- Step 1: Obtain crude estimate $\tilde{\mu} \in [\mu \pm O(\sigma)]$.
- Step 2: Truncate data $X_1, X_2, \cdots, X_n$ to $[\tilde{\mu} \pm O(\sigma\sqrt{\log n})]$.
- Step 3: Add noise to empirical mean with scale $O\left(\frac{\sigma\sqrt{\log n}}{\varepsilon n}\right)$.

- Note $\sqrt{\log n}$ factor comes from Gaussian tail bound.
- This approach doesn't extend well to heavy-tailed distributions.

Our Approach: Trimmed Mean

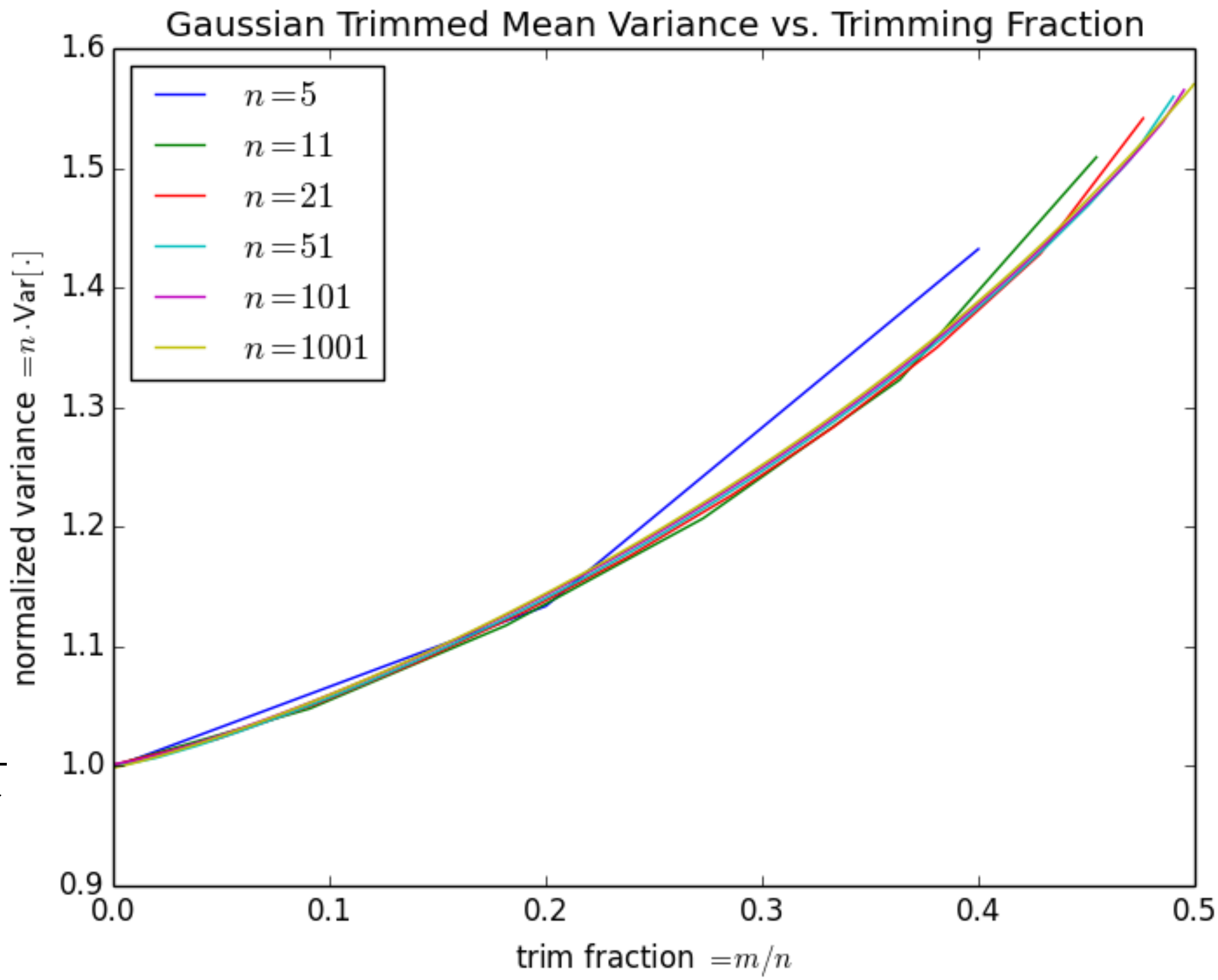
- Intuition: Outlier removal. Remove top m and bottom m.

Define $\mathrm{trim}_m : \mathbb{R}^n \to \mathbb{R}$ by

$$\mathrm{trim}_m(x) = \frac{x_{(m+1)} + x_{(m+2)} + \cdots + x_{(n-m)}}{n - 2m}$$

where $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ is the order statistics of x.

- Interpolates between mean ($m = 0$) and median ($m = \frac{n-1}{2}$).
- Unbiased: $\mathbf{E}[\mathrm{trim}_m(X)] = \mu$ for $X \leftarrow N(\mu, \sigma^2)^n$.
- Variance: $\mathbf{Var}[\mathrm{trim}_m(X)] = \frac{\sigma^2}{n} \cdot \left(1 + O\left(\frac{m}{n}\right)\right)$.
- (This holds for any symmetric distribution.)



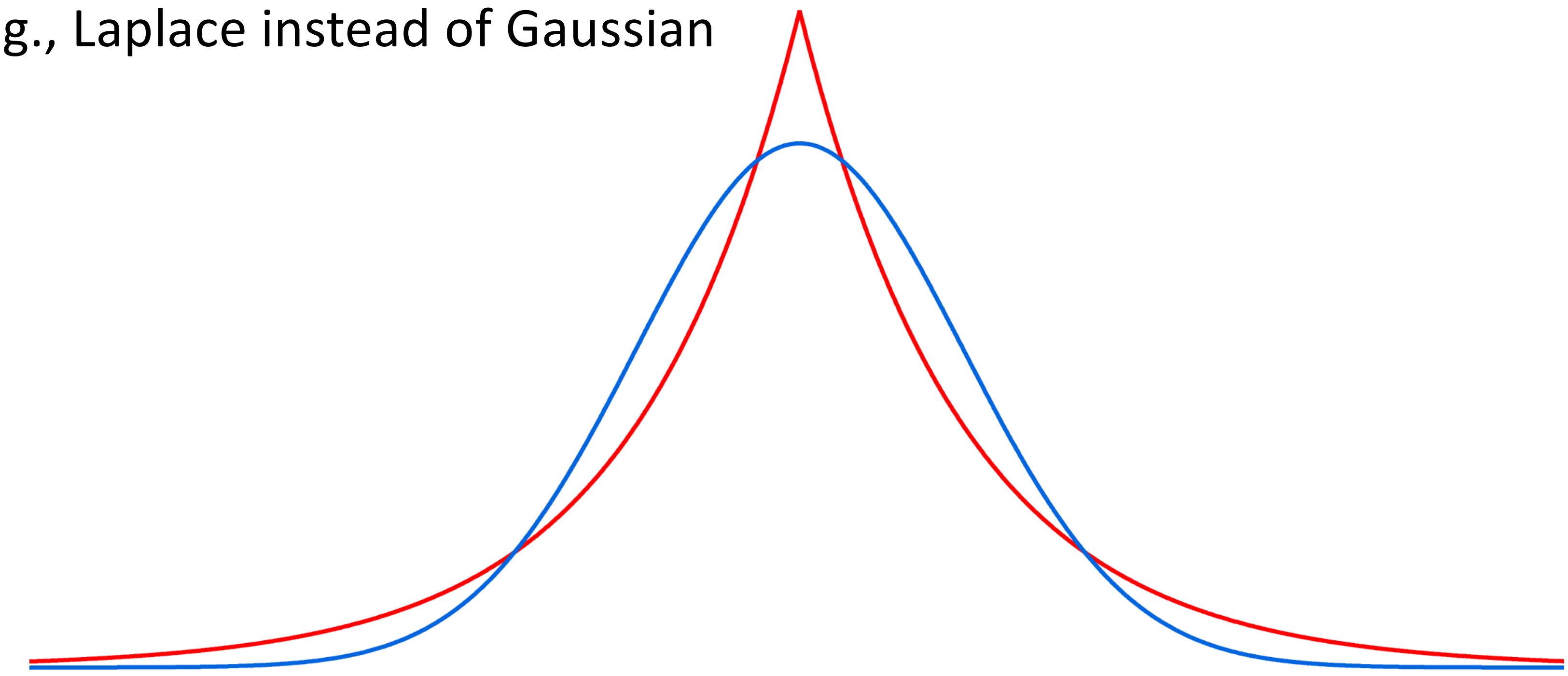
$$\mathbf{Var}[\text{median}] \approx \frac{\pi}{2n}$$

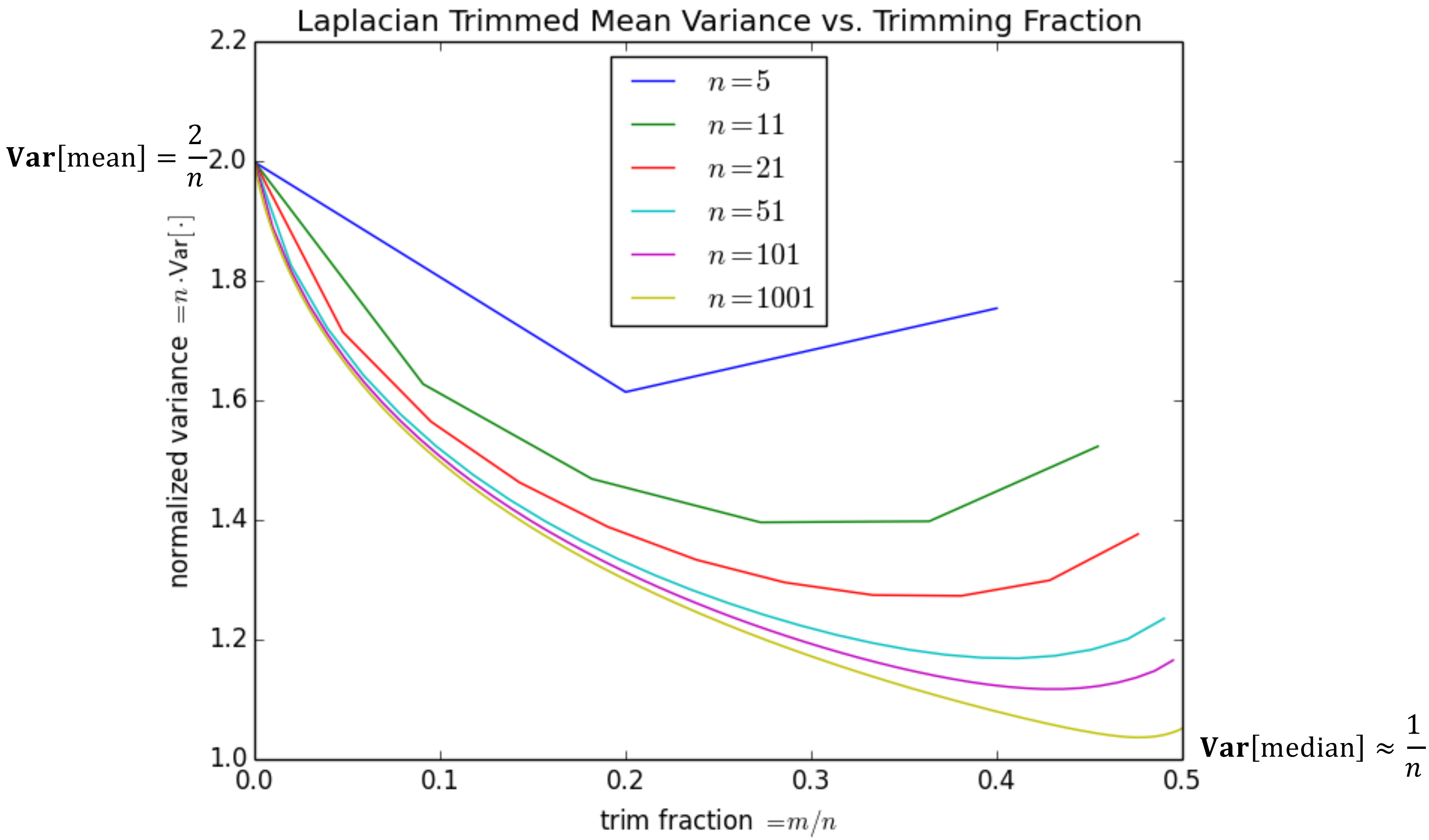
$$\mathbf{Var}[\text{mean}] = \frac{1}{n}$$

Trimmed Mean for Non-Gaussians

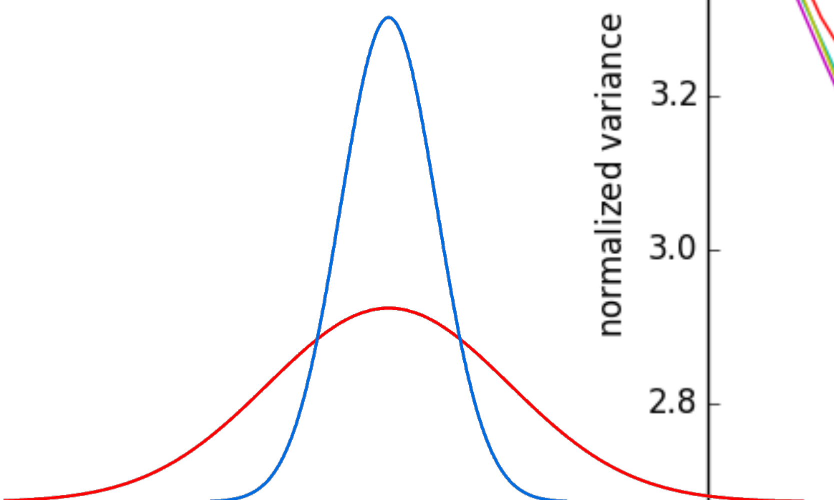
Trimming can actually reduce variance if data is heavy-tailed!

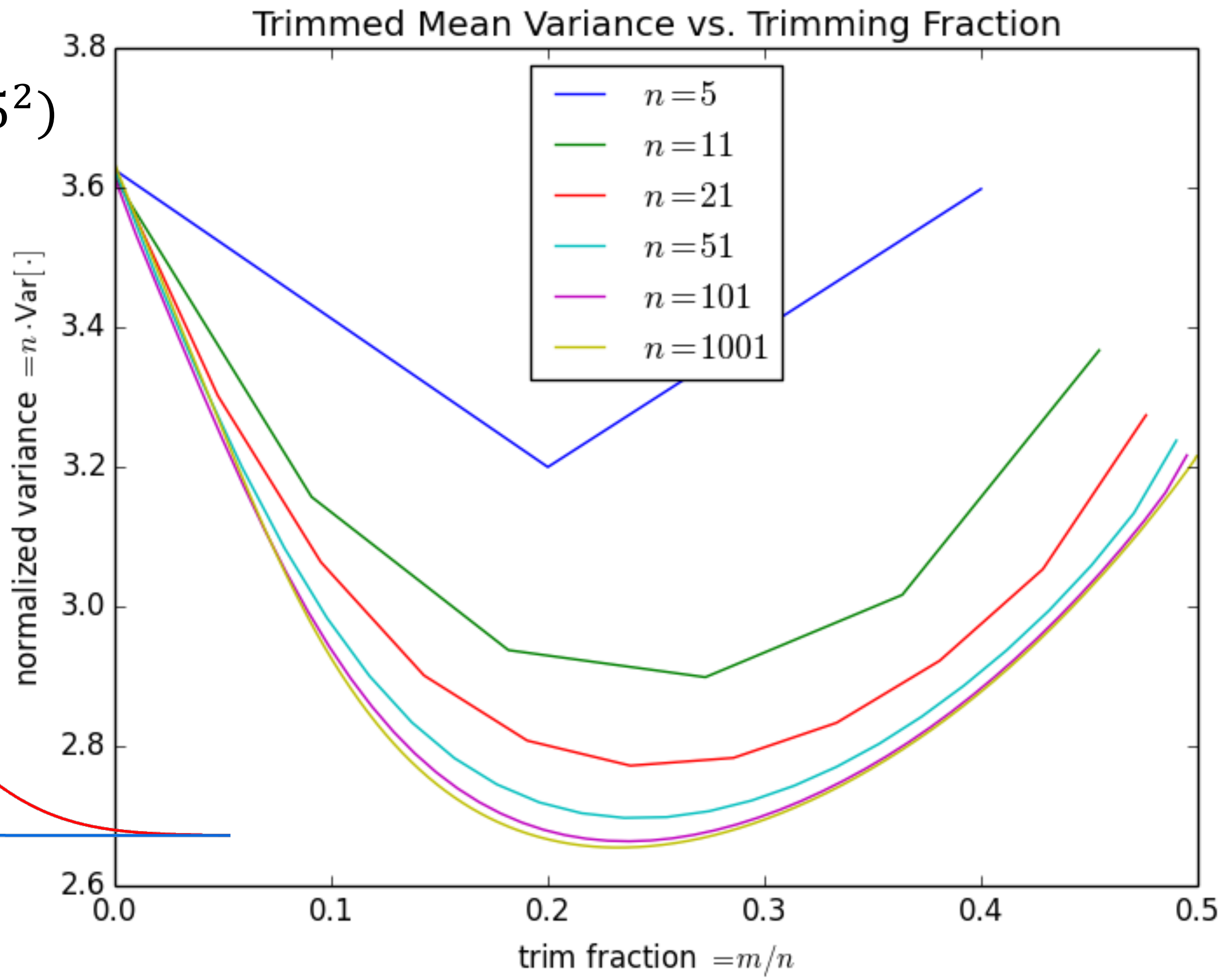
E.g., Laplace instead of Gaussian




Laplacian Trimmed Mean Variance vs. Trimming Fraction
$\mathbf{Var}[\text{mean}] = \frac{2}{n}$
2.2
2.0
1.8
1.6
1.4
1.2
1.0
normalized variance $= n \cdot \mathrm{Var}[\cdot]$
$n=5$
$n=11$
$n=21$
$n=51$
$n=101$
$n=1001$
0.0
0.1
0.2
0.3
0.4
0.5
trim fraction $= m/n$
$\mathbf{Var}[\text{median}] \approx \frac{1}{n}$

$$\frac{1}{2}\mathcal{N}(0,1) + \frac{1}{2}\mathcal{N}(0,2.5^2)$$





Sensitivity of Trimmed Mean?

- Consider large but bounded domain: $\mathrm{trim}_m : [a, b]^n \to [a, b]$

$$\mathrm{trim}_m(x) = \frac{x_{(m+1)} + x_{(m+2)} + \cdots + x_{(n-m)}}{n - 2m}$$

- Global sensitivity: Large, but finite.

$$GS = \max_{x,x'} |\mathrm{trim}_m(x) - \mathrm{trim}_m(x')| = \frac{b - a}{n - 2m}$$

- Local sensitivity: Often much smaller.

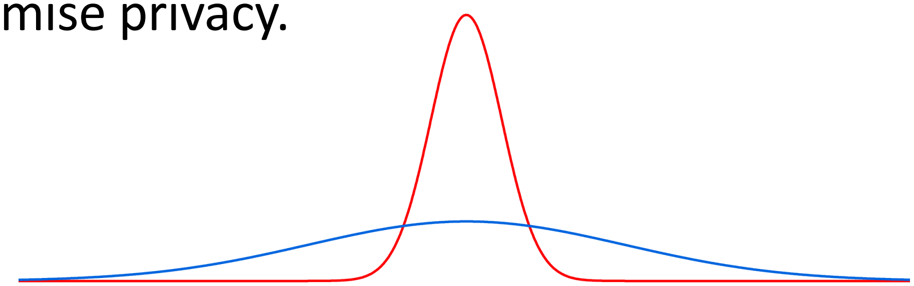
$$LS(x) = \max_{x'} |\mathrm{trim}_m(x) - \mathrm{trim}_m(x')|$$

$$= \frac{\max\{x_{(n-m)} - x_{(m)}, x_{(n-m+1)} - x_{(m+1)}\}}{n - 2m}$$

Smooth Sensitivity

Idea: Smooth Sensitivity [NRS07]

- Can add noise proportional to global sensitivity to attain DP [DMNS06].
- We would like to be able to add noise proportional to local sensitivity.
- Problem: The local sensitivity may itself be high sensitivity. I.e., noise magnitude may compromise privacy.



- Solution: Smooth Sensitivity [Nissim-Raskhodnikova-Smith07]
- Powerful and elegant idea.
- **This work: Getting more mileage out of Smooth Sensitivity.**

Smooth Sensitivity [Nissim-Raskhodnikova-Smith07]

Let $f, g: X^n \to \mathbb{R}$ satisfy, for all neighbouring $x, x' \in X^n$,

$$|f(x) - f(x')| \le g(x) \quad \text{and} \quad e^{-t} g(x) \le g(x') \le e^t g(x).$$

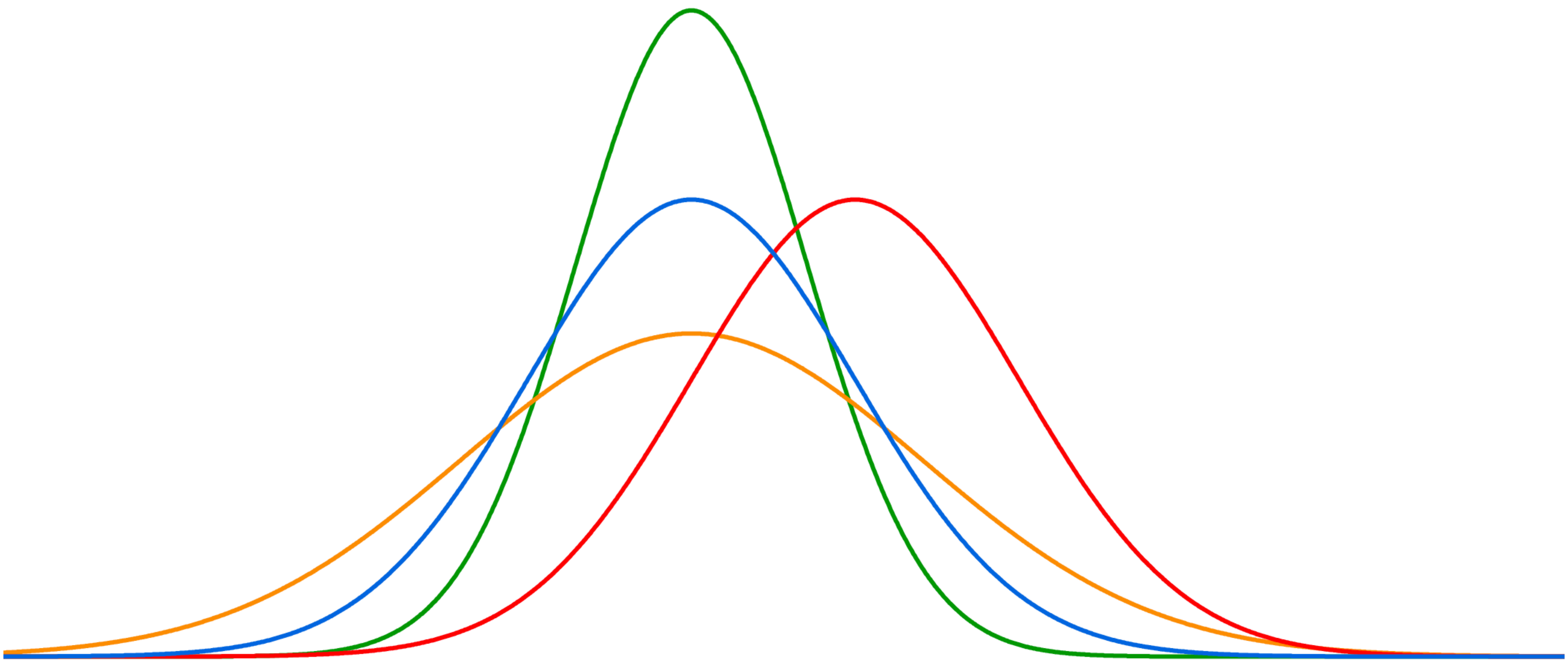
Then we say g is a t-smooth upper bound on the local sensitivity of f.

Smooth sensitivity algorithm:

$$M(x) = f(x) + Z \cdot g(x)$$

- But what noise distribution Z can we use?
- To satisfy ε-DP, we need $Z \approx_\varepsilon Z + s$ (as usual) and also $Z \approx_\varepsilon e^t Z$.

Additive and multiplicative distortions



Smooth Sensitivity Distributions [NRS07]

- Cauchy: density $\propto \frac{1}{1+x^2}$, sample $\mathrm{Z} = \frac{X}{Y}$ for i.i.d. $X, Y \leftarrow N(0,1)$.
  - Provides pure ε-DP.
  - Infinite variance, even mean not well defined!
- More generally: density $\propto \frac{1}{1+|x|^\gamma}$
  - Provides pure ε-DP.
  - $\mathbf{E}[|Z|^p] < \infty$ for all $p < \gamma - 1$.
  - Not all moments exist (inherent for pure ε-DP).
- Laplace, Gaussian
  - Provide approximate (ε, δ)-DP. (Need to pick & pay for δ.)

New Smooth Sensitivity Distributions

- Student's T: density $\propto \left(\frac{1}{d+x^2}\right)^{\frac{d+1}{2}}$, $Z = \frac{X}{\sqrt{Y_1^2+Y_2^2+\cdots+Y_d^2}}$ for $X, Y_1, \cdots, Y_d \leftarrow N(0,1)$.
  - Provides pure ε-DP.
  - $\mathbf{E}[|Z|^p] < \infty$ for all $p < d$.
- Laplace-logNormal: $Z = X \cdot e^{\sigma Y}$ for $X =$ Laplace, $Y =$ Gaussian
  - Provides concentrated DP.
  - $\mathbf{E}[|Z|^p] < \infty$ for all p.
- Uniform-logNormal: $Z = X \cdot e^{\sigma Y}$ for $X =$ Uniform$([-1,1])$, $Y =$ Gaussian
  - Provides concentrated DP. (Analysis not quite as good as Laplace-logNormal.)
  - $\mathbf{E}[|Z|^p] < \infty$ for all p.
- arsinhNormal: $Z = \sinh X$ for $X =$ Gaussian
  - Provides concentrated DP. (Analysis messier than Laplace-logNormal.)
  - $\mathbf{E}[|Z|^p] < \infty$ for all p.

Aside: Concentrated Differential Privacy

- Concentrated DP is a "best of both worlds" between pure ε-DP and approximate (ε, δ)-DP.
  - Advanced composition, no "death and destruction" δ, no superfluous $\log(\frac{1}{\delta})$ factors.
- Several variants [Dwork-Rothblum16,Bun-S.16,Mironov17,Bun-Dwork-Rothblum-S.18], same underlying ideas.
  - Use Rényi divergences from information theory.

M is $\frac{1}{2}\varepsilon^2$-CDP if, for all neighbouring x, x',

$$\forall \alpha > 1 \qquad \mathrm{D}_\alpha(M(x)||M(x')) \leq \frac{1}{2}\varepsilon^2\alpha$$

Laplace-logNormal Privacy Analysis

Laplace-logNormal: $Z = X \cdot e^{\sigma Y}$ for $X =$ Laplace, $Y =$ Gaussian

- Show that Z provides $\frac{1}{2}\varepsilon^2$-CDP with Smooth Sensitivity:

Theorem. $\mathrm{D}_\alpha(Z||e^t Z + s) \leq \frac{1}{2}\varepsilon^2 \alpha$ for all $\alpha > 1$, where

$$\varepsilon = \frac{|t|}{\sigma} + e^{\frac{3}{2}\sigma^2} \cdot |s|$$

- logNormal deals with multiplicative distortion:
$\mathrm{D}_\alpha(Z||e^t Z) = \mathrm{D}_\alpha(X \cdot e^{\sigma Y} \, ||X \cdot e^{\sigma Y + t})$

$$\leq \max_x \mathrm{D}_\alpha(x \cdot e^{\sigma Y} \, ||x \cdot e^{\sigma Y + t}) = \mathrm{D}_\alpha(\sigma Y || \sigma Y + t) = \frac{\alpha t^2}{2\sigma^2}$$

- Get pure DP for additive distortion: $\mathrm{D}_\infty(Z||Z + s) \leq e^{\frac{3}{2}\sigma^2} |s|$
- Apply triangle inequality (a.k.a. group privacy) to combine.

Tails of Smooth Sensitivity + CDP Distributions

All these distributions satisfying concentrated DP – Laplace-logNormal, Uniform-logNormal, & arsinhNormal – have quasi-polynomial tails:

$$\mathbf{P}[|Z| > z] = e^{-\Theta(\log z)^2}$$

Moments: $\mathbf{E}[|Z|^p] = e^{\Theta(p^2)}$. (For pure DP, polynomial tails & infinite moments.)

This is necessary. Lower bound:

- Group privacy: $Z \approx_\varepsilon Z + s \approx_\varepsilon e^t(Z+s) \approx_\varepsilon e^{2t}(Z+s) \approx_\varepsilon \cdots \approx_\varepsilon e^{kt}(Z+s)$
- $\mathrm{p} = \mathbf{P}[Z \geq z] \approx_{k\varepsilon} \mathbf{P}[e^{kt}(Z+s) \geq z] = \mathbf{P}[Z \geq e^{-kt}z - s] \geq \mathbf{P}[Z \geq 0] \geq \frac{1}{2}$
- Can set $k = \frac{\log z - \log s}{t}$ and use group privacy bound:

$$\mathrm{D}_1\left(\frac{1}{2}\middle\|\, p\right) = \frac{1}{2}\log\left(\frac{1}{2p}\right) + \frac{1}{2}\log\left(\frac{1}{2(1-p)}\right) \leq \frac{1}{2}k^2\varepsilon^2$$

- Rearrange: $p \geq p(1-p) \geq \frac{1}{4}e^{-k^2\varepsilon^2} = e^{-O(\log z)^2}$

Smooth Sensitivity
+ Trimmed Mean

Smooth Algorithm for Gaussian Mean

Theorem. Let $n \geq O(\log((b-a)/\sigma)\,/\varepsilon)$. Then there exists a ε-DP (or $\frac{1}{2}\varepsilon^2$-CDP) algorithm $M : \mathbb{R}^n \to \mathbb{R}$ such that, for all $\mu \in [a, b]$, we have

$$\mathbf{E}[(M(X)-\mu)^2] \leq \frac{\sigma^2}{n} + \frac{\sigma^2}{n^2} \cdot O\left(\frac{\log\left(\frac{b-a}{\sigma}\right)}{\varepsilon} + \frac{\log n}{\varepsilon^2}\right)$$

when $X \leftarrow N(\mu, \sigma^2)^n$.

- Matches previous work [Karwa-Vadhan18].
- Unknown $\sigma \in [\sigma_{min}, \sigma_{max}]$: $\log\left(\frac{b-a}{\sigma}\right)$ becomes $\log\left(\frac{b-a}{\sigma_{min}}\right) + \log\left(\frac{\sigma_{max}}{\sigma_{min}}\right)$
- Not specific to Gaussian data. Only use symmetry and tail bound.

Smooth Algorithm for General Means

Theorem. Let $n \geq O(\log(n(b-a)/\sigma)\,/\varepsilon)$. Then there exists a ε-DP (or $\frac{1}{2}\varepsilon^2$-CDP) algorithm $M : \mathbb{R}^n \to \mathbb{R}$ such that, for all $\mu \in [a, b]$, we have

$$\mathbf{E}[(M(X)-\mu)^2] \leq \frac{\sigma^2}{n} \cdot O\left(\frac{\log\left(n\frac{b-a}{\sigma}\right)}{\varepsilon} + \frac{1}{\varepsilon^2}\right)$$

when $X \leftarrow D^n$ and D is any distribution with mean μ and variance σ^2.

- Matches previous work [Feldman-Steinke18].
- This result uses the same algorithm as for Gaussians!
- Algorithm matches distribution and can interpolate between results.

Smooth Sensitivity Algorithm

Let $f, g: X^n \to \mathbb{R}$ satisfy, for all neighbouring $x, x' \in X^n$,

$$|f(x) - f(x')| \leq g(x) \quad \text{and} \quad e^{-t} g(x) \leq g(x') \leq e^{t} g(x).$$

Then we say g is a t-smooth upper bound on the local sensitivity of f.

Smooth sensitivity algorithm:

$$M(x) = f(x) + Z \cdot g(x)$$

For $X \leftarrow N(\mu, \sigma^2)$ and $\mathbf{E}[Z] = 0$,

Non-private error of trimmed mean

$\frac{\sigma^2}{n} \cdot \left(1 + O\left(\frac{m}{n}\right)\right)$

$O\left(\frac{1}{\varepsilon^2}\right)$

$$\mathbf{E}[(M(X) - \mu)^2] = \mathbf{E}[(f(X) - \mu)^2] + \mathbf{E}[g(X)^2] \cdot \mathbf{Var}[Z]$$

???

Smooth Sensitivity of Trimmed Mean

- Consider large but bounded domain: $\mathrm{trim}_m: [a,b]^n \to [a,b]$

$$\mathrm{trim}_m(x) = \frac{x_{(m+1)} + x_{(m+2)} + \cdots + x_{(n-m)}}{n-2m}$$

- Local sensitivity:

$$LS(x) = \frac{\max\{x_{(n-m)} - x_{(m)}, x_{(n-m+1)} - x_{(m+1)}\}}{n-2m}$$

- Smooth sensitivity:

$$g(x) = SS(x,t) = \max_{x'} e^{-t\|x'-x\|} LS(x')$$

$$= \max_{0 \le k \le n} e^{-kt} \max_{0 \le \ell \le \mathrm{k}+1} \frac{x_{(n-m+k+1-\ell)} - x_{(m+1-\ell)}}{n-2m}$$

where $x_{(1-i)} = a$ and $x_{(n+i)} = b$ for $i \ge 1$.

Smooth Sensitivity of Trimmed Mean

Smooth sensitivity:

$$g(x) = SS(x,t) = \max_{x'} e^{-t\|x'-x\|} LS(x')$$

$$= \max_{0 \le k \le n} e^{-kt} \max_{0 \le \ell \le k+1} \frac{x_{(n-m+k+1-\ell)} - x_{(m+1-\ell)}}{n-2m}$$

where $x_{(1-i)} = a$ and $x_{(n+i)} = b$ for $i \geq 1$.

Loose (but sufficient) bound:

$$g(x)^2 \leq \frac{\left(x_{(n)} - x_{(1)}\right)^2 + e^{-mt}(b-a)^2}{(n-2m)^2}$$

Set $m = O\left(\frac{1}{t}\log\left(\frac{b-a}{\sigma}\right)\right)$ and $t = \frac{\varepsilon}{2}$

For $X \leftarrow N(\mu, \sigma^2)$,

$$\mathbf{E}[g(X)^2] \leq \frac{\sigma^2 \cdot \mathrm{O}(\log n) + e^{-mt}(b-a)^2}{(n-2m)^2}$$

Smooth Algorithm for Gaussian Mean

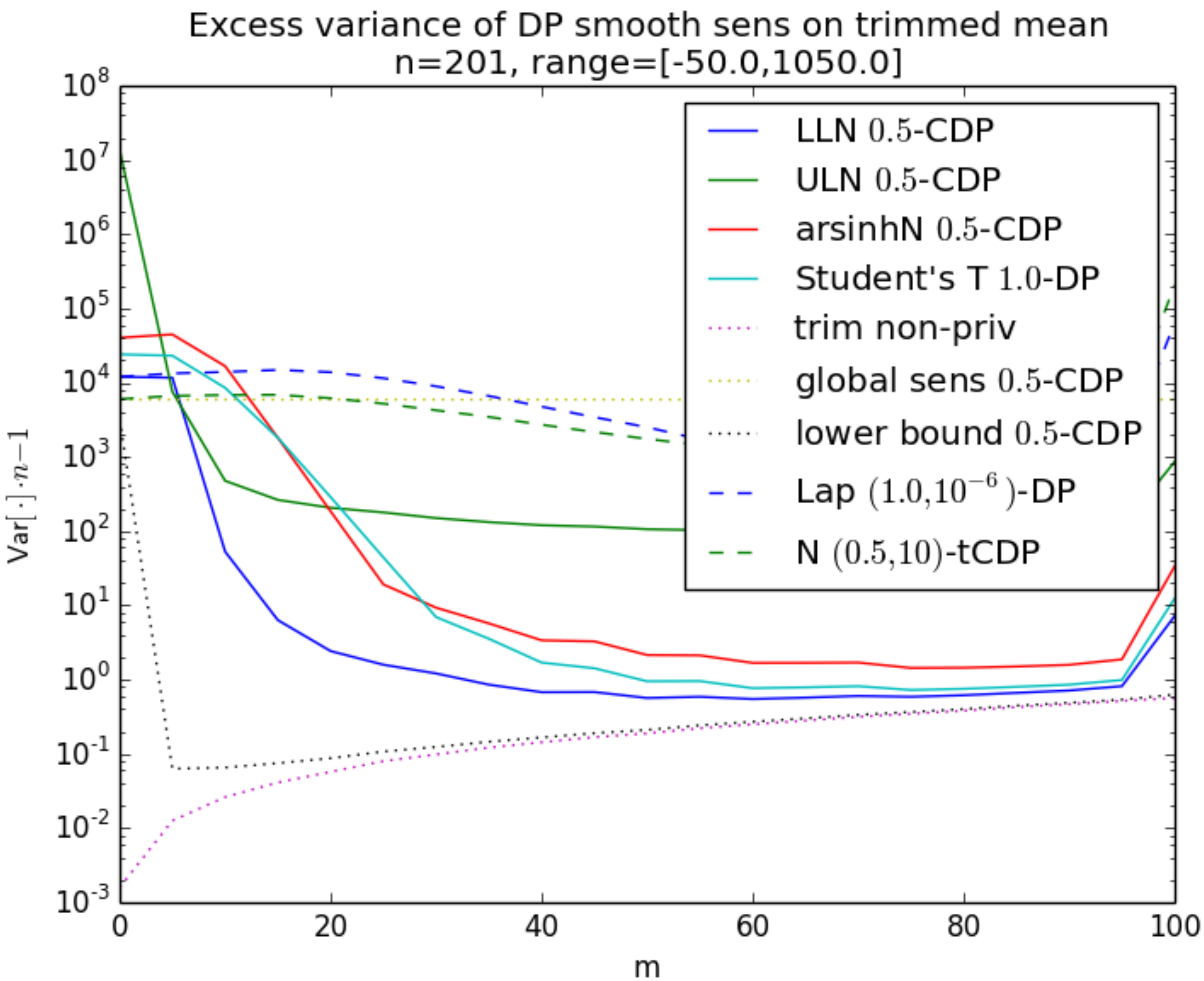
Theorem. Let $n \geq O(\log((b-a)/\sigma)\,/\varepsilon)$. Then there exists a ε-DP (or $\frac{1}{2}\varepsilon^2$-CDP) algorithm $M : \mathbb{R}^n \to \mathbb{R}$ such that, for all $\mu \in [a, b]$, we have

$$\mathbf{E}[(M(X)-\mu)^2] \leq \frac{\sigma^2}{n} + \frac{\sigma^2}{n^2} \cdot O\left(\frac{\log\left(\frac{b-a}{\sigma}\right)}{\varepsilon} + \frac{\log n}{\varepsilon^2}\right)$$

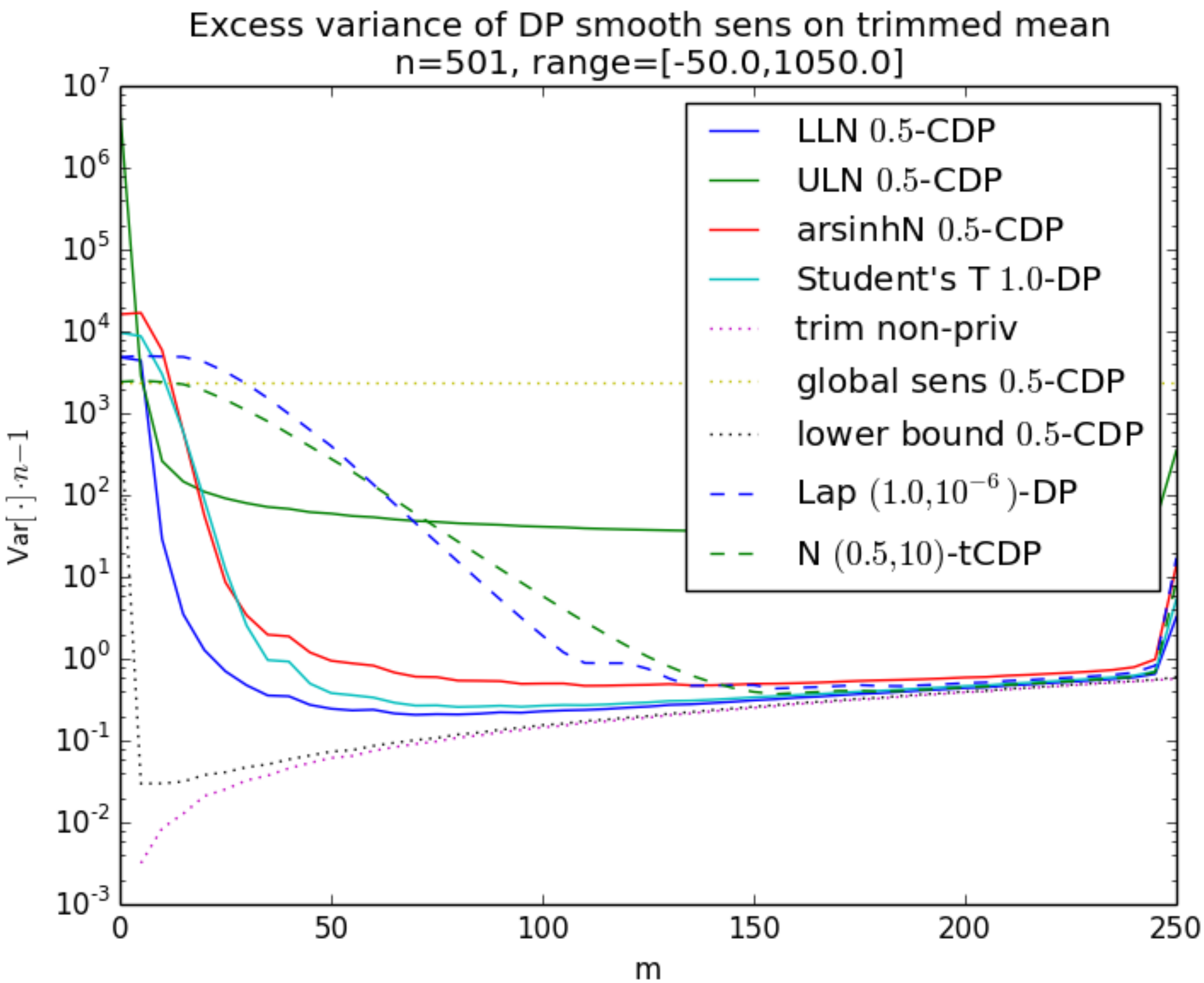
when $X \leftarrow N(\mu, \sigma^2)^n$.

- Matches previous work [Karwa-Vadhan18].
- Unknown $\sigma \in [\sigma_{min}, \sigma_{max}]$: $\log\left(\frac{b-a}{\sigma}\right)$ becomes $\log\left(\frac{b-a}{\sigma_{min}}\right) + \log\left(\frac{\sigma_{max}}{\sigma_{min}}\right)$
- Not specific to Gaussian data. Only use symmetry and tail bound.

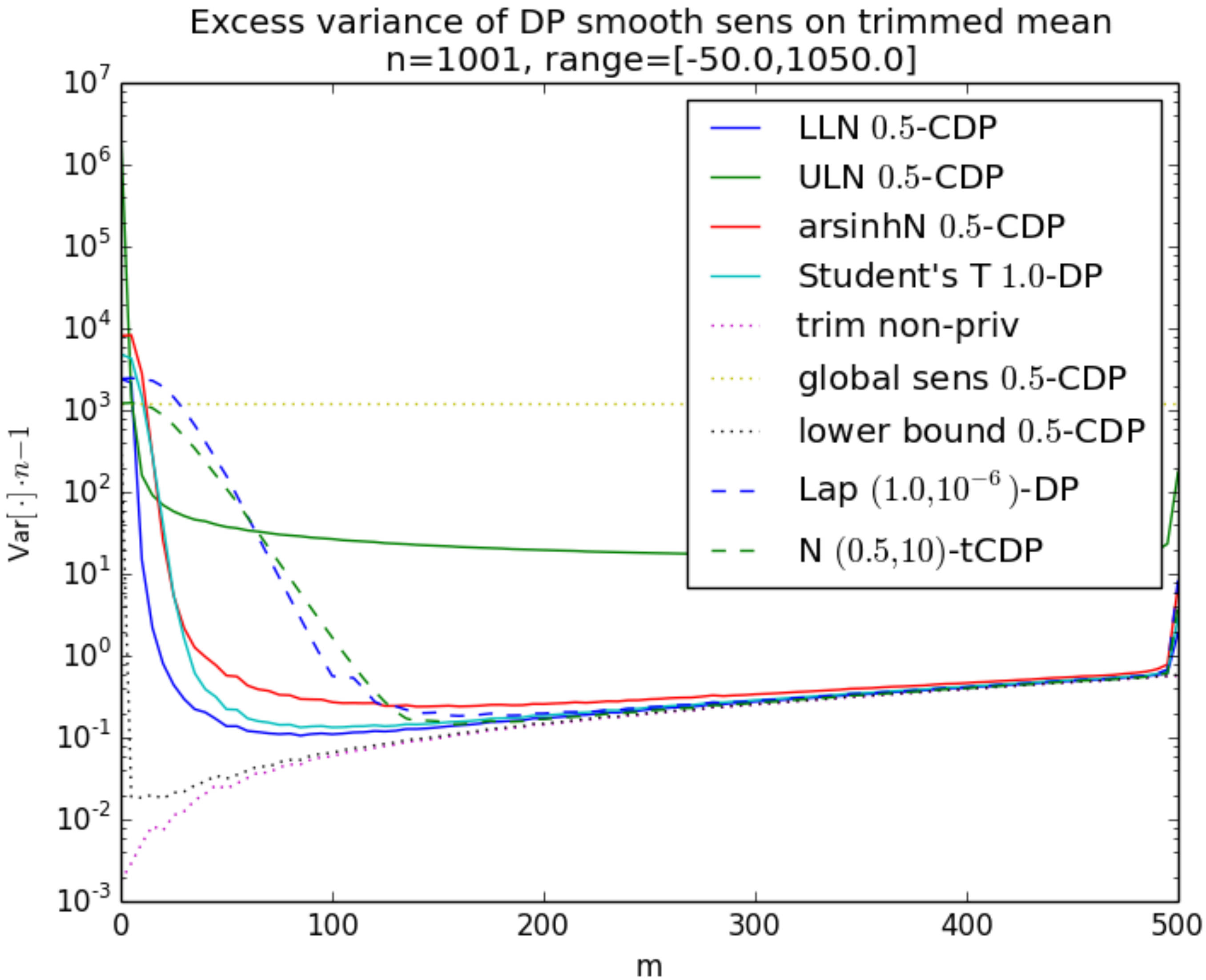
Some Experimental Plots!



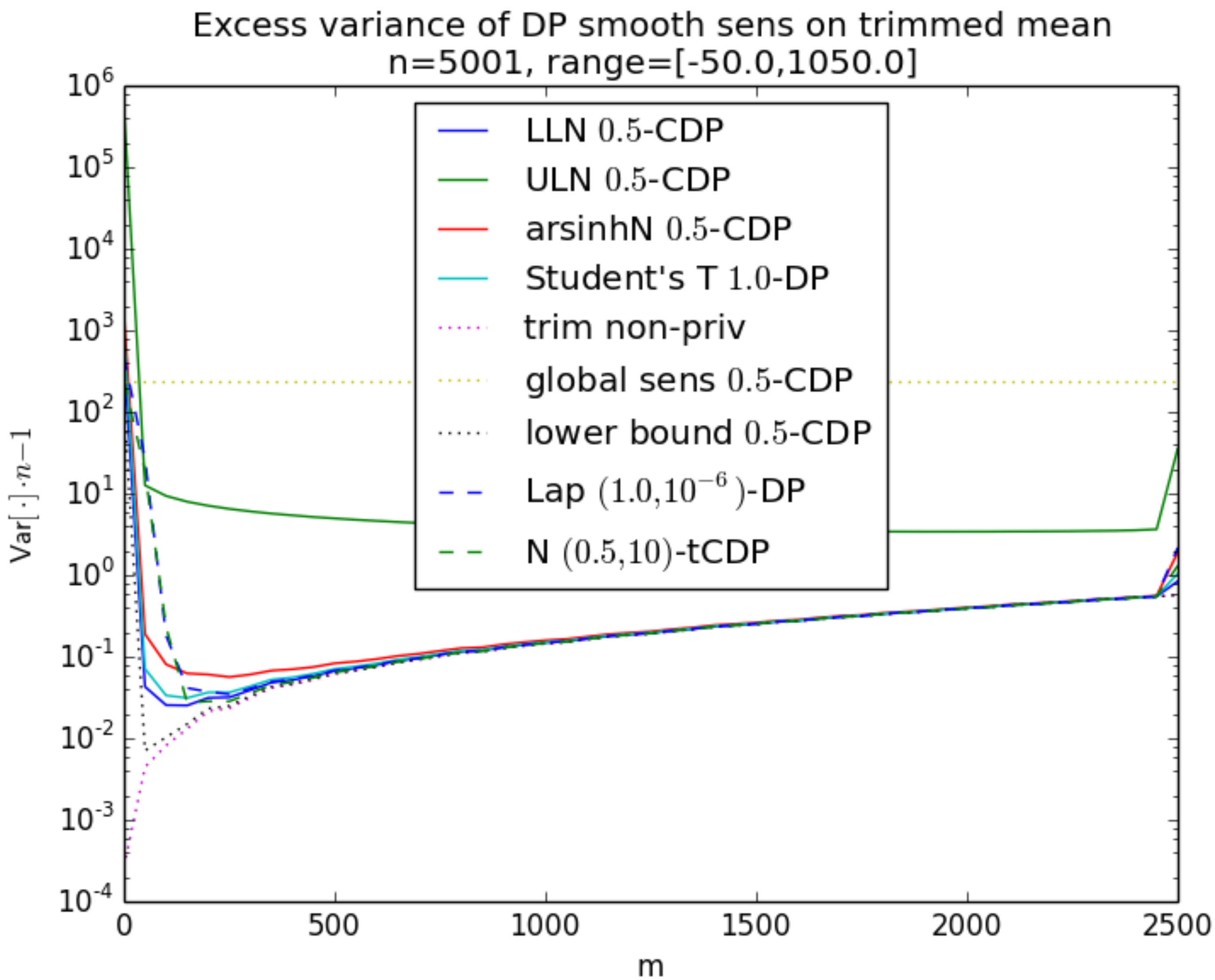
Excess variance of DP smooth sens on trimmed mean
n=201, range=[-50.0,1050.0]
LLN 0.5-CDP
ULN 0.5-CDP
arsinhN 0.5-CDP
Student's T 1.0-DP
trim non-priv
global sens 0.5-CDP
lower bound 0.5-CDP
Lap (1.0,10^-6)-DP
N (0.5,10)-tCDP
Var[·]·n−1
m

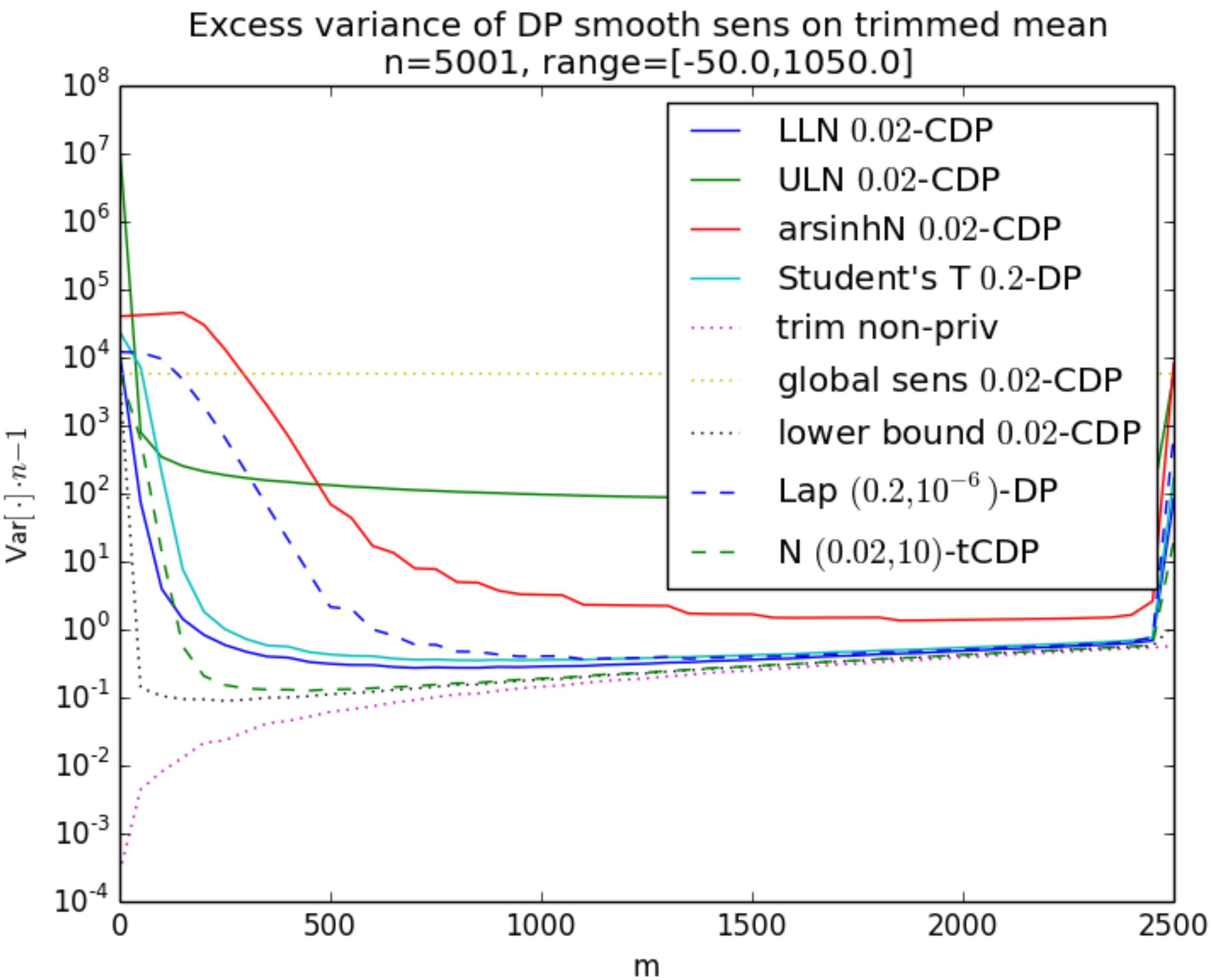


Excess variance of DP smooth sens on trimmed mean
n=501, range=[-50.0,1050.0]
LLN 0.5-CDP
ULN 0.5-CDP
arsinhN 0.5-CDP
Student's T 1.0-DP
trim non-priv
global sens 0.5-CDP
lower bound 0.5-CDP
Lap (1.0,10^-6)-DP
N (0.5,10)-tCDP
Var[·]·n−1
m



Excess variance of DP smooth sens on trimmed mean
n=1001, range=[-50.0,1050.0]
LLN 0.5-CDP
ULN 0.5-CDP
arsinhN 0.5-CDP
Student's T 1.0-DP
trim non-priv
global sens 0.5-CDP
lower bound 0.5-CDP
Lap (1.0,10^-6)-DP
N (0.5,10)-tCDP
Var[·]·n−1
m


Excess variance of DP smooth sens on trimmed mean
n=5001, range=[-50.0,1050.0]
LLN 0.5-CDP
ULN 0.5-CDP
arsinhN 0.5-CDP
Student's T 1.0-DP
trim non-priv
global sens 0.5-CDP
lower bound 0.5-CDP
Lap $(1.0,10^{-6})$-DP
N $(0.5,10)$-tCDP
$\mathrm{Var}[\cdot]\cdot n-1$
m
10^{6}
10^{5}
10^{4}
10^{3}
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0
500
1000
1500
2000
2500



Excess variance of DP smooth sens on trimmed mean
n=5001, range=[-50.0,1050.0]
LLN 0.02-CDP
ULN 0.02-CDP
arsinhN 0.02-CDP
Student's T 0.2-DP
trim non-priv
global sens 0.02-CDP
lower bound 0.02-CDP
Lap $(0.2, 10^{-6})$-DP
N $(0.02, 10)$-tCDP
$\mathrm{Var}[\cdot] \cdot n-1$
m

Conclusion

- Smooth Sensitivity is great!
- New distributions for use with smooth sensitivity.
- Application to mean estimation.

Further work:

- Sharper upper/lower bounds for Gaussian mean estimation?
- Other applications of smooth sensitivity?
- E.g., scale estimation, confidence intervals [Karwa-Vadhan18], model fitting/regression, multivariate distributions
- Other noise distributions (or better analyses of these ones).

Thanks!