Estimating a mean, privately

Also mostly optimally

John Duchi
Based on joint work with Rohith Kuditipudi and Saminul Haque
Stanford University

Problem in this talk: estimating a mean

- Sample $X_i \overset{\text{iid}}{\sim} P$ with mean $\mu = \mathbb{E}[X] \in \mathbb{R}^d$ and covariance $\mathrm{Cov}(X_i) = \Sigma$
- Measure error with respect to norm $\|v\|_\Sigma^2 := v^T \Sigma^{-1} v$ covariance induces
- Sample mean is efficient: for $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$\mathbb{E}\left[\|\widehat{\mu} - \mu\|_\Sigma^2\right] = \mathbb{E}\left[(\widehat{\mu} - \mu)^T \Sigma^{-1} (\widehat{\mu} - \mu)\right] = \frac{d}{n}$$

- Obvious comment: adaptive to covariance, whatever it is

Problem in this talk: estimating a mean

- Sample $X_i \stackrel{\text{iid}}{\sim} P$ with mean $\mu = \mathbb{E}[X] \in \mathbb{R}^d$ and covariance $\mathrm{Cov}(X_i) = \Sigma$
- Measure error with respect to norm $\|v\|_\Sigma^2 := v^T \Sigma^{-1} v$ covariance induces
- Sample mean is efficient: for $\widehat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i$

$$\mathbb{E}\left[\|\widehat{\mu} - \mu\|_\Sigma^2\right] = \mathbb{E}\left[(\widehat{\mu} - \mu)^T \Sigma^{-1} (\widehat{\mu} - \mu)\right] = \frac{d}{n}$$

- Obvious comment: adaptive to covariance, whatever it is

Challenge: no similarly efficient and adaptive estimator under privacy

Differential Privacy

Dwork, McSherry, Nissim, Smith 06 (Dwork et al. 06b)

- Setting: sample *x* of size *n*, and randomized mechanism M for releasing data

Mechanism M is (ε, δ)-differentially private if for all sets *A*

$$\mathbb{P}(\mathsf{M}(x) \in A) \leq e^{\varepsilon}\mathbb{P}(\mathsf{M}(x') \in A) + \delta$$

whenever $x, x' \in \mathcal{X}^n$ differ in only a single element

Differential Privacy

Dwork, McSherry, Nissim, Smith 06 (Dwork et al. 06b)

- Setting: sample *x* of size *n*, and randomized mechanism M for releasing data

Mechanism M is (ε, δ)-differentially private if for all sets *A*

$$\mathbb{P}(\mathsf{M}(x) \in A) \leq e^{\varepsilon}\mathbb{P}(\mathsf{M}(x') \in A) + \delta$$

whenever $x, x' \in \mathcal{X}^n$ differ in only a single element

Said differently: any test of whether data is *x* or *x*’ based on M(*x*) has

$$\mathbb{P}(\text{Type I error}) + \mathbb{P}(\text{Type II error}) \geq \frac{2}{1 + e^{\varepsilon}} - \delta$$

Basic mechanisms

- Have a function $f : \mathcal{X}^n \to \mathbb{R}$ we wish to estimate with *global sensitivity*

$$\mathrm{GS}_f := \sup_{d_{\mathrm{ham}}(x,x') \le 1} |f(x) - f(x')|$$

- Laplace mechanism (Dwork et al. 06):

$$M(x) := f(x) + \frac{\mathrm{GS}_f}{\varepsilon} \mathsf{Lap}(1)$$

Basic mechanisms

- Have a function $f : \mathcal{X}^n \to \mathbb{R}$ we wish to estimate with *global sensitivity*

$$\mathrm{GS}_f := \sup_{d_{\mathrm{ham}}(x,x')\leq 1} |f(x) - f(x')|$$

- Laplace mechanism (Dwork et al. 06):

$$M(x) := f(x) + \frac{\mathrm{GS}_f}{\varepsilon}\mathsf{Lap}(1)$$

Privacy (when sensitivity is 1):

$$\frac{\mathbb{P}(M(x) \in A)}{\mathbb{P}(M(x') \in A)} = \frac{\int_A \exp(-\varepsilon|f(x) - w|)dw}{\int_A \exp(-\varepsilon|f(x') - w|)dw}$$

$$\leq \sup_w \frac{\exp(-\varepsilon|f(x) - w|)}{\exp(-\varepsilon|f(x') - w|)}$$

$$\leq \exp(\varepsilon|f(x) - f(x')|) \leq \exp(\varepsilon)$$

Basic mechanisms

- Have a function $f : \mathcal{X}^n \to \mathbb{R}$ we wish to estimate with *global sensitivity*

$$\mathrm{GS}_f := \sup_{d_{\mathrm{ham}}(x,x') \le 1} |f(x) - f(x')|$$

- Laplace mechanism (Dwork et al. 06):

$$M(x) := f(x) + \frac{\mathrm{GS}_f}{\varepsilon} \mathsf{Lap}(1)$$

Privacy (when sensitivity is 1):

$$\frac{\mathbb{P}(M(x) \in A)}{\mathbb{P}(M(x') \in A)} = \frac{\int_A \exp(-\varepsilon|f(x) - w|)dw}{\int_A \exp(-\varepsilon|f(x') - w|)dw}$$

$$\le \sup_w \frac{\exp(-\varepsilon|f(x) - w|)}{\exp(-\varepsilon|f(x') - w|)}$$

$$\le \exp(\varepsilon|f(x) - f(x')|) \le \exp(\varepsilon)$$

Utility

$$\mathbb{E}[(M(x) - f(x))^2] = \frac{\mathrm{GS}_f^2}{\varepsilon^2}$$

Basic mechanisms: Gaussian

- Have a function $f(x) : \mathcal{X}^n \to \mathbb{R}^d$ we wish to estimate with *global sensitivity*

$$\mathrm{GS}_f := \sup_{d_{\mathrm{ham}}(x,x') \le 1} \|f(x) - f(x')\|_2$$

- Gaussian mechanism (Dwork et al. 06b) is (ε, δ)-differentially private:

$$M(x) := f(x) + \mathcal{N}\left(0, \frac{2\mathrm{GS}_f^2}{\varepsilon^2} \log \frac{1}{\delta} I\right)$$

Utility

$$\mathbb{E}\left[\|M(x) - f(x)\|_2^2\right] = 2\mathrm{GS}_f^2 \cdot \frac{d}{\varepsilon^2} \log \frac{1}{\delta}$$

Note: scaling with dimension is minimax optimal [Steinke/Ullman 15]

Basic mechanisms: 1-dimensional mean

- Suppose data bounded in [-1, 1]

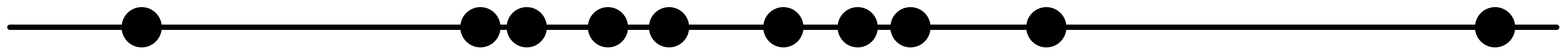
$$f(x) = \overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad\qquad \mathrm{GS}_f = \frac{2}{n}$$

- For either Laplace or Gaussian mechanism,

$$\mathbb{E}[(M(x) - \overline{x}_n)^2] \leq \frac{O(1)}{n^2\varepsilon^2}$$

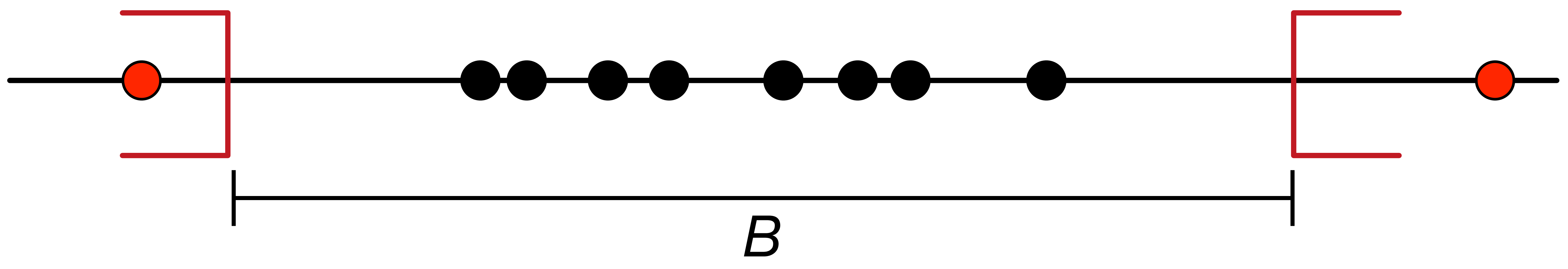
Basic mechanisms: 1-dimensional mean

- If data is unbounded, truncate, *then* apply mechanism



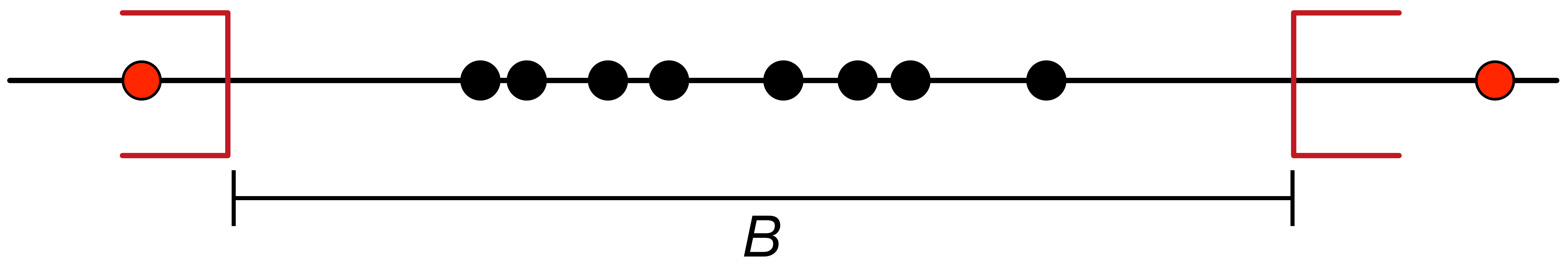
Basic mechanisms: 1-dimensional mean

- If data is unbounded, truncate, *then* apply mechanism



Basic mechanisms: 1-dimensional mean

- If data is unbounded, truncate, *then* apply mechanism

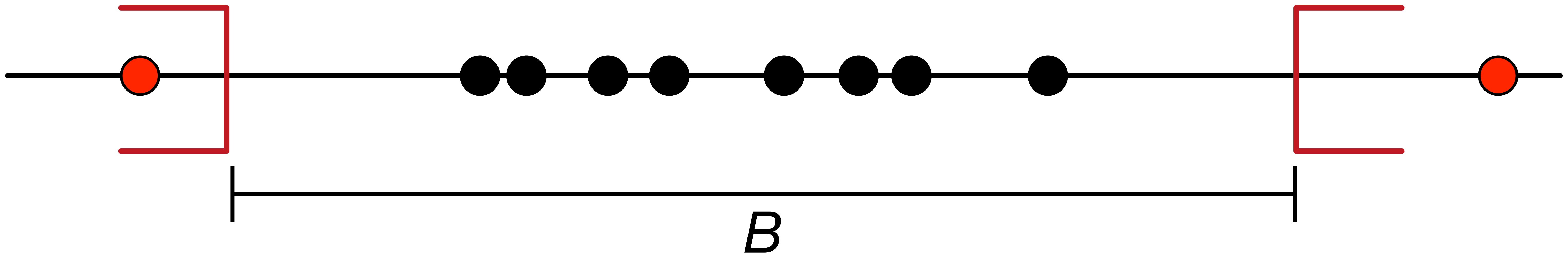


- For either Laplace or Gaussian mechanism, when *X* has *p* moments

$$\mathbb{E}[(M(x) - \overline{x}_n)^2] \leq O(1) \left[\frac{B^2}{n^2 \varepsilon^2} + \frac{1}{B^{2p-2}} \right]$$

Basic mechanisms: 1-dimensional mean

- If data is unbounded, truncate, *then* apply mechanism



- For either Laplace or Gaussian mechanism, when *X* has *p* moments

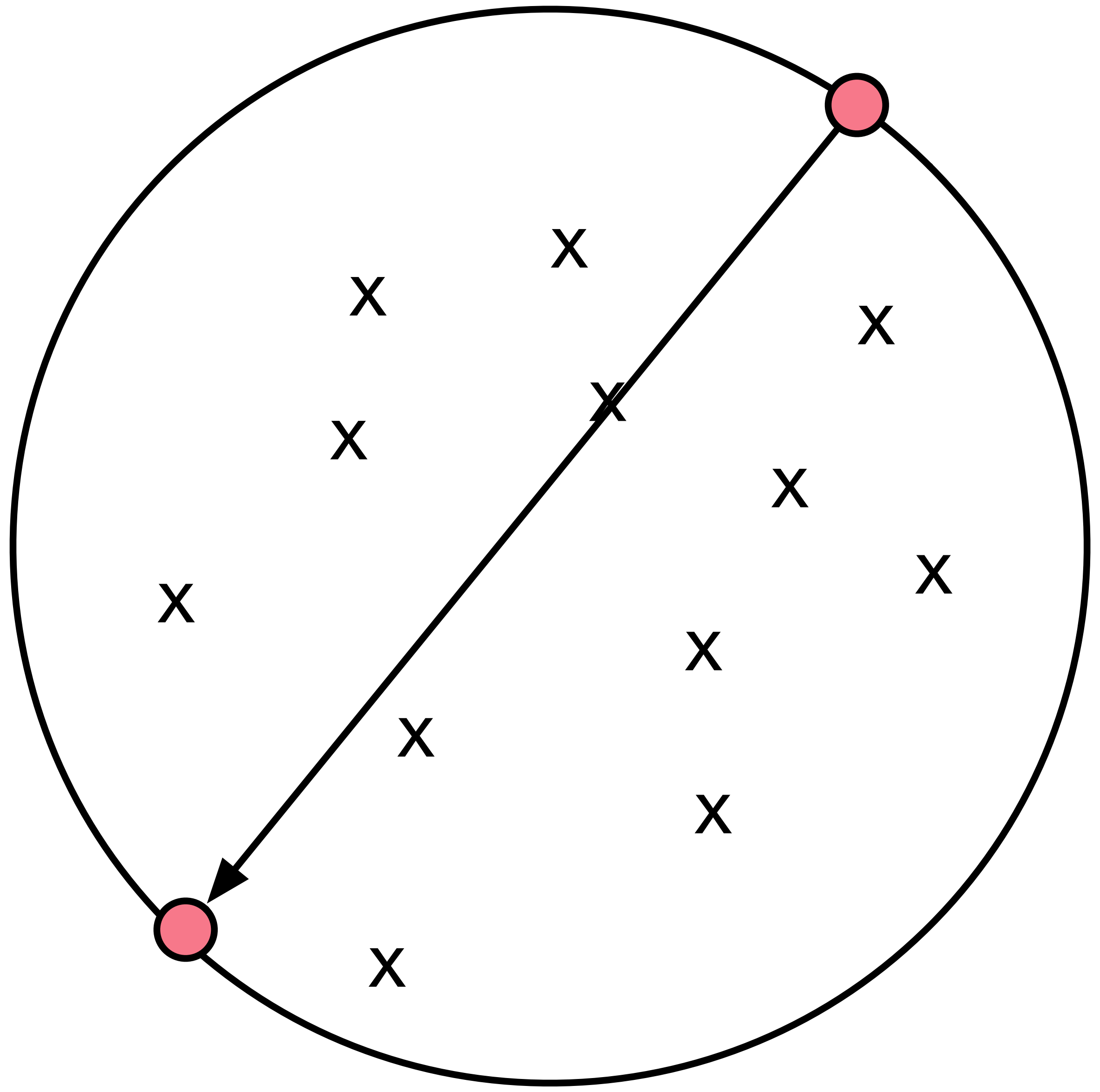
$$\mathbb{E}[(M(x) - \overline{x}_n)^2] \leq O(1) \left[\frac{B^2}{n^2 \varepsilon^2} + \frac{1}{B^{2p-2}} \right]$$

Optimizing for *B*, this is minimax optimal [Barber/Duchi 14]

Basic mechanisms: d-dimensional mean

- Suppose data bounded in an ℓ_2-ball of radius O($\sqrt{d}$) with identity covariance
- Global sensitivity $\mathrm{GS}_f \asymp \frac{\sqrt{d}}{n}$

  and scaling $\|x\|_2 = \sqrt{d}$



- Gaussian mechanism $\mathsf{M}(x) = \overline{x}_n + \mathcal{N}\left(0, \frac{d \log \frac{1}{\delta}}{n^2 \varepsilon^2} I\right)$

Utility

$$\mathbb{E}\left[\|\mathsf{M}(X_1^n) - \mu\|_2^2\right] = \frac{d}{n} + \frac{d^2 \log \frac{1}{\delta}}{n^2 \varepsilon^2}$$

Goal: estimating adaptively to covariance

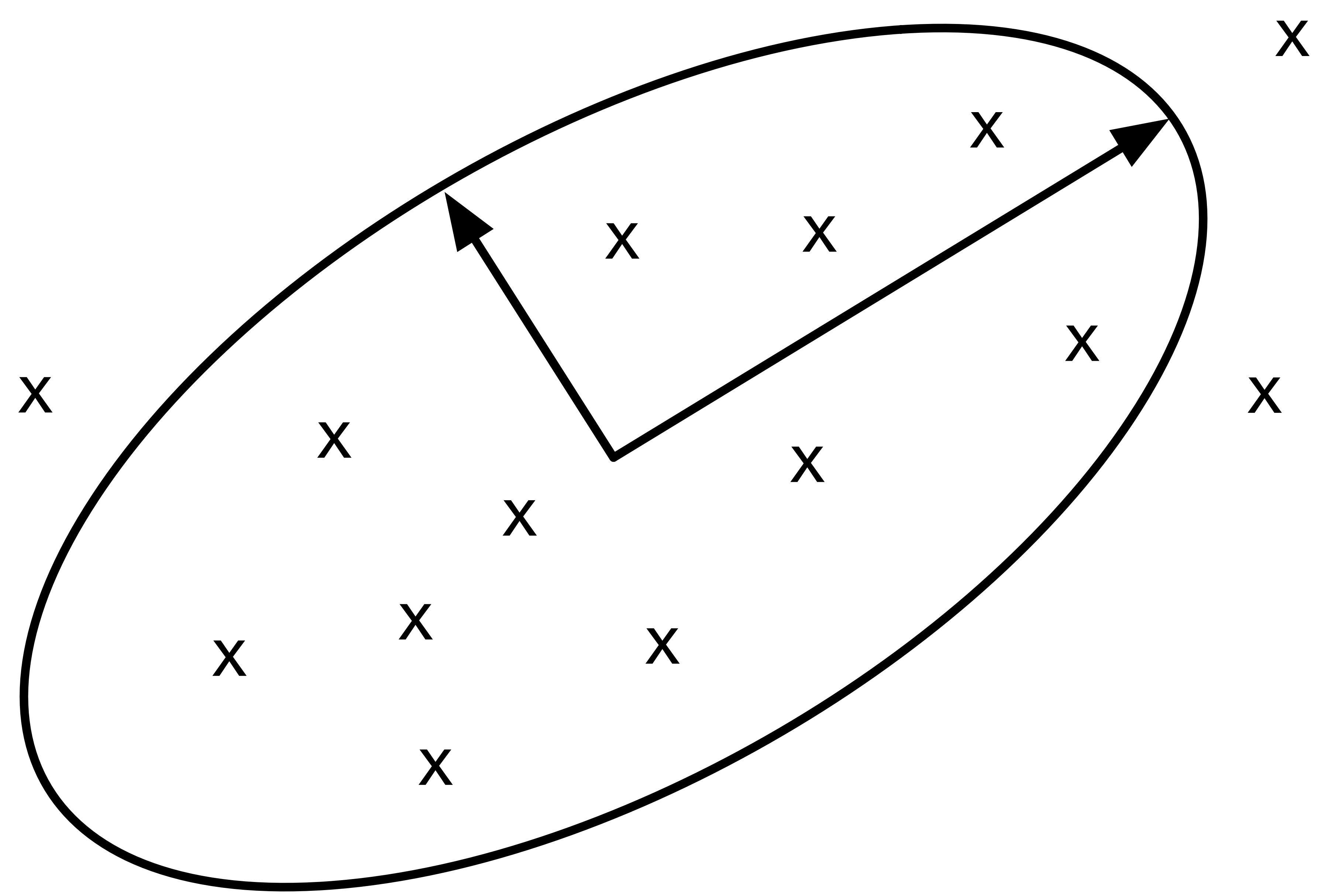
- Sample $X_i \overset{\text{iid}}{\sim} P$ with mean $\mu = \mathbb{E}[X]$ and covariance $\mathrm{Cov}(X_i) = \Sigma$
- Target: a private mechanism $\widetilde{\mu}$ that with high probability achieves

$$\|\widetilde{\mu} - \mu\|_\Sigma^2 \lesssim \boxed{\frac{d}{n}} + \widetilde{O}(1)\boxed{\frac{d^2}{n^2\varepsilon^2}}$$

Statistical efficiency

Minimax privacy lower bound
[Barber & Duchi 14, Steinke & Ullman 15]

Estimating a mean with known covariance

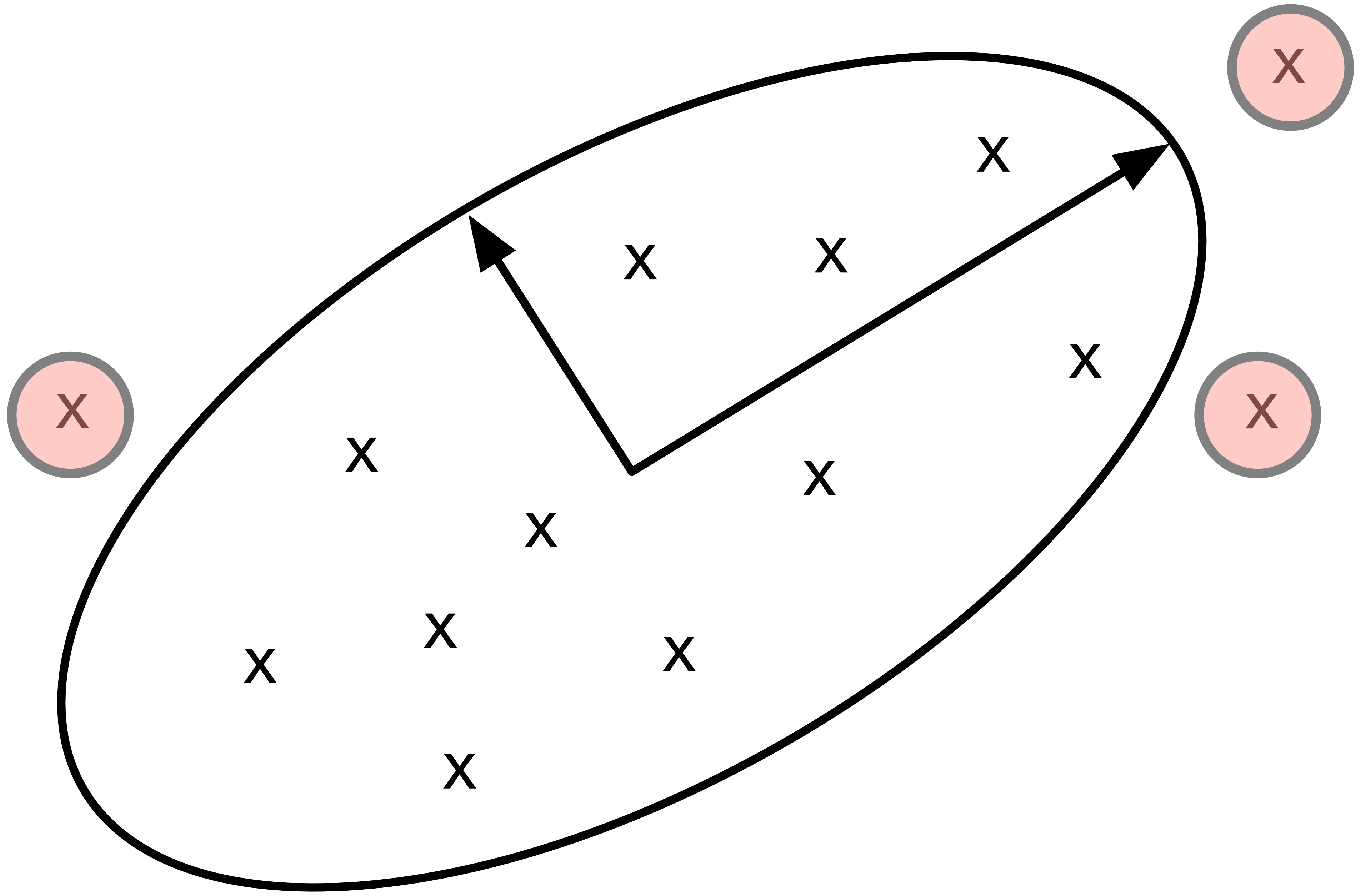


$\{x \mid \|x - \mu\|_\Sigma = \sqrt{d}\}$

- Remove data outside covariance ball
- Estimate truncated mean with sensitivity

$$\|\widehat{\mu}_{\mathrm{tr}}(x) - \widehat{\mu}_{\mathrm{tr}}(x')\|_\Sigma^2 \lesssim \frac{d}{n^2}$$

Estimating a mean with known covariance

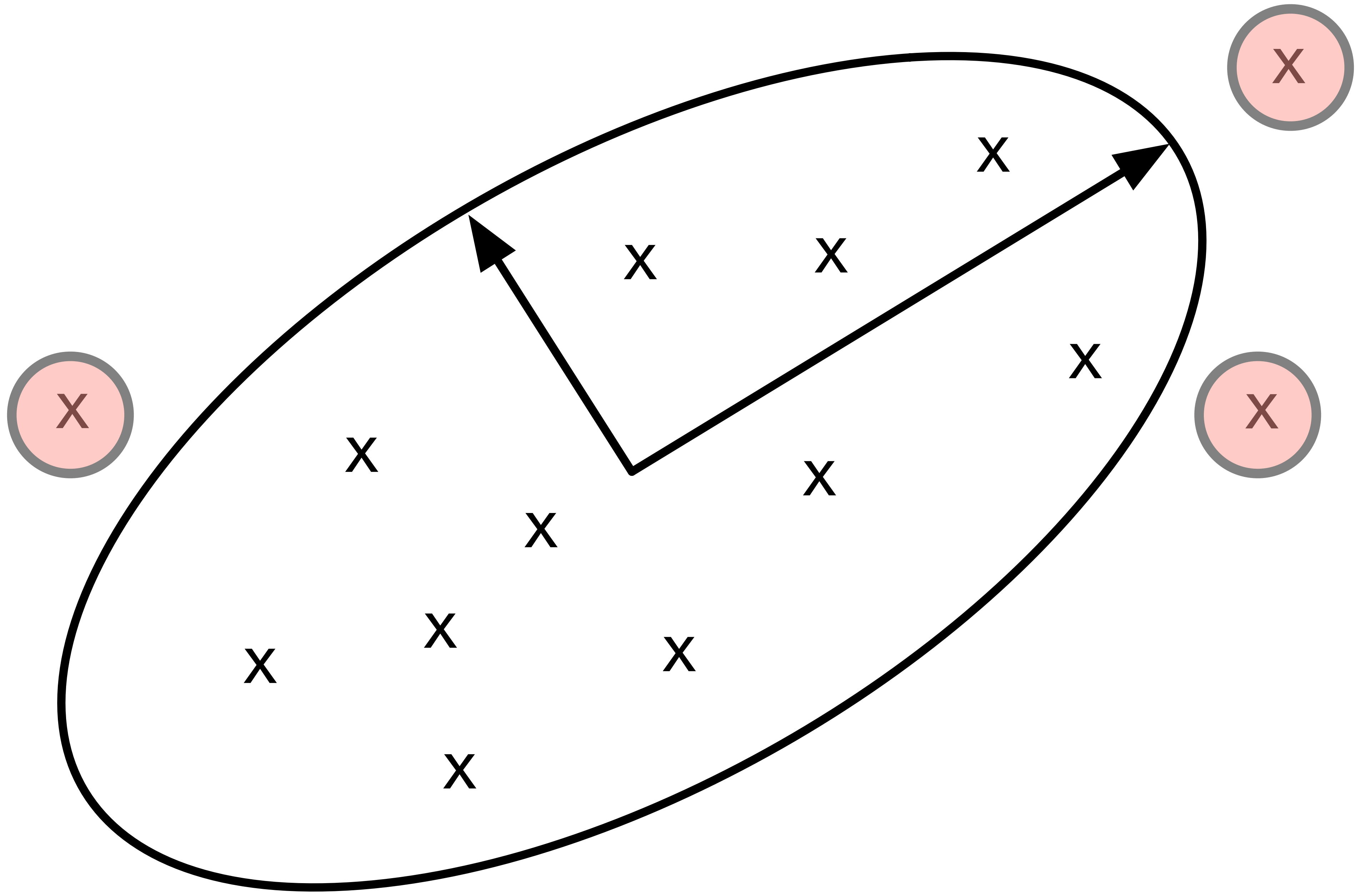


$\{x \mid \|x - \mu\|_\Sigma = \sqrt{d}\}$

- Remove data outside covariance ball
- Estimate truncated mean with sensitivity

$$\|\widehat{\mu}_{\mathrm{tr}}(x) - \widehat{\mu}_{\mathrm{tr}}(x')\|_\Sigma^2 \lesssim \frac{d}{n^2}$$

Estimating a mean with known covariance



$\{x \mid \|x - \mu\|_\Sigma = \sqrt{d}\}$

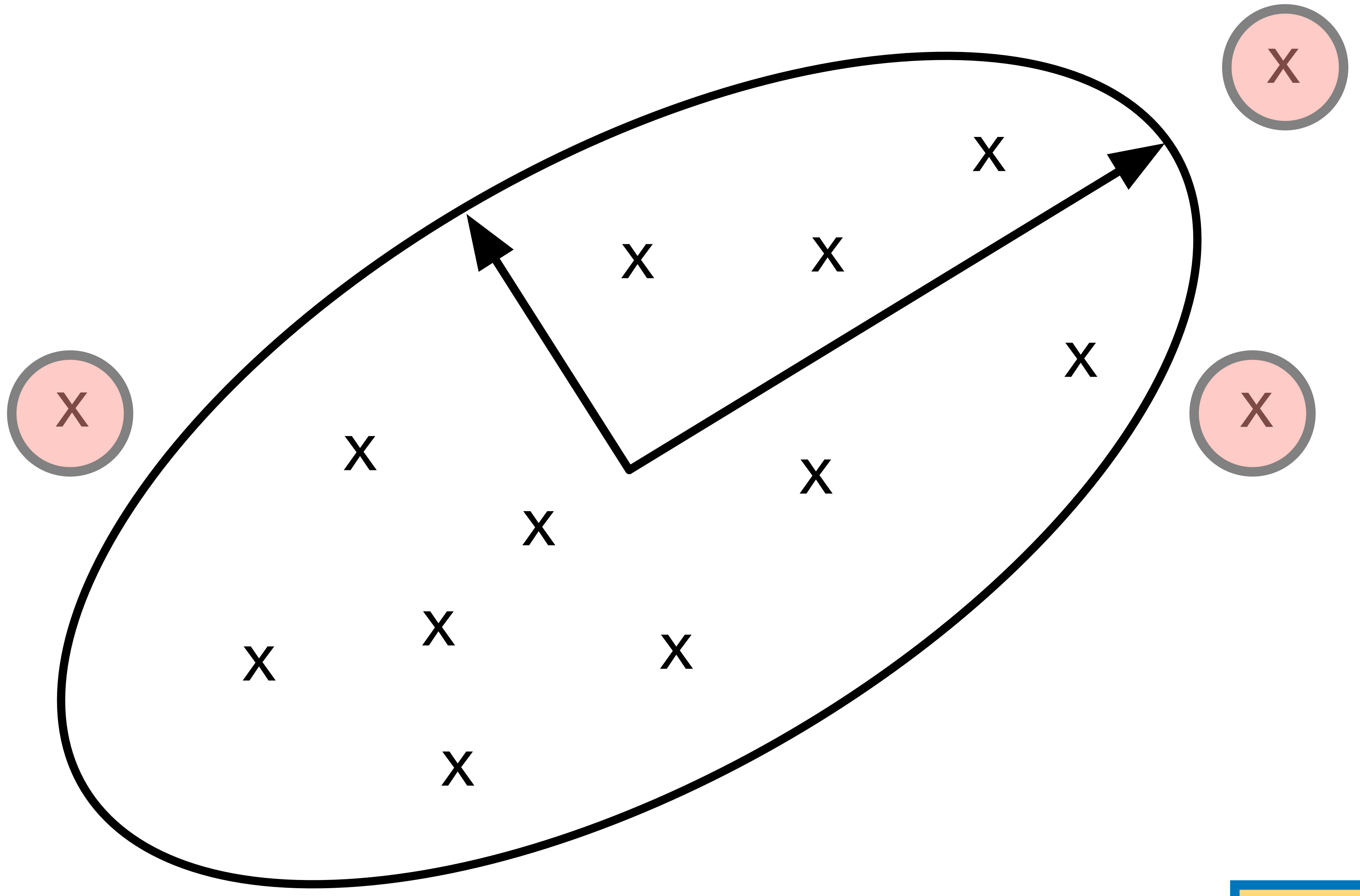
- Remove data outside covariance ball
- Estimate truncated mean with sensitivity

$$\|\widehat{\mu}_{\mathrm{tr}}(x) - \widehat{\mu}_{\mathrm{tr}}(x')\|_\Sigma^2 \lesssim \frac{d}{n^2}$$

- Add normal noise:

$$\mathsf{M}(x) := \widehat{\mu}_{\mathrm{tr}}(x) + \mathcal{N}\left(0, \frac{d \log \frac{1}{\delta}}{n^2} \Sigma\right)$$

Estimating a mean with known covariance



$\{x \mid \|x - \mu\|_\Sigma = \sqrt{d}\}$

- Remove data outside covariance ball
- Estimate truncated mean with sensitivity

$$\|\widehat{\mu}_{\mathrm{tr}}(x) - \widehat{\mu}_{\mathrm{tr}}(x')\|_\Sigma^2 \lesssim \frac{d}{n^2}$$

- Add normal noise:

$$\mathsf{M}(x) := \widehat{\mu}_{\mathrm{tr}}(x) + \mathcal{N}\left(0, \frac{d \log \frac{1}{\delta}}{n^2} \Sigma\right)$$

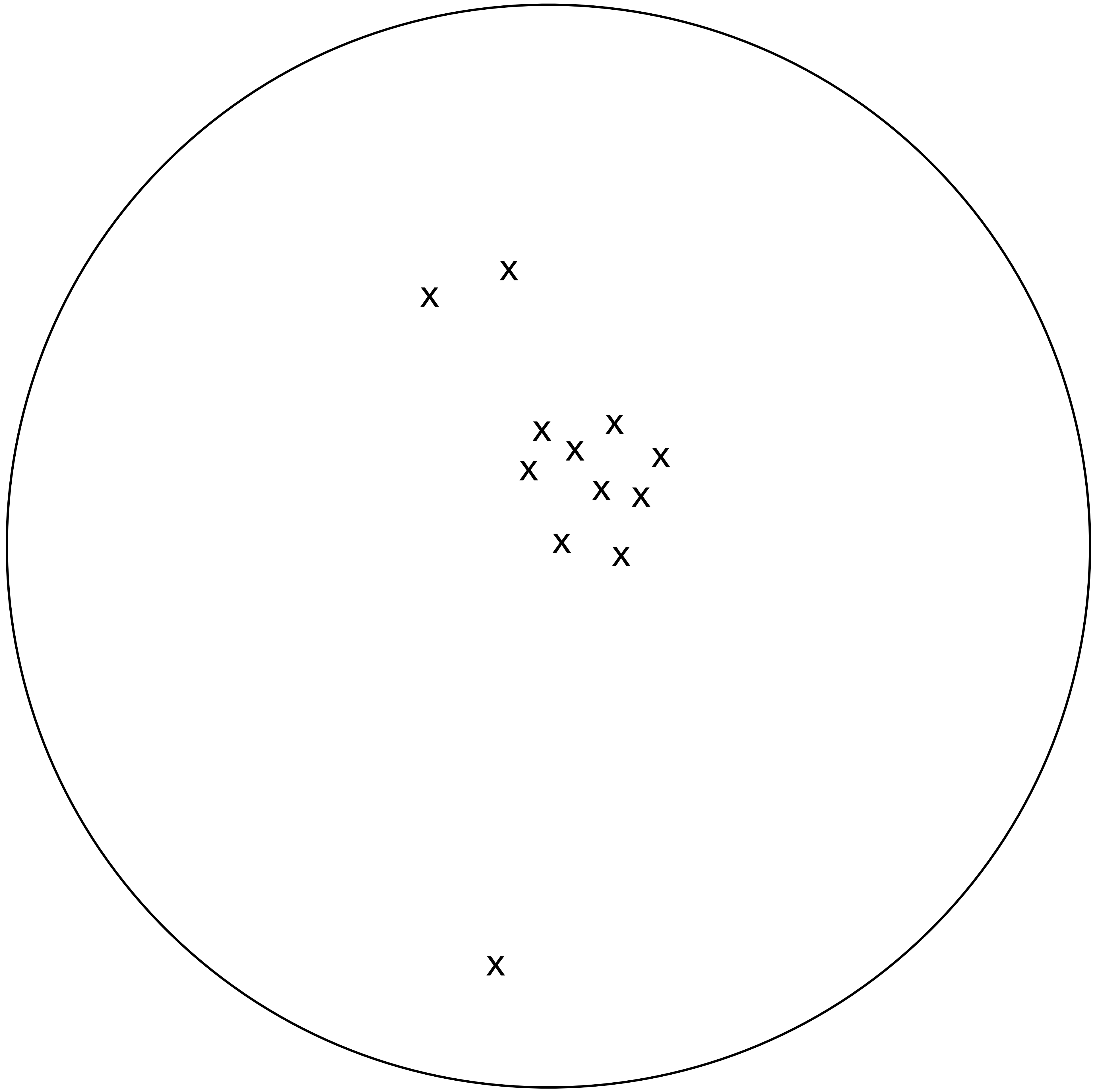
Utility

$$\mathbb{E}\left[\|\mathsf{M}(X) - \mu\|_\Sigma^2\right] = \frac{d}{n} + O(1)\frac{d^2 \log \frac{1}{\delta}}{n^2 \varepsilon^2}$$

CoinPress: adapting to scale

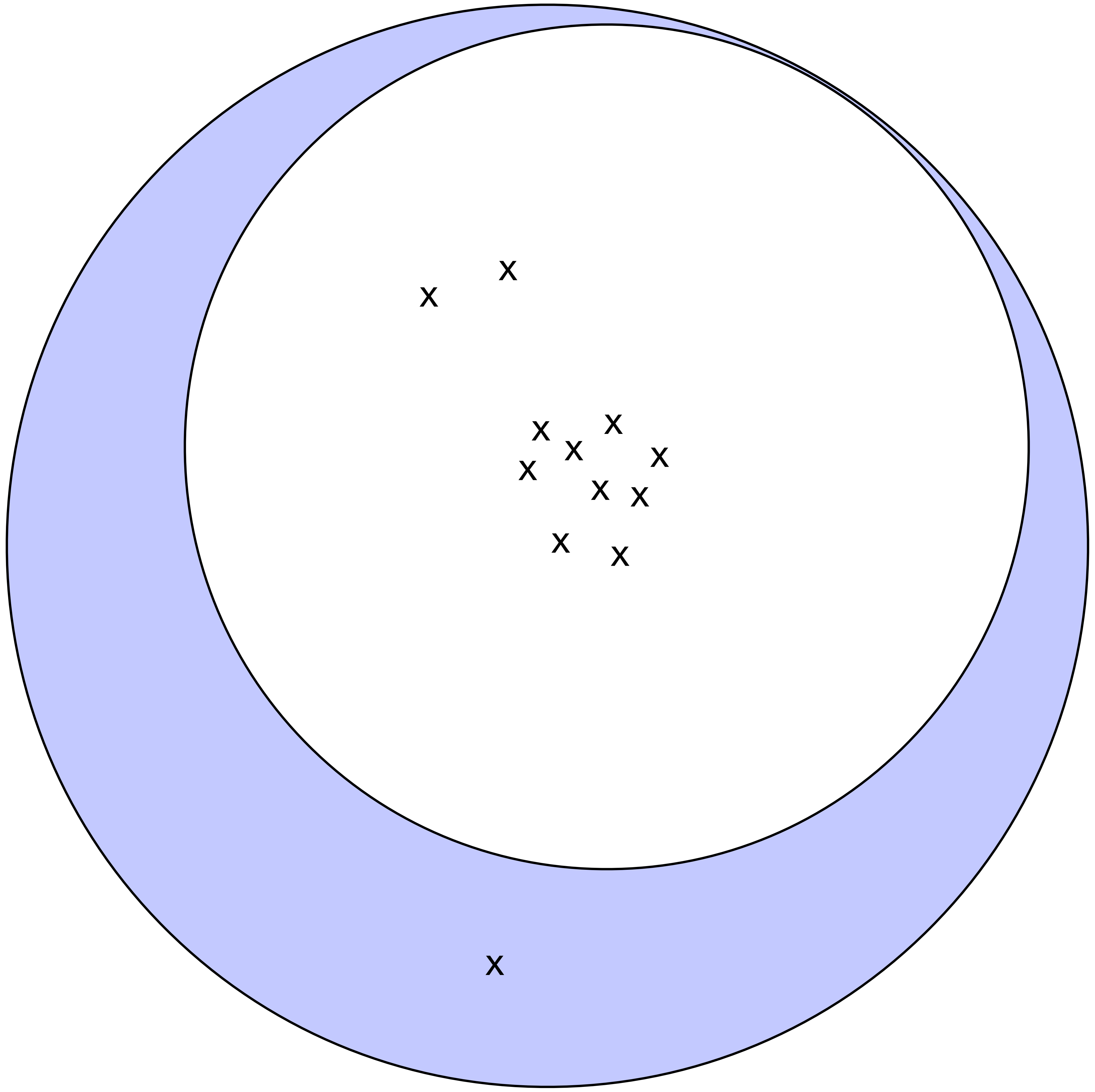
Biswas, Dong, Kamath, Ullman [2020]

- Repeatedly estimate mean in truncated region, shrink region (privately), repeat
- Delightfully practical
- Fails to adapt to covariance



CoinPress: adapting to scale
Biswas, Dong, Kamath, Ullman [2020]

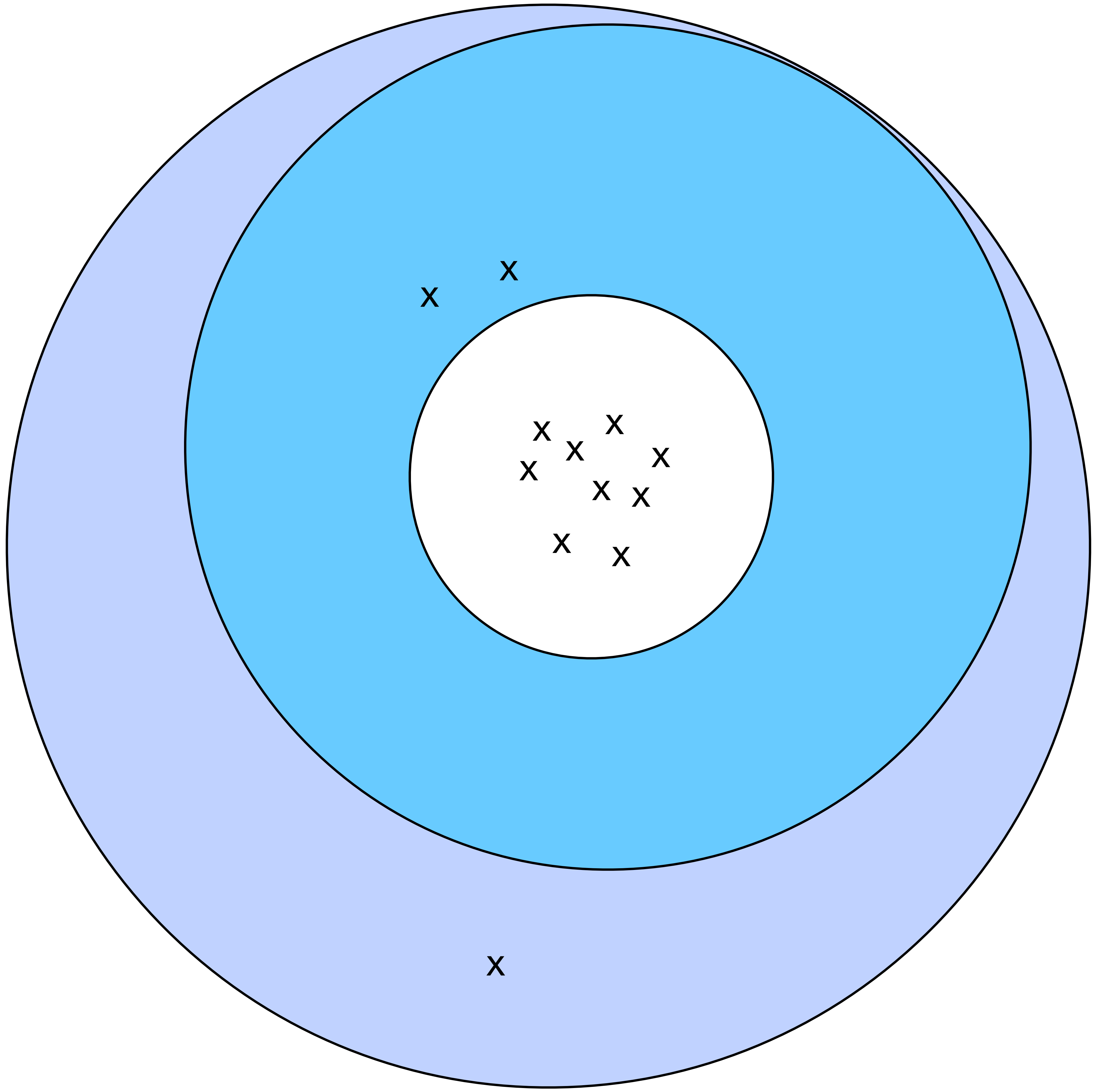
- Repeatedly estimate mean in truncated region, shrink region (privately), repeat
- Delightfully practical
- Fails to adapt to covariance



CoinPress: adapting to scale

Biswas, Dong, Kamath, Ullman [2020]

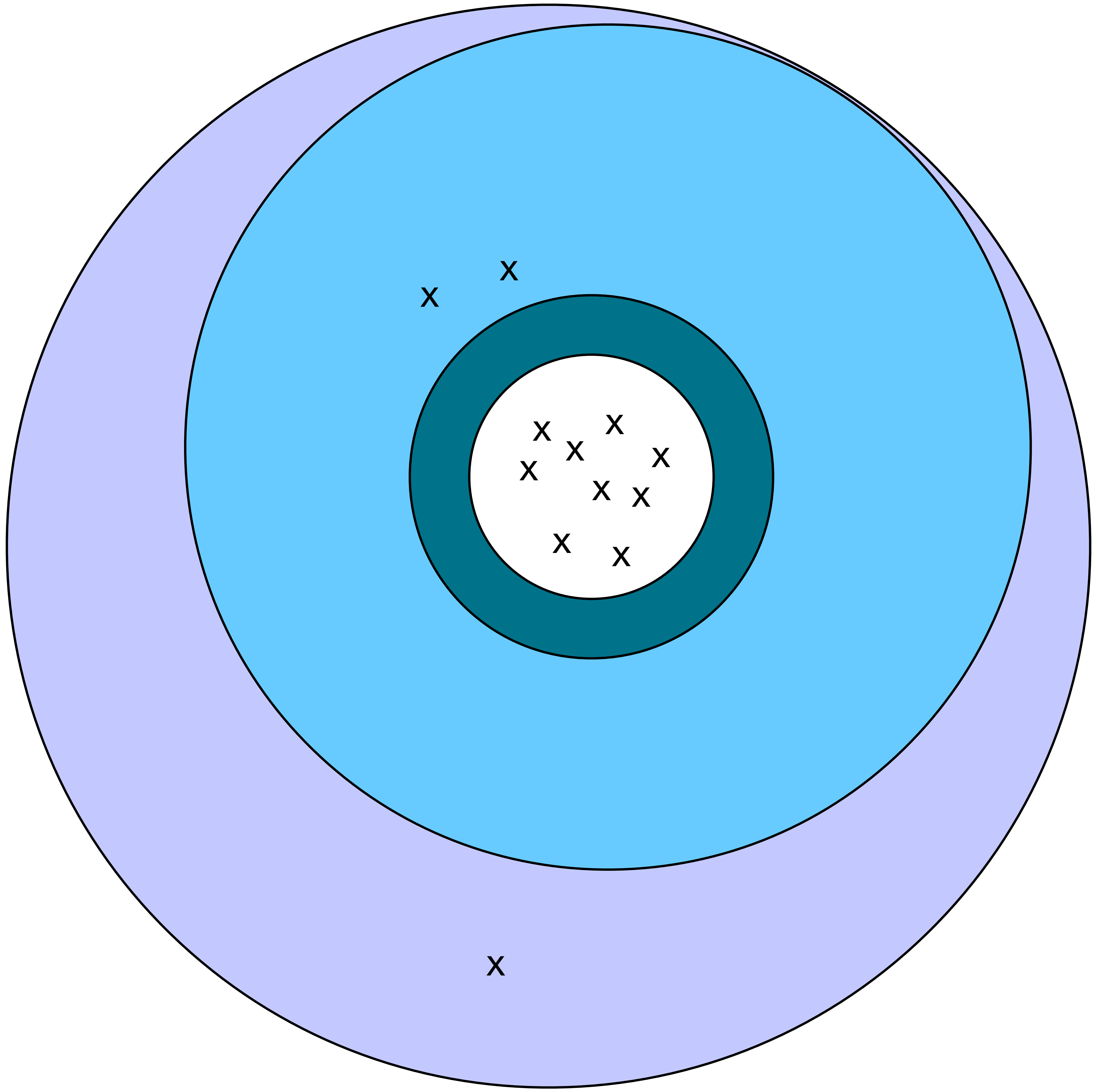
- Repeatedly estimate mean in truncated region, shrink region (privately), repeat
- Delightfully practical
- Fails to adapt to covariance



CoinPress: adapting to scale

Biswas, Dong, Kamath, Ullman [2020]

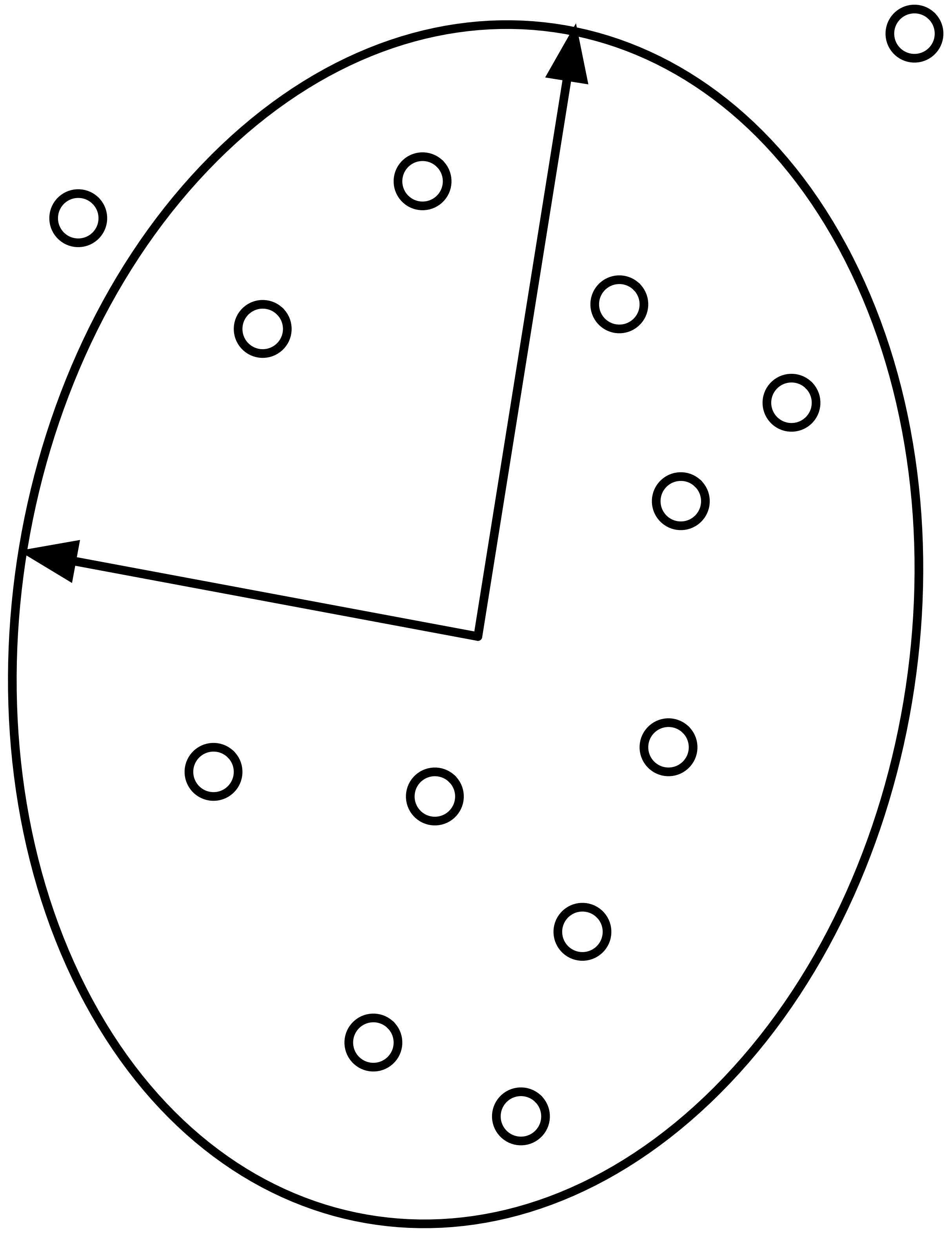
- Repeatedly estimate mean in truncated region, shrink region (privately), repeat
- Delightfully practical
- Fails to adapt to covariance



Safe datasets (and test/release framework)

- "Good" datasets have most data near the mean

$$\mathcal{G}(B) := \{(x_1, \ldots, x_n) \subset \mathbb{R}^d \mid \|\overline{x}_n - x_i\|_{\Sigma_n} \le B, \text{ all } i\}$$



[Dwork & Lei 09, Brown et al. 21, Liu et al. 22]

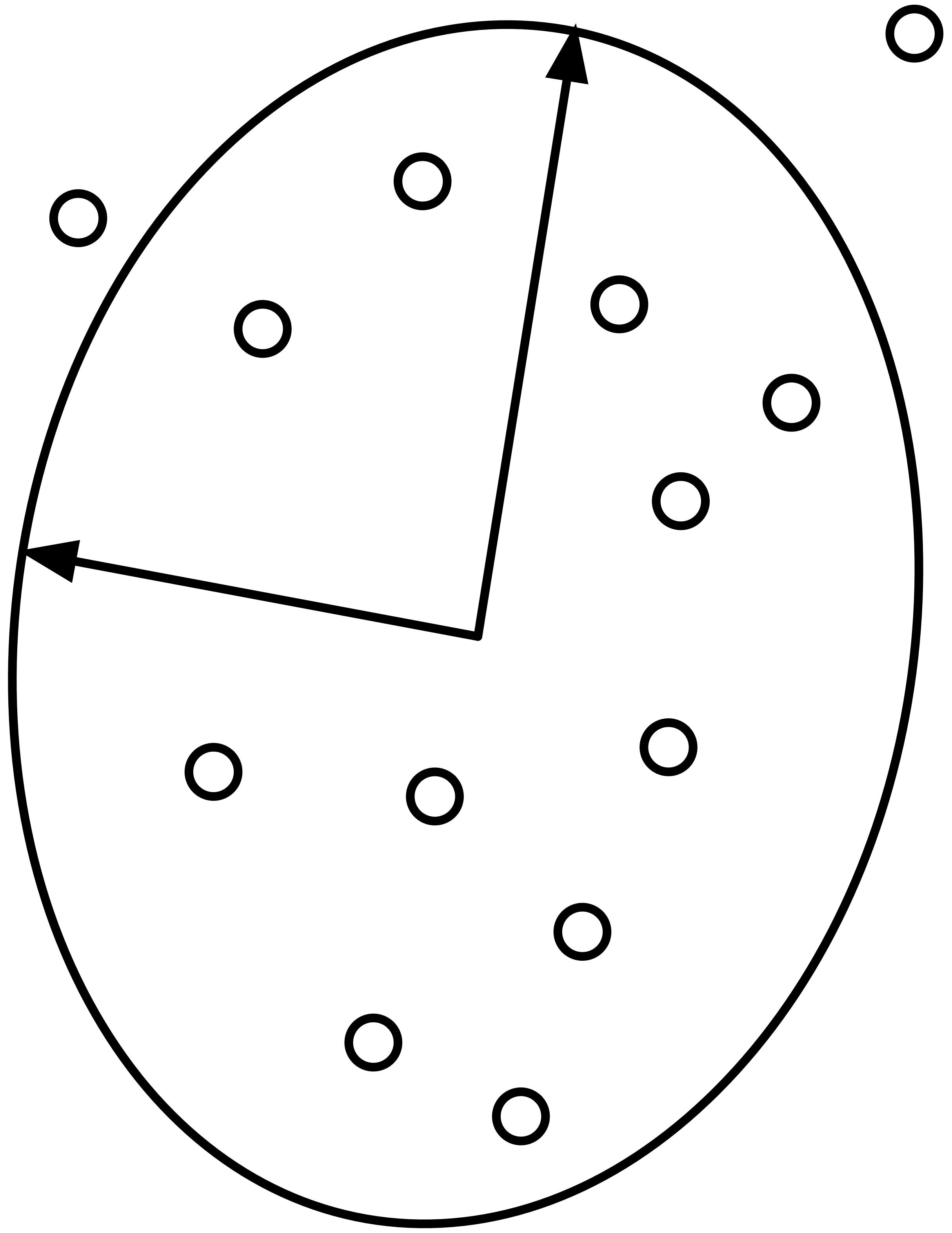
Safe datasets (and test/release framework)

- "Good" datasets have most data near the mean

$$\mathcal{G}(B) := \{(x_1, \ldots, x_n) \subset \mathbb{R}^d \mid \|\overline{x}_n - x_i\|_{\Sigma_n} \leq B, \text{ all } i\}$$

- Hamming distance is 1-Lipschitz, so *T* is private:

$$T = d_{\text{hamming}}\big((x_i)_{i=1}^n, \mathcal{G}(B)\big) + \mathsf{Lap}(1/\varepsilon)$$



[Dwork & Lei 09, Brown et al. 21, Liu et al. 22]

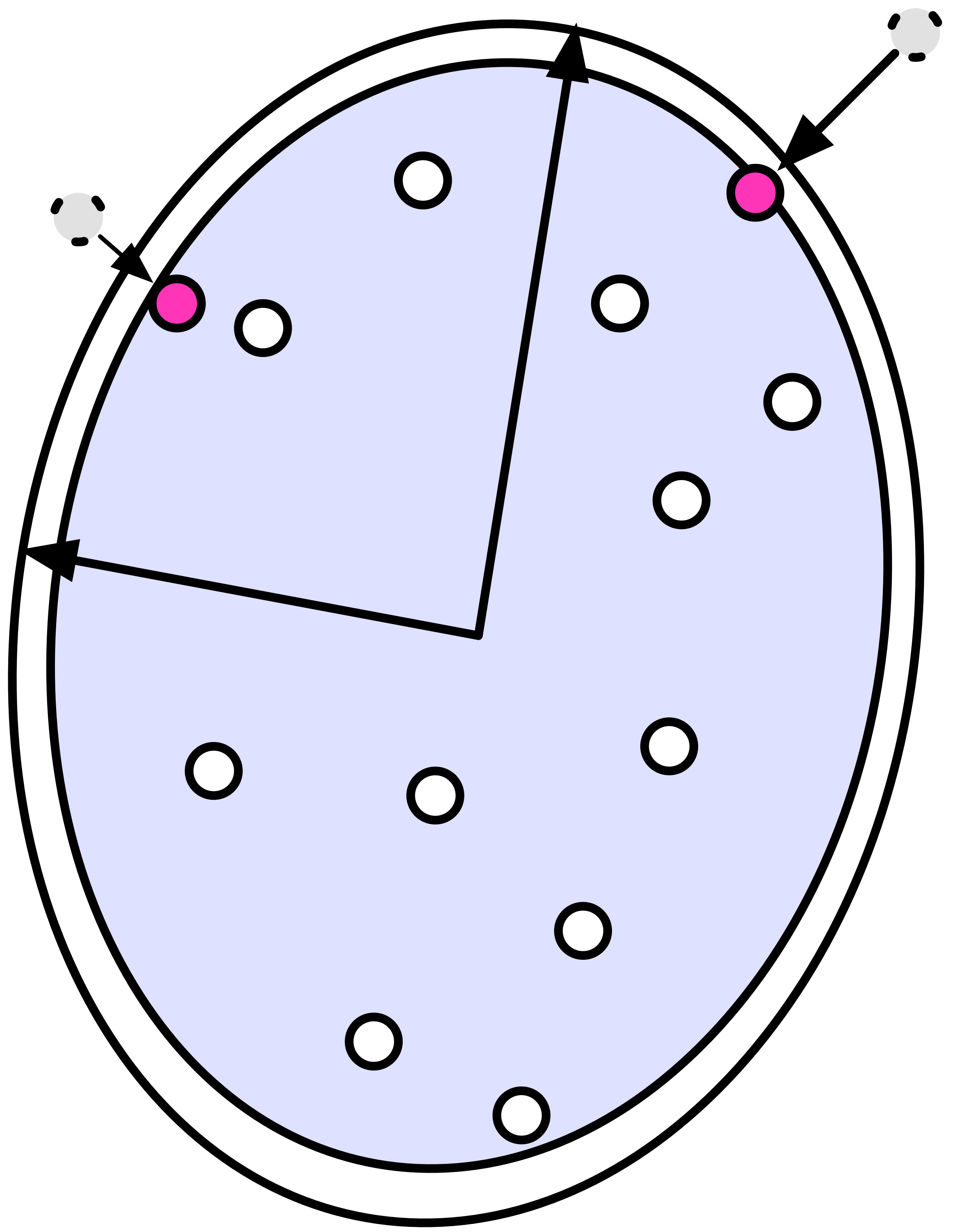
Safe datasets (and test/release framework)

- "Good" datasets have most data near the mean

$$\mathcal{G}(B) \coloneqq \{(x_1, \ldots, x_n) \subset \mathbb{R}^d \mid \|\overline{x}_n - x_i\|_{\Sigma_n} \leq B, \text{ all } i\}$$

- Hamming distance is 1-Lipschitz, so T is private:

$$T = d_{\text{hamming}}\big((x_i)_{i=1}^n, \mathcal{G}(B)\big) + \mathsf{Lap}(1/\varepsilon)$$



[Dwork & Lei 09, Brown et al. 21, Liu et al. 22]

Safe datasets (and test/release framework)

- "Good" datasets have most data near the mean

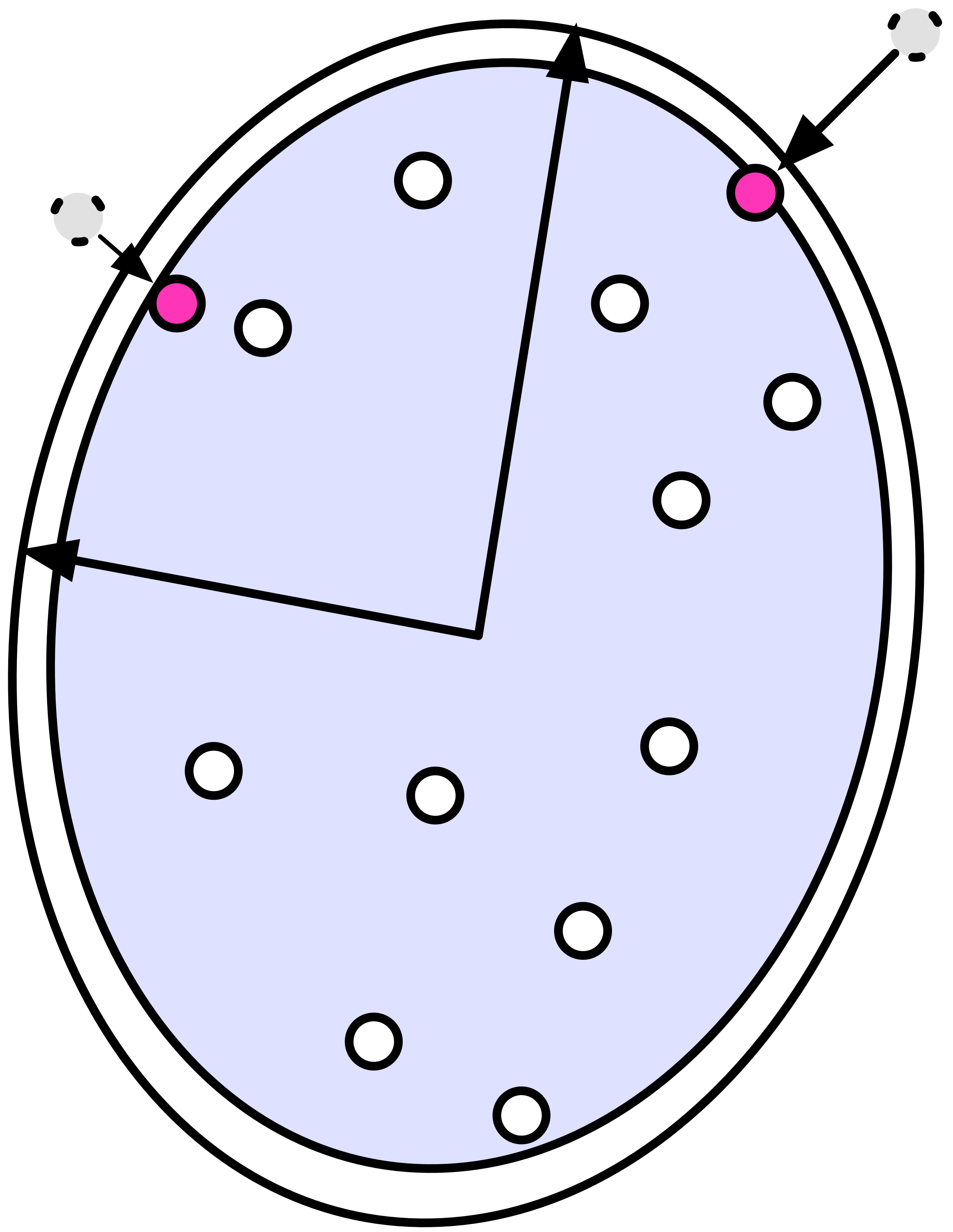
$$\mathcal{G}(B) := \{(x_1, \ldots, x_n) \subset \mathbb{R}^d \mid \|\overline{x}_n - x_i\|_{\Sigma_n} \leq B, \text{ all } i\}$$

- Hamming distance is 1-Lipschitz, so *T* is private:

$$T = d_{\text{hamming}}\big((x_i)_{i=1}^n, \mathcal{G}(B)\big) + \mathsf{Lap}(1/\varepsilon)$$

- Release noisy mean if *T* is small enough:

$$(z_i)_{i=1}^n = \operatorname*{argmin}_{z \in \mathcal{G}(B)} d_{\text{hamming}}(z, x)$$



$d_{\text{ham}}(x, \mathcal{G}(B)) = 2$

[Dwork & Lei 09, Brown et al. 21, Liu et al. 22]

Safe datasets (and test/release framework)

- "Good" datasets have most data near the mean

$$\mathcal{G}(B) := \{(x_1, \ldots, x_n) \subset \mathbb{R}^d \mid \|\overline{x}_n - x_i\|_{\Sigma_n} \le B, \text{ all } i\}$$

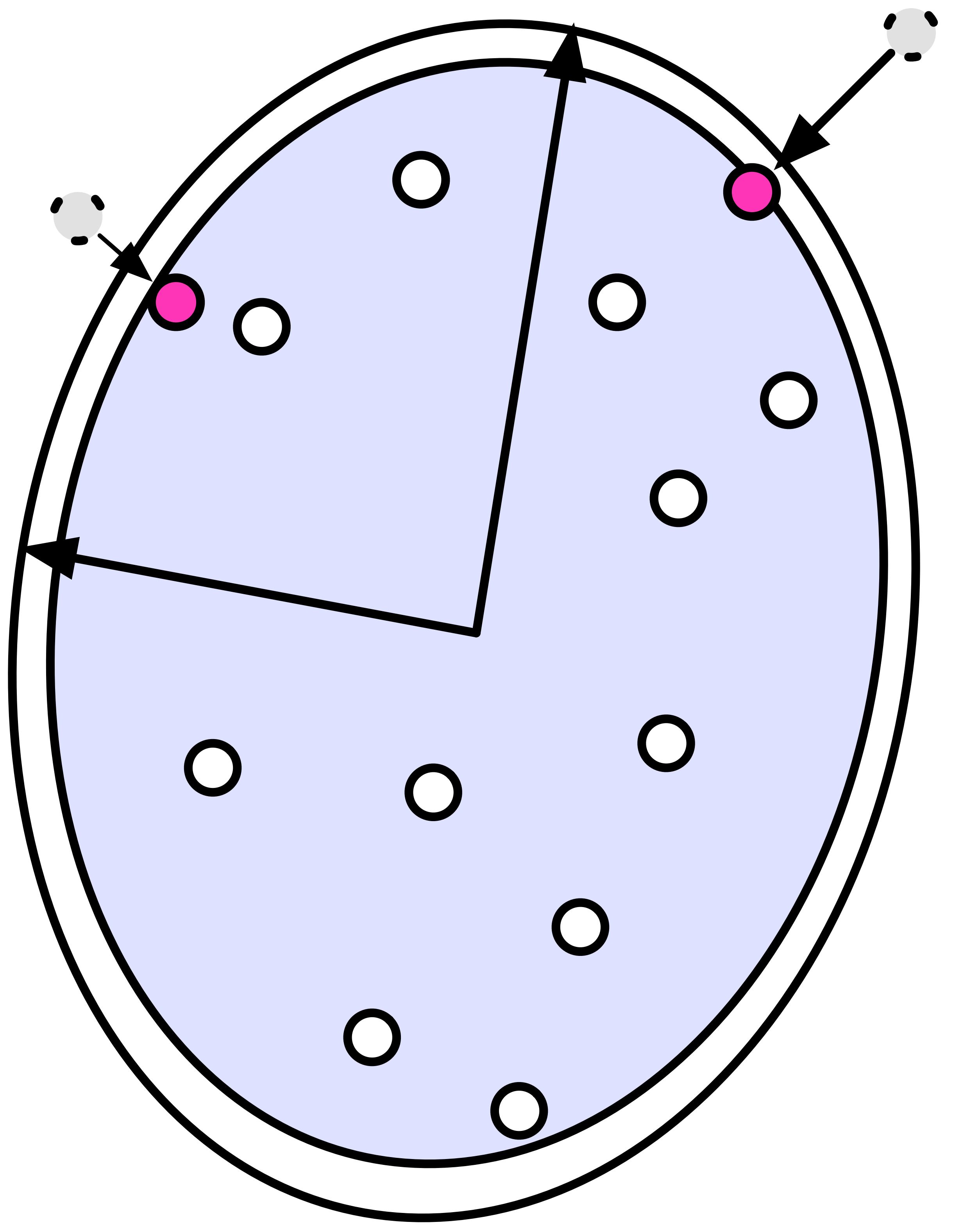
- Hamming distance is 1-Lipschitz, so *T* is private:

$$T = d_{\text{hamming}}\big((x_i)_{i=1}^n, \mathcal{G}(B)\big) + \mathsf{Lap}(1/\varepsilon)$$

- Release noisy mean if *T* is small enough:

$$(z_i)_{i=1}^n = \operatorname*{argmin}_{z \in \mathcal{G}(B)} d_{\text{hamming}}(z, x)$$

$$\widetilde{\mu} = \overline{z}_n + \mathcal{N}\left(0, O(1)\frac{d \log \frac{1}{\delta}}{n\varepsilon^2}\widehat{\text{Cov}}\big((z_i)_{i=1}^n\big)\right)$$



$d_{\text{ham}}(x, \mathcal{G}(B)) = 2$

[Dwork & Lei 09, Brown et al. 21, Liu et al. 22]

Safe datasets (and test/release framework)

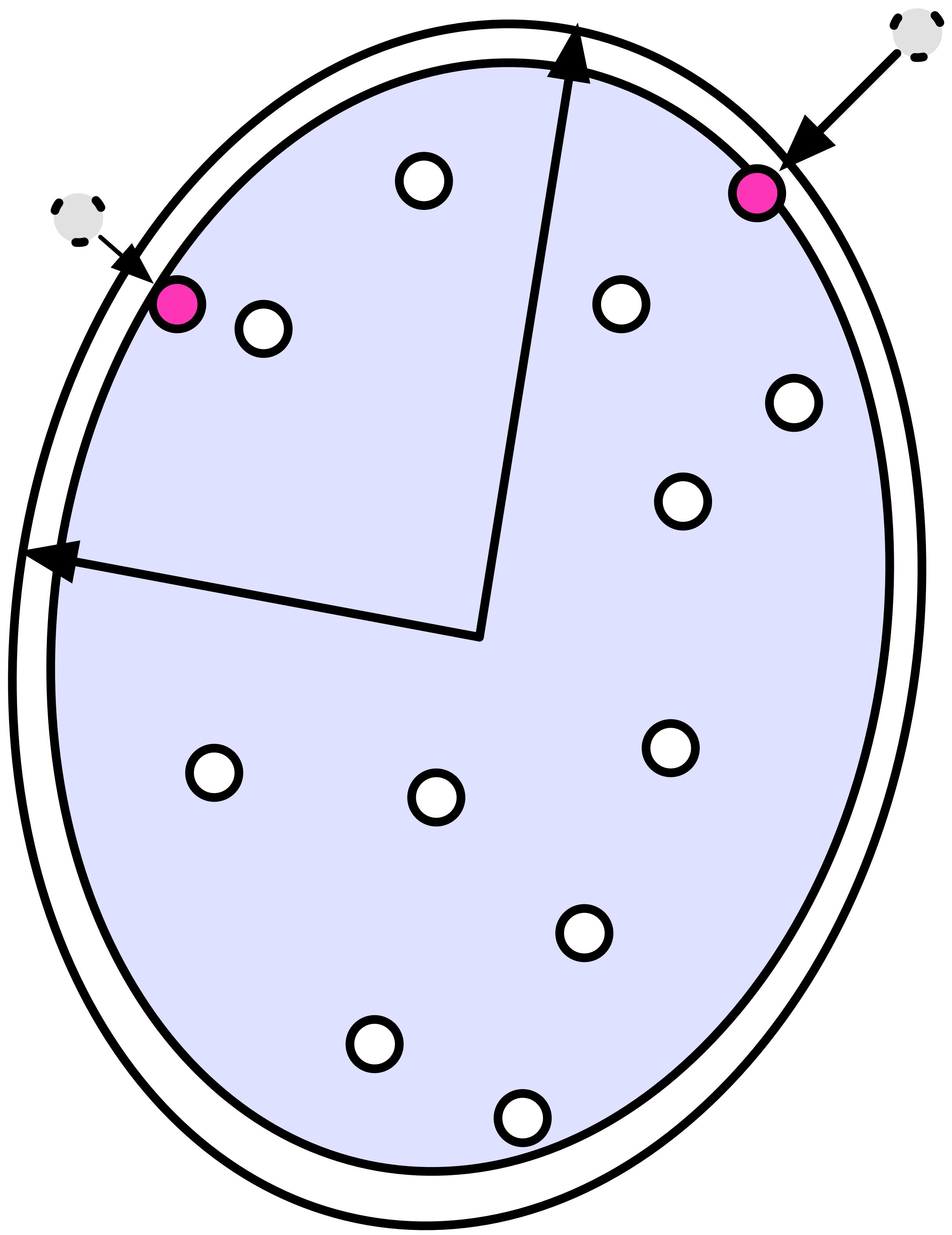
Theorem [Brown et al. 21]

This test/project/release framework, where one adds Gaussian noise to a projected sample mean, achieves

Accuracy: If the data are Gaussian and Σ is full rank,

$$\|\widetilde{\mu} - \mu\|_\Sigma^2 \lesssim \frac{d}{n} + \frac{d^2}{n^2\varepsilon^4} \log^6 \frac{1}{\delta}$$

Privacy: it is always (ε, δ)-differentially private



Safe datasets (and test/release framework)

Theorem [Brown et al. 21]

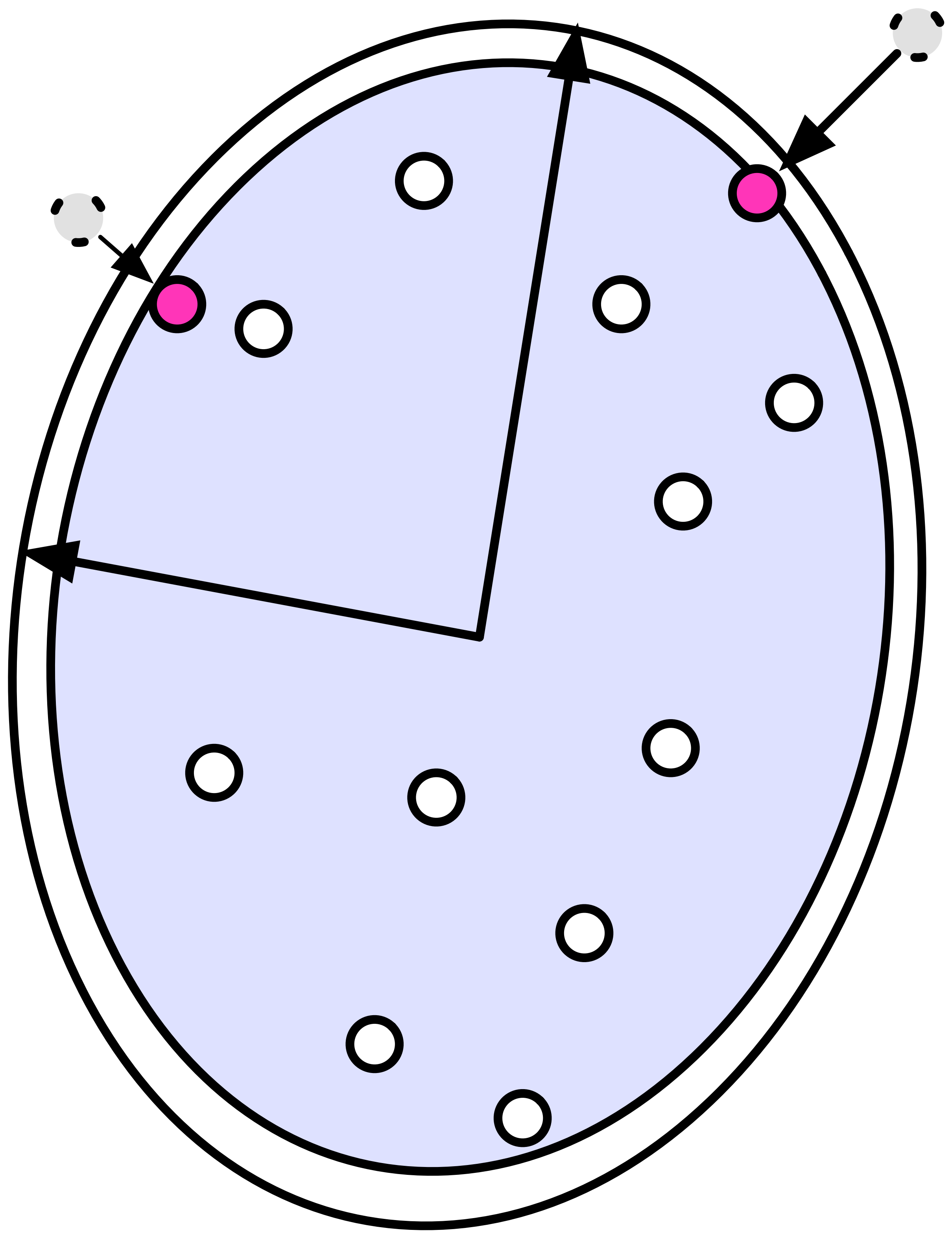
This test/project/release framework, where one adds Gaussian noise to a projected sample mean, achieves

Accuracy: If the data are Gaussian and Σ is full rank,

$$\|\widetilde{\mu} - \mu\|_{\Sigma}^{2} \lesssim \frac{d}{n} + \frac{d^2}{n^2 \varepsilon^4} \log^6 \frac{1}{\delta}$$

Privacy: it is always (ε, δ)-differentially private

(More recent work improves this a bit)



Safe datasets (and test/release framework)

Theorem [Brown et al. 21]

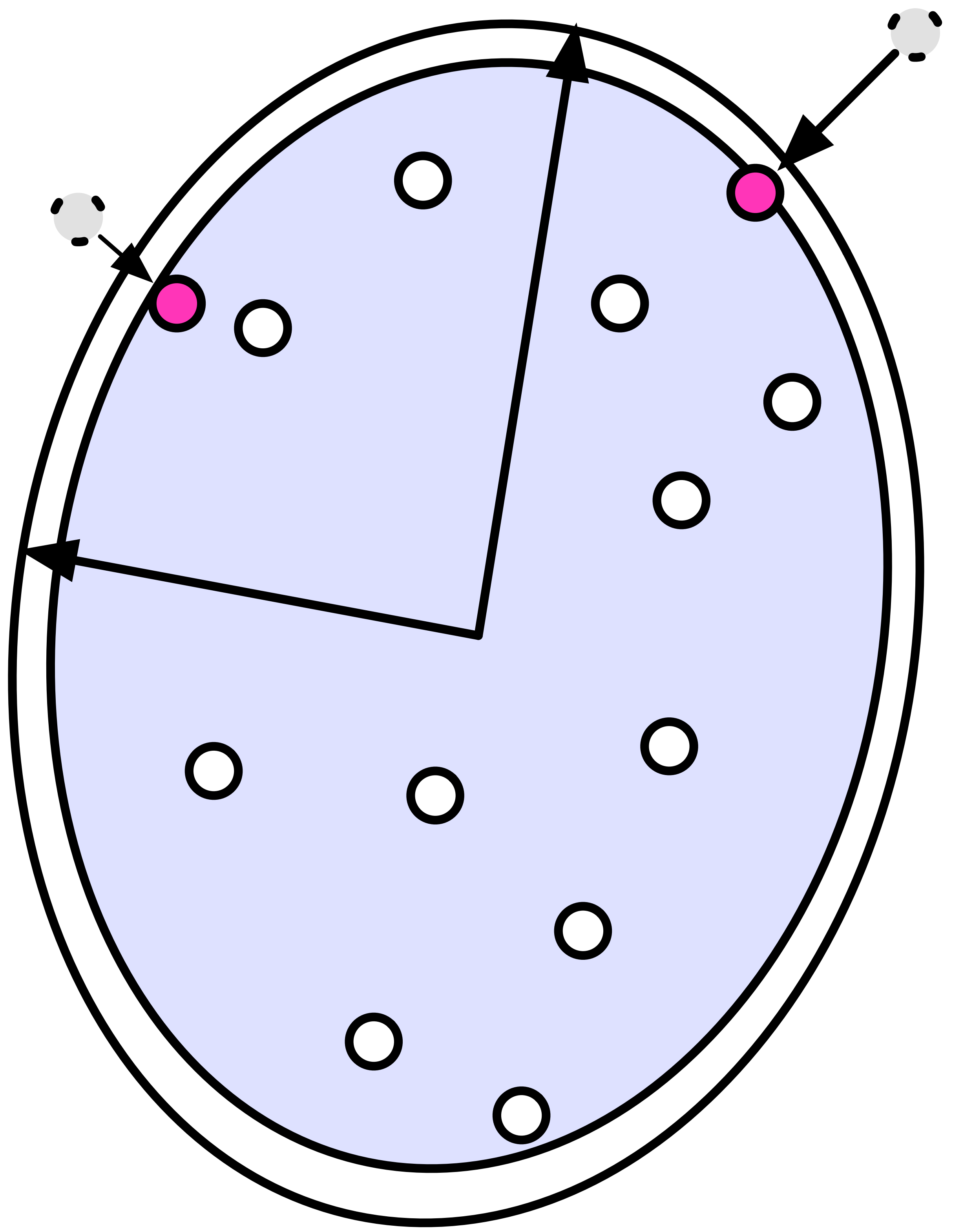
This test/project/release framework, where one adds Gaussian noise to a projected sample mean, achieves

Accuracy: If the data are Gaussian and Σ is full rank,

$$\|\widetilde{\mu} - \mu\|_{\Sigma}^2 \lesssim \frac{d}{n} + \frac{d^2}{n^2 \varepsilon^4} \log^6 \frac{1}{\delta}$$

Privacy: it is always (ε, δ)-differentially private

(More recent work improves this a bit)



- Exponential time algorithm
- (Essentially) requires Gaussianity

Two phase approach: covariance then mean

- Step 1: estimate covariance stably
- Step 2: add Gaussian noise with (estimated) covariance to a trimmed mean

Stable covariance estimation

- Iteratively shrink covariance until it is stable:

$$\widetilde{x}_i \coloneqq x_i - x_{n/2+i}$$

Initialize

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} \widetilde{x}_i \widetilde{x}_i^T \qquad Z_i \overset{\text{iid}}{\sim} \frac{C}{\varepsilon}\mathsf{Laplace}(1)$$

Until no changes occur, do:

for any indices *i* satisfying

$$\widetilde{x}_i^T \Sigma^\dagger \widetilde{x}_i > c \exp(Z_i - Z_0)$$

add them to the removed indices

$$R \leftarrow R \cup \{i\}$$

re-estimate covariance

$$\Sigma \leftarrow \frac{1}{n}\sum_{i \notin R} \widetilde{x}_i \widetilde{x}_i^T$$

Stable covariance estimation

- Iteratively shrink covariance until it is stable:

$$\widetilde{x}_i \coloneqq x_i - x_{n/2+i}$$

Initialize

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} \widetilde{x}_i \widetilde{x}_i^T \qquad Z_i \overset{\text{iid}}{\sim} \frac{C}{\varepsilon}\mathsf{Laplace}(1)$$

Until no changes occur, do:

for any indices *i* satisfying

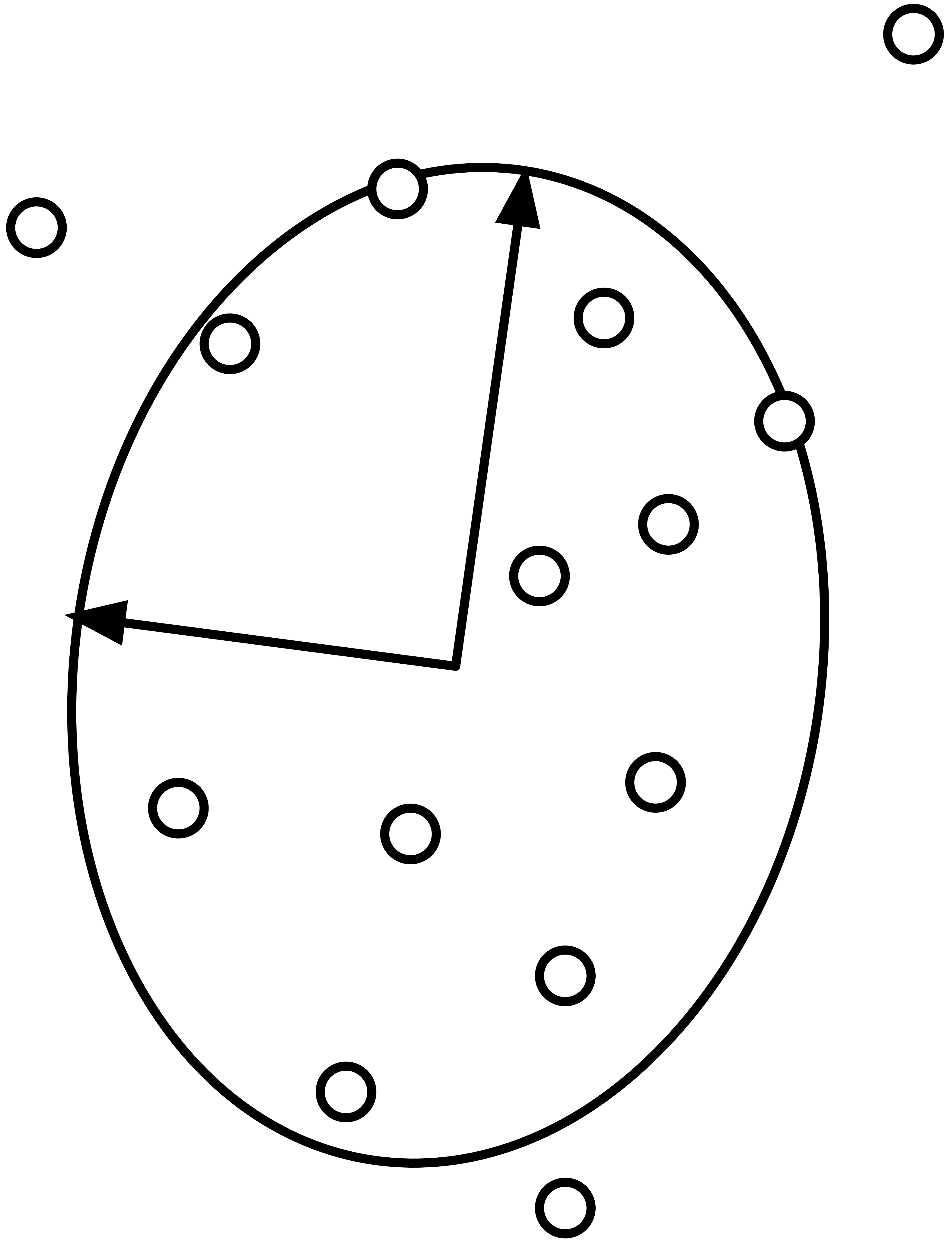
$$\widetilde{x}_i^T \Sigma^\dagger \widetilde{x}_i > c\exp(Z_i - Z_0)$$

add them to the removed indices

$$R \leftarrow R \cup \{i\}$$

re-estimate covariance

$$\Sigma \leftarrow \frac{1}{n}\sum_{i \notin R} \widetilde{x}_i \widetilde{x}_i^T$$



Stable covariance estimation

- Iteratively shrink covariance until it is stable:

$$\widetilde{x}_i := x_i - x_{n/2+i}$$

Initialize

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} \widetilde{x}_i \widetilde{x}_i^T \qquad Z_i \overset{\text{iid}}{\sim} \frac{C}{\varepsilon}\text{Laplace}(1)$$

Until no changes occur, do:

for any indices *i* satisfying

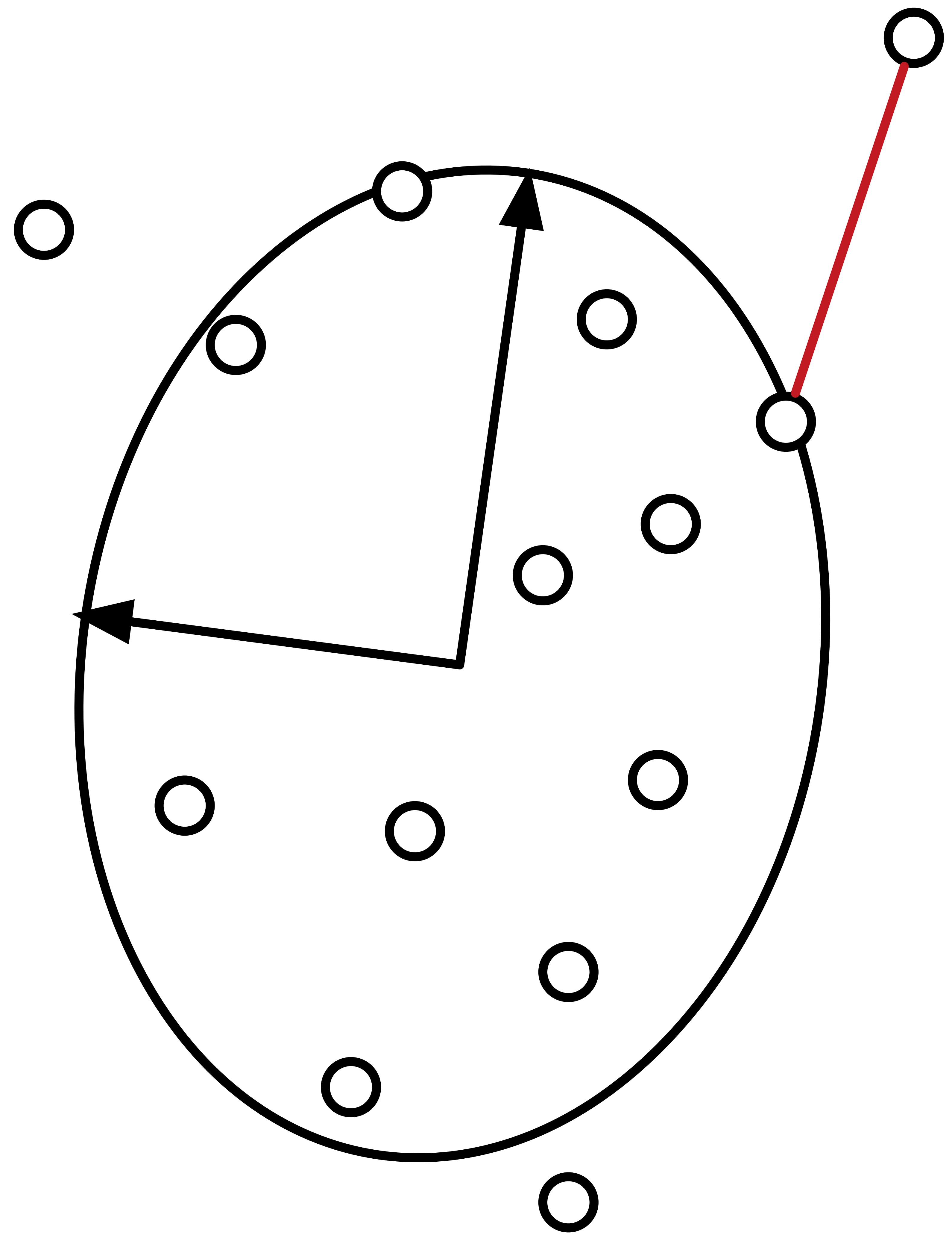
$$\widetilde{x}_i^T \Sigma^\dagger \widetilde{x}_i > c\exp(Z_i - Z_0)$$

add them to the removed indices

$$R \leftarrow R \cup \{i\}$$

re-estimate covariance

$$\Sigma \leftarrow \frac{1}{n}\sum_{i \notin R} \widetilde{x}_i \widetilde{x}_i^T$$



Stable covariance estimation

- Iteratively shrink covariance until it is stable:

$$\widetilde{x}_i := x_i - x_{n/2+i}$$

Initialize

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} \widetilde{x}_i \widetilde{x}_i^T \qquad Z_i \overset{\text{iid}}{\sim} \frac{C}{\varepsilon}\text{Laplace}(1)$$

Until no changes occur, do:

for any indices *i* satisfying

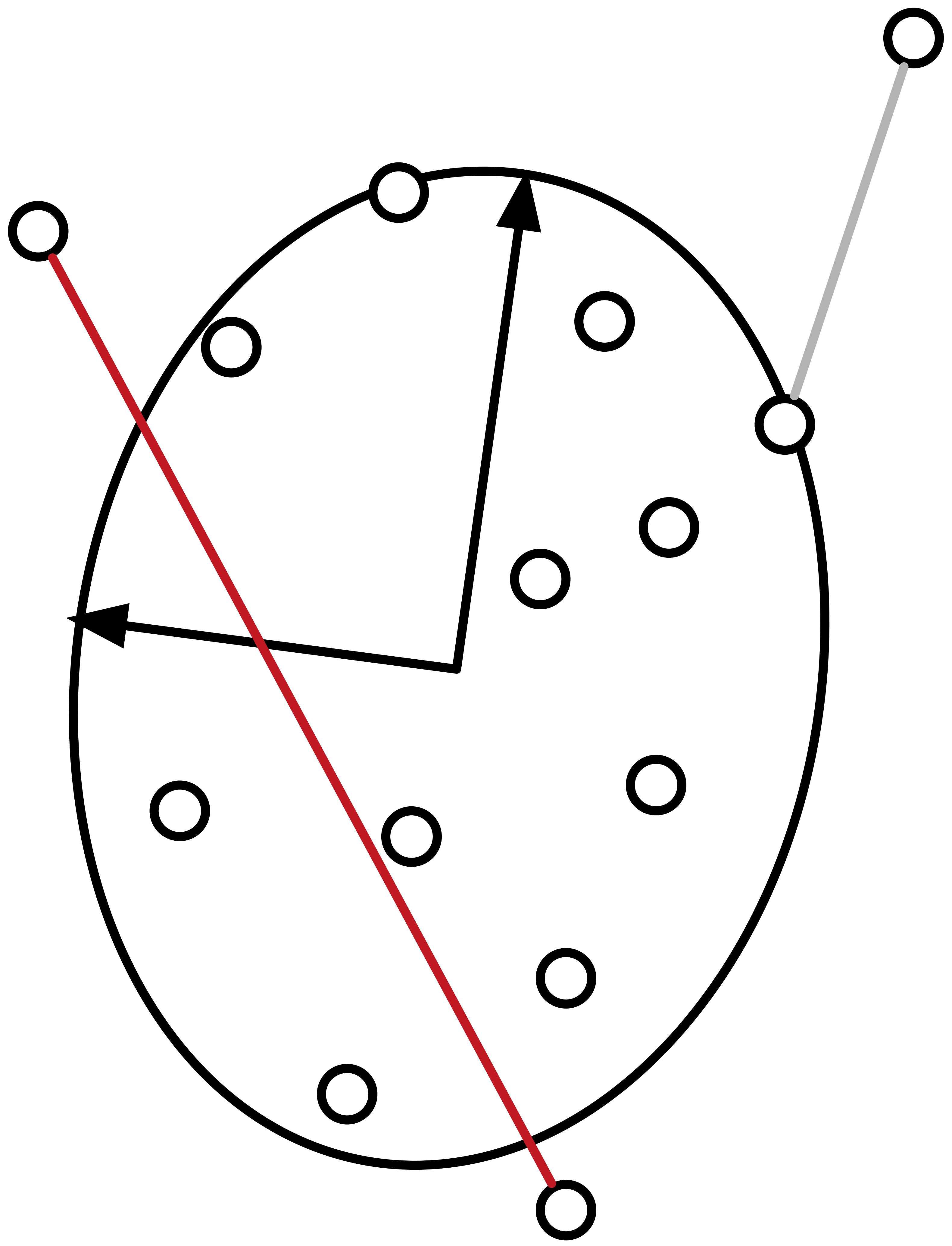
$$\widetilde{x}_i^T \Sigma^\dagger \widetilde{x}_i > c\exp(Z_i - Z_0)$$

add them to the removed indices

$$R \leftarrow R \cup \{i\}$$

re-estimate covariance

$$\Sigma \leftarrow \frac{1}{n}\sum_{i \notin R} \widetilde{x}_i \widetilde{x}_i^T$$



Stable covariance estimation

- Iteratively shrink covariance until it is stable:

$$\widetilde{x}_i \coloneqq x_i - x_{n/2+i}$$

Initialize

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} \widetilde{x}_i \widetilde{x}_i^T \qquad Z_i \overset{\text{iid}}{\sim} \frac{C}{\varepsilon}\text{Laplace}(1)$$

Until no changes occur, do:

for any indices *i* satisfying

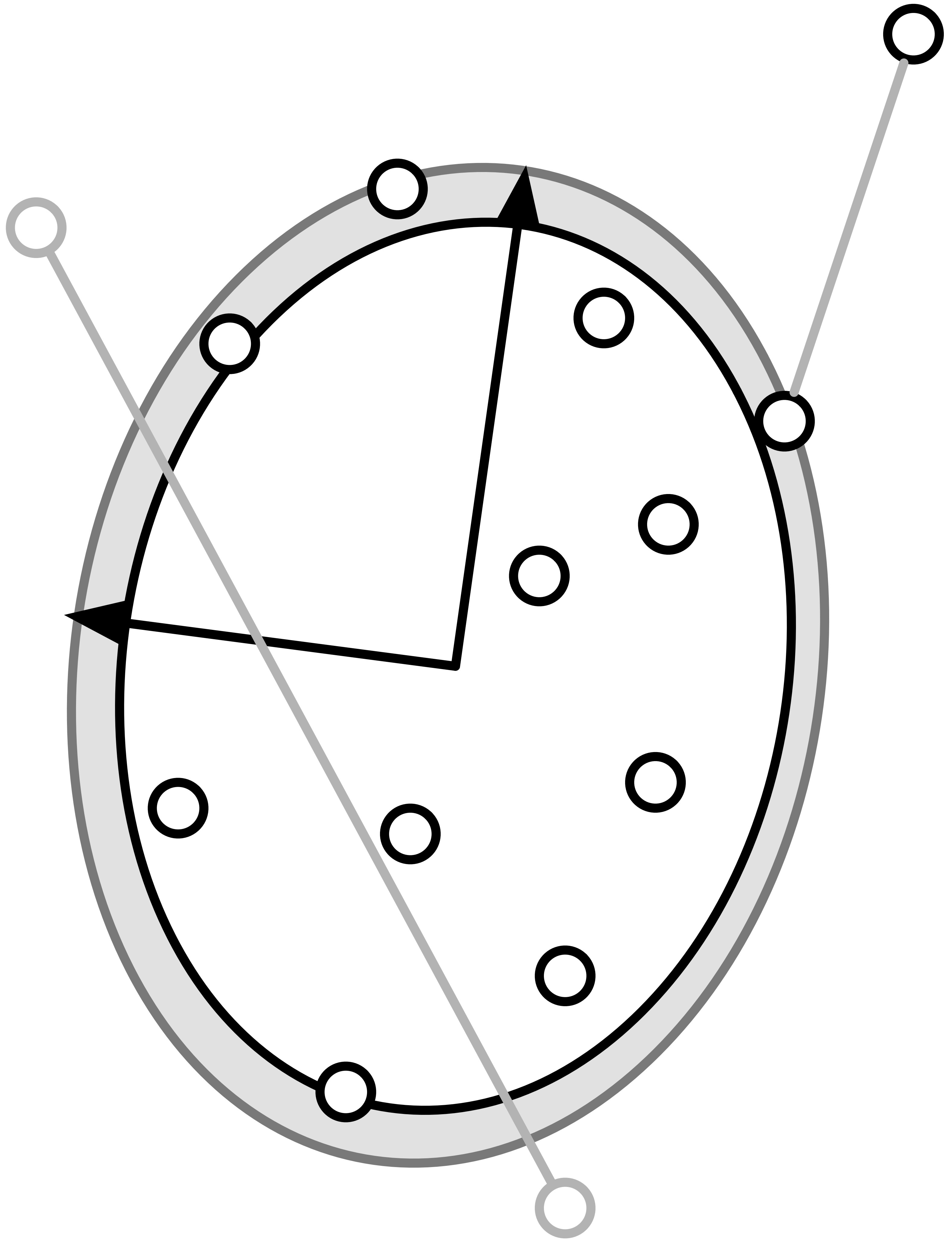
$$\widetilde{x}_i^T \Sigma^\dagger \widetilde{x}_i > c\exp(Z_i - Z_0)$$

add them to the removed indices

$$R \leftarrow R \cup \{i\}$$

re-estimate covariance

$$\Sigma \leftarrow \frac{1}{n}\sum_{i \notin R} \widetilde{x}_i \widetilde{x}_i^T$$



Stable covariance estimation

- Iteratively shrink covariance until it is stable:

$$\widetilde{x}_i := x_i - x_{n/2+i}$$

Initialize

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} \widetilde{x}_i \widetilde{x}_i^T \qquad Z_i \overset{\text{iid}}{\sim} \frac{C}{\varepsilon}\text{Laplace}(1)$$

Until no changes occur, do:

for any indices *i* satisfying

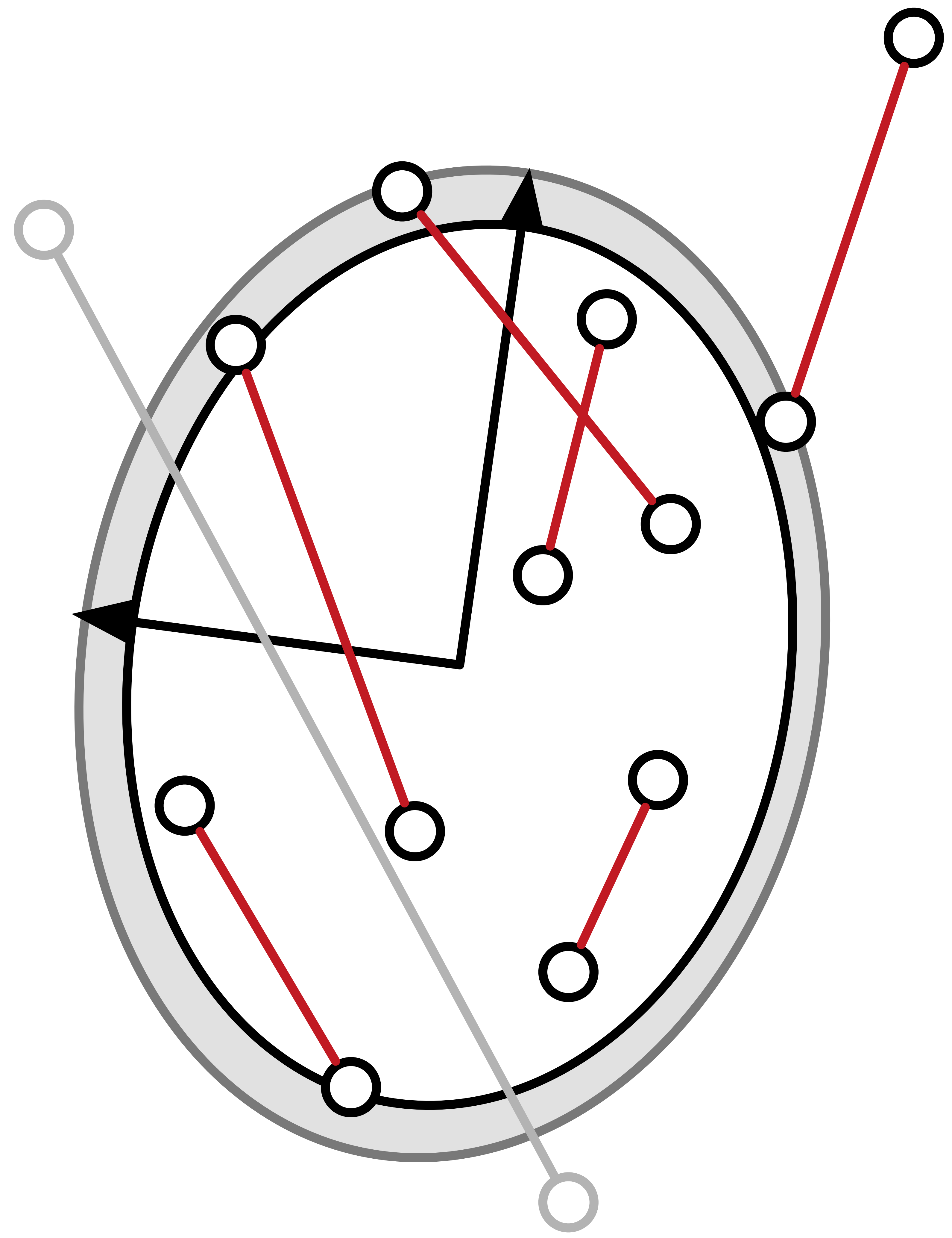
$$\widetilde{x}_i^T \Sigma^\dagger \widetilde{x}_i > c\exp(Z_i - Z_0)$$

add them to the removed indices

$$R \leftarrow R \cup \{i\}$$

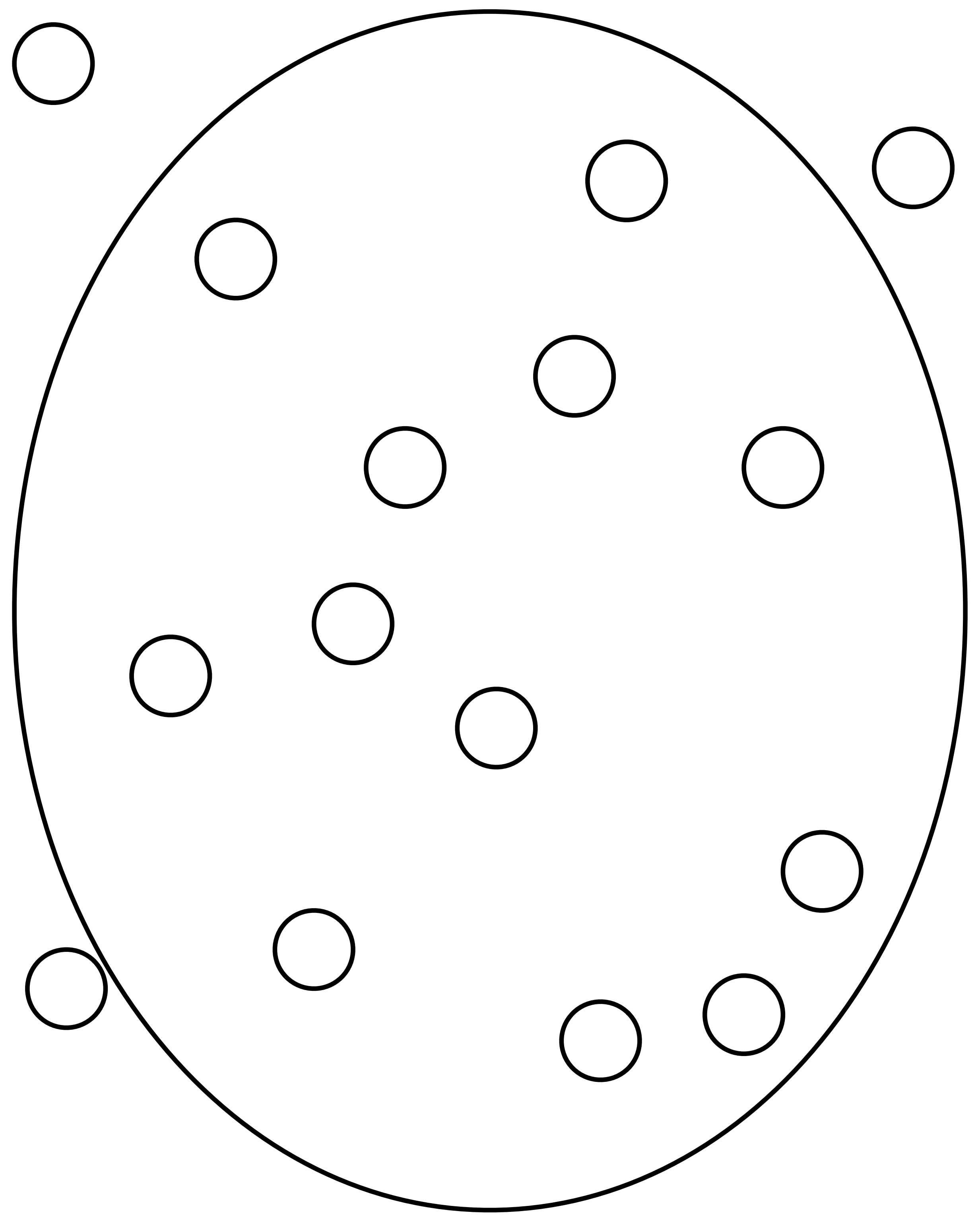
re-estimate covariance

$$\Sigma \leftarrow \frac{1}{n}\sum_{i \notin R} \widetilde{x}_i \widetilde{x}_i^T$$



Stable mean estimation

- Given putative covariance A, consider groups of indices, removing those with large A-diameter, then add noise to the result

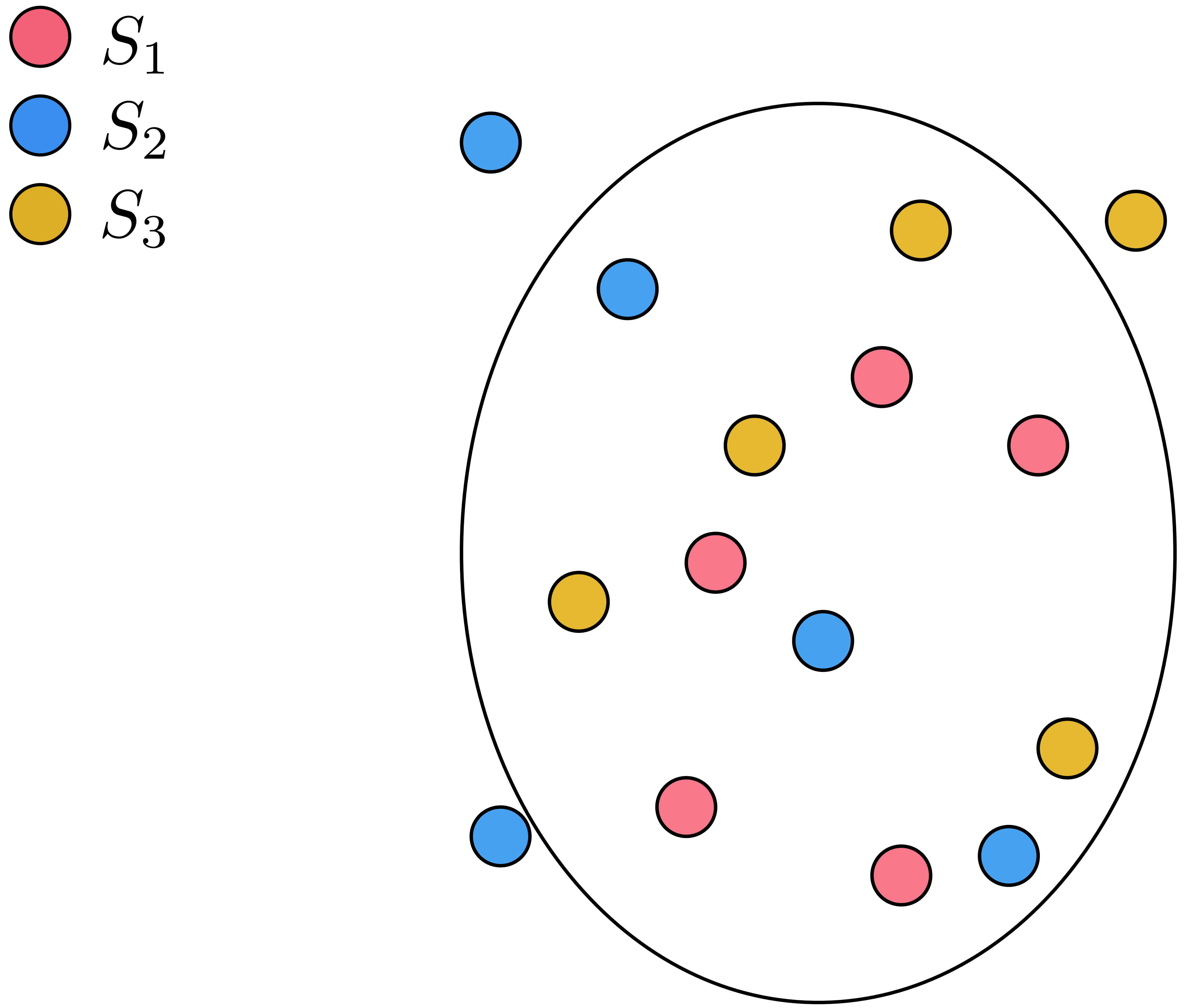


Stable mean estimation

- Given putative covariance *A*, consider groups of indices, removing those with large *A*-diameter, then add noise to the result

Random partition of indices

$$\mathcal{S} = (S_1, \ldots, S_k)$$



Stable mean estimation

- Given putative covariance *A*, consider groups of indices, removing those with large *A*-diameter, then add noise to the result

Random partition of indices

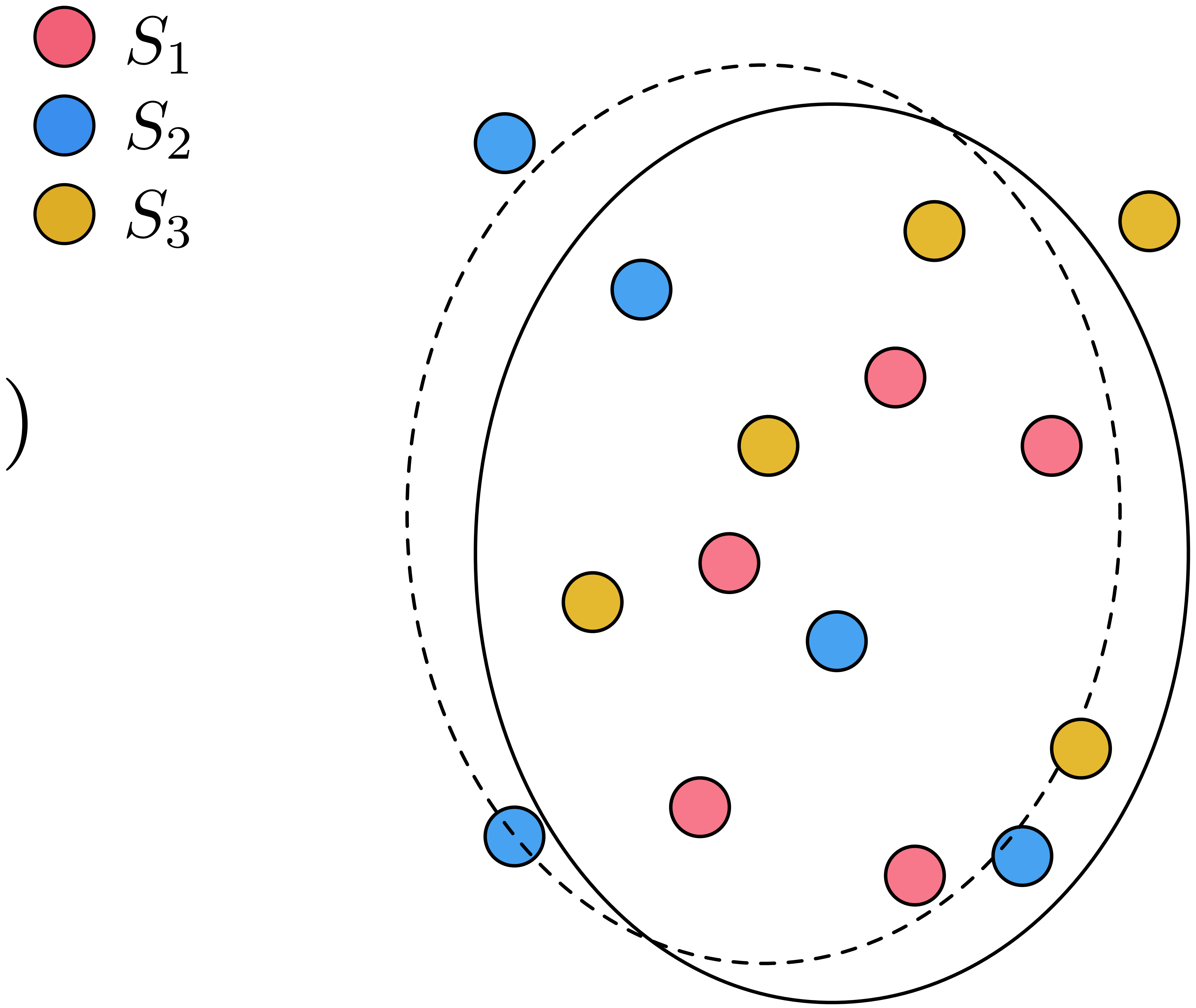
$$\mathcal{S} = (S_1, \ldots, S_k)$$

For each subset *S*, if

$$\max_{i,j \in S} \|x_i - x_j\|_A > c \exp(Z_S)$$

remove *S*:

$$S_{\mathrm{rm}} \leftarrow S_{\mathrm{rm}} \cup S$$



Stable mean estimation

- Given putative covariance A, consider groups of indices, removing those with large A-diameter, then add noise to the result

Random partition of indices

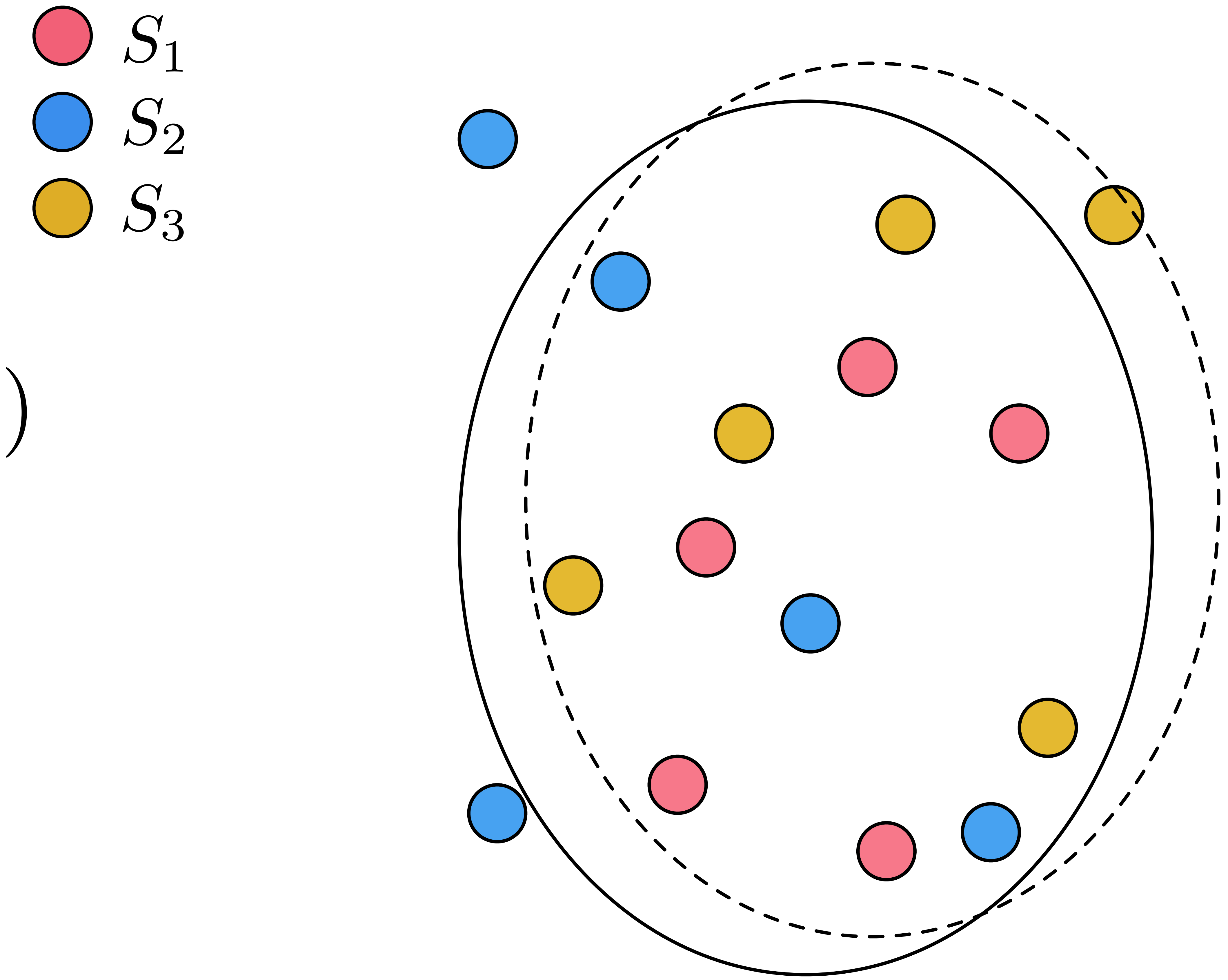
$$\mathcal{S} = (S_1, \dots, S_k)$$

For each subset S, if

$$\max_{i,j \in S} \|x_i - x_j\|_A > c \exp(Z_S)$$

remove S:

$$S_{\text{rm}} \leftarrow S_{\text{rm}} \cup S$$

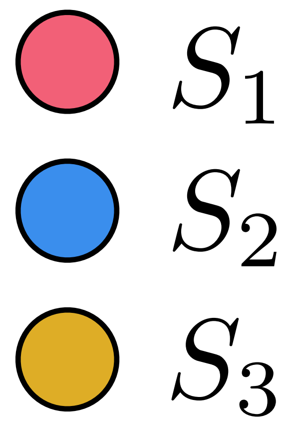


Stable mean estimation

- Given putative covariance *A*, consider groups of indices, removing those with large *A*-diameter, then add noise to the result

Random partition of indices

$$\mathcal{S} = (S_1, \ldots, S_k)$$



For each subset *S*, if

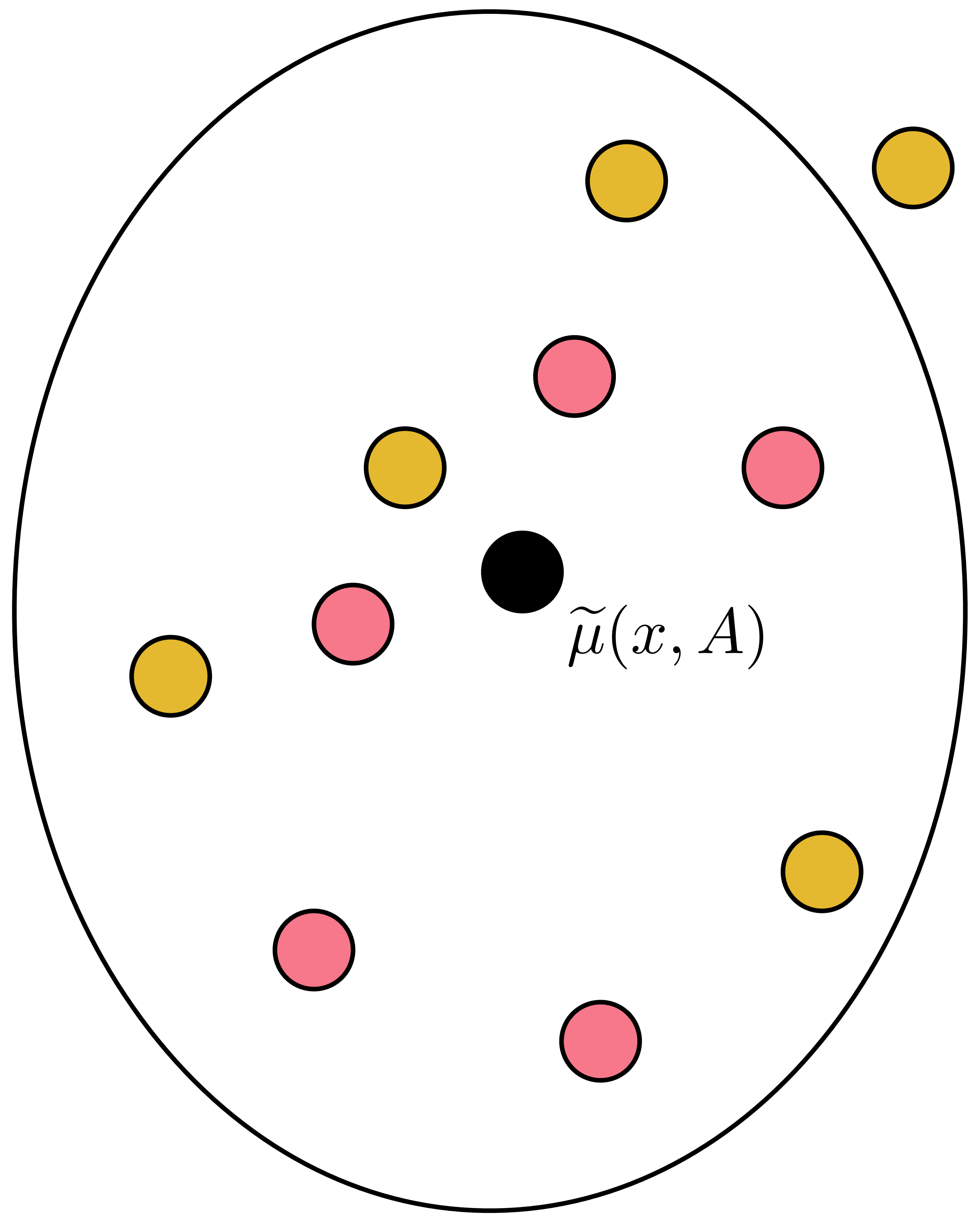
$$\max_{i,j \in S} \|x_i - x_j\|_A > c \exp(Z_S)$$

remove *S*:

$$S_{\mathrm{rm}} \leftarrow S_{\mathrm{rm}} \cup S$$

If number removed is small

$$\widetilde{\mu}(x, A) = \frac{1}{n - \mathrm{card}(S_{\mathrm{rm}})} \sum_{i \in [n] \setminus S_{\mathrm{rm}}} x_i + \mathcal{N}(0, A)$$



Ingredients for analysis

Define distributional closeness: $X \overset{d}{=}_{\varepsilon,\delta} Y$ iff $\begin{array}{l} \mathbb{P}(X \in A) \le e^{\varepsilon}\mathbb{P}(Y \in A) + \delta \\ \mathbb{P}(Y \in A) \le e^{\varepsilon}\mathbb{P}(X \in A) + \delta \end{array}$

Let x, x' be adjacent samples

Lemma

Let R be removed inds, $\widehat{\Sigma}$ covariance.

Then for $\widehat{\Sigma}_{-i} = \widehat{\Sigma} - 1\{i \notin R\}\widetilde{x}_i\widetilde{x}_i^T/n$

$$\|\widehat{\Sigma}_{-i} - \widehat{\Sigma}\| \le \frac{C}{n} \quad \text{w.h.p.}$$

Lemma

If $\|A - B\|_{\mathrm{op}} \le \dfrac{C}{n}$ then

$$\widetilde{\mu}(x, A) \overset{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x, B)$$

Lemma

Compute R' on input x'. Then

$$R \overset{d}{=}_{\varepsilon,\delta} R'$$

Lemma

For any *A*,

$$\widetilde{\mu}(x, A) \overset{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x', A)$$

Lemma

Let R be removed inds, $\widehat{\Sigma}$ covariance.

Then for $\widehat{\Sigma}_{-i} = \widehat{\Sigma} - 1\{i \notin R\}\widetilde{x}_i\widetilde{x}_i^T/n$

$$\|\widehat{\Sigma}_{-i} - \widehat{\Sigma}\| \leq \frac{C}{n} \quad \text{w.h.p.}$$

Lemma

If $\|A - B\|_{\text{op}} \leq \frac{C}{n}$ then

$$\widetilde{\mu}(x, A) \overset{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x, B)$$

Lemma

Compute R' on input x'. Then

$$R \overset{d}{=}_{\varepsilon,\delta} R'$$

Lemma

For any *A*,

$$\widetilde{\mu}(x, A) \overset{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x', A)$$

Privacy guarantees

Theorem (D., Haque, Kuditipudi)

Let $\widetilde{\mu}(x, \widehat{\Sigma})$ be the output of the stable mean procedure with input covariance $\widehat{\Sigma}$ estimated by the stable covariance procedure. Assume

$$n \geq \frac{d}{\varepsilon^2} \log^2 \frac{1}{\delta}$$

Then

$\widetilde{\mu}(x, \widehat{\Sigma})$ is (ε, δ)-differentially private

Lemma

Let R be removed inds, $\widehat{\Sigma}$ covariance.

Then for $\widehat{\Sigma}_{-i} = \widehat{\Sigma} - 1\{i \notin R\}\widetilde{x}_i\widetilde{x}_i^T/n$

$$\|\widehat{\Sigma}_{-i} - \widehat{\Sigma}\| \leq \frac{C}{n} \quad \text{w.h.p.}$$

Lemma

If $\|A - B\|_{\mathrm{op}} \leq \frac{C}{n}$ then

$$\widetilde{\mu}(x, A) \stackrel{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x, B)$$

Lemma

Compute R' on input x'. Then

$$R \stackrel{d}{=}_{\varepsilon,\delta} R'$$

Lemma

For any A,

$$\widetilde{\mu}(x, A) \stackrel{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x', A)$$

Lemma

Let *R* be removed inds, $\widehat{\Sigma}$ covariance.

Then for $\widehat{\Sigma}_{-i} = \widehat{\Sigma} - 1\{i \notin R\}\widetilde{x}_i\widetilde{x}_i^T/n$

$$\|\widehat{\Sigma}_{-i} - \widehat{\Sigma}\| \leq \frac{C}{n} \quad \text{w.h.p.}$$

Lemma

Let R be removed inds, $\widehat{\Sigma}$ covariance.

Then for $\widehat{\Sigma}_{-i} = \widehat{\Sigma} - 1\{i \notin R\}\widetilde{x}_i\widetilde{x}_i^T/n$

$$\|\widehat{\Sigma}_{-i} - \widehat{\Sigma}\| \leq \frac{C}{n} \quad \text{w.h.p.}$$

Lemma

If $\|A - B\|_{\text{op}} \leq \frac{C}{n}$ then

$$\widetilde{\mu}(x, A) \stackrel{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x, B)$$

Lemma

Compute R' on input x'. Then

$$R \stackrel{d}{=}_{\varepsilon,\delta} R'$$

Lemma

For any A,

$$\widetilde{\mu}(x, A) \stackrel{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x', A)$$

Lemma

Compute R' on input x'. Then

$$R \stackrel{d}{=}_{\varepsilon,\delta} R'$$

Lemma

Let *R* be removed inds, $\widehat{\Sigma}$ covariance.

Then for $\widehat{\Sigma}_{-i} = \widehat{\Sigma} - 1\{i \notin R\}\widetilde{x}_i\widetilde{x}_i^T/n$

$$\|\widehat{\Sigma}_{-i} - \widehat{\Sigma}\| \leq \frac{C}{n} \quad \text{w.h.p.}$$

Lemma

If $\|A - B\|_{\text{op}} \leq \frac{C}{n}$ then

$$\widetilde{\mu}(x, A) \stackrel{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x, B)$$

Lemma

Compute R' on input x'. Then

$$R \stackrel{d}{=}_{\varepsilon,\delta} R'$$

Lemma

For any *A*,

$$\widetilde{\mu}(x, A) \stackrel{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x', A)$$

Lemma

If $\|A - B\|_{\mathrm{op}} \leq \dfrac{C}{n}$ then

$$\widetilde{\mu}(x, A) \stackrel{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x, B)$$

Lemma

Let R be removed inds, $\widehat{\Sigma}$ covariance.

Then for $\widehat{\Sigma}_{-i} = \widehat{\Sigma} - 1\{i \notin R\}\widetilde{x}_i\widetilde{x}_i^T/n$

$$\|\widehat{\Sigma}_{-i} - \widehat{\Sigma}\| \leq \frac{C}{n} \quad \text{w.h.p.}$$

Lemma

If $\|A - B\|_{\text{op}} \leq \frac{C}{n}$ then

$$\widetilde{\mu}(x, A) \overset{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x, B)$$

Lemma

Compute R' on input x'. Then

$$R \overset{d}{=}_{\varepsilon,\delta} R'$$

Lemma

For any *A*,

$$\widetilde{\mu}(x, A) \overset{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x', A)$$

Lemma

For any *A*,

$$\widetilde{\mu}(x, A) \overset{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x', A)$$

Accuracy Guarantees

Theorem (D., Haque, Kuditipudi)

Let $\widetilde{\mu}(x, \widehat{\Sigma})$ be the output of the stable mean procedure with input covariance $\widehat{\Sigma}$ estimated by the stable covariance procedure. Assume

$$\max_{i \le n} \|X_i - \mu\|_\Sigma^2 \le M^2$$

with high probability. Then

$$\|\widetilde{\mu}(x, \widehat{\Sigma}) - \mu\|_\Sigma^2 \lesssim \frac{d}{n} + \frac{M^2 d \log^2 \frac{1}{\delta}}{n^2 \varepsilon^2}$$

Corollary. If the data are subgaussian, then w.h.p.

$$\|\widetilde{\mu}(x, \widehat{\Sigma}) - \mu\|_\Sigma^2 \lesssim \frac{d}{n} + \frac{d(d + \log n) \log^2 \frac{1}{\delta}}{n^2 \varepsilon^2}$$

Conclusions, extensions, next steps

- We have a polynomial time private mean estimator adaptive to the covariance with (up to logarithmic factors) minimax optimal covariance
- Algorithm is, unfortunately, still not completely practical
- Can adapt to data with fewer moments, though a bit subtle
- Connections between robustness and differential privacy may offer substantial opportunities for practical (and theoretical) advances