

Theory of Differential Privacy

Lecture 17:

Private mean estimation

EN.601.438/638 Fall 2026

Lydia Zakynthinou

Johns Hopkins University

Oct 26, 2026

Private mean Estimation of Gaussians

$$X_1, \dots, X_n \sim N(\mu, \Sigma) \quad , \mu \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d \times d}, \Sigma \succ 0$$

Goal: Find $\hat{\mu}$: $\|\hat{\mu} - \mu\|_{\Sigma} \leq \alpha$.

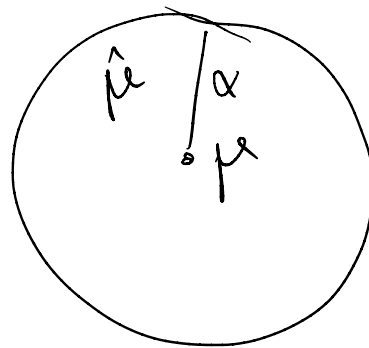
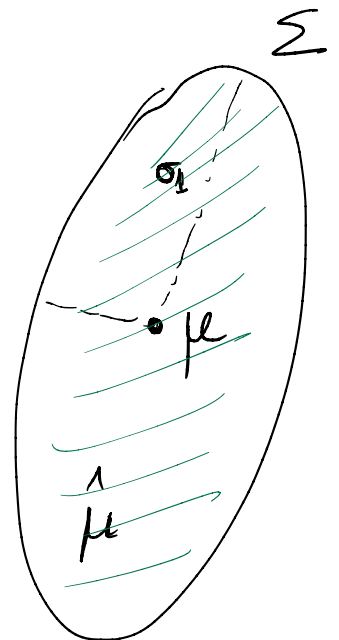
↳ Mahalanobis norm

Emp. mean $\bar{X}$:

$$\|\Sigma^{-1/2} v\|_2 = \|v\|_{\Sigma} = \sqrt{v \Sigma^{-1} v}$$

$$\mathbb{E}[\|\bar{X} - \mu\|_{\Sigma}^2] \leq \frac{d}{n}$$

$$\mathbb{E}[\|\bar{X} - \mu\|_{\Sigma}] = \mathbb{E}[\|\Sigma^{-1/2}(\bar{X} - \mu)\|_2] \\ = \text{tr}(A)$$

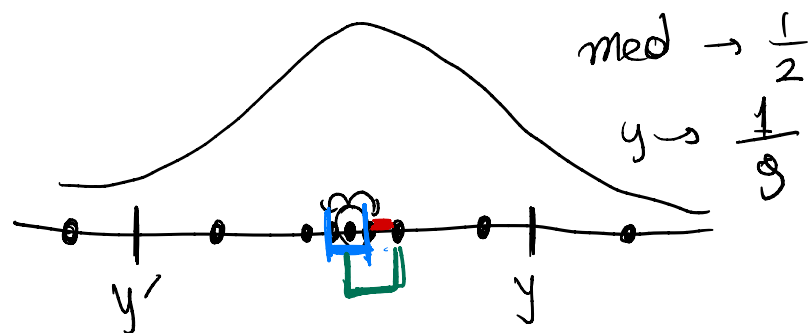


$$\bar{X} \sim N(\mu, \frac{\Sigma}{n}) \quad \text{If } n \geq \frac{d}{\alpha^2} \Rightarrow \mathbb{E}[\|\bar{X} - \mu\|_{\Sigma}] \leq \alpha.$$

$$\Delta \bar{X} = \infty, \quad LS(x) = \infty.$$

if $\varepsilon > \alpha$.

Target theorem $\exists (\varepsilon, \delta)$ -DP w/ $n \geq \frac{d}{\alpha^2} + \frac{d}{\alpha \varepsilon} + \frac{\log(1/\delta)}{\varepsilon}$.

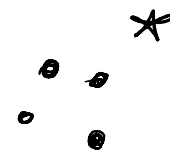


Multivariate Tukey depth:

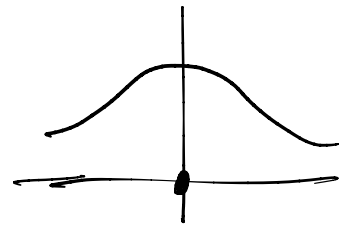
$$T_x(y) = \min_{v \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\langle x_i, v \rangle \geq \langle y, v \rangle\}}.$$

Notice $q(x; y) = n T_x(y)$.

$$\Delta q = \max_{\substack{x \sim x' \\ y}} |q(x'; y) - q(x; y)| = 1$$



Tukey depth & Multivariate



$$T_P(y) = \min_v \Pr[\langle x, v \rangle \geq \langle y, v \rangle] \quad x \sim P$$

for $P = N(\mu, \Sigma)$: $T_P(y) = \Phi(-\|y - \mu\|_\Sigma)$

^
CDF of $N(0, 1)$

Proof sketch:

$$T_P(y) = \min_v \Pr[\langle x, v \rangle \geq \langle y, v \rangle] = *$$

$x \sim N(\mu, \Sigma)$

$$\langle x, v \rangle \sim N(\langle \mu, v \rangle, v^T \Sigma v)$$

$$* = \min_v \Pr \left[\frac{\langle x, v \rangle - \langle \mu, v \rangle}{\sqrt{v^T \Sigma v}} \geq \frac{\langle y, v \rangle - \langle \mu, v \rangle}{\sqrt{v^T \Sigma v}} \right]$$

use that

$$\max_{\substack{u \neq 0 \\ \|u\|_2 = 1}} \langle z, u \rangle = \|z\|_2$$

$$= \min_v \Phi \left(-\frac{\langle y - \mu, v \rangle}{\sqrt{v^T \Sigma v}} \right) = \Phi(-\|y - \mu\|_\Sigma)$$

If $T_P(y) = \frac{1}{2} \Rightarrow y = \mu$.

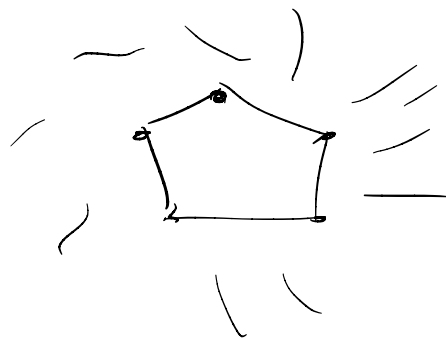
$T_P(y) \geq \frac{1}{2} - \alpha \Rightarrow \|y - \mu\|_\Sigma \leq \Phi^{-1}(\frac{1}{2} + \alpha) \approx \alpha$ for $\alpha \in (0, 1)$.

Typicality, if $n \geq \frac{d}{\alpha^2}$ then $|T_x(y) - T_{N(\mu, \Sigma)}(y)| \leq \alpha \quad \forall y \in \mathbb{R}^d$.

Goal: find y s.t. $T_x(y) \geq \frac{1}{2} - \alpha$.

Exp. Med.

$$p_x(y) = \frac{\exp(\frac{\epsilon}{2} \cdot q(x, y))}{Z_x}, \quad Z_x = \int_{\mathbb{R}^d} \exp(\frac{\epsilon}{2} \cdot q(x, z)) dz.$$

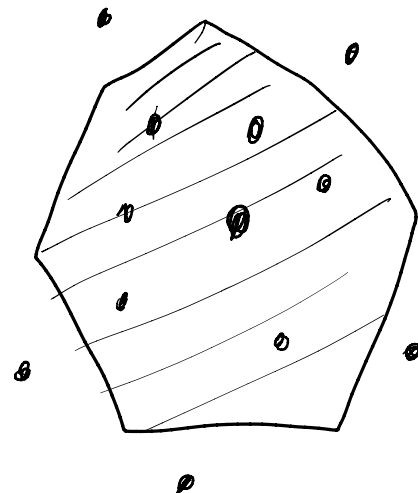
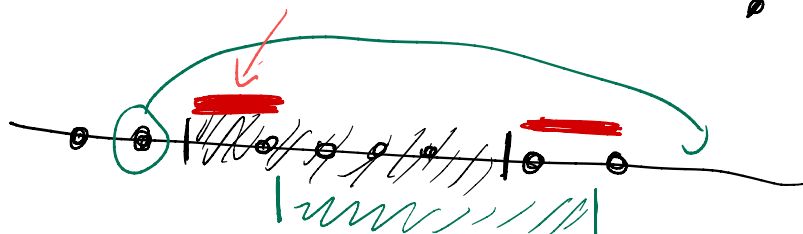


$$\int_{\mathbb{R}^d \setminus \text{conv}(x)} 1 dz = \infty$$

over $y_{t,x} = \{y : q(x,y) \geq t\}$ pick $t = \frac{\eta}{4}$.

$$T_x(y) \geq \frac{t}{n}$$

$M_{\varepsilon,t}(x)$ has density $m_x(y) = \frac{e^{\varepsilon/2 q(x,y)} \cdot \mathbf{1}_{\{y \in y_{t,x}\}}}{Z_{t,x}}$



Define:

$$\text{SAFE}_{\varepsilon,\delta,t} = \{x : M_{\varepsilon,t}(x) \approx_{\varepsilon,\delta} M_{\varepsilon,t}(x') \text{ for every } x' \sim x\}$$

$$\hookrightarrow P \approx_{\varepsilon,\delta} Q \Leftrightarrow \begin{aligned} P(E) &\leq e^\varepsilon Q(E) + \delta \\ Q(E) &\leq e^\varepsilon P(E) + \delta \end{aligned}$$

Restricted E.M. w/ Tukey depth score

② $h = d_H(x, \text{UNSAFE}_{\varepsilon, \delta, t})$, $\text{UNSAFE}_{\varepsilon, \delta, t} = \text{SAFE}_{\varepsilon, \delta, t}^c$.

• If $h + \log\left(\frac{1}{\varepsilon}\right) < \frac{\log(1/(2\delta))}{\varepsilon}$, return $\perp$.
Else run $M_{\varepsilon, t}(x)$.

Thm. REM is $(2\varepsilon, e^{\varepsilon\delta})$ -DP.

$\rightarrow$ If $M_{\varepsilon, t}(x) \not\sim_{\varepsilon\delta} M_{\varepsilon, t}(x')$

$\rightarrow$ If $M_{\varepsilon, t}(x) \approx_{\varepsilon\delta} M_{\varepsilon, t}(x')$