Geometric Lower Bounds for Distributed Parameter Estimation under Communication Constraints

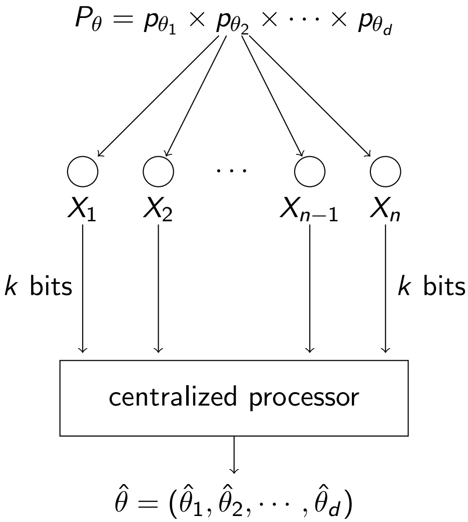
Yanjun Han (Stanford EE)

Joint work with:

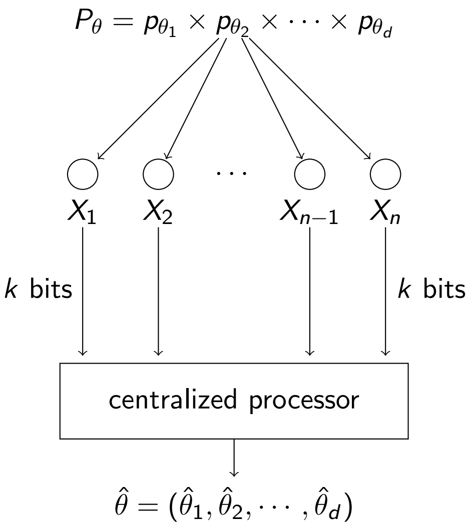
Ayfer Özgür	Stanford EE
Tsachy Weissman	Stanford EE

COLT 2018, Stockholm, Sweden

Distributed Parameter Estimation



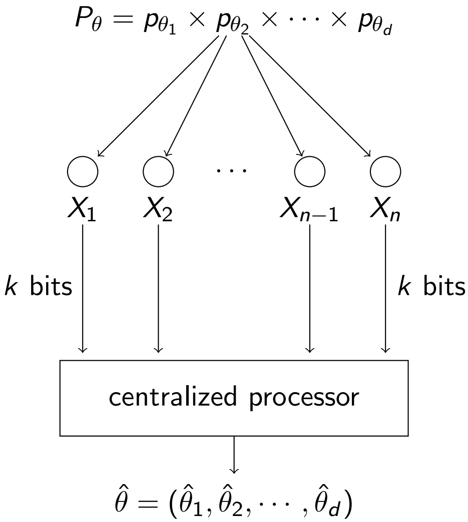
Distributed Parameter Estimation



Parameters:

- n: number of sensors
- k: number of bits
- d: dimensionality

Distributed Parameter Estimation



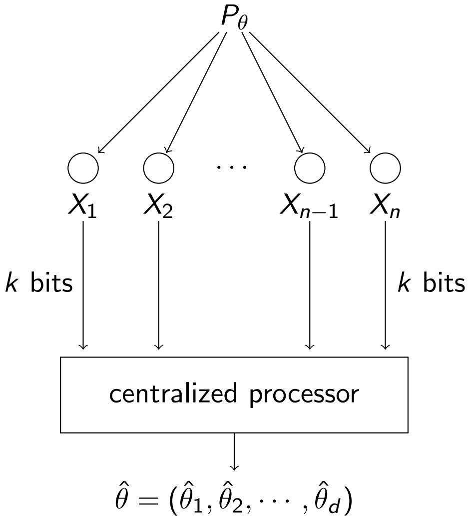
Parameters:

- n: number of sensors
- k: number of bits
- d: dimensionality

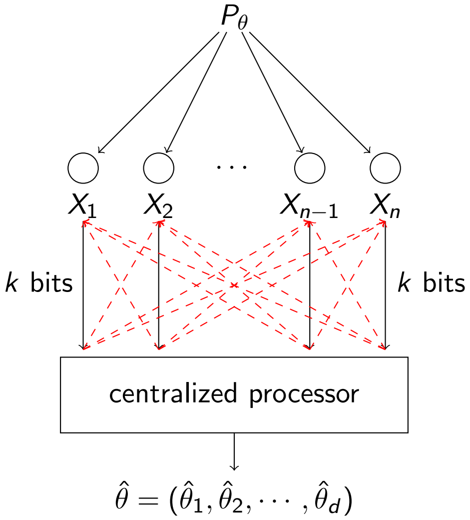
Goal: characterize

$$\inf_{\text{schemes}} \sup_{\theta} \mathbb{E}_\theta \|\hat{\theta} - \theta\|_2^2$$

Blackboard Communication Protocol



Blackboard Communication Protocol



General Lower Bounds

Theorem

Fix any θ_0, let $S(X)$ be the score function of (p_θ) around $\theta = \theta_0$:

$$S(X) = \frac{\partial}{\partial \theta} \log p_\theta(X) \Big|_{\theta=\theta_0}.$$

Assuming mild regularity conditions,

$$\inf_{\textit{schemes}} \sup_{\theta} \mathbb{E}_\theta \|\hat{\theta} - \theta\|_2^2 \gtrsim \frac{d}{n\mathrm{Var}(S(X))} \vee \frac{d^2}{n2^k \mathrm{Var}(S(X))} \vee \frac{d^2}{nk\|S(X)\|_{\psi_2}^2}.$$

Examples

Statistical Model	Centralized MSE	Distributed MSE
$P_\theta = \mathcal{N}(\mu, I_d)$ (ZDJW'13, GMN'14)	$\frac{d}{n}$	$\frac{d}{n} \cdot \frac{d}{k}$
$P_\theta = (\theta_1, \cdots, \theta_d)$ (HMOW'18)	$\frac{1}{n}$	$\frac{1}{n} \cdot \frac{d}{2^k}$
$P_\theta = \prod_{i=1}^d \text{Bern}(\theta_i)$, $\theta_i \in [0,1]$ (ZDJW'13)	$\frac{d}{n}$	$\frac{d}{n} \cdot \frac{d}{k}$
$P_\theta = \prod_{i=1}^d \text{Bern}(\theta_i)$, $\sum_{i=1}^d \theta_i = 1$	$\frac{1}{n}$	$\frac{1}{n} \cdot \frac{d}{2^k}$

Examples

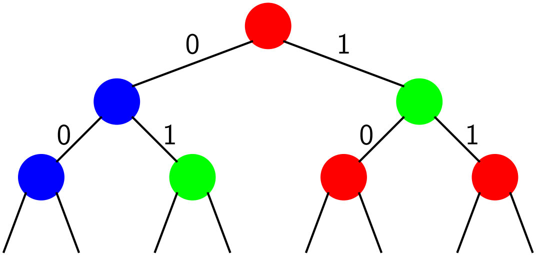
Statistical Model	Centralized MSE	Distributed MSE
$P_\theta = \mathcal{N}(\mu, I_d)$ (ZDJW'13, GMN'14)	$\frac{d}{n}$	$\frac{d}{n} \cdot \frac{d}{k}$
$P_\theta = (\theta_1, \cdots, \theta_d)$ (HMOW'18)	$\frac{1}{n}$	$\frac{1}{n} \cdot \frac{d}{2^k}$
$P_\theta = \prod_{i=1}^d \text{Bern}(\theta_i)$, $\theta_i \in [0,1]$ (ZDJW'13)	$\frac{d}{n}$	$\frac{d}{n} \cdot \frac{d}{k}$
$P_\theta = \prod_{i=1}^d \text{Bern}(\theta_i)$, $\sum_{i=1}^d \theta_i = 1$	$\frac{1}{n}$	$\frac{1}{n} \cdot \frac{d}{2^k}$

Product Bernoulli model:

- $\theta_0 = \frac{1}{2}$: $\frac{d^2}{n}\left(\frac{1}{2^k \cdot 1} \vee \frac{1}{k \cdot 1}\right) = \frac{d^2}{nk}$
- $\theta_0 = \frac{1}{d}$: $\frac{d^2}{n}\left(\frac{1}{2^k \cdot d} \vee \frac{1}{k \cdot d^2}\right) = \frac{d}{n2^k}$

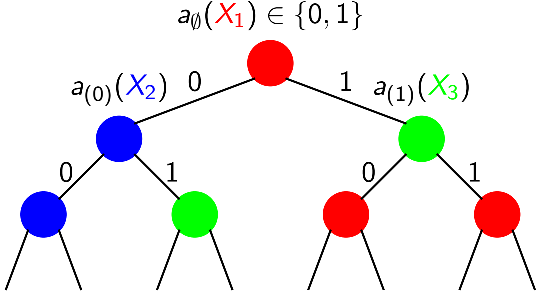
Representation of Protocol

Red - Sensor 1, Blue - Sensor 2, Green - Sensor 3



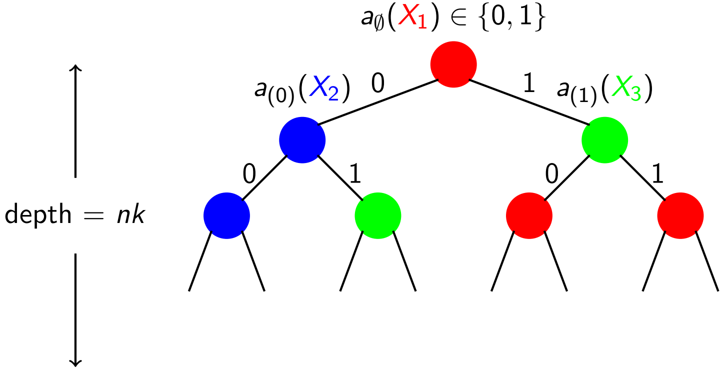
Representation of Protocol

Red - Sensor 1, Blue - Sensor 2, Green - Sensor 3



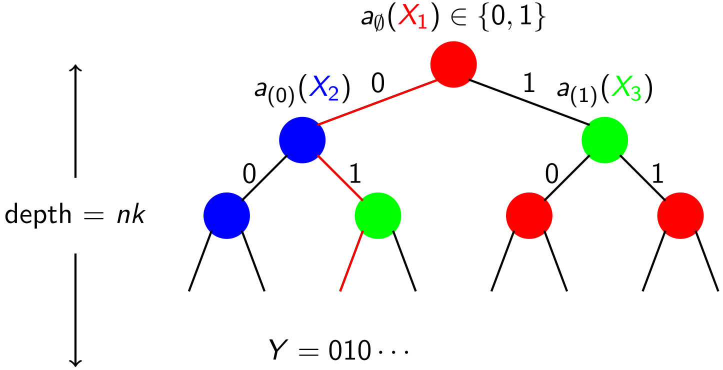
Representation of Protocol

Red - Sensor 1, Blue - Sensor 2, Green - Sensor 3



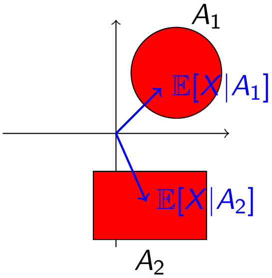
Representation of Protocol

Red - Sensor 1, Blue - Sensor 2, Green - Sensor 3



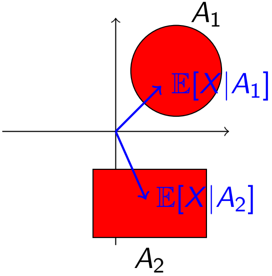
Geometric Inequalities

- Let $X = (X_1, \cdots, X_d)$ be a random vector with independent and zero-mean entries
- Given $\mathbb{P}(A) = t$, aim to maximize $\|\mathbb{E}[X|A]\|_2$



Geometric Inequalities

- Let $X = (X_1, \cdots, X_d)$ be a random vector with independent and zero-mean entries
- Given $\mathbb{P}(A) = t$, aim to maximize $\|\mathbb{E}[X|A]\|_2$



Theorem

- If $\max_{i\in[d]} \mathrm{Var}(X_i) \le \sigma^2$,

$$\|\mathbb{E}[X|A]\|_2^2 \le \sigma^2 \cdot \frac{1 - \mathbb{P}(A)}{\mathbb{P}(A)}$$

- If $\max_{i\in[d]} \|X_i\|_{\psi_2}^2 \le \sigma^2$,

$$\|\mathbb{E}[X|A]\|_2^2 \le C\sigma^2 \cdot \log \frac{1}{\mathbb{P}(A)}$$

Comparison with SDPI

Strong data processing inequality (SDPI):

$$I(\Theta; Y) \le \gamma^*(\Theta, X) I(X; Y)$$

Comparison with SDPI

Strong data processing inequality (SDPI):

$$I(\Theta; Y) \leq \gamma^*(\Theta, X) I(X; Y)$$

- $\Theta - X$ determined by statistical model $X \sim P_\Theta$, $X - Y$ subject to communication constraints

Comparison with SDPI

Strong data processing inequality (SDPI):

$$I(\Theta; Y) \leq \gamma^*(\Theta, X) I(X; Y)$$

- $\Theta - X$ determined by statistical model $X \sim P_\Theta$, $X - Y$ subject to communication constraints
- leads to tight results in Gaussian location model

Comparison with SDPI

Strong data processing inequality (SDPI):

$$I(\Theta; Y) \leq \gamma^*(\Theta, X) I(X; Y)$$

- $\Theta - X$ determined by statistical model $X \sim P_\Theta$, $X - Y$ subject to communication constraints
- leads to tight results in Gaussian location model
- result in linear dependence on k, while tight dependence may be exponential

Comparison with SDPI

Strong data processing inequality (SDPI):

$$I(\Theta; Y) \leq \gamma^*(\Theta, X) I(X; Y)$$

- $\Theta - X$ determined by statistical model $X \sim P_\Theta$, $X - Y$ subject to communication constraints
- leads to tight results in Gaussian location model
- result in linear dependence on k, while tight dependence may be exponential
- unclear operational meaning

Comparison with SDPI

Strong data processing inequality (SDPI):

$$I(\Theta; Y) \leq \gamma^*(\Theta, X) I(X; Y)$$

- $\Theta - X$ determined by statistical model $X \sim P_\Theta$, $X - Y$ subject to communication constraints
- leads to tight results in Gaussian location model
- result in linear dependence on k, while tight dependence may be exponential
- unclear operational meaning

Thank you!