Generalization Bounds for Federated Learning:

Fast Rates, Unparticipating Clients and Unbounded Losses

Xiaolin Hu, Shaojie Li, Yong Liu

Present at ICLR 2023
2023/04/08

Background

- **Federated Learning (FL)**

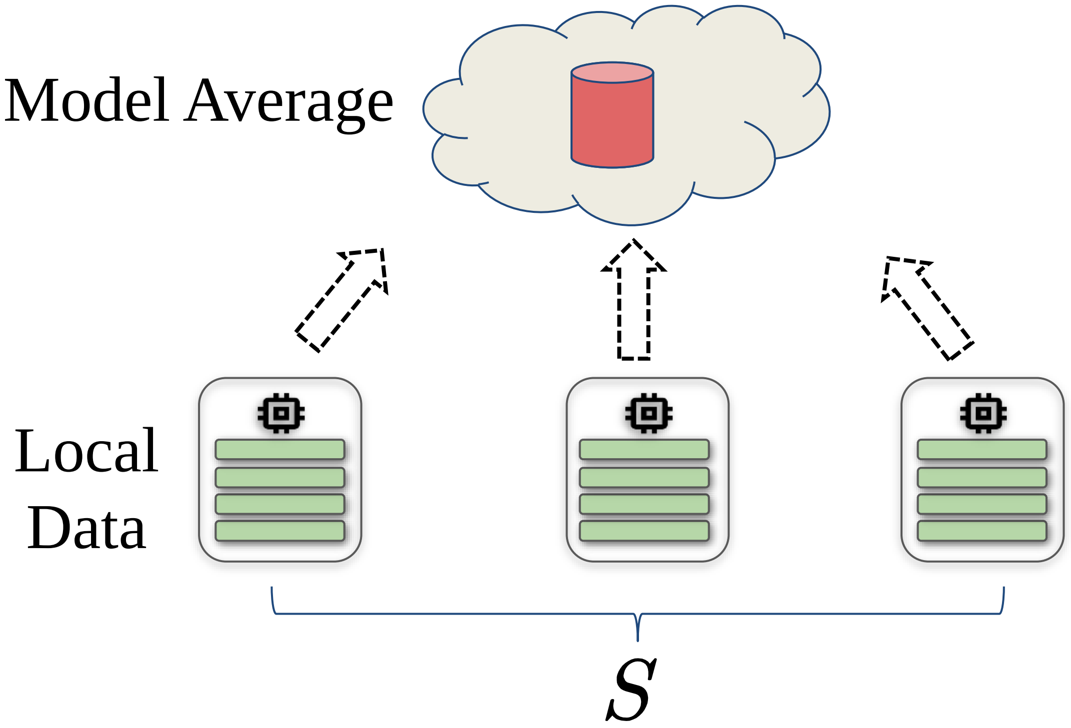


Figure: Federated learning setting[1]

- **FL: Characteristics**
  - **Communication cost:** **Communication cost among clients is high**
  - **Participation rate:** **The realistic participation ratio may be slow**
  - **Data Heterogeneity:** **Data distributions vary across clients**

[1] Figure adapted from https://www.ai4opt.org/sites/default/files/slides/kale.pdf

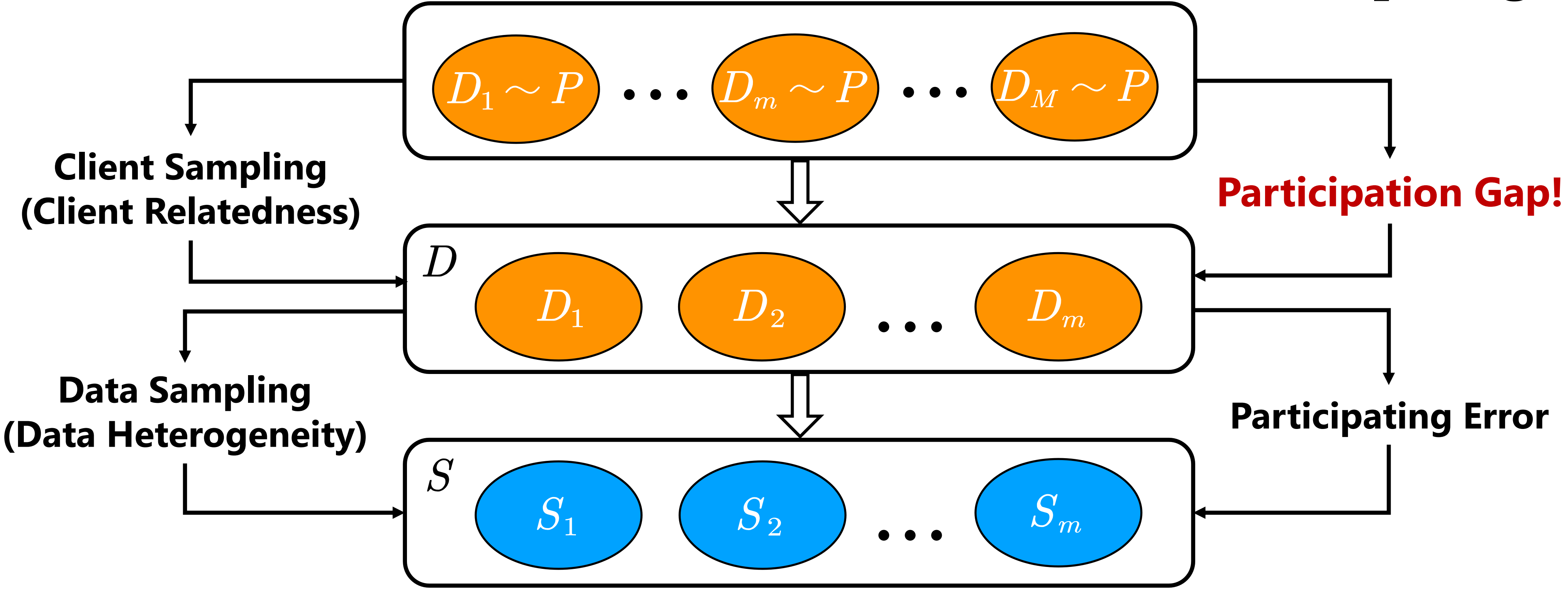
Related Work

- **Related Work**
  - **Optimization: Training Error**
  - **Participating Clients**
  - **Homogeneous Data: i.i.d**
  - **Bounded Losses**
- **Our Work**
  - **Generalization： Testing Error**
  - **Unparticipating Clients**
  - **Heterogeneous Data: Non-i.i.d**
  - **Unbounded Losses**

Two-Level Distribution Framework

- **Modeling the Participating Gap**

M: **Total Clients**
m: **Participating Clients**



Would the unparticipating clients benefit from the model trained by participating clients?

Two-Level Distribution Framework

m: Participating clients
n: Sample size per client

Average over clients — Expectation over data

$$\mathcal{L}_D(h) = \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}_{\boldsymbol{Z}_i \sim \boldsymbol{D}_i}\left[\ell(h, Z_i)\right]$$

Learned model — Loss function

Client Sampling

$\mathcal{L}_P(\hat{\boldsymbol{h}})$?

$$\hat{h} = \operatorname*{argmin}_{h \in \mathcal{H}} \mathcal{L}_S(h)$$

Data Sampling

$$\mathcal{L}_P(h) = \mathop{\mathbb{E}}_{D_i \sim P}\left[\mathop{\mathbb{E}}_{Z \sim D_i}\left[\ell(h, Z)\right]\right]$$

$$\mathcal{L}_S(h) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\ell(h, Z_i^j)$$

Generalization Bounds——Fast Rates

- **Learning Rates for unparticipating Client ——Bounded Losses**

$$\hat{h} = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \mathcal{L}_S(h)$$

$$\mathcal{L}_P(\hat{h}) - \min_{h \in \mathcal{H}} \mathcal{L}_P(h) \leq \mathcal{O}\left(\sqrt{\frac{1}{mn}} + \sqrt{\frac{1}{m}}\right)$$

- **Bernstein Condition:** $\mathbf{E}[\ell(h,Z) - \ell(h^\star,Z)]^2 \leq B\,\mathbf{E}[\ell(h,Z) - \ell(h^\star,Z)]$

- **Learning Rates for unparticipating Client——Bernstein Condition**

$$\mathcal{L}_P(\hat{h}) - \min_{h \in \mathcal{H}} \mathcal{L}_P(h) \leq \mathcal{O}\left(\frac{1}{mn} + \frac{1}{m}\right)$$

Unparticipating clients would benefit from the model trained by participating clients!

Generalization Bounds——Heavy-tail data

■ Heavy-tail Distribution

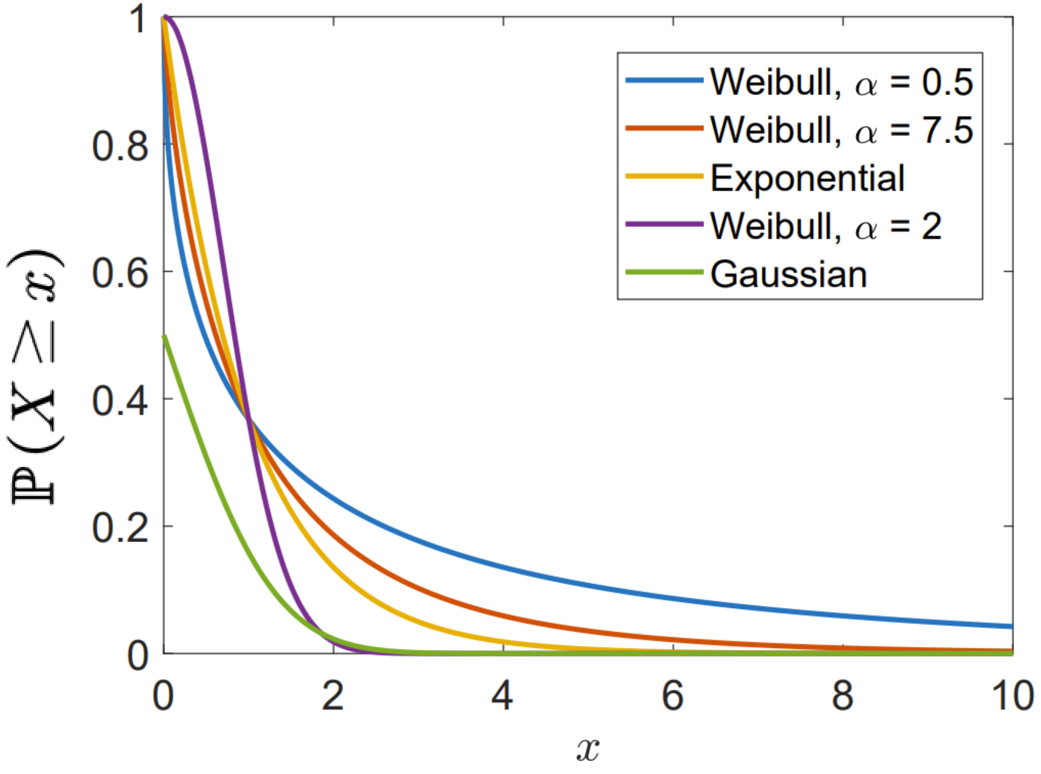


Figure: Illustration of heavy tails[2]

■ Learning Rates for unparticipating Client——Heavy-tail Data

$\eta \in (0, 1/2)$

$$\mathcal{L}_P(\hat{h}) - \min_{h \in \mathcal{H}} \mathcal{L}_P(h) \leq \mathcal{O}\left(\frac{1}{mn^{\frac{1}{2}-\eta}} + \frac{1}{m^{\frac{1}{2}-\eta}}\right)$$

The probability depends on η, m, n

[2] Figure adapted from https://adamwierman.com/wp-content/uploads/2021/05/book-05-11.pdf

THANK YOU!