Research Note

Xiaoyang Song

WN 2022

1 Generalized Latent Factor Model

1.1 Notations

Suppose that we have N people and J items, and each item has C possible categories. We define the following notations:

1. $Y_{ij} \in [0, C-1]$: the response of the ith person on item j, $i \in [0, N-1]$, $j \in [0, J-1]$.
2. p: number of latent factors
3. $\mathbf{x}_i = (\theta_{i1}, ..., \theta_{ip})^\top \in \mathbb{R}^{p\times 1}$: the factor scores for the ith person.
4. $\mathbf{X} \in \mathbb{R}^{N\times p}$: the factor score matrix.
5. $\boldsymbol{\beta}_j = (\beta_{j1}, ..., \beta_{jp})^\top \in \mathbb{R}^{K\times 1}$: the loading vector for item j.
6. $\boldsymbol{\beta} \in \mathbb{R}^{J\times p}$: the loading matrix.

1.2 Model

The generalized latent factor model assumes that Y_{ij} given $\mathbf{x}_i$ and $\boldsymbol{\beta}_j$ is a member of the exponential family with density function:

$$f(Y_{ij} = y \mid \boldsymbol{\beta}_j, \mathbf{x}_i) = \exp\left(\frac{ym_{ij} - b(m_{ij})}{\phi} + c(y, \phi)\right)$$

where $b(.)$ and $c(.)$ are two functions that decide which distribution in the exponential family the model distribution belongs to, ϕ is the scale parameter, and $m_{ij} = \boldsymbol{\beta}_j^\top \mathbf{x}_i$. Correspondingly, the joint likelihood is defined as the following:

$$L(\mathbf{x}_1, ..., \mathbf{x}_N, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_J) = \prod_{i=1}^{N}\prod_{j=1}^{J} f(Y_{ij} = y_{ij} \mid \mathbf{x}_j, \boldsymbol{\beta}_i)$$

2 Multidimensional Graded Model

2.1 Motivation

Motivation: we want to model the probability that the ith respondent chooses the category $k \in [0, C_j - 1]$ for item j using the information of the latent variables called *factors*, where C_j denotes the number of categories for item j and those categories are indexed by integer from 0 to $C_j - 1$ for simplicity.

2.2 Notations

Suppose that we have N people and J items, and each item has C possible categories. We define the following notations:

1. i: denotes the respondent i, $i \in [1, N]$.
2. j: denotes the item j, $j \in [1, J]$
3. C_j: the number of categories of item j.
4. p: number of latent factors
5. $\boldsymbol{\beta}_j$: a $p \times 1$ vector of factor coefficient for item j.
6. $\mathbf{X}$: a $N \times p$ matrix of factor scores of all respondents.
7. $\mathbf{x}_i$: a $p \times 1$ vector of factor scores for respondent i.
8. $\mathbf{Y}$: a $N \times J$ matrix of responses of all respondents.
9. $\mathbf{y}_i$: the response of respondent i, which is a $J \times 1$ vector.
10. y_{ij}: the response of respondent i on item j.
11. $\mathbf{d}_j$: a $(C_j - 1) \times 1$ vector of intercept term for item j.

2.3 MIRT Model

The MIRT model is defined by the following probabilities:

$$\mathbf{P}(y_{ij} \geq 0 \mid \boldsymbol{\beta}_j, \mathbf{d}_j, \mathbf{x}_i) = 1 \tag{1}$$

$$\mathbf{P}(y_{ij} \geq C_j \mid \boldsymbol{\beta}_j, \mathbf{d}_j, \mathbf{x}_i) = 0 \tag{2}$$

$$\mathbf{P}(y_{ij} \geq k \mid \boldsymbol{\beta}_j, \mathbf{d}_j, \mathbf{x}_i) = \frac{1}{1 + \exp\left(-\boldsymbol{\beta}_j^\top \mathbf{x}_i - d_{jk}\right)}, \quad \forall k \in [1, C_j - 1] \tag{3}$$

3 ReBoot Algorithm

Given the MIRT model, we can apply it to the distributed settings and formulate the `ReBoot` algorithm. Similar as the notation that is defined in the preceding section, suppose that we have N respondents and J items. We then distribute the data into m local machines and the estimated loadings and intercepts terms are denoted by $\widehat{\boldsymbol{\beta}}^{(k)}$ and $\widehat{\mathbf{d}}^{(k)}$ for $k \in [m]$, respectively. Let $\widetilde{n}$ denote the bootstrap sample size for each local estimator. The `ReBoot` algorithm can be formulated as the following.

Algorithm 1: `ReBoot` on IRT model

Input: $\{\widehat{\boldsymbol{\beta}}^{(k)}, \widehat{\mathbf{d}}^{(k)}\}_{k=1}^{m}$, $\widetilde{n}$

1: **for** $k = 1, \ldots, m$ **do**

2:   **for** $i = 1, \ldots, \widetilde{n}$ **do**

3:   Draw a Bootstrap feature vector $\widetilde{\mathbf{x}}_i^{(k)}$ from the distribution $f_{\mathbf{x}}(\cdot)$;

4:   Draw a Bootstrap response $\widetilde{Y}_i^{(k)}$ according to $f_{Y|\mathbf{x}}\big(\cdot\,|\widetilde{\mathbf{x}}_i^{(k)}; \widehat{\boldsymbol{\beta}}^{(k)}, \widehat{\mathbf{d}}^{(k)}\big)$;

5:   **end**

6: $\widetilde{\mathcal{D}}^{(k)} \leftarrow \big\{\widetilde{Y}_i^{(k)}\big\}_{i \in [\widetilde{n}]}$;

7: **end**

8: $\widetilde{\mathcal{D}} \leftarrow \cup_{k=1}^{m} \widetilde{\mathcal{D}}^{(k)}$;

9: $\widehat{\boldsymbol{\beta}}^{\mathrm{rb}}, \widehat{\mathbf{d}}^{\mathrm{rb}} \leftarrow \underset{\boldsymbol{\beta}, \mathbf{d} \in \mathcal{B}}{\operatorname{argmin}}\ \ell_{\widetilde{\mathcal{D}}}(\boldsymbol{\beta}, \mathbf{d})$.

Output: $\widehat{\boldsymbol{\beta}}^{\mathrm{rb}}, \widehat{\mathbf{d}}^{\mathrm{rb}}$

4 Rasch Model

Rasch model is a latent factor model for dichotomous data. Suppose that we have N observations and each observation has J item responses. Given the capability of ith person X_i and the difficulty parameter of jth item d_j^*, the probability of a correct response is

$$\mathbb{P}(Y_{ij} = 1 | X_i, d_j^*) = \frac{1}{1 + e^{-(X_i - d_j^*)}}. \tag{4}$$

We are interested in estimating the difficulty parameter for each item, i.e., $\mathbf{d}^* := (d_j^*)_{j \in [J]}$. One typically uses the maximum likelihood estimation(MLE) $\widehat{\mathbf{d}}$ to estimate $\mathbf{d}^*$. Let $\mathbf{Y} := (Y_{ij})_{i \in [N], j \in [J]}$ denote the response matrix and $\mathbf{x} := (X_i)_{i \in [N]}$ denote the capability of N individuals. The negative log-likelihood function is given by

$$\begin{aligned}
\ell(\mathbf{d}) &= -\frac{1}{N}\sum_{i=1}^{N}\log\left\{\int_{-\infty}^{\infty}\prod_{j=1}^{J}\frac{e^{Y_{ij}(x_i-d_j)}}{1+e^{x_i-d_j}}\phi(x_i)dx_i\right\} \\
&\leq -\frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\infty}\log\left\{\prod_{j=1}^{J}\frac{e^{Y_{ij}(X_i-d_j)}}{1+e^{X_i-d_j}}\right\}\phi(X_i)dX_i \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{J}\int_{-\infty}^{\infty}\big\{-Y_{ij}(X_i-d_j)+\log(1+e^{X_i-d_j})\big\}\phi(X_i)dX_i.
\end{aligned}$$

Define the surrogate loss function $\tilde{\ell}(\mathbf{d})$ as

$$\tilde{\ell}(\mathbf{d}) := \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{J}\mathbb{E}_{X_i}\big\{-Y_{ij}(X_i-d_j)+\log(1+e^{X_i-d_j})\big\}.$$

Taking derivative with respect to d_j, we have

$$\begin{aligned}
\nabla\tilde{\ell}_j(d_j) &= \frac{1}{N}\sum_{i=1}^{N}\left\{Y_{ij}-\mathbb{E}_{X_i}\left(\frac{e^{X_i-d_j}}{1+e^{X_i-d_j}}\right)\right\},\ \forall j\in[p] \\
\nabla^2\tilde{\ell}_{jj}(d_j) &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{X_i}\left\{\frac{e^{X_i-d_j}}{(1+e^{X_i-d_j})^2}\right\} \text{ and } \nabla^2\tilde{\ell}_{jl}=0,\ \forall j\neq l, j,l\in[p].
\end{aligned}$$

Our surrogate estimator is defined as $\widehat{\mathbf{d}} := \operatorname{argmin}_{\mathbf{d}}\tilde{\ell}(\mathbf{d})$. By choosing $\tau(\omega)=(3+e^{\omega})^{-1}$, we bound the magnitude of $\mathbf{d}^*-\widehat{\mathbf{d}}$, i.e., $\|\mathbf{d}^*-\widehat{\mathbf{d}}\|_2$ in the following Theorem.

Theorem 4.1. *Let $\kappa := \tau(\sqrt{\log 4}+\|\mathbf{d}^*\|_\infty+\frac{1}{2})/4$. Suppose we have N independent observations $\{\mathbf{y}_i\}_{i\in[N]}$ from the Rasch model* (4) *with $N>4J/\kappa^2$. Then for any $\xi>0$, we have with probability at least $1-2Je^{-\xi}$ that*

$$\|\mathbf{d}^*-\widehat{\mathbf{d}}\|_2 \leq \frac{1}{\kappa}\left(\frac{J\xi}{N}\right)^{1/2}.$$

5 Numerical Study

5.1 Experiments Setup

On the Rasch model, the only parameters of interests are the intercepts terms (i.e. difficulty parameters). We conduct the simulation studies under two regimes:

- Regime I: Fixed global sample size (N).

- Regime II: Fixed local sample size (n).

The dimensionality of the problem, which is controlled by the number of items (i.e. J) also varies among experiments. For the Rasch model, the ground truth parameters are drawn from uniform distribution range from the interval $[-0.2, 0.2]$. The slopes are fixed to be 1 and will not be freely estimated. The local estimator, average estimator, and the `ReBoot` estimator are computed in the experiment.

Due to the concerns about the identifiability problem (i.e. if X_i and d_j shift by the same constant simultaneously, the resulting model will be the same), we fix the intercept of the first item to be the ground truth value during the estimation procedure.

Edge Case Handling: one another potential problem is that there is a possibility that all responses for an item are all from the same category. To resolve this problem, which might occur in reality, we inspect the data before applying the `MH-RM` algorithm and only freely estimate the parameters for those whose responses are valid. For the rest of the items, we simply set its parameters to the LOWER_BOUND or UPPER_BOUND, which we specified beforehand, based on whether its response is 0 or 1, respectively.

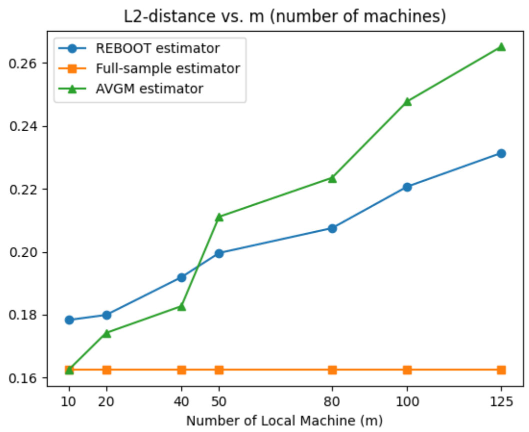

Figure 1: Regime I: $N = 4000$, $J = 20$

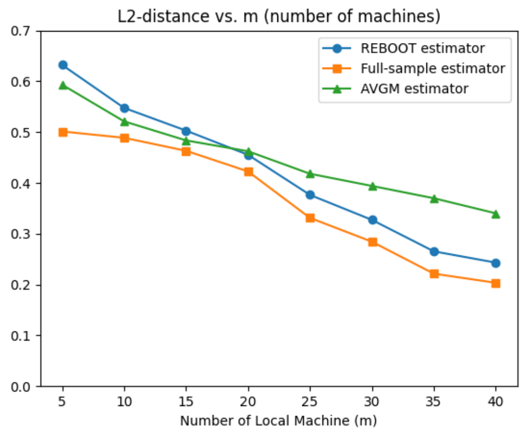

Figure 2: Regime II: $n = 75$, $J = 50$

5.2 Experimental Results

As mentioned in preceding section, we conduct experiments under two regimes. For the first regime, we fixed the global sample size to be $N = 4000$ and $J = 20$ and gradually increases the number of local machines; for the other regime, the local sample size is fixed to be $n = 75$ and $J = 50$. Note that the local sample size is chosen by empirical observation that $n = 75$ enables MH-RM algorithm to converge and give reasonably well estimations on each local machine.

From the simulation results we can observe that the `ReBoot` estimator gradually has its advantages exhibited as the number of machines increases (i.e. the local sample size decreases), which is consistent to the pattern that is observed in simulation study for MIRT model.

5.3 Discussions

- Compared to previous setting with $N = 1000$ and $J = 10$, where the gap between AVGM and ReBoot estimator is in 0.01 magnitude, the new settings in Regime II with more local machines and higher dimension does enlarge the gap to magnitude of 0.1. This improvement is expected.
- Note that for regime I, multiple global sample sizes are tried ($N = 1000, 2500, 4000$) but all giving us the same patterns as expected. For the second regime, the minimum local sample size $n = 75$ is chosen by several trials of experiments and manually inspections.
- Side note: currently working on implementing a complete pipeline that enables people to configure and run `ReBoot` for different IRT models from command line effectively.

6 Proof of Theorem 4.1.

To evaluate local strong convexity of any differential map $\tilde{\ell} : \mathbb{R}^p \to \mathbb{R}$, we define the first-order Taylor remainder of $\tilde{\ell}(\boldsymbol{\beta})$ at $\boldsymbol{\beta}^*$ to be

$$\delta\tilde{\ell}(\boldsymbol{\beta}; \boldsymbol{\beta}^*) := \tilde{\ell}(\boldsymbol{\beta}) - \tilde{\ell}(\boldsymbol{\beta}^*) - \nabla\tilde{\ell}(\boldsymbol{\beta}^*)^\top(\boldsymbol{\beta} - \boldsymbol{\beta}^*).$$

Construct an intermediate estimator $\widehat{\mathbf{d}}_\eta$ between $\widehat{\mathbf{d}}$ and $\mathbf{d}^*$:

$$\widehat{\mathbf{d}}_\eta := \mathbf{d}^* + \eta(\widehat{\mathbf{d}} - \mathbf{d}^*),$$

where $\eta = 1$ if $\|\widehat{\mathbf{d}} - \mathbf{d}^*\|_2 \leq \frac{1}{2}$ and $\eta = 1/(2\|\widehat{\mathbf{d}} - \mathbf{d}^*\|_2)$ if $\|\widehat{\mathbf{d}} - \mathbf{d}^*\|_2 > \frac{1}{2}$. Write $\widehat{\mathbf{d}}_\eta - \mathbf{d}^*$ as $\boldsymbol{\Delta}_\eta$ and $\widehat{\mathbf{d}} - \mathbf{d}^*$ as $\boldsymbol{\Delta}$. By Lemma 7.1, we have

$$\kappa\|\boldsymbol{\Delta}_\eta\|_2^2 \leq \delta\tilde{\ell}(\mathbf{d}; \mathbf{d}^*) \leq -\nabla\ell(\mathbf{d}^*)^\top\boldsymbol{\Delta}_\eta \leq \|\nabla\ell(\mathbf{d}^*)\|_2\|\boldsymbol{\Delta}_\eta\|_2,$$

where $\kappa = \tau(\sqrt{\log 4} + \|\mathbf{d}^*\|_\infty + \frac{1}{2})/4$. This implies that $\|\boldsymbol{\Delta}_\eta\|_2 \leq \|\nabla\ell(\mathbf{d}^*)\|_2/\kappa$. Then we focus on $\|\nabla\ell(\mathbf{d}^*)\|_2$. Let p_{ij} denote the $\mathbb{P}(Y_{ij} = 1|X_i, d_j^*)$. Note that $\mathbb{E}(Y_{ij} - p_{ij}) = 0$. By Hoeffding's inequality, we have with probability at least $1 - 2e^{-\xi}$ that

$$\Big|\frac{1}{N}\sum_{i=1}^{N}(Y_{ij} - p_{ij})\Big| \lesssim \left(\frac{\xi}{N}\right)^{1/2}.$$

It follows that with probability at least $1 - 2e^{-\xi}$ that

$$|\nabla \tilde{\ell}_j(d_j^*)| \leq \mathbb{E}_{\mathbf{x}} \Big| \frac{1}{N} \sum_{i=1}^{N} (Y_{ij} - p_{ij}) \Big| \lesssim \left(\frac{\xi}{N} \right)^{1/2}.$$

Therefore, we have with probability at least $1 - 2Je^{-\xi}$ that

$$\|\mathbf{\Delta}_\eta\|_2 \leq \frac{\|\nabla \ell(\mathbf{d}^*)\|_2}{\kappa J} \leq \frac{1}{\kappa} \left(\frac{J\xi}{N} \right)^{1/2}.$$

Since $\frac{1}{\kappa}\left(\frac{J}{N}\right)^{1/2} < \frac{1}{2}$, we find that $\mathbf{\Delta} = \mathbf{\Delta}_\eta$ according to the construction of $\mathbf{\Delta}_\eta$. The conclusion thus follows.

7 Lemmas

Condition 1. *Suppose the feature vector* $\mathbf{x}$ *is sub-Gaussian with* $\mathbb{E}\mathbf{x} = \mathbf{0}$ *and* $\|\mathbf{x}\|_{\psi_2} \leq K$ *with* $K \geq 1$*.*

Condition 2. $\forall \eta \in \mathbb{R}, 0 < b''(\eta) \leq \infty$*, there exists* $\tau : \mathbb{R}_+ \to \mathbb{R}_+$ *such that for any* $\omega > 0$*,* $b''(\eta) \geq \tau(\omega) > 0$ *whenever* $|\eta| \leq \omega$*.*

Lemma 7.1. *Recall the surrogate loss function*

$$\tilde{\ell}(\mathbf{d}) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{J} \mathbb{E}_{\mathbf{x}_i} \{ -Y_{ij}(X_i - d_j) + b(X_i - d_j) \}.$$

Under Conditions 1 and 2, for any $0 < r < 1$ *and* $t > 0$*, for any* $\mathbf{d} \in \mathbb{R}^p$ *such that* $\|\mathbf{d} - \mathbf{d}^*\|_2 \leq r$*, we have*

$$\delta \tilde{\ell}(\mathbf{d}; \mathbf{d}^*) \geq \kappa \|\mathbf{d} - \mathbf{d}^*\|_2^2, \tag{5}$$

where $\kappa = K\sqrt{\log 4} + \|\mathbf{d}^*\|_\infty + r$*.*

Proof. Given any $\mathbf{\Delta} \in \mathcal{B}(\mathbf{0}, r)$, by the Taylor expansion, we can find $v \in (0, 1)$ such that for any $\omega > 0$

$$\begin{aligned}
\delta \tilde{\ell}(\mathbf{d}^* + \mathbf{\Delta}; \mathbf{d}^*) = \tilde{\ell}(\mathbf{d}^* + \mathbf{\Delta}) - \tilde{\ell}(\mathbf{d}^*) - \nabla \tilde{\ell}(\mathbf{d}^*)^\top \mathbf{\Delta} &= \frac{1}{2} \mathbf{\Delta}^\top \nabla^2 \tilde{\ell}(\mathbf{d}^* + v\mathbf{\Delta}) \mathbf{\Delta} \\
&= \frac{1}{2N} \sum_{i=1}^{N} \mathbb{E}_{X_i} \left\{ \sum_{j=1}^{J} b''(X_i - d_j) \Delta_j^2 \right\} \\
&\geq \frac{1}{2N} \sum_{i=1}^{N} \mathbb{E}_{X_i} \left[\sum_{j=1}^{J} b''(X_i - d_j) \Delta_j^2 \mathbb{1}_{\{\max(|X_i + \|\mathbf{d}\|_\infty|, |X_i - \|\mathbf{d}\|_\infty|) \leq \omega\}} \right] \\
&\geq \frac{\tau(\omega)}{2} \|\mathbf{\Delta}\|_2^2 \, \mathbb{P}\{\max(|X_i + \|\mathbf{d}\|_\infty|, |X_i - \|\mathbf{d}\|_\infty|) \leq \omega\}.
\end{aligned}$$

By Proposition 2.5.2 in [1], we have

$$\mathbb{P}\{\max(|X_i + \|\mathbf{d}\|_\infty|, |X_i - \|\mathbf{d}\|_\infty|) \le \omega\} \le 2\exp\left\{ - \frac{\min(\omega + \|\mathbf{d}\|_\infty, \omega - \|\mathbf{d}\|_\infty)^2}{K^2} \right\}.$$

Choose $\omega = K\sqrt{\log 4} + \|\mathbf{d}^*\|_\infty + r$. We can deduce that for all $\boldsymbol{\Delta} \in \mathbb{R}^p$ such that $\|\boldsymbol{\Delta}\|_2 \le r$,

$$\delta\tilde{\ell}(\mathbf{d}^* + \boldsymbol{\Delta}; \mathbf{d}^*) \ge \frac{\tau(K\sqrt{\log 4} + \|\mathbf{d}^*\|_\infty + r)}{4}\|\boldsymbol{\Delta}\|_2^2.$$

□

A DistributedPCA Algorithm

For MIRT model, to aggregate the loading estimators from m local machines, we can also adopt the `DistributedPCA` algorithm by slightly modifying it. The `DistributedPCA` algorithm for MIRT model is formulated below, where $\widehat{\boldsymbol{\beta}}^{(k)}$ is the local estimator for loadings of the kth machine.

Algorithm 2: DistributedPCA

Input: $\{\widehat{\boldsymbol{\beta}}^{(k)}\}_{k=1}^m$

1: **for** $k = 1, \ldots, m$ **do**

2: Compute $\widehat{\boldsymbol{\Sigma}}^{(k)} \leftarrow \widehat{\boldsymbol{\beta}}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}$;

3: Compute the eigenvectors $\widehat{\mathbf{V}}^{(k)}$ of $\widehat{\boldsymbol{\Sigma}}^{(k)}$ and send it to the central server.

4: **end**

5: On the central server, compute $\widetilde{\boldsymbol{\Sigma}} \leftarrow \frac{1}{m}\sum_{k=1}^m \widehat{\mathbf{V}}^{(k)}\widehat{\mathbf{V}}^{(k)\top}$ and its eigenvector $\widetilde{\mathbf{V}}$;

6: **for** $k = 1, \ldots, m$ **do**

7: Compute the eigenvalues $\widehat{\boldsymbol{\Lambda}}^{(k)} \leftarrow \mathrm{diag}(\widetilde{\mathbf{V}}^\top \widehat{\boldsymbol{\Sigma}}^{(k)} \widetilde{\mathbf{V}})$ locally and send it to the central server;

8: **end**

9: On the central server, compute $\widetilde{\boldsymbol{\Lambda}} \leftarrow \frac{1}{m}\Sigma_{k=1}^m \widehat{\boldsymbol{\Lambda}}^{(k)}$;

10: $\widehat{\boldsymbol{\beta}}^{\mathrm{dpca}} \leftarrow \widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Lambda}}^{\frac{1}{2}}$.

Output: $\widehat{\boldsymbol{\beta}}^{\mathrm{dpca}}$

B MIRT model simulation

B.1 Experiments Setup

The simulation study is conducted on the MIRT model that is introduced in the preceding section, where the ground truth parameters for slopes (i.e. loadings) and intercepts (i.e. difficulty parameters) are borrowed from Cai's (2010) paper. Specifically, two regimes of experiments are conducted:

- Fixed global sample size ($N = 1000$ and $N = 2500$) and gradually increase the number of machines.
- Fixed local sample size ($n = 125$) and gradually increase the number of machines.

Under both regimes, the local estimator, average estimator (i.e. the arithmetic mean of the local estimators), `ReBoot` Estimator and full-sample estimator are computed and evaluated.

Due to the concerns about the identifiability problem of the estimated loading matrix, the Positive-Diagonal-Lower-Triangular (PLT) constraint is enforced on the parameter matrix before the estimation.

B.2 Experimental Results

As for statistical error metrics, the SinTheta distance is used to measure the error for the loading matrix, while the Frobenius norm is used for intercepts. The experimental results are provided below:

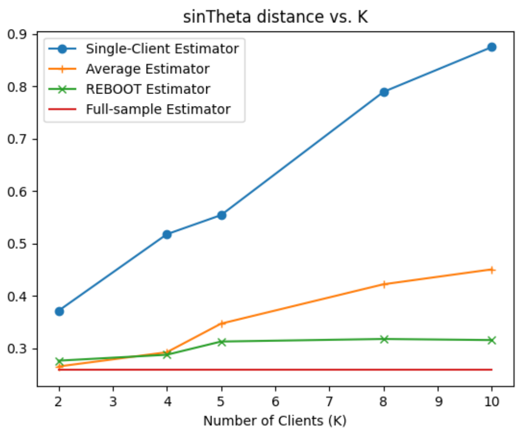

Figure 3: $N = 1000$

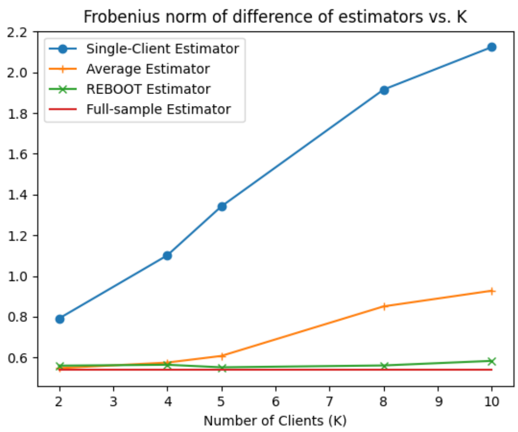

Figure 4: $N = 1000$

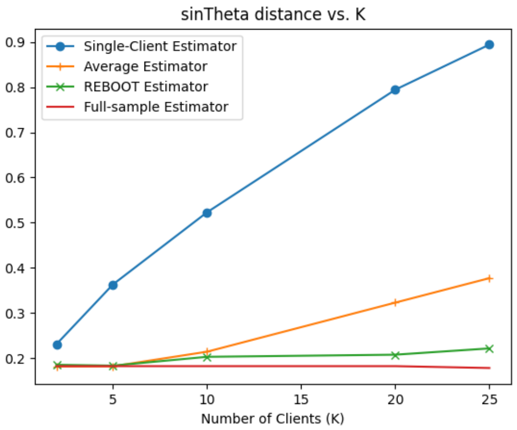

Figure 5: $N = 2500$

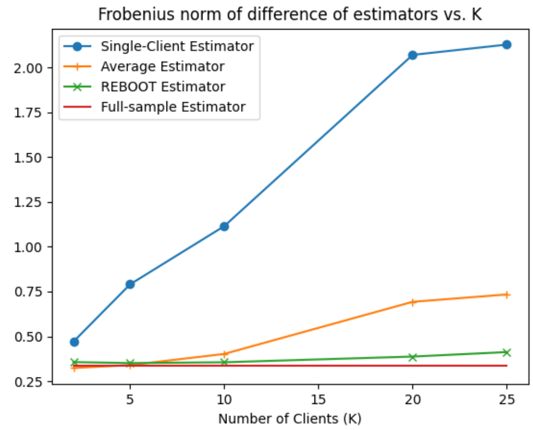

Figure 6: $N = 2500$

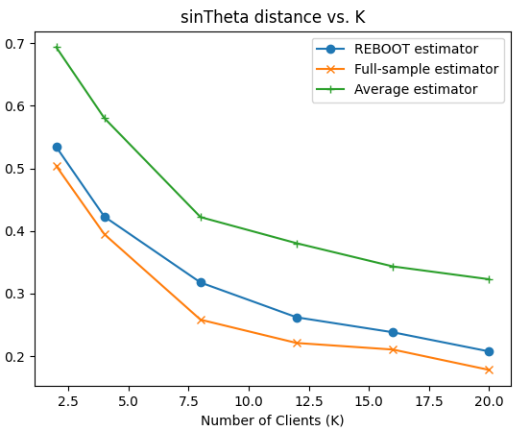

Figure 7: $n = 125$

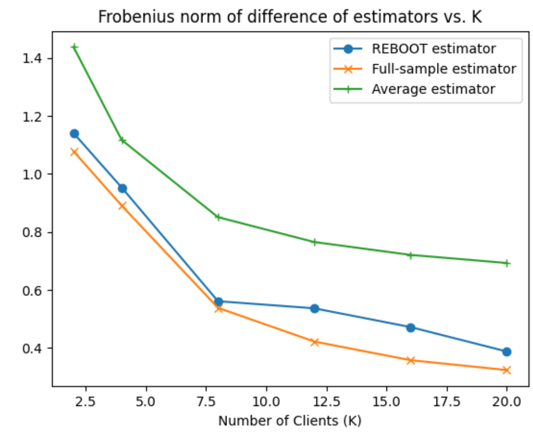

Figure 8: $n = 125$

B.3 Discussions

From the experimental results, the `ReBoot` estimator tends to give the best estimation for both loadings and intercepts for the most times. When the number of machines is small, implying that the local sample size is sufficient, the average estimator tends to have similar statistical error as the `ReBoot` estimator. For all cases, full-sample estimator gives the best estimator, which is consistent with the theoretical expectation, and the local estimator has the worst performance.

References

[1] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

Experiment setup: fixed global sample size: $N = 1000$

K	2	4	5	8	10
Single client estimator	0.3715	0.5177	0.5542	0.7896	0.8752
Average estimator	0.2651	0.2923	0.3469	0.4221	0.4506
`REBOOT` estimator	0.2762	0.2875	0.3127	0.3176	0.3155
`MIRT` estimator	0.2582				

Table 1: sinTheta distance for estimators of slopes

Experiment setup: fixed global sample size: $N = 1000$

K	2	4	5	8	10
Single client estimator	0.7915	1.1012	1.3416	1.9153	2.1240
Average estimator	0.5475	0.5751	0.6078	0.8510	0.9279
`REBOOT` estimator	0.5599	0.5644	0.5518	0.5610	0.5835
`MIRT` estimator	0.5383				

Table 2: Frobenius norm of difference of intercept estimators

Experiment setup: fixed global sample size: $N = 2500$

K	2	5	10	20	25
Single client estimator	0.2307	0.3630	0.5228	0.7943	0.8942
Average estimator	0.1816	0.1818	0.2141	0.3228	0.3770
`REBOOT` estimator	0.1851	0.1832	0.2028	0.2075	0.2214
`MIRT` estimator	0.1780				

Table 3: sinTheta distance for estimators of slopes

Experiment setup: fixed global sample size: $N = 2500$

K	2	5	10	20	25
Single client estimator	0.4710	0.7902	1.1151	2.0704	2.1280
Average estimator	0.3235	0.3398	0.4021	0.6928	0.7339
`REBOOT` estimator	0.3565	0.3508	0.3554	0.3874	0.4124
`MIRT` estimator	0.3233				

Table 4: Frobenius norm of difference of intercept estimators

Experiment setup: fixed local sample size: $n = 125$

K	2	4	8	12	16	20
Average estimator	0.6935	0.5803	0.4221	0.3803	0.3433	0.3228
`REBOOT` estimator	0.5343	0.4229	0.3176	0.2620	0.2381	0.2075
`MIRT` estimator	0.5038	0.3948	0.2582	0.2210	0.2105	0.1780

Table 5: sinTheta distance for estimators of slopes

Experiment setup: fixed local sample size: $n = 125$

K	2	4	8	12	16	20
Average estimator	1.4382	1.1175	0.8510	0.7652	0.7208	0.6928
`REBOOT` estimator	1.1412	0.9534	0.5610	0.5365	0.4717	0.3874
`MIRT` estimator	1.0784	0.8910	0.5383	0.4218	0.3575	0.3233

Table 6: Frobenius norm of difference of intercept estimators