Supplementary Material for "A Tensor-EM Method for Large-Scale Latent Class Analysis with Binary Responses"

In this supplementary material, the proof of Theorem 3 (clustering consistency) is presented in Appendix 1. Then Corollary 2 (consistency of item parameters) is proved in Appendix 2. In Appendix 3, more simulation results are provided to show the good performance of the proposed EM-tensor method.

Appendix 1: Proof of Theorem 3

Outline of proof idea. The proof follows the following 8 steps.

Step 1: Express $\ell(\mathbf{R};\mathbf{Z}) - \bar{\ell}(\mathbf{Z})$ in terms of $\sum_a n_a \sum_j D(\widehat{\theta}_{j,a}\|\bar{\theta}_{j,a}) + X - \mathbb{E}(X)$, where X is a random variable depending on $\mathbf{R}$ and $\bar{\boldsymbol{\theta}}^{(\mathbf{Z})}$ under $\mathbf{Z}$, and $n_a = \sum_{i=1}^N Z_{i,a}$.

Step 2: Bound the first term $\sum_j \sum_a n_{j,a} D(\widehat{\theta}_{j,a}\|\bar{\theta}_{j,a})$ in the above display uniformly over all possible $\mathbf{Z}$.

Step 3: Bound the second term $X - \mathbb{E}(X)$ using Bernstein type inequality. Combine this and Step 2 to obtain a bound for $\sup_{\mathbf{Z}} |\ell(\mathbf{R};\mathbf{Z}) - \bar{\ell}(\mathbf{Z})|$.

Step 4: (Denote the true latent class memberships by $\mathbf{Z}^0$ and joint MLE by $\widehat{\mathbf{Z}}$.) Establish $\bar{\ell}(\mathbf{Z}^0) \geq \bar{\ell}(\mathbf{Z})$ for all $\mathbf{Z}$. Use triangle inequality to upper-bound the non-negative quantity $\bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\widehat{\mathbf{Z}})$.

$$0 \leq \bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\widehat{\mathbf{Z}}) \leq [\bar{\ell}(\mathbf{Z}^0) - \ell(\mathbf{R};\mathbf{Z}^0)] + [\ell(\mathbf{R};\mathbf{Z}^0) - \ell(\mathbf{R};\widehat{\mathbf{Z}})] + [\ell(\mathbf{R};\widehat{\mathbf{Z}}) - \bar{\ell}(\widehat{\mathbf{Z}})]$$

Since in the above display the middle group of terms $[\ell(R;\, \mathbf{Z}^0) - \ell(R;\, \widehat{\mathbf{Z}})] \leq 0$, we have $0 \leq \bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\widehat{\mathbf{Z}}) \leq 2\sup_{\mathbf{Z}} |\ell(\mathbf{R};\, \mathbf{Z}) - \bar{\ell}(\mathbf{Z})|$.

Step 5: Introduce the notion of partitions and generalize $\bar{\ell}(\mathbf{Z})$ to $\bar{\ell}(\Pi)$.

Step 6: Show that a refined partition increases $\bar{\ell}(\cdot)$. To be concrete, let Π^* be a refined partition of Π, then we have $\bar{\ell}(\Pi^*) \geq \bar{\ell}(\Pi)$.

Step 7: Show that for any latent class assignment $\mathbf{Z}$, we can find a partition Π^* that refines $\Pi^{\mathbf{Z}}$ and $\bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\Pi^*) \geq \frac{1}{2} J\beta_J N_e(\mathbf{z})$.

Step 8: Apply results in step 6 and step 7 to MLE $\widehat{\mathbf{z}}$, we have

$$\text{bound in step } 4 \geq \bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\Pi^{\widehat{\mathbf{Z}}}) \geq \bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\Pi^*) \geq \frac{1}{2} J\beta_J N_e(\widehat{\mathbf{z}}).$$

Now we formally begin the proof of Theorem 3 in the above several steps. In derivations below we abbreviate $\bar{\boldsymbol{\theta}}^{(\mathbf{Z})}$ as $\bar{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}^{(\mathbf{Z})}$ as $\widehat{\boldsymbol{\theta}}$ to simplify notations.

Step 1. Define $D(p\|q) = p\log(p/q) + (1-p)\log((1-p)/(1-q))$, the Kullback-Leibler divergence of a Bernoulli distribution with parameter p from that with parameter q. In this step we prove a lemma as follows.

Lemma 1. *Let $(R_{i,j};\, 1 \leq N, 1 \leq J)$ denote independent Bernoulli trials with parameters $(P_{i,j};\, 1 \leq N, 1 \leq J)$. Under a general latent class model, given an arbitrary $\mathbf{Z}$, there is*

$$\begin{aligned} &\sup_{\boldsymbol{\theta}} \ell(\mathbf{R};\, \mathbf{Z}, \boldsymbol{\theta}) - \sup_{\boldsymbol{\theta}} \mathbb{E}[\ell(\mathbf{R};\, \mathbf{Z}, \boldsymbol{\theta})] \\ =& \sum_{a=1}^{L} n_a \sum_j D(\widehat{\theta}_{j,a}\|\bar{\theta}_{j,a}) + \sum_i \sum_j (R_{i,j} - P_{i,j}) \log\left(\frac{\bar{\theta}_{j,z_i}}{1-\bar{\theta}_{j,z_i}}\right) \\ =& \sum_{a=1}^{L} n_a \sum_j D(\widehat{\theta}_{j,a}\|\bar{\theta}_{j,a}) + X - \mathbb{E}X, \end{aligned} \tag{1}$$

where $X = \sum_i \sum_j R_{i,j} \log\left(\frac{\bar{\theta}_{j,z_i}}{1-\bar{\theta}_{j,z_i}}\right)$ is random variable depending on $\mathbf{Z}$ and

$$\widehat{\theta}_{j,a} = \frac{\sum_i Z_{i,a} R_{i,j}}{\sum_i Z_{i,a}}, \quad \bar{\theta}_{j,a} = \frac{\sum_i Z_{i,a} P_{i,j}}{\sum_i Z_{i,a}} \tag{2}$$

Given a fixed $\mathbf{Z}$, denote $n_a^{(\mathbf{Z})} = \sum_{i=1}^N Z_{i,a}$.The maximizing properties of $\widehat{\theta}_{j,a}$ and $\bar{\theta}_{j,a}$ in 2 imply that

$$n_a \widehat{\theta}_{j,a} = \sum_i Z_{i,a} R_{i,j}, \quad n_a \bar{\theta}_{j,a} = \sum_i Z_{i,a} P_{i,j}. \tag{3}$$

Using (3), we have the following,

$$\begin{aligned}
& \ell(\mathbf{R}; \mathbf{Z}) - \bar{\ell}(\mathbf{Z}) \\
= & \sum_i \sum_j \sum_{a=1}^L Z_{i,a}[R_{i,j} \log \widehat{\theta}_{j,a} + (1 - R_{i,j}) \log(1 - \widehat{\theta}_{j,a})] \\
& \qquad - \sum_i \sum_j \sum_{a=1}^L Z_{i,a}[P_{i,j} \log \bar{\theta}_{j,a} + (1 - P_{i,j}) \log(1 - \bar{\theta}_{j,a})] \\
= & \sum_j \sum_{a=1}^L n_a[\widehat{\theta}_{j,a} \log \widehat{\theta}_{j,a} + (1 - \widehat{\theta}_{j,a}) \log(1 - \widehat{\theta}_{j,a})] - \sum_j \sum_{a=1}^L n_a[\bar{\theta}_{j,a} \log \bar{\theta}_{j,a} + (1 - \bar{\theta}_{j,a}) \log(1 - \bar{\theta}_{j,a})] \\
= & \sum_j \sum_{a=1}^L \Big\{ n_a[\widehat{\theta}_{j,a} \log \widehat{\theta}_{j,a} + (1 - \widehat{\theta}_{j,a}) \log(1 - \widehat{\theta}_{j,a})] - n_a[\widehat{\theta}_{j,a} \log \bar{\theta}_{j,a} + (1 - \widehat{\theta}_{j,a}) \log(1 - \bar{\theta}_{j,a})] \Big\} \\
& + \sum_j \sum_{a=1}^L \Big\{ n_a[\widehat{\theta}_{j,a} \log \bar{\theta}_{j,a} + (1 - \widehat{\theta}_{j,a}) \log(1 - \bar{\theta}_{j,a})] - n_a[\bar{\theta}_{j,a} \log \bar{\theta}_{j,a} + (1 - \bar{\theta}_{j,a}) \log(1 - \bar{\theta}_{j,a})] \Big\} \\
= & \sum_{a=1}^L n_a \sum_j D(\widehat{\theta}_{j,a} \| \bar{\theta}_{j,a}) + \sum_i \sum_j \Big\{ [R_{i,j} \log \bar{\theta}_{j,z_i} + (1 - R_{i,j}) \log(1 - \bar{\theta}_{j,z_i})] \\
& \qquad\qquad - [P_{i,j} \log \bar{\theta}_{j,z_i} + (1 - P_{i,j}) \log(1 - \bar{\theta}_{j,z_i})] \Big\} \\
= & \sum_{a=1}^L n_a \sum_j D(\widehat{\theta}_{j,a} \| \bar{\theta}_{j,a}) + \sum_i \sum_j R_{i,j} \log \Big(\frac{\bar{\theta}_{j,z_i}}{1 - \bar{\theta}_{j,z_i}} \Big) - \sum_i \sum_j P_{i,j} \log \Big(\frac{\bar{\theta}_{j,z_i}}{1 - \bar{\theta}_{j,z_i}} \Big).
\end{aligned}$$

Define the random variable

$$X = \sum_i \sum_j R_{i,j} \log(\bar{\theta}_{j,z_i}/(1 - \bar{\theta}_{j,z_i})), \tag{4}$$

then X depends on $\mathbf{Z}$ and the above display becomes the summation of $\sum_{a=1}^L n_a \sum_j D(\widehat{\theta}_{j,a} \| \bar{\theta}_{j,a})$ and $X - \mathbb{E}[X]$. This establishes (1) in Lemma 1. In the following, we bound the first term $\sum_{a=1}^L n_a \sum_j D(\widehat{\theta}_{j,a} \| \bar{\theta}_{j,a})$ and the second term $X - \mathbb{E}[X]$ in the above display uniformly over all possible $\mathbf{Z}$, respectively in Step 2 and Step 3.

Step 2. In this step we prove the following theorem.

Theorem 1. *The following event happens with probability at least* $1-\delta$,

$$\max_{\mathbf{Z}} \left\{ \sum_j \sum_a n_a D(\widehat{\theta}^{\mathbf{Z}}_{j,a} \| \bar{\theta}^{\mathbf{Z}}_{j,a}) \right\} < N \log L + JL \log\left(\frac{N}{L}+1\right) - \log \delta.$$

Given any fixed latent class memberships $\mathbf{Z}$, every $\widehat{\theta}_{j,a}$ is an average of n_a independent Bernoulli random variables $R_{1,j}, \ldots, R_{N,j}$ with mean $\bar{\theta}_{j,a}$. We apply the Chernoff-Hoeffding theorem to obtain

$$\mathbb{P}(\widehat{\theta}_{j,a} \geq \bar{\theta}_{j,a} + t) \leq e^{-n_a D(\bar{\theta}_{j,a}+t \| \bar{\theta}_{j,a})}, \quad \mathbb{P}(\widehat{\theta}_{j,a} \leq \bar{\theta}_{j,a} - t) \leq e^{-n_a D(\bar{\theta}_{j,a}-t \| \bar{\theta}_{j,a})}. \tag{5}$$

Note that given a fixed $\mathbf{Z}$, each $\widehat{\theta}_{j,a}$ can take values only in the finite set $\{0, 1/n_a, 2/n_a, \ldots, 1\}$ of cardinality $n_a + 1$. We denote this range of $\widehat{\theta}_{j,a}$ by $\widehat{\Theta}^{j,a}$. Then $\mathbb{P}(\widehat{\theta}_{j,a} = \vartheta) \leq \exp\{-n_a D(\vartheta \| \bar{\theta}_{j,a})\}$ for any $\vartheta \in \widehat{\Theta}^{j,a}$. Now consider the cardinality of the set $\widehat{\Theta}$ given $\mathbf{Z}$. Since for each of the $J \times L$ entries in $\widehat{\boldsymbol{\theta}}$, $\widehat{\theta}_{j,a}$ can independently take on $n_a + 1$ different values, there is $|\widehat{\Theta}| = [\prod_a (n_a+1)]^J$. Considering the natural constraint $\sum_{a=1}^L n_a = N$, we have

$$|\widehat{\Theta}| = \Big[\prod_{a=1}^{L} (n_a+1)\Big]^J \leq \Big[\Big(\frac{N}{L}+1\Big)^L\Big]^J. \tag{6}$$

Define $\widehat{\Theta}_\epsilon = \{\widetilde{\boldsymbol{\theta}} \in \widehat{\Theta} : \sum_j \sum_a n_a D(\widetilde{\theta}_{j,a} \| \bar{\theta}_{j,a}) \geq \epsilon\}$, then $\widehat{\Theta}_\epsilon \subseteq \widehat{\Theta}$. Note that the components of $\widehat{\boldsymbol{\theta}}$ depend on different components of $\{R_{i,j}, i \in [N], j \in [J]\}$ and thus are independent. We have

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{j=1}^{J}\sum_{a=1}^{L} n_a D(\widehat{\theta}_{j,a} \| \bar{\theta}_{j,a}) \geq \epsilon\Big) \\
= &\sum_{\widetilde{\boldsymbol{\theta}} \in \widehat{\Theta}_\epsilon} \mathbb{P}\Big(\widehat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}\Big) \\
\leq &\sum_{\widetilde{\boldsymbol{\theta}} \in \widehat{\Theta}_\epsilon} \prod_j \prod_a \exp\{-n_a D(\widetilde{\theta}_{j,a} \| \bar{\theta}_{j,a})\} \\
\leq &\sum_{\widetilde{\boldsymbol{\theta}} \in \widehat{\Theta}_\epsilon} \exp\{-\sum_j \sum_a n_a D(\widetilde{\theta}_{j,a} \| \bar{\theta}_{j,a})\} \\
\leq &\sum_{\widetilde{\boldsymbol{\theta}} \in \widehat{\Theta}_\epsilon} \exp\{-\epsilon\} \\
\leq &|\widehat{\Theta}_\epsilon| e^{-\epsilon} \leq |\widehat{\Theta}| e^{-\epsilon} \leq \Big(\frac{N}{L}+1\Big)^{JL} e^{-\epsilon}.
\end{aligned}$$

The above result holds for fixed $\mathbf{Z}$, so applying a union bound over all the L^N possible assignment $\mathbf{Z}$, there is

$$\mathbb{P}\Big(\max_{\mathbf{Z}}\left\{\sum_j\sum_a n_a D(\widehat{\theta}_{j,a}\|\bar{\theta}_{j,a})\right\}\geq\epsilon\Big)\leq L^N\Big(\frac{N}{L}+1\Big)^{JL}e^{-\epsilon}.$$

Now take $\delta = L^N\Big(\frac{N}{L}+1\Big)^{JL}e^{-\epsilon}$, then $\epsilon = N\log L + JL\log(\frac{N}{L}+1)-\log\delta$. Therefore the following event happens with probability at least $1-\delta$,

$$\max_{\mathbf{Z}}\left\{\sum_j\sum_a n_a D(\widehat{\theta}_{j,a}\|\bar{\theta}_{j,a})\right\}<\epsilon = N\log L + JL\log\left(\frac{N}{L}+1\right)-\log\delta.$$

This concludes the proof of Theorem 1.

Step 3. In this step we bound $|X-\mathbb{E}[X]|$, with X defined in (4). Denote $X_{i,j}=R_{i,j}\log(\bar{\theta}_{j,z_i}/(1-\bar{\theta}_{j,z_i}))$, then $X=\sum_i\sum_j X_{i,j}$. Under Assumption 1, we have $|X_{i,j}|\leq\gamma\log J$. Then we have $\sum_i\sum_j\mathbb{E}[X_{i,j}^2]\leq\sum_i\sum_j\mathbb{P}(R_{i,j}=1)\gamma^2(\log J)^2=\gamma^2\sum_i\sum_j P_{i,j}(\log J)^2=\gamma^2 MNJ(\log J)^2$. Applying the Bernstein's inequality to the sum of independent bounded random variables, we have the following holds for any fixed $\mathbf{Z}$,

$$\begin{aligned}\mathbb{P}(|X-\mathbb{E}[X]|\geq\epsilon) &\leq 2\exp\left\{-\frac{(1/2)\epsilon^2}{\sum_i\sum_j\mathbb{E}[X_{i,j}^2]+(1/3)\gamma\log(J)\epsilon}\right\}\\ &\leq 2\exp\left\{-\frac{(1/2)\epsilon^2}{\gamma^2 MNJ(\log J)^2+(1/3)\gamma\log(J)\epsilon}\right\}.\end{aligned}$$

We next prove the following theorem.

Theorem 2. *Under the following scaling (as $N,J\to\infty$),*

$$\frac{MJ}{\log L}\to\infty\,,\ \frac{N}{L}\to\infty, \tag{7}$$

$$\sqrt{\frac{M}{J}}\left(\frac{N}{L}\right)^{1-\xi}\to\infty \text{ for some small } \xi>0,$$

we have

$$\frac{1}{NJ}\max_{\mathbf{Z}}|\ell(\mathbf{R};\mathbf{Z})-\bar{\ell}(\mathbf{Z})| = o_P\left(\frac{\sqrt{M\log L}}{\sqrt{J}}(\log J)^{1+\eta}\right)$$

for any $\eta>0$.

We need to bound $|\ell(\mathbf{R};\, \mathbf{Z}) - \bar{\ell}(\mathbf{Z})|$ uniformly over all the $\mathbf{Z}$. Combining the results of Step 2 and Step 3, since there are L^N possible assignments of $\mathbf{Z}$, we apply the union bound to obtain

$$
\begin{aligned}
&\mathbb{P}(\max_{\mathbf{Z}} |\ell(\mathbf{R};\, \mathbf{Z}) - \bar{\ell}(\mathbf{Z})| \geq 2\epsilon\delta_{NJ}) \\
\leq\ & L^N \mathbb{P}\left[\left\{\sum_j \sum_a n_a D(\widehat{\theta}_{j,a} \| \bar{\theta}_{j,a}) \geq \epsilon\delta_{NJ}\right\} \cup \{|X - \mathbb{E}[X]| \geq \epsilon\delta_{NJ}\}\right] \\
\leq\ & \exp\left\{N \log L + JL \log\left(\frac{N}{L} + 1\right) - \epsilon\delta_{NJ}\right\} \\
&+ 2\exp\left\{N \log L - \frac{(1/2)\epsilon^2\delta_{NJ}^2}{\gamma^2 MNJ(\log J)^2 + (1/3)\gamma\log(J)\epsilon\delta_{NJ}}\right\}.
\end{aligned} \tag{8}
$$

In order for the term on the right hand side of the above display to go to zero, the following of δ_{NJ} would suffice,

$$\delta_{NJ} = N\sqrt{MJ\log L}(\log J)^{1+\eta}. \tag{9}$$

for a small positive constant η. Then the right hand side of (8) goes to zero as N, J go large and hence the scaling of J described in the theorem yields $\mathbb{P}(\max_{\mathbf{Z}} |\ell(\mathbf{R};\, \mathbf{Z}) - \bar{\ell}(\mathbf{Z})| \geq 2\epsilon\delta_{NJ}) = o(1)$, which implies

$$\frac{1}{NJ}\max_{\mathbf{Z}} |\ell(\mathbf{R};\, \mathbf{Z}) - \bar{\ell}(\mathbf{Z})| = o_P\left(\frac{\sqrt{M\log L}}{\sqrt{J}}(\log J)^{1+\eta}\right). \tag{10}$$

This proves Theorem 2.

Step 4. Denote the true class assignments by $\mathbf{Z}^0$. We first establish

$$\bar{\ell}(\mathbf{Z}^0) \geq \bar{\ell}(\mathbf{Z}), \quad \text{for all } \mathbf{Z}. \tag{11}$$

First note that $\theta^0_{j,z_i^0} = P_{i,j}$, and

$$\bar{\theta}_{j,z_i^0} = \frac{\sum_{m=1}^N Z^0_{m,z_i^0} P_{m,j}}{\sum_{m=1}^N Z^0_{m,z_i^0}} = \frac{\sum_{m=1}^N Z^0_{m,z_i^0} P_{i,j}}{\sum_{m=1}^N Z^0_{m,z_i^0}} = P_{i,j}.$$

The difference $\bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\mathbf{Z})$ can be written as

$$\bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\mathbf{Z}) = \sum_j \sum_i [P_{i,j}\log\left(\frac{\bar{\theta}^0_{j,z_i^0}}{\bar{\theta}^{\mathbf{Z}}_{j,z_i}}\right) + (1 - P_{i,j})\log\left(\frac{1 - \bar{\theta}^0_{j,z_i^0}}{1 - \bar{\theta}^{\mathbf{Z}}_{j,z_i}}\right)]$$

$$= \sum_j \sum_i [P_{i,j} \log\left(\frac{P_{i,j}}{\bar{\theta}^{\mathbf{Z}}_{j,z_i}}\right) + (1-P_{i,j}) \log\left(\frac{1-P_{i,j}}{1-\bar{\theta}^{\mathbf{Z}}_{j,z_i}}\right)] = \sum_i \sum_j D(P_{i,j} \| \bar{\theta}^{\mathbf{Z}}_{j,z_i}) \geq 0,$$

therefore establishing (11). Since the above holds for every $\mathbf{Z}$, it also holds for the maximum likelihood estimator $\widehat{\mathbf{Z}}$. We further upper bound $\bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\mathbf{Z})$ from above as follows,

$$0 \leq \bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\widehat{\mathbf{Z}}) \leq [\bar{\ell}(\mathbf{Z}^0) - \ell(\mathbf{R};\, \mathbf{Z}^0)] + \underbrace{[\ell(\mathbf{R};\, \mathbf{Z}^0) - \ell(\mathbf{R};\, \widehat{\mathbf{Z}})]}_{\leq 0} + [\ell(\mathbf{R};\, \widehat{\mathbf{Z}}) - \bar{\ell}(\widehat{\mathbf{Z}})],$$

where $[\ell(\mathbf{R};\, \mathbf{Z}^0) - \ell(\mathbf{R};\, \widehat{\mathbf{Z}})] \leq 0$ results from the definition of $\widehat{\mathbf{Z}}$ as the MLE, that is $\mathbf{Z}$ maximizes the $\ell(\mathbf{R};\, \mathbf{Z}, \widehat{\boldsymbol{\theta}}^{\mathbf{Z}})$. Therefore

$$\begin{aligned} 0 \leq \bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\widehat{\mathbf{Z}}) &\leq [\bar{\ell}(\mathbf{Z}^0) - \ell(\mathbf{R};\, \mathbf{Z}^0)] + [\ell(\mathbf{R};\, \widehat{\mathbf{Z}}) - \bar{\ell}(\widehat{\mathbf{Z}})] \\ &\leq 2 \sup_{\mathbf{Z}} |\bar{\ell}(\mathbf{Z}) - \ell(\mathbf{R};\, \mathbf{Z})| \\ &= o_p(\delta_{NJ}). \end{aligned}$$

Step 5. To establish the consistency of MLE in clustering subjects into latent classes, we need to introduce the notion of partitions. First we observe that any latent class assignment $\mathbf{Z}$ defines a partition on $[N]$ into T subsets $(S_1, \ldots, S_T)$ via mapping $\Pi^{\mathbf{Z}}$ from $[N]$ to $[T]$ such that for any subject we have $\theta^0_{j,z_i} = \theta^0_{j,\Pi^{\mathbf{Z}}_i}$ for all j. We now generalize this notion. For any partition on $[N]$, define

$$\bar{\theta}^{\Pi}_{j,a} = \frac{1}{|S_a|} \sum_{i=1}^{N} \theta^0_{j,z^0_i} I(i \in S_a) = \frac{1}{|S_a|} \sum_{i=1}^{N} P_{i,j} I(i \in S_a)$$

as the average over all i in the subset S_a indexed by $\Pi_i = a$. We then define generalization of $\bar{\ell}(\mathbf{Z})$ as

$$\bar{\ell}(\Pi) = \sum_i \sum_j [P_{i,j} \log(\bar{\theta}^{\Pi}_{j,\Pi_i}) + (1-P_{i,j}) \log(1-\bar{\theta}^{\Pi}_{j,\Pi_i})].$$

Note that $\bar{\theta}^{\Pi^{\mathbf{Z}}}_{j,a} = \bar{\theta}^{\mathbf{Z}}_{j,a}$ and hence $\bar{\ell}(\Pi^{\mathbf{Z}}) = \bar{\ell}(\mathbf{Z})$ when the partition $\Pi^{\mathbf{Z}}$ is induced by latent class assignment $\mathbf{Z}$.

We will proceed as follows: in step 6 we show a refined partition increases $\bar{\ell}(\cdot)$. We then construct a refined partition Π^* for every partition $\Pi^{\mathbf{Z}}$ induced by $\mathbf{Z}$ and prove $\bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\Pi^*) \geq \frac{1}{2} N_e(\mathbf{z})\beta_J$ in step 7. Finally we apply the results to MLE $\widehat{\mathbf{Z}}$ and obtain the desired results in step 8.

Step 6. We prove the following lemma:

Lemma 2. *Let Π^* be a refinement of any partition Π of $[N]$, then we have $\bar{\ell}(\Pi^*) \geq \bar{\ell}(\Pi)$.*

Given $a \in [T^*]$ indexing S_a^* in Π^*, since $S_a^* \subseteq S_b$ for some S_b in Π, let $F(a)$ denote its index under Π (i.e. b). We have

$$\begin{aligned}
\bar{\ell}\left(\Pi^*\right) &= \sum_{a=1}^{T^*} |S_a^*| \sum_{j=1}^{J} \left\{ \bar{\theta}_{j,a}^{\Pi^*} \log \bar{\theta}_{j,a}^{\Pi^*} + \left(1 - \bar{\theta}_{j,a}^{\Pi^*}\right) \log\left(1 - \bar{\theta}_{j,a}^{\Pi^*}\right) \right\} \\
&\geq \sum_{a=1}^{T^*} |S_a^*| \sum_{j=1}^{J} \left\{ \bar{\theta}_{j,a}^{\Pi^*} \log \bar{\theta}_{j,F(a)}^{\Pi} + \left(1 - \bar{\theta}_{j,a}^{\Pi^*}\right) \log\left(1 - \bar{\theta}_{j,F(a)}^{\Pi}\right) \right\} \\
&= \sum_{b=1}^{T} |S_b| \sum_{j=1}^{J} \left\{ \bar{\theta}_{j,b}^{\Pi} \log \bar{\theta}_{j,b}^{\Pi} + \left(1 - \bar{\theta}_{j,b}^{\Pi}\right) \log\left(1 - \bar{\theta}_{j,b}^{\Pi}\right) \right\} = \bar{\ell}(\Pi).
\end{aligned}$$

The first equality is obtained by rewriting $\bar{\ell}(\Pi)$ in terms of subsets. The inequality follows from non-negativity of K-L distance. Then we combine terms in same class under Π and obtain the second equality.

Step 7. Now we prove a result on refinement.

Lemma 3. *For any latent class assignment $\mathbf{Z}$, there exists a partition Π^* that refines $\Pi^{\mathbf{Z}}$ and*

$$\bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\Pi^*) \geq \frac{1}{2} N_e(\mathbf{z}) J \beta_J$$

For a given $\mathbf{Z}$, partition each latent class assigned by $\mathbf{Z}$ into sub-classes according to true assignments $\mathbf{Z}^0$ of each sample. For each sample i_1 that is incorrectly assigned by $\mathbf{Z}$ (by definition this means its true class under $\mathbf{Z}^0$ is not in the majority within its estimated class under $\mathbf{Z}$), we find another sample i_2 assigned to same class under $\mathbf{Z}$ but i_1 and i_2 belong to different class under

$\mathbf{Z}^0$ and make these two samples (i_1, i_2) a pair. We allow two misclassified samples to form a pair. Note that since incorrectly assigned samples are not in the majority of that class, we can find a pair for each of them.

Here is a simple example. Suppose in one class of $\mathbf{Z}$, we have 7 samples and $\mathbf{Z}^0$ (true latent class assignments) assigns them as three sub-classes $\{1,2,3,4\},\{5,6\},\{7\}$. In this example samples indexed by 5,6 and 7 are misclassified. We can find pairs $(4,5),(6,7)$.

The refined partition Π^* contains all such pairs and remaining correctly assigned samples in all classes assigned by $\mathbf{Z}$. So for the above example, the refined subset for that class is $\{1,2,3\},\{4,5\},\{6,7\}$. Let $e(\mathbf{z})$ by the set of incorrectly assigned sample. Clearly Π^* is a refinement and we have

$$\begin{aligned}
\bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\Pi^*) &= \sum_i \sum_j D(P_{i,j} \| \bar{\theta}^{\Pi^*}_{j,\Pi^*_i}) \\
&\geq \sum_{i \in e(\mathbf{z})} \sum_j D(P_{i,j} \| \bar{\theta}^{\Pi^*}_{j,\Pi^*_i}) \\
&= \sum_{i \in e(\mathbf{z})} \sum_j D(P_{i,j} \| \frac{P_{i,j} + P_{i',j}}{2})
\end{aligned}$$

where i and i' are in different classes under $\mathbf{Z}^0$ while in same subset under Π^* by definition. Apply Pinsker's inequality we have

$$\begin{aligned}
D(P_{i,j} \| \frac{P_{i,j} + P_{i',j}}{2}) &\geq \frac{1}{2}\left[|P_{i,j} - \frac{P_{i,j} + P_{i',j}}{2}| + |1 - P_{i,j} - (1 - \frac{P_{i,j} + P_{i',j}}{2})| \right]^2 \\
&= \frac{1}{2}(P_{i,j} - P_{i',j})^2 \\
&= \frac{1}{2}(\theta^0_{j,z^0_i} - \theta^0_{j,z^0_{i'}})^2
\end{aligned}$$

Hence we have

$$\begin{aligned}
\bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\Pi^*) &\geq \sum_{i \in e(\mathbf{z})} \sum_j \frac{1}{2}(\theta^0_{j,z^0_i} - \theta^0_{j,z^0_{i'}})^2 \\
&\geq \sum_{i \in e(\mathbf{z})} \frac{1}{2} \| \boldsymbol{\theta}^0_{\cdot,z^0_i} - \boldsymbol{\theta}^0_{\cdot,z^0_{i'}} \|^2 \\
&\geq \frac{1}{2} N_e(\mathbf{z}) J \beta_J
\end{aligned}$$

Step 8. Apply Lemma 3 to MLE $\widehat{\mathbf{z}}$, there exists a refinement of $\Pi^{\widehat{\mathbf{Z}}}$ denoted as Π^* such that

$$\bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\Pi^*) \geq \frac{1}{2} N_e(\mathbf{z}) J \beta_J$$

By Lemma 2 we have $\bar{\ell}(\Pi^*) \geq \bar{\ell}(\Pi^{\widehat{\mathbf{Z}}})$. So we conclude that

$$\begin{aligned} o_P(\delta_{NJ}) &= \bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\widehat{\mathbf{Z}}) \\ &\geq \bar{\ell}(\mathbf{Z}^0) - \bar{\ell}(\Pi^*) \\ &\geq \frac{1}{2} N_e(\mathbf{z}) J \beta_J \end{aligned}$$

which completes the proof.

Appendix 2: Proof of Corollary 2

Recall $m_a = \arg\max_{l \in [L]} \sum_{i \in \widehat{C}_a} Z^0_{i,l}$ is the class index under $\mathbf{Z}^0$ for cluster $\widehat{C}_a$. For any $0 < \epsilon < \tau$, define the following event

$$A^\epsilon_N = \{N_e(\widehat{\mathbf{z}})/N \leq \epsilon\}.$$

On the event A^ϵ_N, for any $l \in [L]$, since we assume $n^0_l/N \geq \tau > 0$, we claim that there is exactly one $a \in [L]$ such that $m_a = l$, i.e. the $a-$th cluster represents the $l-$th class. To see the existence of such a, assume by contradiction that for some l there is no a such that $m_a = l$, then all subjects in class l are misclassified and we have

$$N_e(\widehat{\mathbf{z}})/N \geq n^0_l/N \geq \tau > \epsilon,$$

a contradiction. Since for each $l-$th class we can find $a - th$ cluster to represent it and there are exactly L clusters $\widehat{C}_1, \ldots, \widehat{C}_L$, such a must be unique for all the L classes. Note that $\mathbb{P}(A^\epsilon_N) \to 1$, the first statement in the corollary is proved.

For any $\epsilon \in (0, \tau)$, from the argument above, on the event A^ϵ_N for each l we can find exactly one

$a \in [L]$ such that $m_a = l$, then the joint MLE for $\theta^0_{j,l}$ is

$$\widehat{\theta}_{j,a} = \widehat{\theta}^{(\widehat{\mathbf{z}})}_{j,a} = \frac{\sum_{i=1}^N \widehat{Z}_{i,a} R_{i,j}}{\sum_{i=1}^N \widehat{Z}_{i,a}}.$$

Recall we can rewrite $\theta^0_{j,l}$ as

$$\theta^0_{j,l} = \frac{\sum_{i=1}^N Z^0_{i,l}\theta^0_{j,l}}{\sum_{i=1}^N Z^0_{i,l}} = \frac{\sum_{i=1}^N Z^0_{i,l} P_{i,j}}{\sum_{i=1}^N Z^0_{i,l}}.$$

By triangle inequality we have

$$\begin{aligned}
&\max_j |\widehat{\theta}_{j,a} - \theta^0_{j,l}| \\
&= \max_j \left| \frac{\sum_{i=1}^N \widehat{Z}_{i,a} R_{i,j}}{\sum_{i=1}^N \widehat{Z}_{i,a}} - \frac{\sum_{i=1}^N Z^0_{i,l} P_{i,j}}{\sum_{i=1}^N Z^0_{i,l}} \right| \\
&\leq \max_j \left| \frac{\sum_{i=1}^N \widehat{Z}_{i,a} R_{i,j}}{\sum_{i=1}^N \widehat{Z}_{i,a}} - \frac{\sum_{i=1}^N \widehat{Z}_{i,a} R_{i,j}}{\sum_{i=1}^N Z^0_{i,l}} \right| + \max_j \left| \frac{\sum_{i=1}^N \widehat{Z}_{i,a} R_{i,j}}{\sum_{i=1}^N Z^0_{i,l}} - \frac{\sum_{i=1}^N Z^0_{i,l} R_{i,j}}{\sum_{i=1}^N Z^0_{i,l}} \right| \\
&+ \max_j \left| \frac{\sum_{i=1}^N Z^0_{i,l} R_{i,j}}{\sum_{i=1}^N Z^0_{i,l}} - \frac{\sum_{i=1}^N Z^0_{i,l} P_{i,j}}{\sum_{i=1}^N Z^0_{i,l}} \right| \\
&\equiv I_1 + I_2 + I_3.
\end{aligned}$$

We then analyze these three terms.

$$I_1 \leq \max_j \sum_i \widehat{Z}_{i,a} R_{i,j} \frac{\sum_i |\widehat{Z}_{i,a} - Z^0_{i,l}|}{n^0_l \sum_i \widehat{Z}_{i,a}} \leq \frac{\sum_i |\widehat{Z}_{i,a} - Z^0_{i,l}|}{n^0_l}.$$

There are two cases in which $|\widehat{Z}_{i,a} - Z^0_{i,l}| = 1$:

- $\widehat{Z}_{i,a} = 1, Z^0_{i,l} = 0$, i.e. subject i is in cluster a but not in class l. Since $m_a = l$, subject i is misclassified and counted in $N_e(\widehat{\mathbf{z}})$.

- $\widehat{Z}_{i,a} = 0, Z^0_{i,l} = 1$, i.e. subject i is in class l but not in cluster a. Since cluster a is the only cluster that represents class l, subject i must be misclassified and counted in $N_e(\widehat{\mathbf{z}})$.

By clustering consistency we have

$$I_1 \leq \frac{N_e(\widehat{\mathbf{z}})}{n^0_l} \leq \frac{N_e(\widehat{\mathbf{z}})}{\tau N} \xrightarrow{P} 0.$$

For the second term we have

$$I_2 = \frac{\max_j |\sum_i R_{i,j}(\widehat{Z}_{i,a} - Z^0_{i,l})|}{n^0_l} \le \frac{\sum_i |\widehat{Z}_{i,a} - Z^0_{i,l}|}{n^0_l} \le \frac{N_e(\widehat{\mathbf{z}})}{n^0_l} \xrightarrow{P} 0.$$

For the third term, we apply Hoeffding's inequality and obtain

$$\mathbb{P}(I_3 \ge \delta) = \mathbb{P}\left(\max_j \frac{|\sum_i Z^0_{i,l}(R_{i,j} - P_{i,j})|}{n^0_l} \ge \delta\right) \le J\exp(-2n^0_l\delta^2) \le J\exp(-2\tau N\delta^2) \to 0$$

where in the last step we use the scaling condition $\sqrt{\frac{M}{J}}\left(\frac{N}{L}\right)^{1-\xi} \to \infty$. This shows

$$I_3 \xrightarrow{P} 0.$$

Note that on $(A^\epsilon_N)^c$ we may not be able to define $\widehat{\theta}_{j,a}$ since the first statement in the corollary may not hold (we may not find the $a-th$ cluster for each $l-$th class). Mathematically we can arbitrarily define any $\widehat{\theta}_{j,a}$ as long as it is in $[0,1]$. Since $\mathbb{P}(A^\epsilon_N) \to 1$, we only need to focus on the situation on A^ϵ_N. We then have

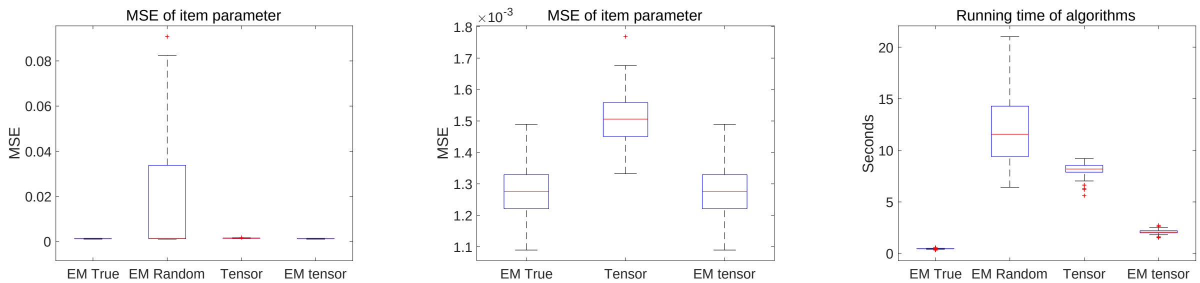
$$\begin{aligned}
&\mathbb{P}(\max_j |\widehat{\theta}_{j,a} - \theta^0_{j,l}| \ge \epsilon) \\
\le &\mathbb{P}(\max_j |\widehat{\theta}_{j,a} - \theta^0_{j,l}| I_{(A^\epsilon_N)^c} \ge \epsilon/2) + \mathbb{P}(\max_j |\widehat{\theta}_{j,a} - \theta^0_{j,l}| I_{A^\epsilon_N} \ge \epsilon/2) \\
\le &\mathbb{P}((A^\epsilon_N)^c) + \mathbb{P}(I_1 I_{A^\epsilon_N} \ge \epsilon/6) + \mathbb{P}(I_2 I_{A^\epsilon_N} \ge \epsilon/6) + \mathbb{P}(I_3 I_{A^\epsilon_N} \ge \epsilon/6) \to 0.
\end{aligned}$$

This completes the proof.

Appendix 3: More simulation results

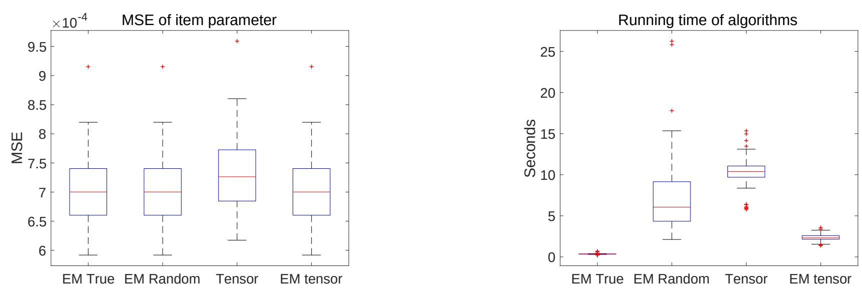
Random-effect LCM

We first present more simulation results for random-effect LCM.



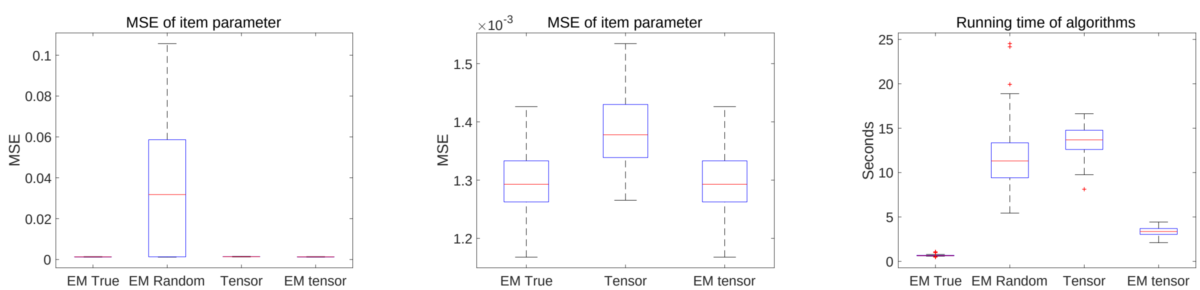
(a) MSE of item parameters (b) MSE without EM-random (c) Running time of the algorithms

Figure 1: $N = 1000, J = 100, L = 10$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



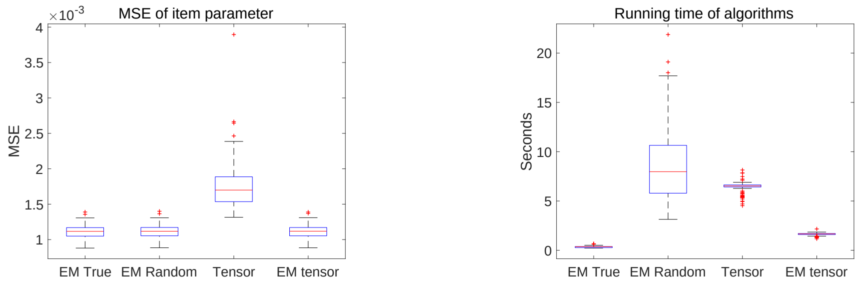
(a) MSE of item parameters (b) Running time of the algorithms

Figure 2: $N = 1000, J = 200, L = 5$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



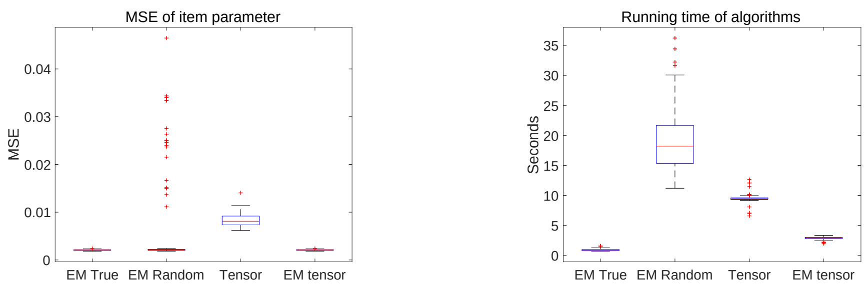
(a) MSE of item parameters (b) MSE without EM-random (c) Running time of the algorithms

Figure 3: $N = 1000, J = 200, L = 10$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



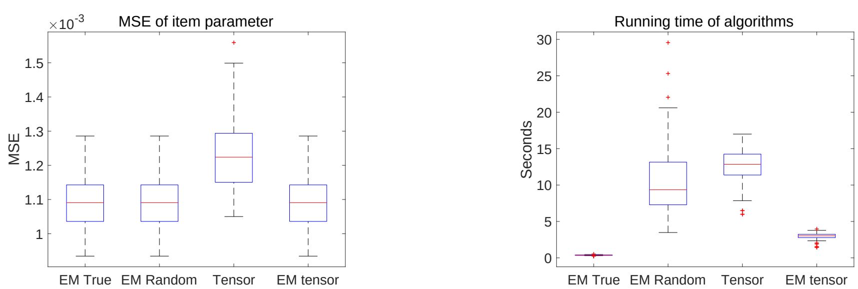
(a) MSE of item parameters (b) Running time of the algorithms

Figure 4: $N = 1000, J = 100, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



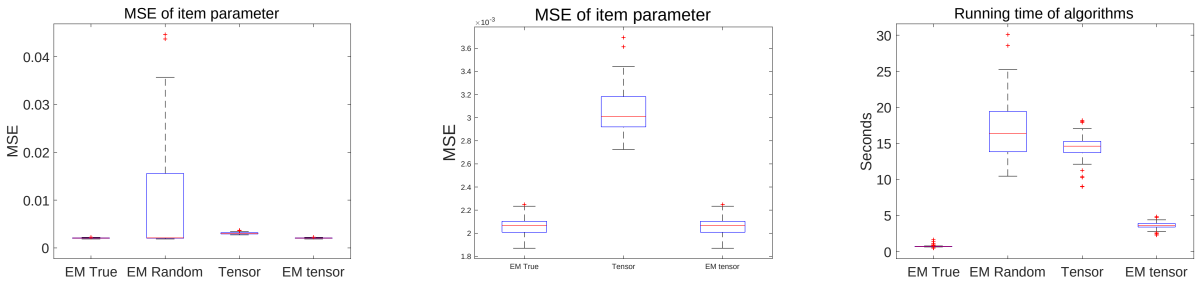
(a) MSE of item parameters (b) Running time of the algorithms

Figure 5: $N = 1000, J = 100, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



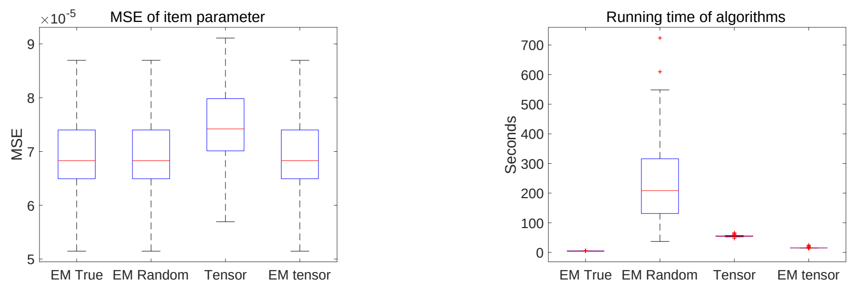
(a) MSE of item parameters (b) Running time of the algorithms

Figure 6: $N = 1000, J = 200, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



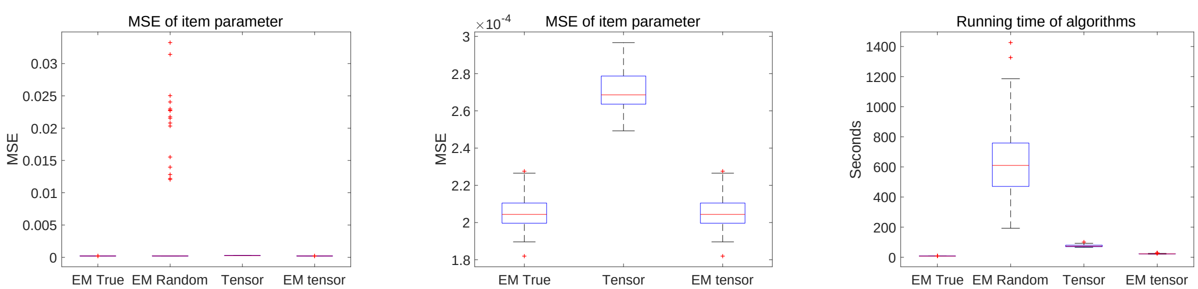
(a) MSE of item parameters (b) MSE without EM-random (c) Running time of the algorithms

Figure 7: $N = 1000, J = 200, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



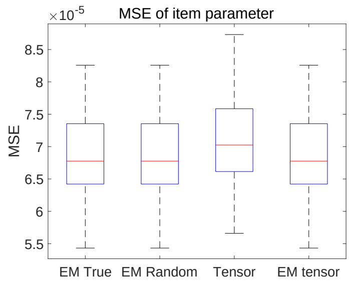
(a) MSE of item parameters (b) Running time of the algorithms

Figure 8: $N = 10000, J = 100, L = 5$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$

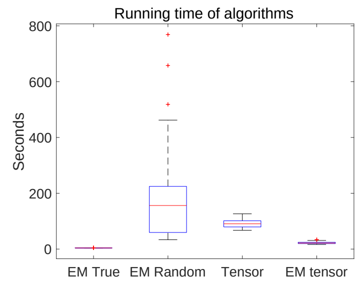


(a) MSE of item parameters (b) MSE without EM-random (c) Running time of the algorithms

Figure 9: $N = 10000, J = 200, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

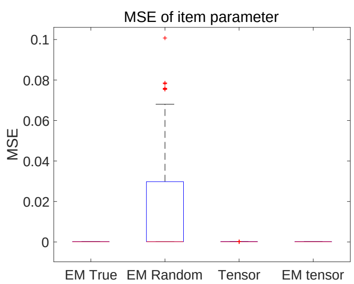


(a) MSE of item parameters

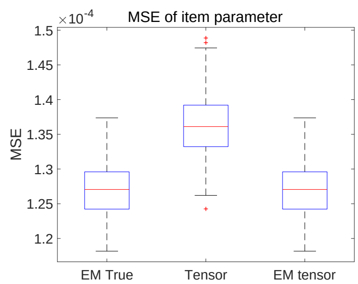


(b) Running time of the algorithms

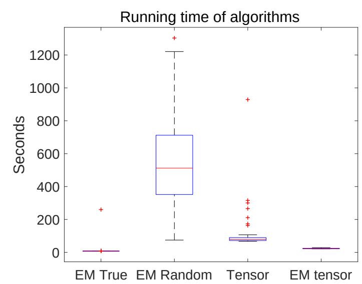
Figure 10: $N = 10000, J = 200, L = 5$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



(a) MSE of item parameters

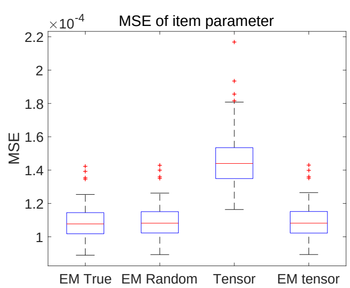


(b) MSE without EM-random

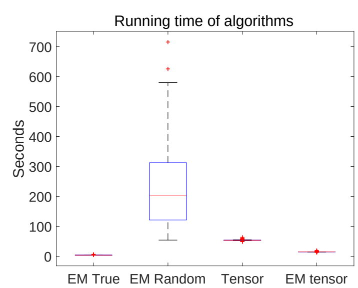


(c) Running time of the algorithms

Figure 11: $N = 10000, J = 200, L = 10$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$

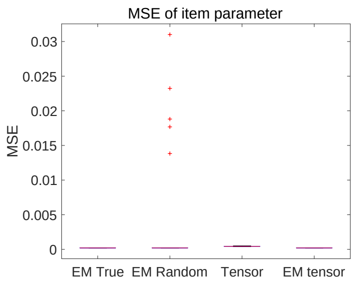


(a) MSE of item parameters

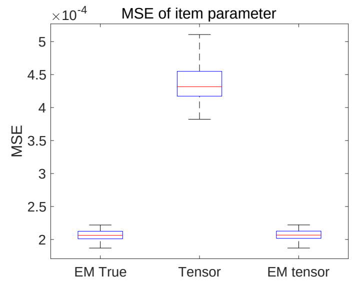


(b) Running time of the algorithms

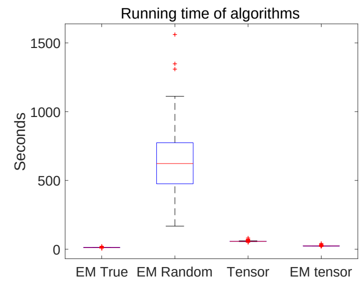
Figure 12: $N = 10000, J = 100, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



(a) MSE of item parameters

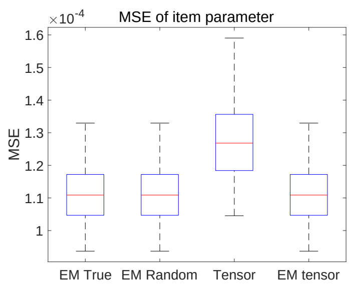


(b) MSE without EM-random

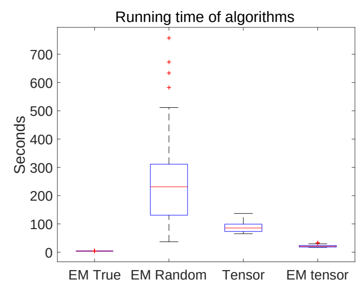


(c) Running time of the algorithms

Figure 13: $N = 10000, J = 100, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

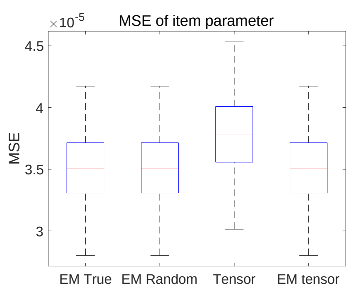


(a) MSE of item parameters

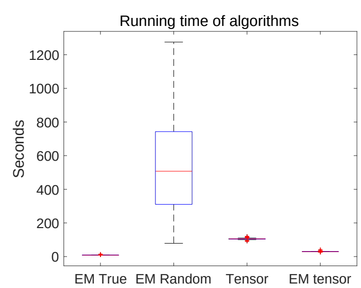


(b) Running time of the algorithms

Figure 14: $N = 10000, J = 200, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

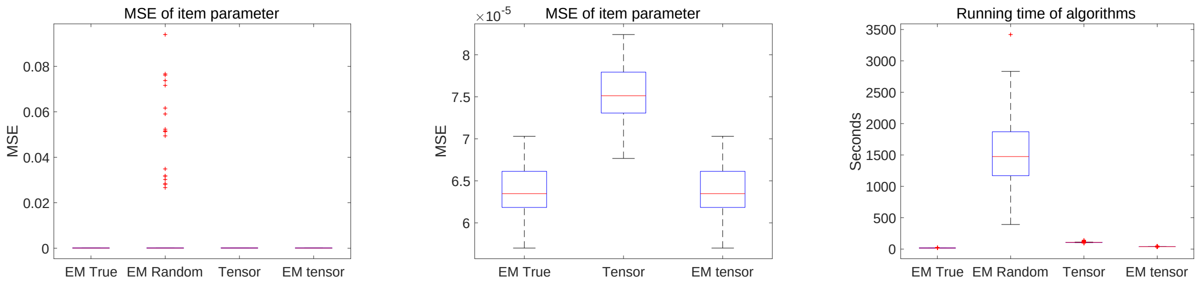


(a) MSE of item parameters



(b) Running time of the algorithms

Figure 15: $N = 20000, J = 100, L = 5$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$

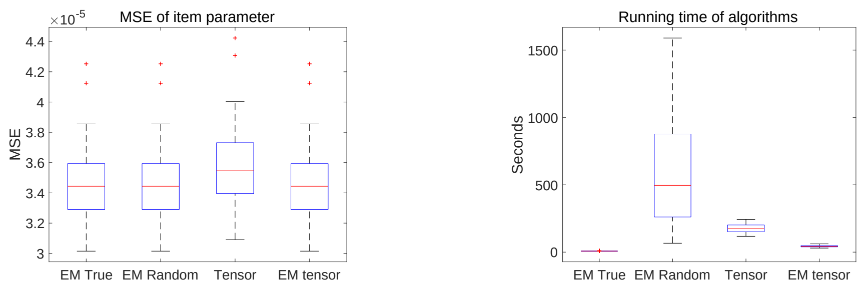


(a) MSE of item parameters

(b) MSE without EM-random

(c) Running time of the algorithms

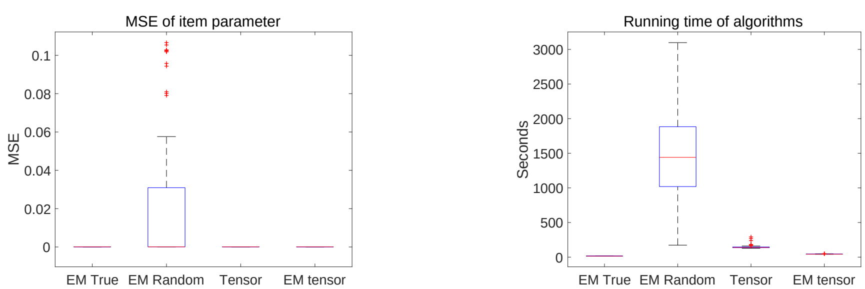
Figure 16: $N = 20000, J = 100, L = 10$,item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



(a) MSE of item parameters

(b) Running time of the algorithms

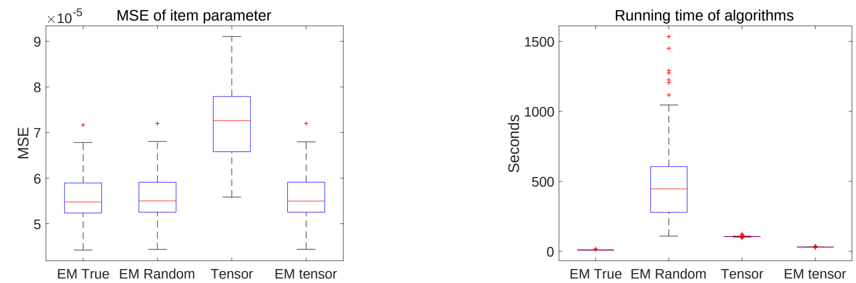
Figure 17: $N = 20000, J = 200, L = 5$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



(a) MSE of item parameters

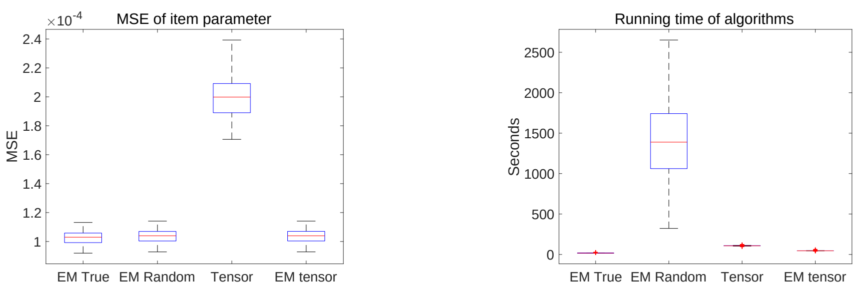
(b) Running time of the algorithms

Figure 18: $N = 20000, J = 200, L = 10$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



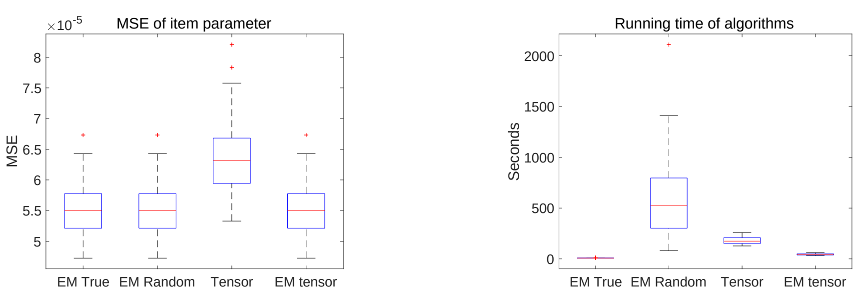
(a) MSE of item parameters (b) Running time of the algorithms

Figure 19: $N = 20000, J = 100, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



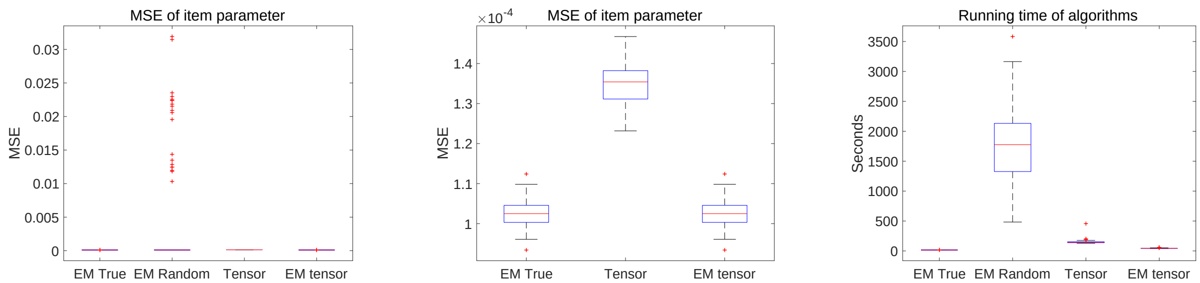
(a) MSE of item parameters (b) Running time of the algorithms

Figure 20: $N = 20000, J = 100, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



(a) MSE of item parameters (b) Running time of the algorithms

Figure 21: $N = 20000, J = 200, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

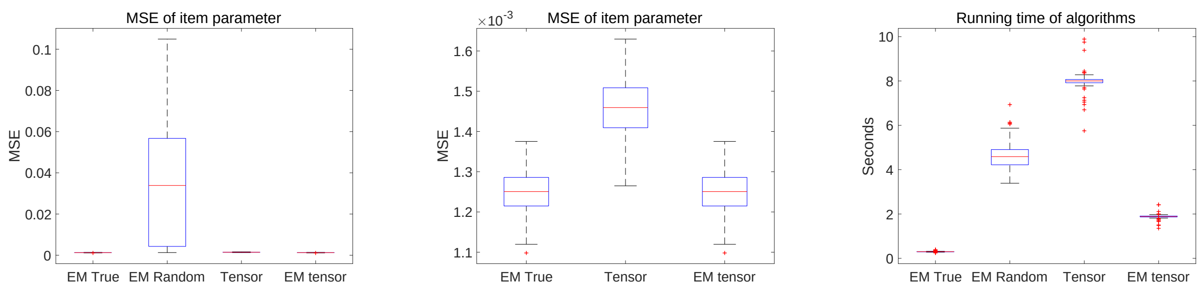


(a) MSE of item parameters (b) MSE without EM-random (c) Running time of the algorithms

Figure 22: $N = 20000, J = 200, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

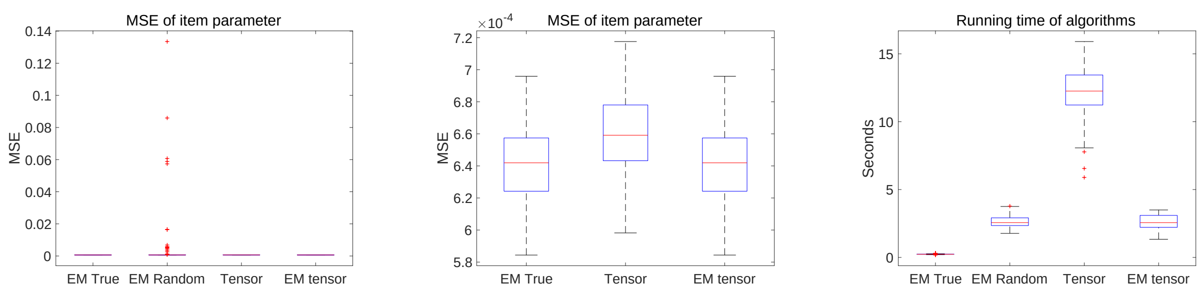
Fixed-effect LCM

Then the simulation results of fixed-effect LCM are presented.



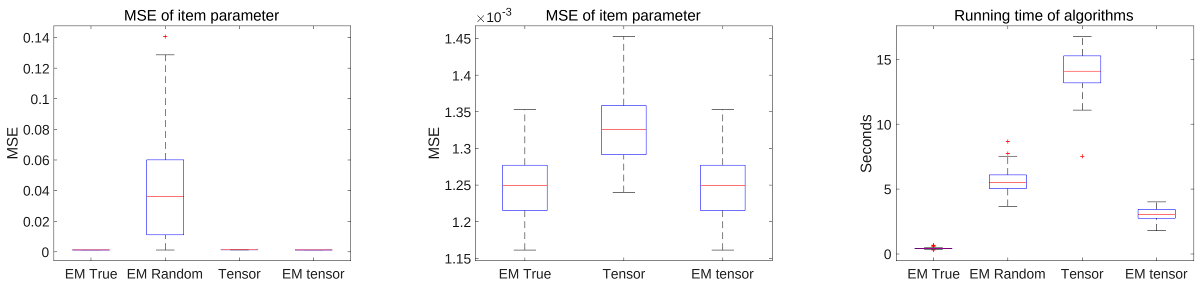
(a) MSE of item parameters (b) MSE without EM-random (c) Running time of the algorithms

Figure 23: $N = 1000, J = 100, L = 10$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



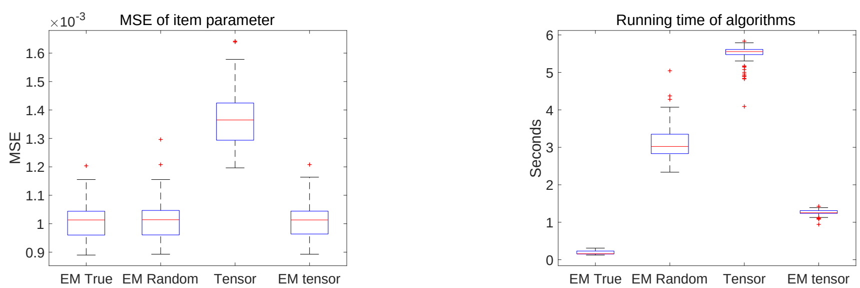
(a) MSE of item parameters (b) MSE without EM-random (c) Running time of the algorithms

Figure 24: $N = 1000, J = 200, L = 5$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



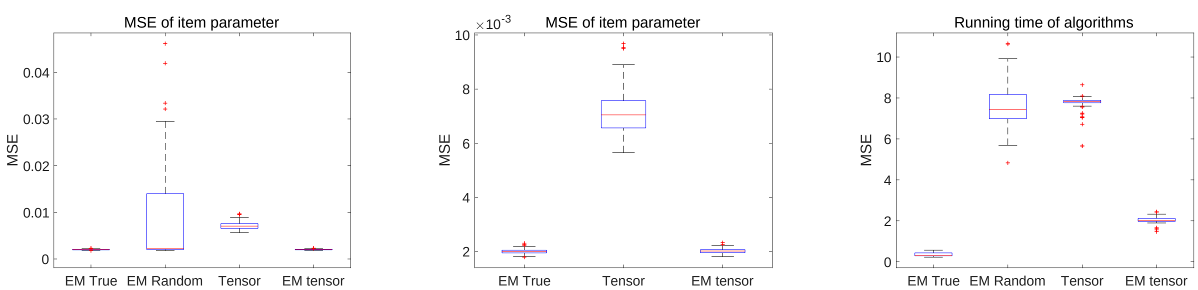
(a) MSE of item parameters (b) MSE without EM-random (c) Running time of the algorithms

Figure 25: $N = 1000, J = 200, L = 10$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



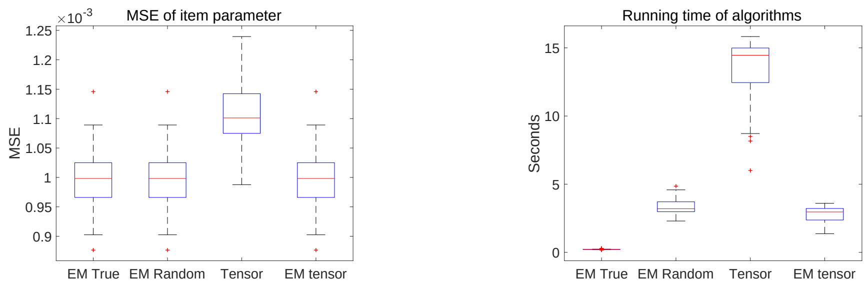
(a) MSE of item parameters (b) Running time of the algorithms

Figure 26: $N = 1000, J = 100, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



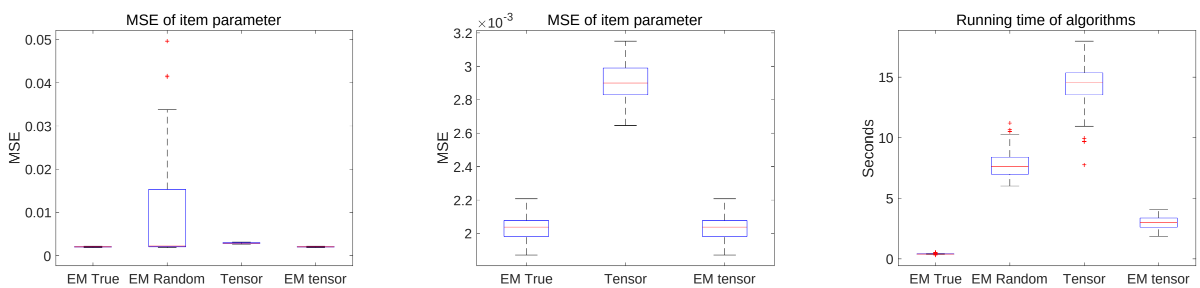
(a) MSE of item parameters (b) MSE without EM-random (c) Running time of the algorithms

Figure 27: $N = 1000, J = 100, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



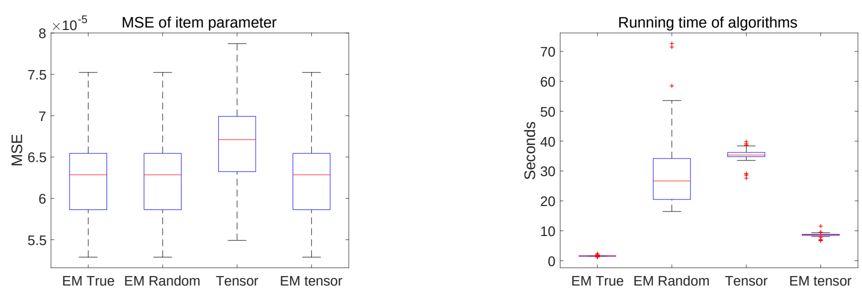
(a) MSE of item parameters (b) Running time of the algorithms

Figure 28: $N = 1000, J = 200, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



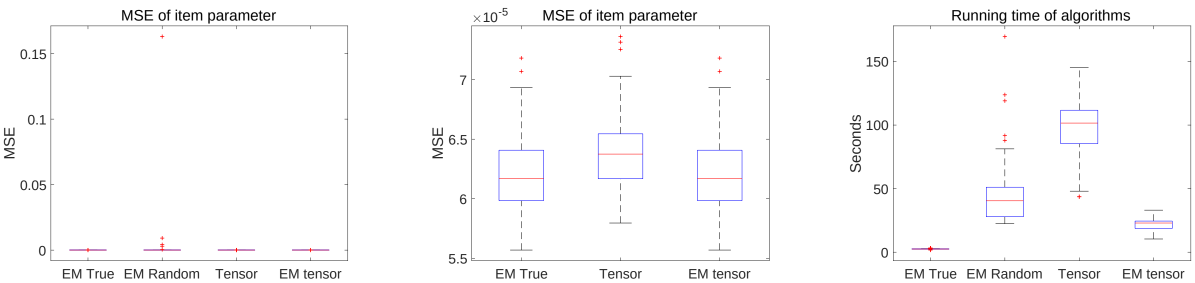
(a) MSE of item parameters (b) MSE without EM-random (c) Running time of the algorithms

Figure 29: $N = 1000, J = 200, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



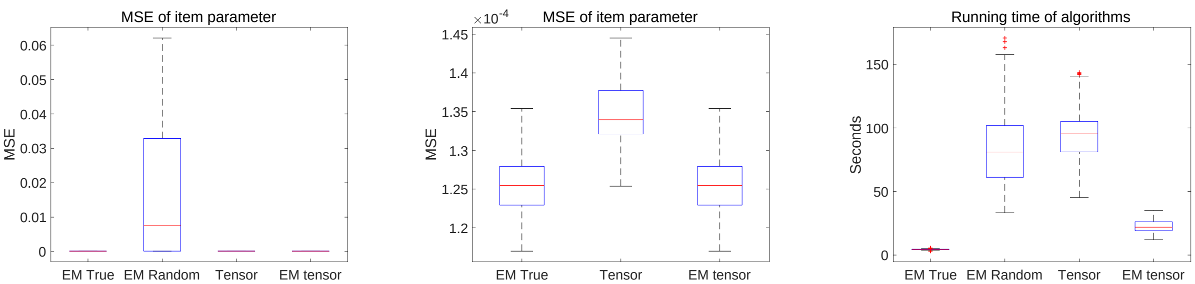
(a) MSE of item parameters (b) Running time of the algorithms

Figure 30: $N = 10000, J = 100, L = 5$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



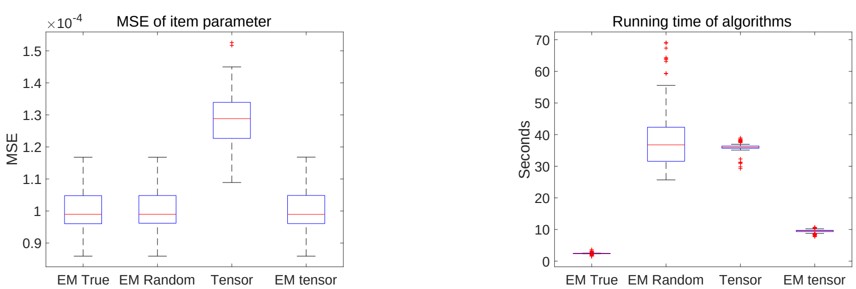
(a) MSE of item parameters (b) MSE without EM-random (c) Running time of the algorithms

Figure 31: $N = 10000, J = 200, L = 5$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



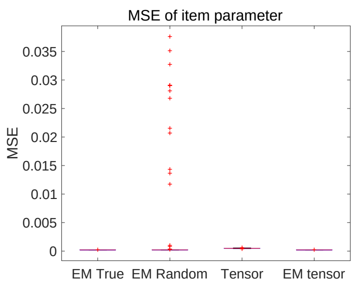
(a) MSE of item parameters (b) MSE without EM-random (c) Running time of the algorithms

Figure 32: $N = 10000, J = 200, L = 10$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$

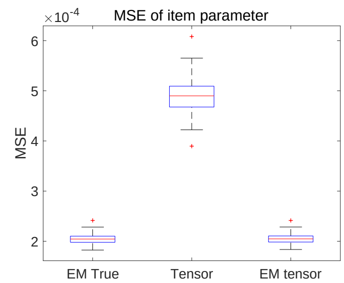


(a) MSE of item parameters (b) Running time of the algorithms

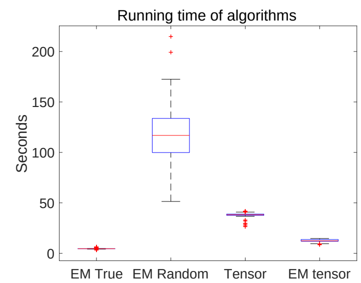
Figure 33: $N = 10000, J = 100, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



(a) MSE of item parameters

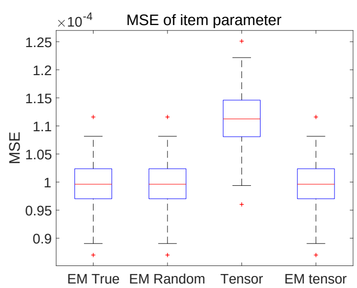


(b) MSE without EM-random

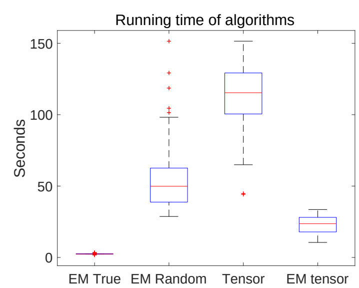


(c) Running time of the algorithms

Figure 34: $N = 10000, J = 100, L = 10$,item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

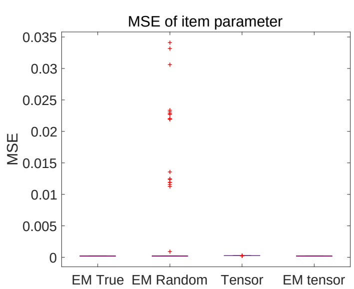


(a) MSE of item parameters

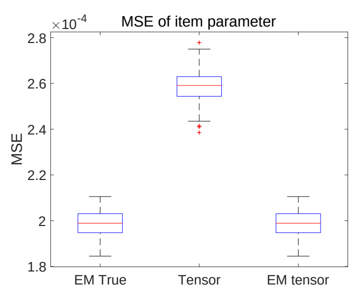


(b) Running time of the algorithms

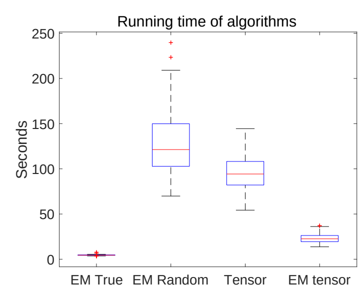
Figure 35: $N = 10000, J = 200, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



(a) MSE of item parameters

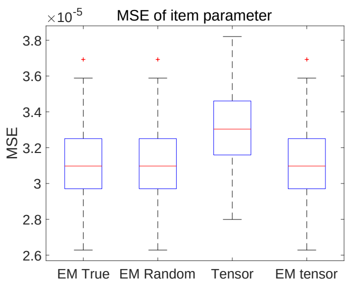


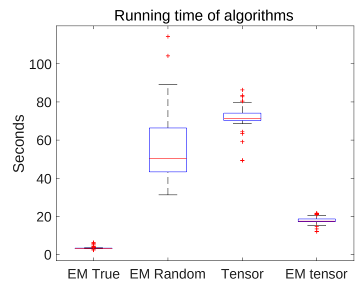
(b) MSE without EM-random



(c) Running time of the algorithms

Figure 36: $N = 10000, J = 200, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

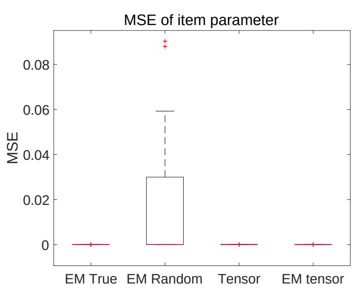


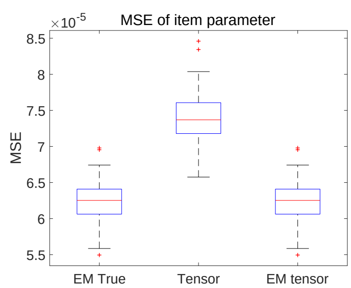


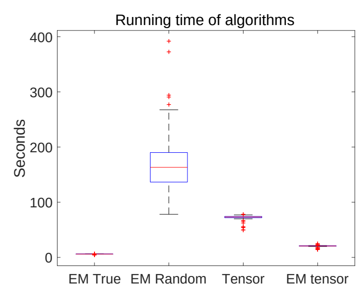
(a) MSE of item parameters

(b) Running time of the algorithms

Figure 37: $N = 20000, J = 100, L = 5$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$





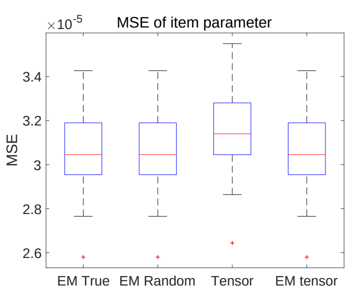


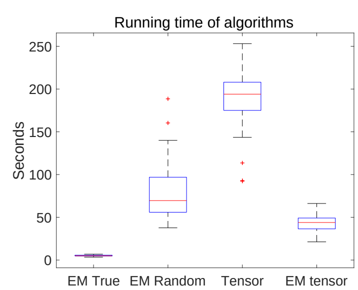
(a) MSE of item parameters

(b) MSE without EM-random

(c) MSE without EM-random

Figure 38: $N = 20000, J = 100, L = 10$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$

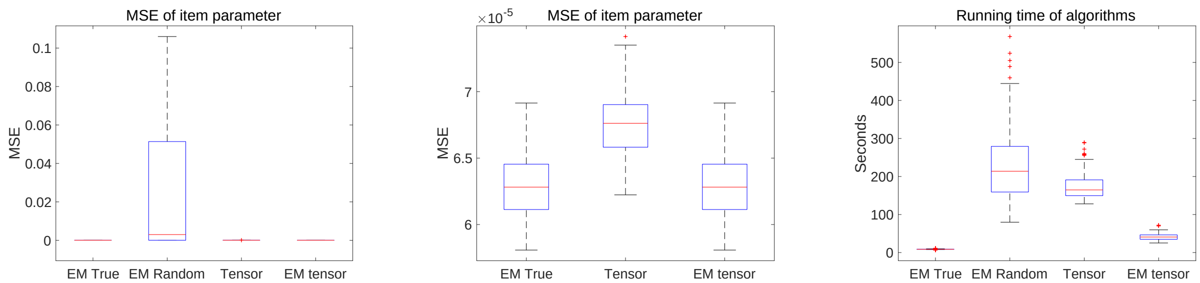




(a) MSE of item parameters

(b) Running time of the algorithms

Figure 39: $N = 20000, J = 200, L = 5$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$

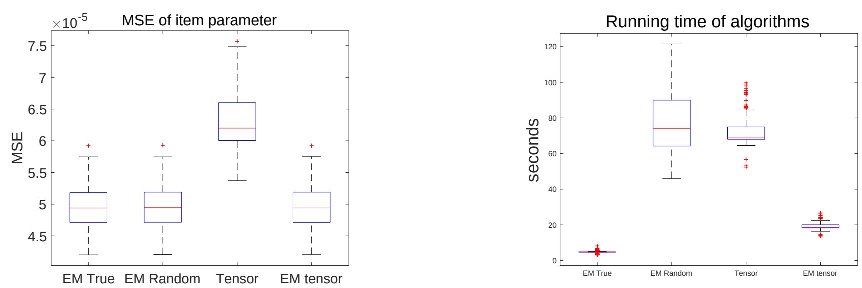


(a) MSE of item parameters

(b) MSE without EM-random

(c) Running time of the algorithms

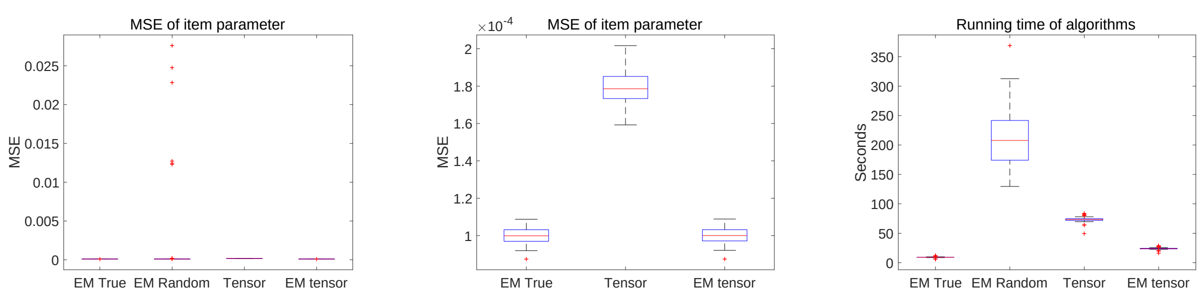
Figure 40: $N = 20000, J = 200, L = 10$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



(a) MSE of item parameters

(b) Running time of the algorithms

Figure 41: $N = 20000, J = 100, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

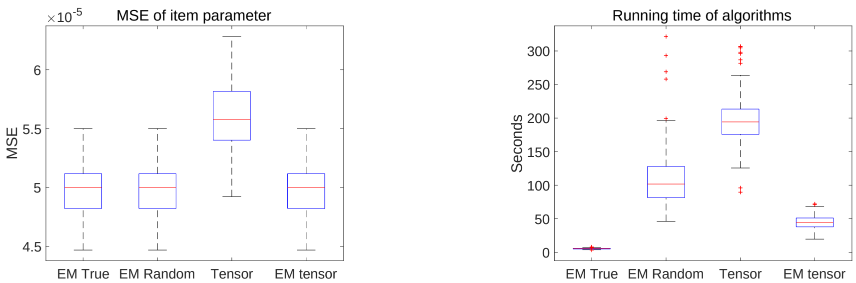


(a) MSE of item parameters

(b) MSE without EM-random

(c) Running time of the algorithms

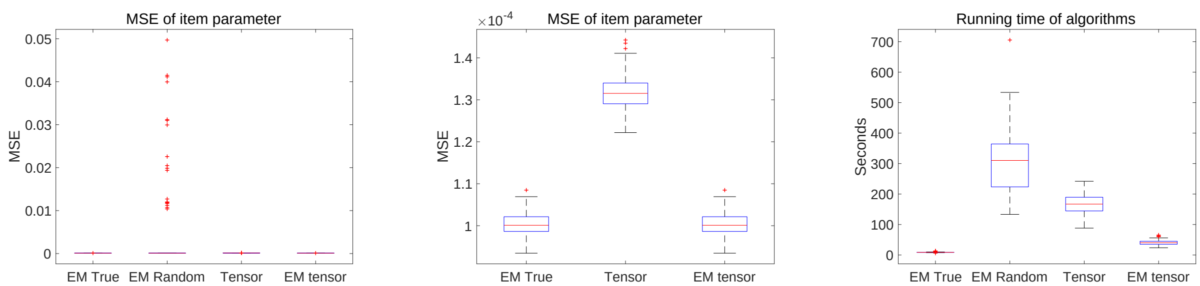
Figure 42: $N = 20000, J = 100, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



(a) MSE of item parameters

(b) Running time of the algorithms

Figure 43: $N = 20000, J = 200, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



(a) MSE of item parameters

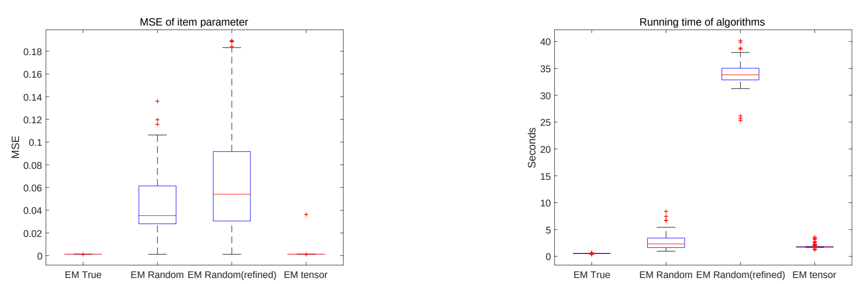
(b) MSE without EM-random

(c) Running time of the algorithms

Figure 44: $N = 20000, J = 200, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

EM-random and Tensor-EM with same initializations

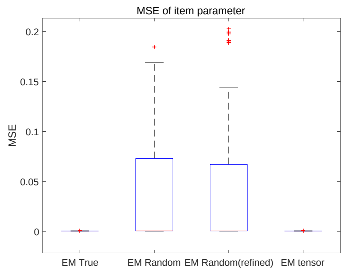
We then present the simulation results when EM-random and Tensor-EM are implemented with same initializations together with the "smarter" version of EM-random.



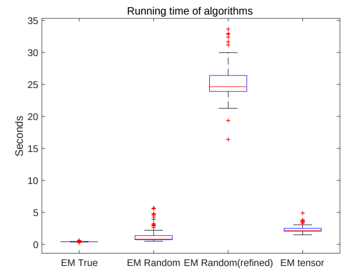
(a) MSE of item parameters

(b) Running time of the algorithms

Figure 45: $N = 1000, J = 100, L = 10$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$

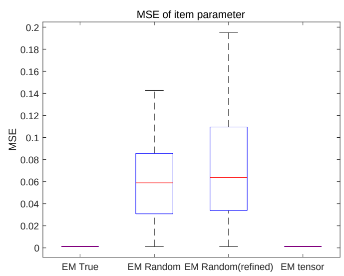


(a) MSE of item parameters

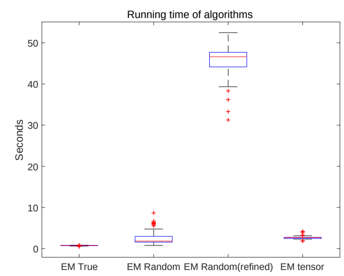


(b) Running time of the algorithms

Figure 46: $N = 1000, J = 200, L = 5$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$

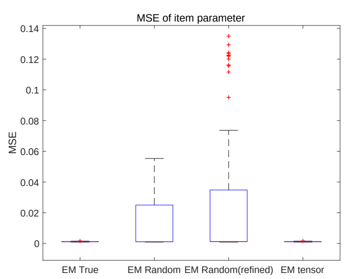


(a) MSE of item parameters

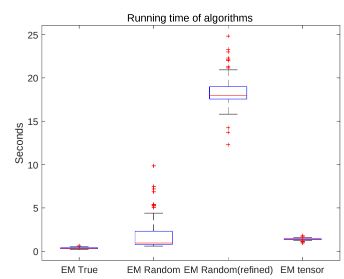


(b) Running time of the algorithms

Figure 47: $N = 1000, J = 200, L = 10$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$

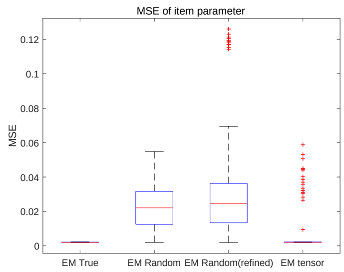


(a) MSE of item parameters

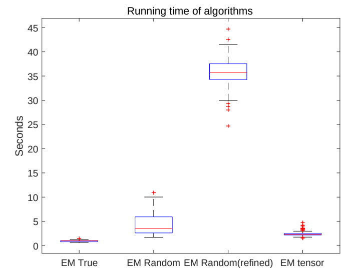


(b) Running time of the algorithms

Figure 48: $N = 1000, J = 100, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

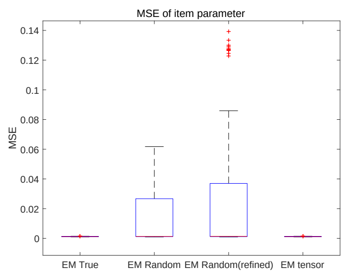


(a) MSE of item parameters

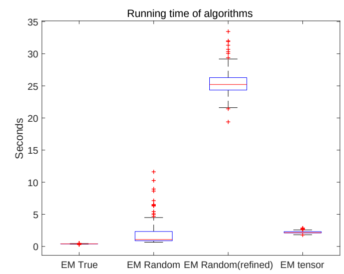


(b) Running time of the algorithms

Figure 49: $N = 1000, J = 100, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

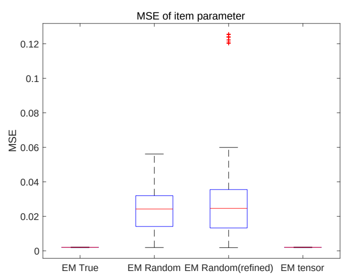


(a) MSE of item parameters

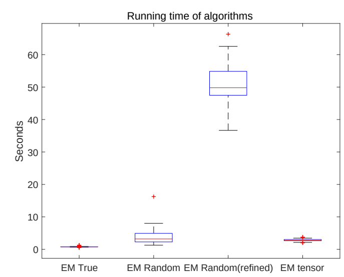


(b) Running time of the algorithms

Figure 50: $N = 1000, J = 200, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

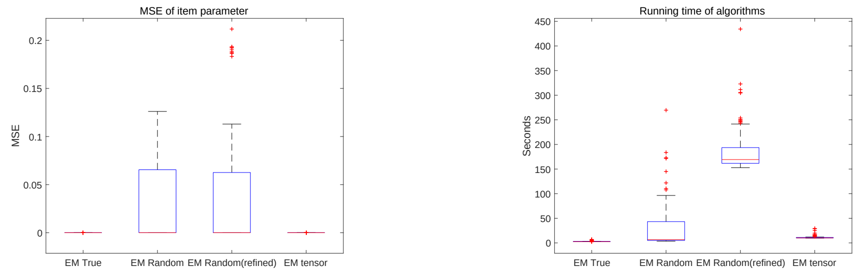


(a) MSE of item parameters



(b) Running time of the algorithms

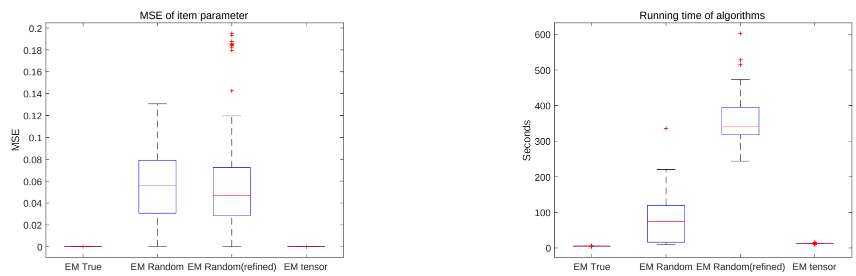
Figure 51: $N = 1000, J = 200, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$



(a) MSE of item parameters

(b) Running time of the algorithms

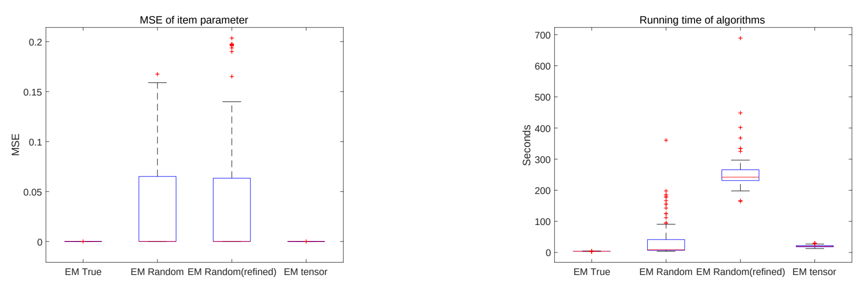
Figure 52: $N = 10000, J = 100, L = 5$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



(a) MSE of item parameters

(b) Running time of the algorithms

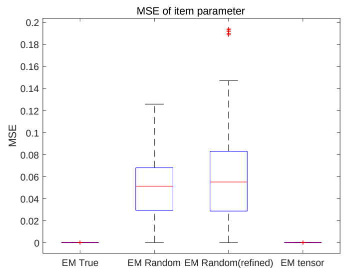
Figure 53: $N = 10000, J = 100, L = 10$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$



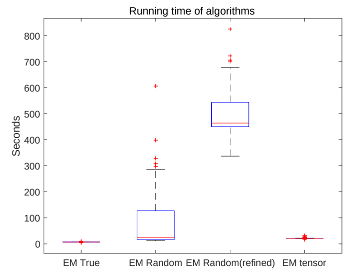
(a) MSE of item parameters

(b) Running time of the algorithms

Figure 54: $N = 10000, J = 200, L = 5$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$

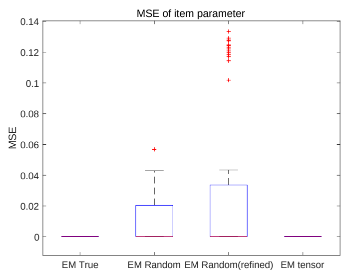


(a) MSE of item parameters

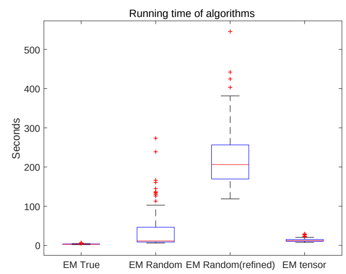


(b) Running time of the algorithms

Figure 55: $N = 10000, J = 200, L = 10$, item parameters $\in \{0.1, 0.2, 0.8, 0.9\}$

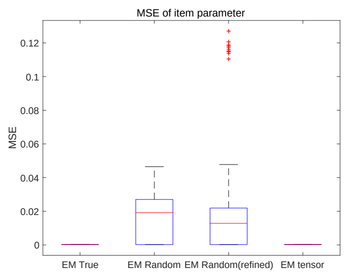


(a) MSE of item parameters

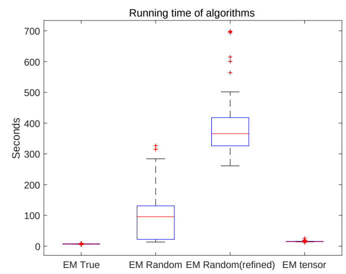


(b) Running time of the algorithms

Figure 56: $N = 10000, J = 100, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

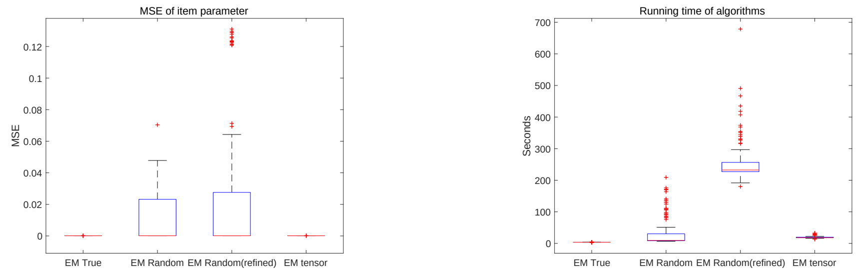


(a) MSE of item parameters



(b) Running time of the algorithms

Figure 57: $N = 10000, J = 100, L = 10$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

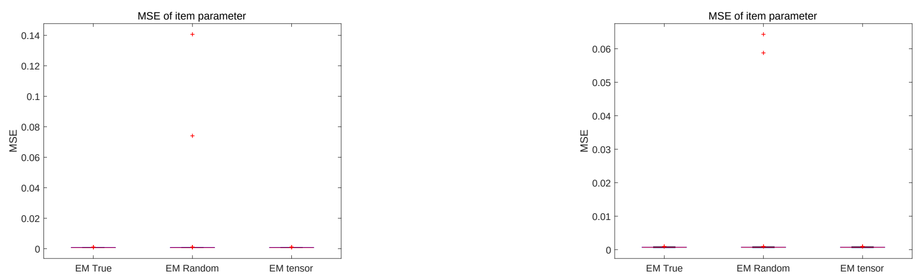


(a) MSE of item parameters

(b) Running time of the algorithms

Figure 58: $N = 10000, J = 200, L = 5$, item parameters $\in \{0.2, 0.4, 0.6, 0.8\}$

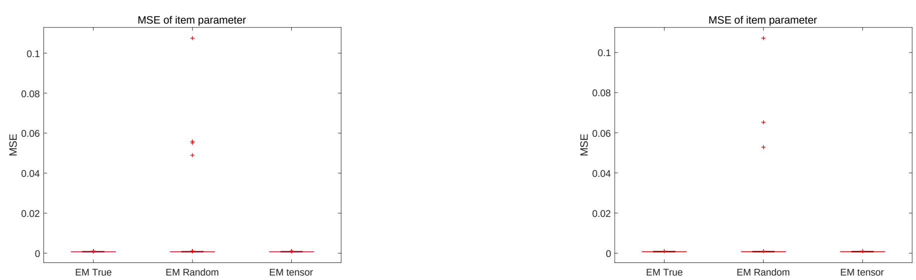
Simulations under local dependence



(a) MSE of item parameters $\rho = 0.3$

(b) MSE of item parameters $\rho = 0.7$

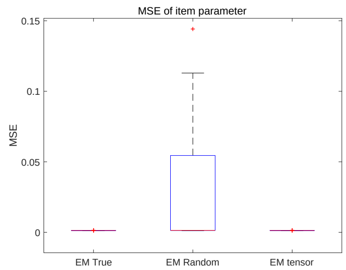
Figure 59: Random-effect LCM, $N = 1000, J = 100, L = 5, \theta_{j,a} \in \{0.1, 0.2, 0.8, 0.9\}$



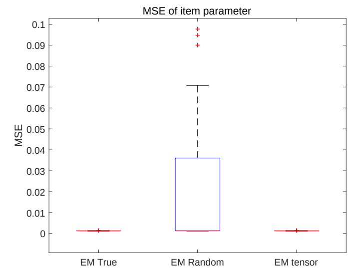
(a) MSE of item parameters $\rho = 0.3$

(b) MSE of item parameters $\rho = 0.7$

Figure 60: Random-effect LCM, $N = 1000, J = 200, L = 5, \theta_{j,a} \in \{0.1, 0.2, 0.8, 0.9\}$

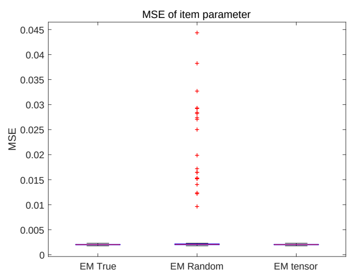


(a) MSE of item parameters $\rho = 0.3$

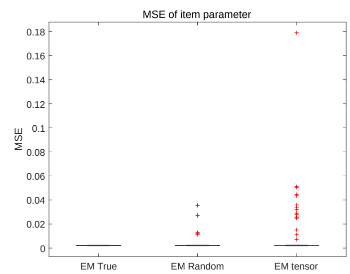


(b) MSE of item parameters $\rho = 0.7$

Figure 61: Random-effect LCM, $N = 1000, J = 200, L = 10, \theta_{j,a} \in \{0.1, 0.2, 0.8, 0.9\}$

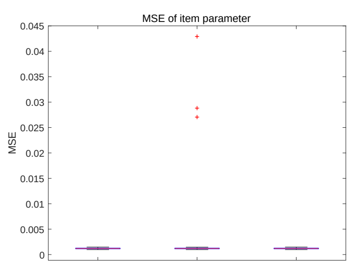


(a) MSE of item parameters $\rho = 0.3$

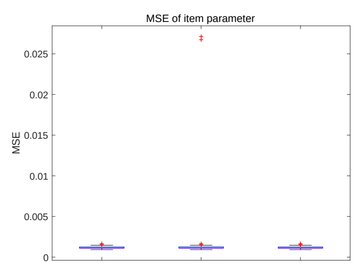


(b) MSE of item parameters $\rho = 0.7$

Figure 62: Random-effect LCM, $N = 1000, J = 100, L = 10, \theta_{j,a} \in \{0.2, 0.4, 0.6, 0.8\}$

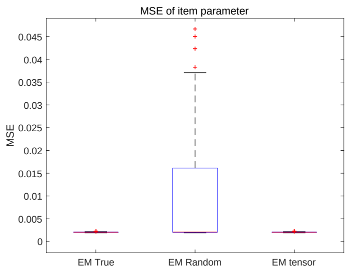


(a) MSE of item parameters $\rho = 0.3$

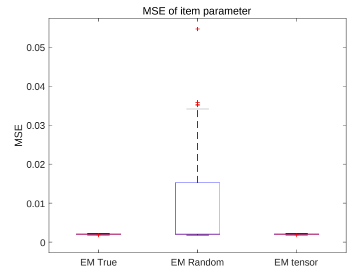


(b) MSE of item parameters $\rho = 0.7$

Figure 63: Random-effect LCM, $N = 1000, J = 200, L = 5, \theta_{j,a} \in \{0.2, 0.4, 0.6, 0.8\}$



(a) MSE of item parameters $\rho = 0.3$



(b) MSE of item parameters $\rho = 0.7$

Figure 64: Random-effect LCM, $N = 1000, J = 200, L = 10, \theta_{j,a} \in \{0.2, 0.4, 0.6, 0.8\}$