APPENDIX

A. Proof of Alleviator

1) Assumption 1: All session data are i.i.d. (i.e., independent and identically distributed) samples affected by random user's behavior under some uniform distribution $\mathbb{P}_{\mathbf{s}}$.

i.e. For a set of collected sampled sessions: $\{\boldsymbol{s}_i\}_{i=1}^m$, where $\boldsymbol{s}_i = [s_{i,1}, s_{i,2}, \cdots, s_{i,l_i}]$ as we defined in Section III-A. For all i, we have: $\boldsymbol{s}_i - \epsilon_i \sim \mathbb{P}_s$ where ϵ_i is the random user's behavior, which can be intuitively understood as the user's random behavior

2) Assumption 2 (Empirical Bayes Assumption I): The encoded value of a session (i.e., z_{ij} in Section IV-A1) is not the real value, just an observation affected by random user's behavior:

i.e., for a not affected session $\boldsymbol{s}_i - \epsilon_i$; the real encoded latent variable is $\boldsymbol{\mu}_i = [\mu_{i1}, \cdots, \mu_{in'}]^T$; For a fixed $j = 1, 2, \cdots, n'$ (where n' is the dim for latent variable), the $\{\mu_{ij}\}_{i=1}^m$ and follows distribution $\mathbb{P}_s^{(j)}$.

That means the real value for $j-th$ propriety of the $i-th$ session should be μ_{ij}; But due to the effect of random user's behavior, the encoded result we observe from the session-encoder is z_{ij}

3) Assumption 3 (Empirical Bayes Assumption II): The observed value of the encoded session z_{ij} follows the distribution of $\mathcal{N}(\mu_{ij}, \sigma_j^2)$ and $\sigma_j \geqslant 1$ (if this assumption is not met, we can always do batch normalization to make the σ_j not far from 1).

The Normal distribution assumption comes from the statistic common that if a distribution is affected by extremely complex factors, like the random user's behavior. The safest way is to assume that they are normally distributed. Since all numbers in $\{z_{ij}\}_{i=1}^m$ represent the same factor of the session (the $j-th$ encoded factor), it is reasonable to assume they have the same and relatively large variance.

4) Target: In high-level understanding, what we observed in the real data is not the full fact but noisy data that have information of the underlying true value. Our goal is to obtain the underlying true value (i.e., μ_{ij} in our case) through observed values (the z_{ij} in our case).

The rigorous definition of the target is: given a batched, observed encoded result: $\boldsymbol{Z} \in \mathbb{R}^{m \times n'}$ and its corresponding underlying true value $\boldsymbol{\mu} \in \mathbb{R}^{m \times n'}$

For a fixed $j \in [1, n']$, get an estimator $\hat{\mu}_{ij}|\boldsymbol{\xi}_j$ for μ_{ij} s.t. $\mathbb{E}[\sum_{i=1}^m (\hat{\mu}_{ij} - \mu_{ij})^2]$ is small.

Consider the Max likelihood estimator $\hat{\mu}_{ij}^{(MLE)} = z_{ij}$, and $\hat{\mu}_{(JS)}^{ij} = (1 - \frac{m-2}{\|\boldsymbol{\xi}_j\|^2}) z_{ij}$.

Claim that: $\mathbb{E}[\sum_{i=1}^m (\mu_{ij} - \hat{\mu}_{ij}^{(JS)})^2] \leqslant \mathbb{E}[\sum_{i=1}^m (\mu_{ij} - \hat{\mu}_{ij}^{(MLE)})^2]$

5) Proof of claim: $\mathbb{E}[\sum_{i=1}^m (\mu_{ij} - \hat{\mu}_{ij})^2] = \sum_{i=1}^m \mathbb{E}[(z_{ij} - \hat{\mu}_{ij})^2 - (z_{ij} - \mu_{ij})^2 + 2(\hat{\mu}_{ij} - \mu_{ij})(z_{ij} - \mu_{ij})] = \sum_{i=1}^m \mathbb{E}[(z_{ij} - \hat{\mu}_{ij})^2] - m \cdot \sigma_j + 2\sum_{i=1}^m \mathbb{E}[(\hat{\mu}_{ij} - \mu_{ij})(z_{ij} - \mu_{ij})]$

Consider distribution function for z_{ij} as: $\varphi(z_{ij}|\mu_{ij}, \sigma_j) = \frac{1}{\sqrt{2\pi} \cdot \sigma_j} \exp(-\frac{(z_{ij} - \mu_{ij})^2}{2\sigma_j^2})$. Therefore, $(z_{ij} - \mu_{ij})(\hat{\mu}_{ij} - \mu_{ij}) = -\sigma_j^2 \frac{\partial}{\partial z_{ij}} \varphi(z_{ij}|\mu_{ij}, \sigma_j)$

Therefore, for any continuous, differentiable, and $|f(z)| < \infty$, function $f: \mathbb{R} \to \mathbb{R}$. For simplicity, denote $\varphi(z_{ij}|\mu_{ij}, \sigma_j)$ as $\varphi(z_{ij})$ we have:

$$
\begin{aligned}
&\mathbb{E}[(z_{ij} - \mu_{ij}) f(z_{ij})] \\
=&\int_{-\infty}^{+\infty} (z_{ij} - \mu_{ij}) f(z_{ij}) \varphi(z_{ij}) dz_{ij} \\
=&(-\sigma_j^2) \cdot (\int_{-\infty}^{+\infty} (\frac{\partial}{\partial z_{ij}} \varphi(z_{ij})) f(z_{ij}) dz_{ij}) \\
=&(-\sigma_j^2) \cdot \varphi(z_{ij}) \cdot f(z_{ij})|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} f'(z_{ij}) \varphi(z_{ij}) dz_{ij} \\
=&\sigma_j^2 \cdot (\int_{-\infty}^{+\infty} f'(z_{ij}) \varphi(z_{ij}) dz_{ij}) \\
=&\sigma_j^2 \mathbb{E}[\frac{\partial}{\partial z_{ij}} f(z_{ij})]
\end{aligned}
$$

Therefore, we have: $\mathbb{E}[(z_{ij} - \mu_{ij})(\hat{\mu}_{ij} - \mu_{ij})] = \mathbb{E}[\frac{\partial \hat{\mu}_{ij}}{\partial z_{ij}}] \cdot \sigma_j^2$. Therefore, when $\hat{\mu}_{ij}$ is MLE: $\mathbb{E}[\sum_{i=1}^m (\mu_{ij} - \hat{\mu}_{ij}^{(MLE)})^2] = 0 - m \cdot \sigma_j^2 + 2m \cdot \sigma_j^2 = m \cdot \sigma_j^2$ When $\hat{\mu}_{ij}$ is $\hat{\mu}_{ij}^{(JS)}$. We have: $\mathbb{E}[\frac{\partial \hat{\mu}_{ij}^{(JS)}}{\partial z_{ij}}] = 1 - \frac{m-2}{\|\boldsymbol{\xi}_j\|^2} + \frac{2(m-2) z_{ij}^2}{\|\boldsymbol{\xi}_j\|^4}$

Therefore, $\sum_{i=1}^m \mathbb{E}[(\hat{\mu}_{ij}^{(JS)} - \mu_{ij})(z_{ij} - \mu_{ij})] = m - \mathbb{E}[\frac{(m-2)^2}{\|\boldsymbol{\xi}_j\|^2}]$

With $\mathbb{E}[(z_{ij} - \hat{\mu}_{ij}^{(JS)})^2] = \mathbb{E}[\frac{(m-2)^2}{\|\boldsymbol{\xi}_j\|^2} z_{ij}^2]$, we have: $\mathbb{E}[\sum_{i=1}^m (\mu_{ij} - \hat{\mu}_{ij}^{(JS)})^2] = m \cdot \sigma_j^2 + m(1 - 2\sigma_j^2)\mathbb{E}[\frac{(m-2)^2}{\|\boldsymbol{\xi}_j\|^2}]$

Since in our assumption $\sigma_j^2 \geqslant 1$, we have:

$$
\mathbb{E}[\sum_{i=1}^m (\mu_{ij} - \hat{\mu}_{ij}^{(JS)})^2] \leqslant \mathbb{E}[\sum_{i=1}^m (\mu_{ij} - \hat{\mu}_{ij}^{(MLE)})^2] \tag{21}
$$