Supplements for "Generalized tensor-response model with multi-sided covariates"

1 Proofs

Theorem 1.1. *Consider a generalized tensor regression model with multi-sided covariates* $\mathcal{X} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K\}$*. Suppose the entries in* $\mathcal{Y}$ *are independent realizations of an exponential family distribution, and* $\mathbb{E}(\mathcal{Y}|\mathcal{X})$ *follows the low-rank tensor regression model (6). Under Assumption 1, there exist two absolute constants* $C_1, C_2 > 0$*, such that, with probability at least* $1-\exp(-C_1 \sum_k p_k)$,

$$Loss(\mathcal{B}_{true},\ \hat{\mathcal{B}}) \leq C_3 \sum_k p_k,$$

where $C_3 = C_3(\boldsymbol{r}) = \frac{1}{C_2^{2K} U} \frac{\prod_k r_k}{\max_k r_k} > 0$ *is a constant that does not depend on the dimensions* $\{d_k\}$ *and* $\{p_k\}$.

Proof of Theorem 1. Let $\ell(\mathcal{B}) = \mathbb{E}(\mathcal{L}_{\mathcal{Y}}(\mathcal{B}))$, where the expectation is take with respect to $\mathcal{Y} \sim \mathcal{B}_{\text{true}}$ under the true parameter. We show that

C1. The stochastic deviation $\mathcal{L}_{\mathcal{Y}}(\mathcal{B}) - \ell(\mathcal{B})$ is uniformly small for all $\mathcal{B} \in \mathcal{P}$,

C2. There exist two positive constants $c_1, c_2 > 0$ such that

$$c_1 \|\hat{\mathcal{B}} - \mathcal{B}_{\text{true}}\|_F^2 \leq \ell(\hat{\mathcal{B}}) - \ell(\mathcal{B}_{\text{true}}) \leq c_2 \|\hat{\mathcal{B}} - \mathcal{B}_{\text{true}}\|_F^2.$$

To prove C1, note that

$$\begin{aligned} \mathcal{L}_{\mathcal{Y}}(\mathcal{B}) - \ell(\mathcal{B}) &= \langle \mathcal{Y} - \mathbb{E}(\mathcal{Y}|\mathcal{X}),\ \Theta(\mathcal{B}) \rangle \\ &= \langle \mathcal{Y} - b'(\Theta_{\text{true}}),\ \Theta \rangle \\ &= \langle \mathcal{E} \times_1 \boldsymbol{X}_1^T \times_2 \cdots \times_K \boldsymbol{X}_K^T,\ \mathcal{B} \rangle, \end{aligned}$$

where $\mathcal{E} = [\![\varepsilon_{i_1,\ldots,i_K}]\!] \stackrel{\text{def}}{=} \mathcal{Y} - b'(\Theta_{\text{true}})$. Based on Assumption A1, $\mathcal{E}$ is a sub-Gaussian tensor with parameter bounded by $C_1 = \phi U$. Therefore, $\check{\mathcal{E}} \stackrel{\text{def}}{=} \mathcal{E} \times_1 \boldsymbol{X}_1^T \times_2 \cdots \times_K \boldsymbol{X}_K^T$ is a $(p_1, \ldots, p_K)$-dimensional sub-Gaussian with parameter bounded by $C_2 = \phi U c_2^{2K}$. By Cauchy-Schwarz inequality,

$$|\mathcal{L}_{\mathcal{Y}}(\mathcal{B}) - \ell(\mathcal{B})| \leq \left\|\check{\mathcal{E}}\right\|_2 \|\mathcal{B}\|_* .$$

where $\|\cdot\|_2$ denotes the tensor spectral norm and $\|\cdot\|_*$ denotes the tensor nuclear norm.

We have that $\|\mathcal{B}\|_* \leq \frac{\prod_k r_k}{\max_k r_k} \|\mathcal{B}\|_F$ by [?, ?]. Moreover, the Gaussian tensor theory [?] shows that $\left\|\check{\mathcal{E}}\right\|_2 \leq C_1 \sum_k p_k$ with probability at least $1 - \exp(-C_2 \sum_k p_k)$

To prove C2, we note that

$$\ell(\mathcal{B}) = \ell(\mathcal{B}_{\text{true}}) - \frac{1}{2} \text{vec}(\mathcal{B} - \mathcal{B}_{\text{true}})^T \mathbb{E}(\mathcal{H}_{\mathcal{Y}}(\check{\mathcal{B}})) \text{vec}(\mathcal{B} - \mathcal{B}_{\text{true}}), \tag{1}$$

where $\mathcal{H}_{\mathcal{Y}}(\check{\mathcal{B}})$ is the Hession of $\frac{\partial \ell^2(\mathcal{B})}{\partial^2 \mathcal{B}}$ evaluated at $\check{\mathcal{B}} = \alpha \text{vec}(\alpha \mathcal{B} + (1-\alpha)\mathcal{B}_{\text{true}})$ for some $\alpha \in [0, 1]$. Recall that $b''(\theta) = \text{Var}(y|\theta)$ if $y \in \mathbb{R}$ follows the exponential family distribution with function $b(\cdot)$. Therefore, the equation (1) can be written as

$$\ell(\mathcal{B}) - \ell(\mathcal{B}_{\text{true}}) = -\frac{1}{2} \sum_{i_1,\ldots,i_K} b''(\check{\theta}_{i_1,\ldots,i_K})(\theta_{i_1,\ldots,i_K} - \theta_{\text{true},i_1,\ldots,i_K})^2 \leq -\frac{L}{2} \|\Theta - \Theta_{\text{true}}\|_F^2,$$

holds for all $\mathcal{B} \in \mathcal{P}$, provided that $\min_{|\theta|\le\alpha} |b''(\theta)| \ge L > 0$.

Now we consider the constrained MLE $\hat{\mathcal{B}}$. By definition, $\mathcal{L}_{\mathcal{Y}}(\hat{\mathcal{B}}) - \mathcal{L}_{\mathcal{Y}}(\mathcal{B}_{\text{true}}) \ge 0$. This implies that

$$
\begin{aligned}
0 &\le \mathcal{L}_{\mathcal{Y}}(\hat{\mathcal{B}}) - \mathcal{L}_{\mathcal{Y}}(\mathcal{B}_{\text{true}}) \\
&\le \left(\mathcal{L}_{\mathcal{Y}}(\hat{\mathcal{B}}) - \ell(\hat{\mathcal{B}})\right) - \left(\mathcal{L}_{\mathcal{Y}}(\mathcal{B}_{\text{true}}) - \ell(\mathcal{B}_{\text{true}})\right) + \left(\ell(\hat{\mathcal{B}}) - \ell(\mathcal{B}_{\text{true}})\right) \\
&\le 2 \sup_{\mathcal{B}\in\mathcal{P}} |\mathcal{Y}| - \frac{L}{2}\|\hat{\Theta} - \Theta_{\text{true}}\|_F^2 \\
&\le 2 \sup_{\mathcal{B}} |\mathcal{L}_{\mathcal{Y}}(\mathcal{B}) - \ell(\mathcal{B})| - \frac{L}{2}\|\hat{\Theta} - \Theta_{\text{true}}\|_F^2
\end{aligned}
$$

Therefore, the statement

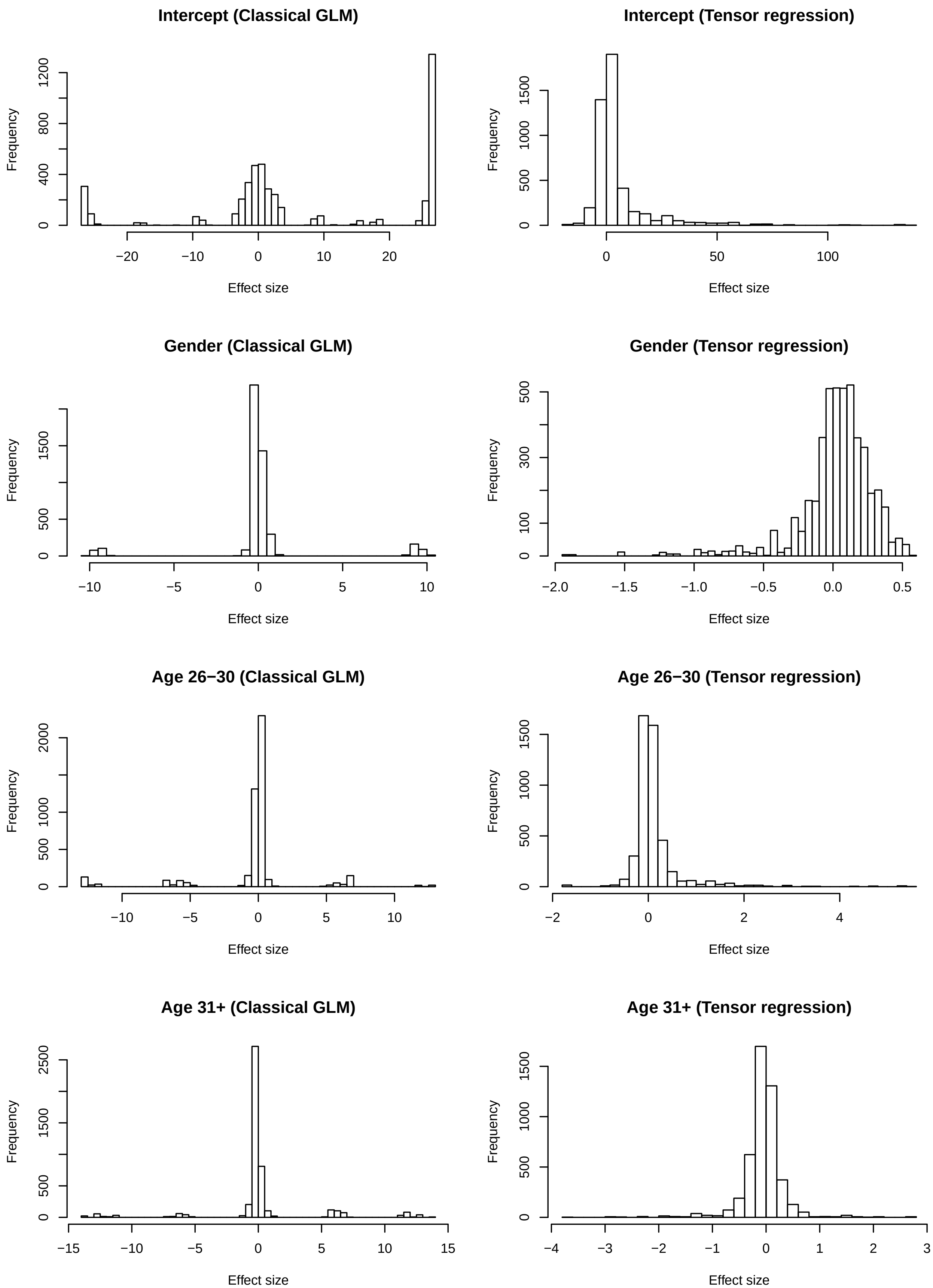
$$
\begin{aligned}
\|\hat{\Theta} - \Theta_{\text{true}}\|_F &\le \frac{2}{L}\left\langle \mathcal{E},\ \frac{\hat{\Theta} - \Theta_{\text{true}}}{\|\hat{\Theta} - \Theta_{\text{true}}\|_F} \right\rangle \qquad (2) \\
&\le \frac{2}{L} \sup_{\Theta:\|\Theta\|_F=1, \Theta=\mathcal{B}\times_1 \boldsymbol{X}_1 \times_2 \cdots \times_K \boldsymbol{X}_K} \langle \mathcal{E},\ \Theta\rangle \\
&\le \frac{2}{L} \sup_{\mathcal{B}\in\mathcal{P}:\|\mathcal{B}\|_F \le \prod_k \sigma_{\min}^{-1}(\boldsymbol{X}_k)} \langle \mathcal{E} \times_1 \boldsymbol{X}_1^T \times_2 \cdots \times_K \boldsymbol{X}_K^T,\ \mathcal{B}\rangle.
\end{aligned}
$$

Combining (2) with C1 yields the final conclusion.

□

2 Additional results for real data analysis

1. commonbloc0, blockpositionindex
2. officialvisits, violentactions, militaryactions, duration, negativebehavior, boycottembargo, aidenemy, negativecomm, accusation, protestsunoffialacts, nonviolentbehavior, emigrants,r elexports, timesincewar, commonbloc2, rintergovorgs3, relintergovorgs, intergovorgs
3. economicaid, releconomicaid, conferences, booktranslations, relbooktranslations, severdiplomatic, expeldiplomats, attackembassy, unweightedunvote, tourism, reltourism, tourism3, relemigrants, emigrants3, students, relstudents, exports, exports3, lostterritory, dependent, militaryalliance, warning
4. treaties, reltreaties, exportbooks, relexportbooks, weightedunvote, ngo, relngo, ngoorgs3, embassy, reldiplomacy, timesinceally, independence, commonbloc1


Intercept (Classical GLM)
Frequency
0
400
800
1200
−20
−10
0
10
20
Effect size
Intercept (Tensor regression)
Frequency
0
500
1000
1500
0
50
100
Effect size
Gender (Classical GLM)
Frequency
0
500
1500
−10
−5
0
5
10
Effect size
Gender (Tensor regression)
Frequency
0
100
300
500
−2.0
−1.5
−1.0
−0.5
0.0
0.5
Effect size
Age 26−30 (Classical GLM)
Frequency
0
500
1000
2000
−10
−5
0
5
10
Effect size
Age 26−30 (Tensor regression)
Frequency
0
500
1000
1500
−2
0
2
4
Effect size
Age 31+ (Classical GLM)
Frequency
0
500
1500
2500
−15
−10
−5
0
5
10
15
Effect size
Age 31+ (Tensor regression)
Frequency
0
500
1000
1500
−4
−3
−2
−1
0
1
2
3
Effect size