Thought about STD

Jiaxin Hu

May 19, 2021

1 Equivalent model of STD

1.1 Claim 1

Claim 1: the new noise tensor consists of i.i.d. (sub-)Gaussian entries if and only if all $\boldsymbol{X}_k, k \in [3]$ has orthogonal columns and the diagonal elements of $(\boldsymbol{X}_k^T \boldsymbol{X}_k)^{-1}$ keep the same for $k \in [3]$.

Proof. Let $\mathcal{E}' = \mathcal{E} \times_1 \boldsymbol{M}_1 \times_2 \boldsymbol{M}_2 \times_3 \boldsymbol{M}_3$, where $\mathcal{E}$ has independent sub-Gaussian-σ entries. By the Prop 1 in the supplementary of STD paper, $\mathcal{E}'$ is a sub-Gaussian tensor with a different parameter σ', where σ' depends on $\boldsymbol{M}_k$. Now, consider the case that $\mathcal{E}$ has i.i.d. Gaussian entries with variance 1. We only need to show that $\operatorname{cov}(\mathcal{E}'_{i,j,k}, \mathcal{E}'_{i',j',k'}) = 0$ and $\operatorname{cov}(\mathcal{E}'_{i,j,k}, \mathcal{E}'_{i,j,k}) = \sigma^2$, for arbitrary (i,j,k), $(i',j',k') \neq (i,j,k)$, and for a constant σ^2.

Note that for arbitrary (i,j,k) we have

$$\mathcal{E}'_{i,j,k} = \sum_{a\in[d_1], b\in[d_2], c\in[d_3]} \mathcal{E}_{abc} \boldsymbol{M}_{1,ia} \boldsymbol{M}_{2,jb} \boldsymbol{M}_{3,kc}.$$

Then, the covariance is equal to

$$\begin{aligned}
\operatorname{cov}(\mathcal{E}'_{i,j,k}, \mathcal{E}'_{i',j',k'}) &= \sum_{a\in[d_1], b\in[d_2], c\in[d_3]} \boldsymbol{M}_{1,ia} \boldsymbol{M}_{2,jb} \boldsymbol{M}_{3,kc} \boldsymbol{M}_{1,i'a} \boldsymbol{M}_{2,j'b} \boldsymbol{M}_{3,k'c} \\
&= \sum_{b\in[d_2], c\in[d_3]} \boldsymbol{e}_{1,i}^T \boldsymbol{M}_1 \boldsymbol{M}_1^T \boldsymbol{e}_{1,i'} \boldsymbol{M}_{2,jb} \boldsymbol{M}_{3,kc} \boldsymbol{M}_{2,j'b} \boldsymbol{M}_{3,k'c} \\
&= [\boldsymbol{e}_{1,i}^T (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} \boldsymbol{e}_{1,i'}][\boldsymbol{e}_{2,j}^T (\boldsymbol{X}_2^T \boldsymbol{X}_2)^{-1} \boldsymbol{e}_{2,j'}][\boldsymbol{e}_{3,k}^T (\boldsymbol{X}_3^T \boldsymbol{X}_3)^{-1} \boldsymbol{e}_{3,k'}], \qquad (1)
\end{aligned}$$

where $\boldsymbol{e}_{k,i} \in \mathbb{R}^{p_k}$ whose i-th elements is 1 and the other entries remain 0, for $k \in [3], i \in [p_k]$. The equation (1) implies that the covariance equal to 0 for all pairs $(i',j',k') \neq (i,j,k)$ if and only if all $\boldsymbol{X}_k, k \in [3]$ should have orthogonal columns.

Next, note the variance of $\mathcal{E}_{i,j,k}$ is

$$\begin{aligned}
\operatorname{cov}(\mathcal{E}'_{i,j,k}, \mathcal{E}'_{i,j,k}) &= [\boldsymbol{e}_{1,i}^T (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} \boldsymbol{e}_{1,i}][\boldsymbol{e}_{2,j}^T (\boldsymbol{X}_2^T \boldsymbol{X}_2)^{-1} \boldsymbol{e}_{2,j}][\boldsymbol{e}_{3,k}^T (\boldsymbol{X}_3^T \boldsymbol{X}_3)^{-1} \boldsymbol{e}_{3,k}] \\
&= (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1}_{ii} (\boldsymbol{X}_2^T \boldsymbol{X}_2)^{-1}_{jj} (\boldsymbol{X}_3^T \boldsymbol{X}_3)^{-1}_{kk}. \qquad (2)
\end{aligned}$$

The equation (2) implies that all entries of $\mathcal{E}'$ have equal variance if and only if the diagonal elements of $(\boldsymbol{X}_k^T \boldsymbol{X}_k)^{-1}$ should be the same for $k \in [3]$.

□

Alternative Way: Using Tensor Covariance

Proof of the Claim 1: Let $\mathcal{E}' = \mathcal{E} \times \boldsymbol{M}_1 \times_2 \boldsymbol{M}_2 \times_3 \boldsymbol{M}_3$. We only need to find the sufficient and necessary condition such that $\mathrm{Cov}(\mathrm{vec}(\mathcal{E}'))$ is a diagonal matrix. By the Prop 2.1 of Hoff et al. (2011), the covariance matrix for $\mathcal{E}'$ is

$$\mathrm{Cov}(\mathrm{vec}(\mathcal{E}')) = \boldsymbol{M}_3\boldsymbol{M}_3^T \otimes \boldsymbol{M}_2\boldsymbol{M}_2^T \otimes \boldsymbol{M}_2\boldsymbol{M}_1^T = (\boldsymbol{X}_3^T\boldsymbol{X}_3)^T \otimes (\boldsymbol{X}_2^T\boldsymbol{X}_2)^T \otimes (\boldsymbol{X}_1^T\boldsymbol{X}_1)^T,$$

which implies that the covariance matrix is a diagonal matrix if and only if all $\boldsymbol{X}_k$s have orthogonal columns. □

Remark 1. The kronecker product $\otimes$ is not exchangeable. The order of $\boldsymbol{M}_k\boldsymbol{M}_k^T$ depends on the order of vectorization. For example, the above vectorization order is *mode* 1 $\rightarrow$ *mode* 2 $\rightarrow$ *mode* 3.

Remark 2. The "separable" covariance structure of a multiway array $\mathcal{Y} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is defined as

$$\mathrm{Cov}(\mathcal{Y}) = \Sigma_1 \circ \cdots \circ \Sigma_K, \quad \mathrm{Cov}(\mathrm{vec}(\mathcal{Y})) = \Sigma_K \otimes \cdots \otimes \Sigma_1,$$

where $\circ$ denotes the outer product and the covariance $\mathrm{Cov}(\mathcal{Y}) \in \mathbb{R}^{d_1 \times d_1 \times d_2 \times d_2 \times \cdots \times d_K \times d_K}$.

The "separable" structure is a special structure of tensor covariance, which substantially reduces the number of parameters for the covariance from $d_1^2 \cdots d_K^2$ to $\sum_k d_k^2$. Besides, the "separable" structure is closely related to Tucker product, and a family of Gaussian tensor with separable covariance structure can be generated via the tucker production of an i.i.d. Gaussian tensor and several matrices. See Hoff et al. (2011), and the Section 7 of our STD paper.

1.2 Claim 2:

Claim 2: Consider the QR decomposition of the feature matrix $\boldsymbol{X}_k = \boldsymbol{Q}_k\boldsymbol{R}_k, k \in [K]$, where $\boldsymbol{Q}_k$ has orthogonal columns. Then the new noise tensor $\mathcal{E} \times \{\boldsymbol{Q}_1^T, ..., \boldsymbol{Q}_K^T\}$ consists of i.i.d. Gaussian entries.

Proof. Note that

$$\mathcal{B} \times \{\boldsymbol{X}_1, ..., \boldsymbol{X}_K\} = (\mathcal{B} \times \{\boldsymbol{R}_1, ..., \boldsymbol{R}_K\}) \times \{\boldsymbol{Q}_1, ..., \boldsymbol{Q}_K\}.$$

Let $\boldsymbol{M}_k = (\boldsymbol{Q}_k^T\boldsymbol{Q}_k)^{-1}\boldsymbol{Q}_k^T = \boldsymbol{Q}_k^T$. By Claim 1, we know that the new noise tensor $\mathcal{E}' = \mathcal{E} \times \{\boldsymbol{M}_1, ..., \boldsymbol{M}_K\} = \mathcal{E} \times \{\boldsymbol{Q}_1^T, ..., \boldsymbol{Q}_K^T\}$ has i.i.d. Gaussian entries. □

References

Hoff, P. D. et al. (2011). Separable covariance arrays via the tucker product, with applications to multivariate relational data. *Bayesian Analysis*, 6(2):179–196.