Nonparametric Tensor Model via Hypergraphons

Miaoyan Wang, Dec 12, 2020

1 Set-up

Let $\mathcal{Y} = [\![\mathcal{Y}(\boldsymbol{i})]\!] \in \mathbb{R}^{d_1\times\cdots\times d_K}$ be an order-K $(d_1,\ldots,d_K)$-dimensional data tensor, where $\boldsymbol{i} = (i_1,\ldots,i_K)$ is the K-way index. We propose a two-stage generative process for the tensor observation. First, we draw a collection of i.i.d. random variables, $x_k(i) \in \mathcal{X}_k$, from some probability measure $(\mathcal{X}_k, \mu_{\mathcal{X}_k})$, for all $i \in [d_k]$, $k \in [K]$. The Cartesian product of the random variables, denoted $(x_1(i_1),\ldots,x_K(i_K))$, represents the latent features at position $\boldsymbol{i} = (i_1,\ldots,i_K)$ of the tensor. Second, conditional on the latent features, the tensor entries are drawn independently with mean $f((x_1(i_1),\ldots,x_K(i_K))$.

- Nonparametric mean: there exists a unknown function $f\colon \mathcal{X}_1\times\cdots\times\mathcal{X}_K \mapsto [0,1]$ such that

$$\mathbb{E}\mathcal{Y}(\boldsymbol{i}) = f(x_1(i_1),\ldots,x_K(i_K)), \quad \text{for all } \boldsymbol{i} = (i_1,\ldots,i_K) \in [d_1]\times\cdots\times[d_K]. \tag{1}$$

  Here $x_k(i_k) \in \mathcal{X}_k$ denotes the latent feature associated with the i_k-th entry along the k-th mode of the tensor, where $k \in [K]$.

- Latent features: for each mode k, the latent features $\{x_k(i_k)\colon i_k \in [d_k]\}$ are sampled independently (identical??) from the probability measure $(\mathcal{X}_k, \mu_{\mathcal{X}_k})$.

- Conditionally independence: conditional on the latent features $x_k(i_k)$, the tensor entries $\mathcal{Y}(\boldsymbol{i})$ are independent, sub-Gaussian random variables; i.e. $\mathbb{E}\exp[t(\mathcal{Y}(\boldsymbol{i}) - \mathbb{E}(\mathcal{Y}(\boldsymbol{i}))] \leq \exp(t^2\sigma^2/2)$ for all $\boldsymbol{i} \in [d_1]\times\cdots\times[d_K]$ and $t \in \mathbb{R}$.

- For simplicity, we set $\mathcal{X}_k = [0,1]$ and $\mu_{\mathcal{X}_k}$ the Lebesgue measure over $[0,1]$ for all $k \in [K]$. Furthermore, we assume the latent features are mutually independent across K modes.

- Regularity conditions on f. Two possible options: 1. stepwise functions, 2. Holder smooth functions.

We call the model (1) the nonparametric tensor model because the function f is unknown and to be estimated.

In the case of binary tensor observations, our nonparametric model is closely connected to the hypergraphon model in the graphical literature. Specifically, let $\mathcal{G} = (V,E)$ be a K-uniform hypergraph, where $V = [d]$ is the node set and $E \subset V^{\otimes K}$ is the hyperedge set with each hyperedge connecting precisely K nodes, $K \leq d$. The hypergraphon model assumes that the hyperedges are generated through a symmetric, measurable function $f\colon [0,1]^K \to [0,1]$,

$$\mathbb{1}\{\boldsymbol{i} \in E\} \sim \text{Bernoulli}(f(x(i_1),\ldots,x(i_K)), \quad \text{for all } \boldsymbol{i} \in [d]^K,$$

where $\{x(i)\colon i \in [d]\}$ is an i.i.d. random sample from $U[0,1]$, and the events $\mathbb{1}\{\boldsymbol{i} \in E\}$ are mutually independent conditional on $\{x(i)\}$. The function f is referred to as the K-uniform hypergraphon. Our nonparametric tensor model (1) generalizes the hypergraphon model by allowing more flexible observations with mode-specific latent features?? and asymmetric latent function f. For this reason, we adopt the terminology and call the function f a hypergraphon.

We use $(\mathcal{X}_k, \mu_{\mathcal{X}_k}, f)$ to denote the sampling scheme for the latent features and the hypergraphon associated with our nonparametric tensor model (1). By specializing the latent features in $\mathcal{X}_k$ and the function f, the conditional mean model (1) incorporates several common previously-studied tensor models as special cases.

2 Examples

Low-rank model. Let $\mathcal{X}_k \subset \mathbb{R}^{r_k}$ be a bounded close set, and $\mu_{\mathcal{X}_k}$ a probability measure over $\mathcal{X}_k$. Consider a multilinear hypergraphon

$$\begin{aligned} f\colon \mathcal{X}_1 \times \cdots \times \mathcal{X}_K &\to \mathbb{R} \\ (x_1, \ldots, x_K) &\mapsto \mathcal{C} \times_1 x_1^T \times_2 \cdots \times_K x_K^T, \end{aligned} \tag{2}$$

where $\mathcal{C} \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ is a fixed coefficient tensor. Let $x_k(i_k) \in \mathcal{X}_k$ be the realization of the mode-k latent feature at index $i_k \in [d_k]$, and $\boldsymbol{X}_k = [x_k(1)|\ldots|x_k(d_k)] \in \mathbb{R}^{r_k \times d_k}$ the corresponding feature matrix. Then, model (1) induces a rank-$(r_1, \ldots, r_K)$ Tucker model:

$$\mathbb{E}\left(\mathcal{Y}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K\right) = \mathcal{C} \times_1 \boldsymbol{X}_1^T \times_2 \cdots \times_K \boldsymbol{X}_K^T.$$

Similarly, our model incorporates the CP tensor model by setting $r_1 = \cdots = r_K = r$ and a super-diagonal core tensor $\mathcal{C}$.

Nonlinear single-index model. Consider the same setting as in Example 1. Let $f' = g \circ f$, where f is defined as in (2), $g\colon \mathbb{R} \to \mathbb{R}$ is a nonlinear monotonic function, and $\circ$ denotes the function composition. Then the model (1) induces a nonlinear single-index model:

$$\mathbb{E}(\mathcal{Y}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K) = g(\mathcal{C} \times_1 \boldsymbol{X}_1^T \times_2 \cdots \times_K \boldsymbol{X}_K^T).$$

Here the function g could be either parametric such as a logistic function as in Bradly-Terry model, or nonparametric such as a monotonic, Lipschitz function as in [**?**]. Note that, with the nonlinear transformation, the data tensor is likely to have full rank in expectation.

Stochastic transitivity model. Let $\mathcal{X}_k \subset \mathbb{R}$ be a bounded close set, and $\mu_{\mathcal{X}_k}$ a probability measure over $\mathcal{X}_k$. Consider a monotonic hypergraphon $f\colon \mathbb{R}^K \to \mathbb{R}$ in that

$$f(x_1,\ldots,x_K) \le f(x_1',\ldots,x_K')$$

whenever $x_k \le x_k'$ for all $k \in [K]$. Then, model (1) reduces to the strong stochastic transitivity model; i.e., there exist a set of permutations $\sigma_k\colon [d_k] \to [d_k]$ such that the entries are monotonically increasing along the permuted indices:

$$\mathbb{E}\mathcal{Y}(\sigma_1(i_1),\ldots,\sigma_K(i_K)) \le \mathbb{E}\mathcal{Y}(\sigma_1(i_1'),\ldots,\sigma_K(i_K')), \tag{3}$$

whenever $\sigma_k(i_k) \le \sigma_k(i_k')$ for all $k \in [K]$. The strong stochastic transitivity (3) is also known as rank-1 permutation model; it was initially proposed for the matrix case $K = 2$. Our formulation extends the model to higher-order cases. More generally, by setting f as a mixture of shape-constrained functions over multivariate latent features $\mathcal{X}_k = \mathbb{R}^r$, our model encompasses the more general low permutation-rank models and statistical seriation models.

Stochastic block model. Let $\mathcal{X}_k = [0,1]$ and $\mu_{\mathcal{X}_k}$ the Lebesque measure over $[0,1)$. For each k, write $\mathcal{X}_k = [0,1/r_k) \cup [2/r_k, 3/r_k) \cup \cdots \cup [(r_k-1)/r_k, 1)$ as a disjoint union of r_k equal-sized intervals. Define a piecewise constant hypergraphon

$$\begin{aligned} f\colon [0,1]^K &\to \mathbb{R} \\ (x_1,\ldots,x_K) &\mapsto \sum_{j_1,\ldots,j_K} c_{j_1,\ldots,j_K} \mathbb{1}\left\{x_k \in \left[\frac{j_k-1}{r_k},\ \frac{j_k}{r_k}\right),\ \text{for all } k \in [K]\right\}, \end{aligned}$$

where $\mathcal{C} = [\![c_{j_1,\ldots,j_K}]\!] \in \mathbb{R}^{r_1\times\cdots\times r_k}$ is a fixed tensor specifying the block means. Let $x_k(i_k) \in \mathcal{X}_k$ be the realization of the mode-k latent feature at index $i_k \in [d_k]$. Then model (1) reduces to a stochastic block model,

$$\mathbb{E}\mathcal{Y}(\boldsymbol{i})|\{x_k(i_k)\} = c_{j_1,\ldots,j_K},$$

where $j_k \in [r_k]$ is the mode-k block index for which $(j_k-1)/r_k \le x_k(i_k) \le j_k/r_k$, $k \in [K]$.