Hypergraphon estimation error 2

Chanwoo Lee
March 1, 2021

1 Notation and problem setting

Let E be a set of possible m-uniform hyperedges from n vertices without diagonal entries,

$$E = \{(i_1, \ldots, i_m) \in [n]^m \colon |\{i_1, \ldots, i_m\}| = m\}.$$

We denote an index of m-uniform hyperedges as $\omega = (\omega_1, \ldots, \omega_m) \in [n]^m$ and a membership vector of m-vertices as $a = (a_1, \ldots, a_m) \in [k]^m$. Let $z \colon [n] \to [k]$ be a membership function. For a given membership function z and a membership vector $a \in [k]^m$, define $E_{Z^{-1}(a)}$ as a set of m-uniform hyperedges whose clustering group belongs to a i.e.,

$$E_{z^{-1}(a)} = \{\omega \in E \colon z(\omega_\ell) = a_\ell \text{ for all } \ell \in [m]\}.$$

We define a block average on a set $E_{z^{-1}(a)}$ for a given membership function z, a membership vector a, and a tensor $\Theta \in ([n])^{\otimes m}$ as

$$\bar{\Theta}_a(z) = \frac{1}{|E_{z^{-1}(a)}|} \sum_{\omega \in E_{z^{-1}(a)}} \Theta_\omega.$$

Now we consider an undirected m-uniform hypergraph. The connectivity is encoded by an adjacency tensor $\{\mathcal{A}_\omega\}_{\omega \in E}$ which takes values in $\{0, 1\}$. We assume that $\mathcal{A}_\omega \sim$ Bernoulli(Θ_ω), where

$$\Theta_\omega = f(\xi_{\omega_1}, \ldots, \xi_{\omega_m}), \text{ for all } \omega = (\omega_1, \ldots, \omega_m) \in E, \tag{1}$$

where $f \colon [0,1]^m \to [0,1]$ is a symmetric function called graphon such that $f(\xi_{\omega_1}, \ldots, \xi_{\omega_m}) = f(\xi_{\sigma(\omega_1)}, \ldots, \xi_{\sigma(\omega_m)})$ for all permutation $\sigma \colon [n] \to [n]$. Conventionally, we set $\Theta_\omega = 0$ for all $\omega \in [n]^m \setminus E$. In addition, we further assume that a graphon f is α-Hölder continuous with a constant L.

Definition 1. For $\alpha \in (0, 1]$ and $L > 0$, a function $f \colon [0,1]^m \to [0,1]$ is a α-Hölder continuous with a constant L, denoted as $f \in \mathcal{H}(\alpha, L)$, if

$$|f(x) - f(y)| \leq L\|x - y\|_\alpha^\alpha \quad \text{for all } x, y \in [0,1]^m, \tag{2}$$

where the norm $\|x\|_p^p := \sum_{i=1}^m |x_i|^p$ for $x \in \mathbb{R}^m$.

1-norm vs p norm.
Which is more convenient for our case?

2 Probability matrix estimation

2.1 Under hypergraphon model

Lemma 1 (Block approximation). Suppose the true parameter Θ admits the form (1) with $f \in \mathcal{H}(\alpha, L)$. For every integer $k \leq n$, there exists $z^* \colon [n] \to [k]$, satisfying

$$\frac{1}{|E|} \sum_{a \in [k]^m} \sum_{\omega \in E_{(z^*)^{-1}(a)}} (\Theta_\omega - \bar{\Theta}_a(z^*))^2 \leq CL^2 \left(\frac{m^2}{k^2}\right)^\alpha.$$

Proof. Define $z^* \colon [n] \to [k]$ by

$$(z^*)^{-1}(\ell) = \left\{ i \in [n] \colon \xi_i \in \left[\frac{\ell - 1}{k}, \frac{\ell}{k}\right) \right\}, \quad \text{for each } \ell \in [k].$$

By the construction of z^* for $\xi_{\omega_\ell} \in [(a_{\ell-1}-1)/k, a_\ell/k]$,

$$\begin{aligned}
|f(\xi_{\omega_1},\dots,\xi_{\omega_m}) - \bar{\Theta}_a(z^*)| &= \left| f(\xi_{\omega_1},\dots,\xi_{\omega_m}) - \frac{1}{|E_{z^{-1}(a)}|} \sum_{(\omega_1',\dots,\omega_m')\in E_{z^{-1}(a)}} f(\xi_{\omega_1'},\dots,\xi_{\omega_m'}) \right| \\
&\le \frac{1}{|E_{z^{-1}(a)}|} \sum_{(\omega_1',\dots,\omega_m')\in E_{z^{-1}(a)}} |f(\xi_{\omega_1},\dots,\xi_{\omega_m}) - f(\xi_{\omega_1'},\dots,\xi_{\omega_m'})| \\
&\le \frac{1}{|E_{z^{-1}(a)}|} \sum_{(\omega_1,\dots,\omega_m)\in E_{z^{-1}(a)}} L\|(\xi_{\omega_1},\dots,\xi_{\omega_m}) - (\xi_{\omega_1'},\dots,\xi_{\omega_m'})\|_\alpha^\alpha \\
&\le CL\left(\frac{m}{k}\right)^\alpha.
\end{aligned}$$

□

Let $\tilde{\Theta}$ be a minimizer of the least square error from the adjacency tensor $\mathcal{A}$,

$$\tilde{\Theta} = \operatorname*{arg\,min}_{\Theta\in\mathcal{P}_k} \sum_{\omega\in E} (\mathcal{A}_\omega - \Theta_\omega)^2,$$

where

$$\mathcal{P}_k = \{\Theta \in ([0,1]^n)^{\otimes m} \colon \Theta = \mathcal{C} \times_2 \boldsymbol{M} \times_2 \cdots \times_m \boldsymbol{M}, \text{ with a membership matrix } \boldsymbol{M} \text{ and a core tensor } \mathcal{C} \in ([0,1]^k)^{\otimes m}\}.$$

We estimate the probability tensor by $\hat{\Theta} = \mathrm{cut}(\tilde{\Theta})$ such that

$$\mathrm{cut}(\Theta_\omega) = \begin{cases} \Theta_\omega & \text{if } \omega \in E, \\ 0 & \text{if } \omega \in [n]^m \setminus E. \end{cases} \tag{3}$$

Notice $\|\mathcal{A} - \hat{\Theta}\|_F^2 \le \|\mathcal{A} - \Theta\|_F^2$ for any k block tensor $\Theta \in \mathrm{cut}(\mathcal{P}_k)$.

Theorem 2.1 (Hypergraphon model)**.** Suppose the true parameter Θ admits the form (1) with $f \in \mathcal{H}(\alpha, L)$ Let $\hat{\Theta}$ be the estimator from (3). Then, there exist two constants $C_1, C_2 > 0$ such that,

$$\frac{1}{n^m}\|\hat{\Theta} - \Theta\|_F^2 \le C_1\left(n^{\frac{-2m\alpha}{m+2\alpha}} + \frac{m \log n}{n^{m-1}}\right),$$

with probability at least $1 - \exp\left(-C_2\left(n\log n + n^{\frac{m^2}{m+2\alpha}}\right)\right)$ uniformly over $f \in \mathcal{H}(\alpha, L)$.

Remark 1. There are two ingredients in the rate of convergence, the nonparametric rate $n^{-2m\alpha/(m+2\alpha)}$ and the clustering rate $\log n/n^{m-1}$. Depending on constants m and α, convergence rate becomes

$$n^{\frac{-2m\alpha}{m+2\alpha}} + \frac{\log n}{n^{m-1}} \asymp \begin{cases} n^{\frac{-2\alpha}{1+\alpha}} & m = 2, \alpha \in (0,1), \\ \log n/n & m = 2, \alpha = 1, \\ n^{\frac{-2m\alpha}{m+2\alpha}} & m > 2. \end{cases}$$

m=1?

Notice that the nonparametric rate dominates the clustering rate when $m > 2$. The intuitive explanation for this phenomenon is that the number of parameters increases exponentially due to the possible m-combinations from k clusters (which contributes to the nonparametric rate) while the possible number of k-cluster of n-vertices remains the same (which contributes to the clustering rate) when m increases under the choice of $k = \lceil n^{\frac{m}{m+2\alpha}} \rceil$. That means that the hypergraphon becomes much harder to estimate while we have more information about clusters because there are large blocks of the tensors when $m > 2$. When $m = 2$, Theorem 2.1 reduces to the results in Gao et al. [2015].

more block means to estimate?

Proof. First, we can find a block tensor Θ^* close to true Θ by Lemma 1. By triangular inequality,

$$\|\hat{\Theta}-\Theta\|_F^2 \leq 2\underbrace{\|\hat{\Theta}-\Theta^*\|_F^2}_{\text{(i)}} + 2\underbrace{\|\Theta^*-\Theta\|_F^2}_{\text{(ii)}}. \tag{4}$$

Since we have already shown the error bound of (ii) in Lemma 1, we bound the error from (i). From the definition of $\hat{\Theta}$, we have

$$\|\hat{\Theta}-\mathcal{A}\|_F^2 \leq \|\Theta^*-\mathcal{A}\|_F^2. \tag{5}$$

Combining (5) with the fact

$$\begin{aligned}\|\hat{\Theta}-\mathcal{A}\|_F^2 &= \|\hat{\Theta}-\Theta^*+\Theta^*-\mathcal{A}\|_F^2\\ &= \|\hat{\Theta}-\Theta^*\|_F^2 + \|\Theta^*-\mathcal{A}\|_F + 2\langle\hat{\Theta}-\Theta^*, \Theta^*-\mathcal{A}\rangle,\end{aligned}$$

yields

$$\begin{aligned}\|\hat{\Theta}-\Theta^*\|_F^2 &\leq 2\langle\hat{\Theta}-\Theta^*, \mathcal{A}-\Theta^*\rangle\\ &= 2\left(\langle\hat{\Theta}-\Theta^*, \mathcal{A}-\Theta\rangle + \langle\hat{\Theta}-\Theta^*, \Theta-\Theta^*\rangle\right)\\ &\leq 2\|\hat{\Theta}-\Theta^*\|_F\left(\left\langle\frac{\hat{\Theta}-\Theta^*}{\|\hat{\Theta}-\Theta^*\|_F}, \mathcal{A}-\Theta\right\rangle + \|\Theta-\Theta^*\|_F\right).\end{aligned}$$

Let $\mathcal{M} = \{\boldsymbol{M}\colon \boldsymbol{M}$ is the collection of membership matrices$\}$. Then,

$$\begin{aligned}\left\langle\frac{\hat{\Theta}-\Theta^*}{\|\hat{\Theta}-\Theta^*\|_F}, \mathcal{A}-\Theta\right\rangle &\leq \sup_{\Theta'\in\mathcal{P}_k}\sup_{\Theta''\in\mathcal{P}_k}\left\langle\frac{\text{cut}(\Theta')-\text{cut}(\Theta'')}{\|\text{cut}(\Theta')-\text{cut}(\Theta'')\|_F}, \mathcal{A}-\Theta\right\rangle\\ &\leq \sup_{\boldsymbol{M},\boldsymbol{M}'\in\mathcal{M}}\sup_{\mathcal{C},\mathcal{C}'\in([0,1]^k)^{\otimes m}}\left\langle\frac{\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))-\text{cut}(\Theta(\boldsymbol{M}',\mathcal{C}'))}{\|\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))-\text{cut}(\Theta(\boldsymbol{M}',\mathcal{C}'))\|_F}, \mathcal{A}-\Theta\right\rangle.\end{aligned}$$

Notice that $\mathcal{A}-\Theta$ is sub-Gaussian with proxy parameter $\sigma^2 = 1/4$. By union bound and the property of sub-Gaussian, we have, for any $t>0$.

$$\begin{aligned}&\mathbb{P}\left(\sup_{\boldsymbol{M},\boldsymbol{M}'\in\mathcal{M}}\sup_{\mathcal{C},\mathcal{C}'\in([0,1]^k)^{\otimes m}}\left|\left\langle\frac{\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))-\text{cut}(\Theta(\boldsymbol{M}',\mathcal{C}'))}{\|\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))-\text{cut}(\Theta(\boldsymbol{M}',\mathcal{C}'))\|_F}, \mathcal{A}-\Theta\right\rangle\right| \geq t\right)\\ &\leq \mathbb{P}\left(\sup_{\boldsymbol{M},\boldsymbol{M}'\in\mathcal{M}}\sup_{\mathcal{C},\mathcal{C}'\in([0,1]^k)^{\otimes m}}\left|\left\langle\frac{\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}-\mathcal{C}'))+\text{cut}(\Theta(\boldsymbol{M}-\boldsymbol{M}',\mathcal{C}'))}{\|\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}-\mathcal{C}'))+\text{cut}(\Theta(\boldsymbol{M}-\boldsymbol{M}',\mathcal{C}'))\|_F}, \mathcal{A}-\Theta\right\rangle\right| \geq t\right)\\ &\leq \mathbb{P}\left(\sup_{\boldsymbol{M}\in\mathcal{M}}\sup_{\mathcal{C}\in([-1,1]^k)^{\otimes m}}\left|\left\langle\frac{\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))}{\|\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))\|_F}, \mathcal{A}-\Theta\right\rangle\right| \geq \frac{t}{2}\right)\\ &\qquad + \mathbb{P}\left(\sup_{\boldsymbol{M},\boldsymbol{M}'\in\mathcal{M}}\sup_{\mathcal{C}\in([-1,1]^k)^{\otimes m}}\left|\left\langle\frac{\text{cut}(\Theta(\boldsymbol{M}-\boldsymbol{M}',\mathcal{C}))}{\|\text{cut}(\Theta(\boldsymbol{M}-\boldsymbol{M}',\mathcal{C}))\|_F}, \mathcal{A}-\Theta\right\rangle\right| \geq \frac{t}{2}\right)\\ &= 2\exp\left(-C_1t^2 + C_2k^m + 2n\log k\right),\end{aligned}$$

where $C_1, C_2 > 0$ are universal constants. The last line follows from Lemma 6. Choosing $t = C\sqrt{n\log k + k^m}$ yields

$$\text{(ii)} \leq C_1\left(\left(\frac{m}{k}\right)^{2\alpha} + \left(\frac{k}{n}\right)^m + \frac{\log k}{n^{m-1}}\right),$$

with probability at least $1 - \exp\left(-C_2(n \log k + k^m)\right)$. Combinations of two error bounds in (4) and setting $k = \lceil n^{\frac{m}{m+2\alpha}} \rceil$, completes the theorem. $\square$

2.2 Under stochastic block model

Theorem 2.2 (Stochastic block model)**.** Let $\hat{\Theta}$ be the estimator from (3). Suppose true probability tensor $\Theta \in \mathrm{cut}(\mathcal{P}_k)$ for fixed block size k. Then, there exists two constants $C_1, C_2 > 0$, such that

$$\frac{1}{n^m}\|\hat{\Theta} - \Theta\|_F^2 \leq C_1 \left(\frac{k}{n}\right)^m + m\frac{\log k}{n^{m-1}},$$

with probability at least $1 - \exp\left(-C_2(n \log k + k^m)\right)$. In particular, suppose $k \asymp n^{\delta}$ for some $\delta \in [0, 1]$. Then, the convergence rate becomes

$$\left(\frac{k}{n}\right)^m + \frac{\log k}{n^{m-1}} \asymp \begin{cases} n^{-m} & k = 1, \\ n^{-m+1} & \delta = 0, k \geq 2, \\ n^{-m+1}\log(n) & \delta \in (0, 1/m], \\ n^{-m(1-\delta)} & \delta \in (1/m, 1]. \end{cases}$$

Remark 2. The first part of error is the nonparametric rate, which is determined by the number of parameters and the number of observation of the model. For the stochastic block model with k-clusters, the number of parameters is $\mathcal{O}(k^m)$ while the number of observation is $\mathcal{O}(n^m)$. The proportion of two numbers is absorbed in the nonparametric rate. The second part of error is the clustering rate. The number of possible k-clusters of n-vertices is $\mathcal{O}(k^n)$. We can check that the the possible number of clusters is reflected by logarithm in the clustering rate.

Proof. By similar way in the proof of Theorem 2.1, we have

$$\begin{aligned}
\|\hat{\Theta} - \Theta\|_F^2 &\leq 2\langle \hat{\Theta} - \Theta, \mathcal{A} - \Theta \rangle \\
&= 2\|\hat{\Theta} - \Theta\|_F \left\langle \frac{\hat{\Theta} - \Theta}{\|\hat{\Theta} - \Theta\|_F}, \mathcal{A} - \Theta \right\rangle \\
&\leq 2\|\hat{\Theta} - \Theta\|_F \sup_{\Theta' \in \mathcal{P}_k} \sup_{\Theta'' \in \mathcal{P}_k} \left\langle \frac{\mathrm{cut}(\Theta') - \mathrm{cut}(\Theta'')}{\|\mathrm{cut}(\Theta') - \mathrm{cut}(\Theta'')\|_F}, \mathcal{A} - \Theta \right\rangle.
\end{aligned}$$

Notice the last inequality holds because $\Theta \in \mathrm{cut}(\mathcal{P}_k)$. Therefore, we have the result following the proof of Theorem 2.1. $\square$

3 Hypergraphon estimation

3.1 Hölder continuous graphon

For a given probability tensor Θ, define the empirical hypergraphon $f_\Theta \colon [0,1]^m \to [0,1]$ as the following piecewise constant function:

$$\tilde{f}_\Theta(x_1, \ldots, x_m) = \Theta_{\lfloor x_1 \rfloor, \ldots, \lfloor x_m \rfloor}.$$

For any hypergraphon estimator $\hat{f}$, we define the squared error

$$\delta^2(\hat{f}, f) := \inf_{\tau \in \mathcal{T}} \int_{(0,1)^m} |f(\tau(x)) - \hat{f}(x)|^2 dx,$$

where $\mathcal{T}$ is the set of all measure-preserving bijection $\tau \colon [0,1] \to [0,1]$.

Our goal is to construct the upper bound of error $\mathbb{E}\left[\delta^2(f_{\hat{\Theta}}, f)\right]$. By triangular inequality, we have

$$\mathbb{E}\left[\delta^2(f_{\hat{\Theta}}, f)\right] \leq \underbrace{\frac{2}{n^m}\mathbb{E}\|\hat{\Theta}-\Theta\|_F^2}_{\text{(i)}} + 2\underbrace{\mathbb{E}\left[\delta^2(f_{\Theta}, f)\right]}_{\text{(ii)}}.$$

The following two lemmas show the upper bounds of (i) and (ii).

Lemma 2 (Estimation error for Hölder continuous hypergraphon)**.** Suppose the true parameter Θ admits the form (1) with $f \in \mathcal{H}(\alpha, L)$. Let $\hat{\Theta}$ be the estimator from (3). Then

$$\frac{1}{n^m}\mathbb{E}\|\hat{\Theta}-\Theta\|_F^2 \leq C\left(n^{\frac{-2m\alpha}{m+2\alpha}} + \frac{\log n}{n^{m-1}}\right),$$

for some constant C only depending proportionally on L^2.

Proof. Expectation is with respect to Bernoulli distribution $\mathcal{A}$ and uniform distributions $\boldsymbol{\xi} := \{(\xi_{\omega_1}, \ldots, \xi_{\omega_m})\}_{\omega \in E}$. Notice that

$$\frac{1}{n^m}\mathbb{E}_{\mathcal{A},\boldsymbol{\xi}}\|\hat{\Theta}-\Theta\|_F^2 = \frac{1}{n^m}\mathbb{E}_{\boldsymbol{\xi}}\left[\mathbb{E}_{\mathcal{A}}\left[\|\hat{\Theta}-\Theta\|_F^2 \Big| \boldsymbol{\xi}\right]\right]$$

Similar to the proof of Theorem 2.1, let a block tensor Θ^* be a block tensor satisfying Lemma 1. Given $\boldsymbol{\xi}$, we have the following by triangular inequality,

$$\|\hat{\Theta}-\Theta\|_F^2 \leq \underbrace{\|\hat{\Theta}-\Theta^*\|_F^2}_{\text{(i)}} + \underbrace{\|\Theta^*-\Theta\|_F^2}_{\text{(ii)}}. \tag{6}$$

Since we have already shown the error bound of (ii) in Lemma 1 given $\boldsymbol{\xi}$, we bound the error from (i) given $\boldsymbol{\xi}$. From the definition of $\hat{\Theta}$, we have

why not directly use Thm 2.2

$$\|\hat{\Theta}-\Theta^*\|_F^2 \leq \left(2\left\langle \frac{\hat{\Theta}-\Theta^*}{\|\hat{\Theta}-\Theta^*\|_F}, \mathcal{A}-\Theta\right\rangle + 2\|\Theta-\Theta^*\|_F\right)^2 \leq 4\left(\left\langle \frac{\hat{\Theta}-\Theta^*}{\|\hat{\Theta}-\Theta^*\|_F}, \mathcal{A}-\Theta\right\rangle^2 + \|\Theta-\Theta^*\|_F^2\right). \tag{7}$$

Notice $\|\Theta^*-\Theta\|_F$ is a constant with respect to $\mathcal{A}$. Therefore, only random variable related to $\mathcal{A}$ in $\|\hat{\Theta}-\Theta^*\|_F$ is the inner product term. Let $\mathcal{M} = \{\boldsymbol{M}\colon \boldsymbol{M} \text{ is the collection of membership matrices}\}$. Then,

$$\begin{aligned}
\left\langle \frac{\hat{\Theta}-\Theta^*}{\|\hat{\Theta}-\Theta^*\|_F}, \mathcal{A}-\Theta\right\rangle^2 &\leq \sup_{\Theta' \in \mathcal{P}_k}\sup_{\Theta'' \in \mathcal{P}_k}\left\langle \frac{\text{cut}(\Theta')-\text{cut}(\Theta'')}{\|\text{cut}(\Theta')-\text{cut}(\Theta'')\|_F}, \mathcal{A}-\Theta\right\rangle^2 \\
&\leq \sup_{\boldsymbol{M},\boldsymbol{M}' \in \mathcal{M}}\sup_{\mathcal{C},\mathcal{C}' \in ([0,1]^k)^{\otimes m}}\left\langle \frac{\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))-\text{cut}(\Theta(\boldsymbol{M}',\mathcal{C}'))}{\|\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))-\text{cut}(\Theta(\boldsymbol{M}',\mathcal{C}'))\|_F}, \mathcal{A}-\Theta\right\rangle^2 \\
&\leq \sup_{\boldsymbol{M},\boldsymbol{M}' \in \mathcal{M}}\sup_{\mathcal{C},\mathcal{C}' \in ([0,1]^k)^{\otimes m}}\left\langle \frac{\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}-\mathcal{C}'))+\text{cut}(\Theta(\boldsymbol{M}-\boldsymbol{M}',\mathcal{C}'))}{\|\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}-\mathcal{C}'))+\text{cut}(\Theta(\boldsymbol{M}-\boldsymbol{M}',\mathcal{C}'))\|_F}, \mathcal{A}-\Theta\right\rangle^2 \\
&\leq 2\sup_{\boldsymbol{M} \in \mathcal{M}}\sup_{\mathcal{C} \in ([-1,1]^k)^{\otimes m}}\left\langle \frac{\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}-\mathcal{C}'))}{\|\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}-\mathcal{C}'))\|_F}, \mathcal{A}-\Theta\right\rangle^2 \\
&\qquad + 2\sup_{\boldsymbol{M},\boldsymbol{M}' \in \mathcal{M}}\sup_{\mathcal{C} \in ([-1,1]^k)^{\otimes m}}\left\langle \frac{\text{cut}(\Theta(\boldsymbol{M}-\boldsymbol{M}',\mathcal{C}))}{\|\text{cut}(\Theta(\boldsymbol{M}-\boldsymbol{M},\mathcal{C}))\|_F}, \mathcal{A}-\Theta\right\rangle^2.
\end{aligned}$$

Notice that $\mathcal{A} - \Theta$ is sub-Gaussian with proxy parameter $\sigma^2 = 1/4$. By Lemma 7, we have

$$\mathbb{E}_{\mathcal{A}}\left[\left\langle \frac{\hat{\Theta} - \Theta^*}{\|\hat{\Theta} - \Theta^*\|_F}, \mathcal{A} - \Theta \right\rangle^2 \middle| \boldsymbol{\xi} \right] \leq C\left(k \log n + k^m\right), \tag{8}$$

for some constant $C > 0$.

Combination of (6),(7),(8), and Lemma 1 yields

$$\frac{1}{n^m}\mathbb{E}_{\mathcal{A}}\left[\|\hat{\Theta} - \Theta\|_F^2 \middle| \boldsymbol{\xi}\right] \leq C\left(\left(\frac{m}{k}\right)^{2\alpha} + \left(\frac{k}{n}\right)^m + \frac{\log k}{n^{m-1}}\right), \tag{9}$$

for some constant $C > 0$. Setting $k = \lceil n^{\frac{m}{m+2\alpha}} \rceil$ and the fact that the inequality (9) is not depending on the random variable $\boldsymbol{\xi}$, completes the proof as follows.

$$\frac{1}{n^m}\mathbb{E}_{\mathcal{A},\boldsymbol{\xi}}\|\hat{\Theta} - \Theta\|_F^2 = \frac{1}{n^m}\mathbb{E}_{\boldsymbol{\xi}}\left[\mathbb{E}_{\mathcal{A}}\left[\|\hat{\Theta} - \Theta\|_F^2 \middle| \boldsymbol{\xi}\right]\right] \leq C\left(n^{\frac{-2m\alpha}{m+2\alpha}} + \frac{\log n}{n^{m-1}}\right).$$

□

Lemma 3 (Agnostic error for Hölder continuous hypergraphon)**.** Suppose $f \in \mathcal{H}(\alpha, L)$. Then

$$\mathbb{E}\left[\delta^2(f_\Theta, f)\right] \leq \frac{C}{n^\alpha},$$

for some constant $C > 0$ only depending proportionally on m^2L^2.

Proof. By triangular inequality, we have

$$\mathbb{E}\left[\delta^2(f_\Theta, f)\right] \leq 2\mathbb{E}\left[\delta^2(f_\Theta, f_{\Theta'})\right] + 2\mathbb{E}\left[\delta^2(f_{\Theta'}, f)\right],$$

where $\Theta' \in ([0,1]^n)^{\otimes m}$ such that $\Theta'_\omega = f(\xi_{\omega_1}, \ldots, \xi_{\omega_m})$ for all $\omega \in [n]^m$. Notice $\Theta'_\omega = \Theta_\omega$ for $\omega \in E$ but $\Theta_\omega = 0$ for $\omega \in [n]^m \setminus E$. By definition of Θ',

$$\mathbb{E}\left[\delta^2(f_\Theta, f_{\Theta'})\right] = \int_{[0,1]^m} |f_\Theta(x) - f_{\Theta'}(x)| dx < \frac{C}{n},$$

for some $C > 0$ only depending on m. This is because $|f_\Theta(x) - f_{\Theta'}(x)| = 0$ outside of a set of measure $\left(n^m - \binom{n}{m}n!\right)/n^m = C/n$. Notice $C = \mathcal{O}(m^2)$. Hence it suffices to prove that

$$\mathbb{E}\left[\delta^2(f_{\Theta'}, f)\right] \leq \frac{C}{n^\alpha}.$$

We have

$$\delta^2(f_{\Theta'}, f) = \inf_{\tau \in \mathcal{T}} \sum_{i_1, \ldots, i_m = 1}^{n} \int_{(i_1-1)/n}^{i_1/n} \cdots \int_{(i_m-1)/n}^{i_m/n} |f(\tau(x_1), \ldots, \tau(x_m)) - \Theta'_{i_1,\ldots,i_m}|^2 dx_1 \cdots dx_m$$

The infimum over all measure-preserving bijection is smaller than the minimum over the subclass of measure-preserving bijection τ such that

$$\int_{(i_1-1)/n}^{i_1/n} \cdots \int_{(i_m-1)/n}^{i_m/n} f(\tau(x_1), \ldots \tau(x_m)) dx_1 \cdots dx_m = \int_{(\sigma(i_1)-1)/n}^{\sigma(i_1)/n} \cdots \int_{(\sigma(i_m)-1)/n}^{\sigma(i_m)/n} f(x_1, \ldots x_m) dx_1 \cdots dx_m$$

for some permutation σ. For $x \in \prod_{\ell=1}^m [(\sigma(i_\ell)-1)/n, \sigma(i_\ell)/n]$,

$$\begin{aligned}|f(x_1,\ldots,x_m) - f(\xi_1,\ldots,\xi_m)|^2 \le & 2\left|f(x_1,\ldots,x_m) - f\left(\frac{\sigma(i_1)}{n+1},\cdots\frac{\sigma(i_m)}{n+1}\right)\right|^2 \\ & + 2\left|f\left(\frac{\sigma(i_1)}{n+1},\cdots\frac{\sigma(i_m)}{n+1}\right) - f(\xi_{(\sigma(i_1))},\ldots,\xi_{(\sigma(i_m))})\right|^2 \\ & + 2\left|f(\xi_{(\sigma(i_1))},\ldots,\xi_{(\sigma(i_m))}) - f(\xi_{i_1},\ldots,\xi_{i_m})\right|^2,\end{aligned} \tag{10}$$

where $\xi_{(\ell)}$ denotes the ℓ-th largest element of the set $\{\xi_1,\ldots,\xi_n\}$. Choose random permutation σ such that $\xi_{\sigma^{-1}(1)} \le \xi_{\sigma^{-1}(2)} \cdots \le \xi_{\sigma^{-1}(n)}$. Then the third summand in (10) is 0 almost surely.

For the first summand in (10), notice $(\sigma(i_1)/(n+1),\ldots\sigma(i_m)/(n+1)) \in \prod_{\ell=1}^m[(\sigma(i_\ell)-1)/n, \sigma(i_\ell)/n]$. From (2), we obtain

$$\left|f(x_1,\ldots,x_m) - f\left(\frac{\sigma(i_1)}{n+1},\cdots\frac{\sigma(i_m)}{n+1}\right)\right|^2 \le m^2M^2\left(\frac{1}{n}\right)^{2\alpha}.$$

Integrating and taking expectation on the first summand yields,

$$\mathbb{E}\left[\sum_{i_1,\ldots,i_m=1}^n \int_{(\sigma(i_1)-1)/n}^{\sigma(i_1)/n}\cdots\int_{(\sigma(i_m)-1)/n}^{\sigma(i_m)/n}\left|f(x_1,\ldots,x_m) - f\left(\frac{\sigma(i_1)}{n+1},\cdots\frac{\sigma(i_m)}{n+1}\right)\right|^2 dx_1\cdots dx_m\right] \le m^2L^2\left(\frac{1}{n}\right)^{2\alpha}. \tag{11}$$

With (2), the second summand on (10) is bounded,

$$\begin{aligned}\left|f\left(\frac{\sigma(i_1)}{n+1},\cdots\frac{\sigma(i_m)}{n+1}\right) - f(\xi_{(\sigma(i_1))},\ldots,\xi_{(\sigma(i_m))})\right|^2 &\le \left(L\sum_{\ell=1}^m\left|\frac{\sigma(i_\ell)}{n+1} - \xi_{(\sigma(i_\ell))}\right|^\alpha\right)^2 \\ &\le 2L^2\sum_{\ell=1}^m\left|\frac{\sigma(i_\ell)}{n+1} - \xi_{(\sigma(i_\ell))}\right|^{2\alpha}.\end{aligned}$$

Integrating and taking expectation on the second summand yields,

$$\begin{aligned}&\mathbb{E}\left[\sum_{i_1,\ldots,i_m=1}^n \int_{(\sigma(i_1)-1)/n}^{\sigma(i_1)/n}\cdots\int_{(\sigma(i_m)-1)/n}^{\sigma(i_m)/n}\left|f\left(\frac{\sigma(i_1)}{n+1},\cdots\frac{\sigma(i_m)}{n+1}\right) - f(\xi_{(\sigma(i_1))},\ldots,\xi_{(\sigma(i_m))})\right|^2 dx_1\cdots dx_m\right] \\ &\le \left[\sum_{i_1,\ldots,i_m=1}^n \int_{(\sigma(i_1)-1)/n}^{\sigma(i_1)/n}\cdots\int_{(\sigma(i_m)-1)/n}^{\sigma(i_m)/n} 2L^2\sum_{\ell=1}^m\left|\frac{\sigma(i_\ell)}{n+1} - \xi_{(\sigma(i_\ell))}\right|^{2\alpha} dx_1\cdots dx_m\right] \\ &= 2mL^2\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n\left|\frac{i}{n+1} - \xi_{(i)}\right|^{2\alpha}\right] \\ &\le 2mL^2\max_{i=1,\ldots,n}\mathbb{E}\left[\left|\frac{i}{n+1} - \xi_{(i)}\right|^{2\alpha}\right] \\ &\le 2mL^2\max_{i=1\ldots,n}\left[\mathrm{Var}(\xi_{(i)})\right]^\alpha \le CmL^2\left(\frac{1}{n}\right)^\alpha,\end{aligned} \tag{12}$$

where we have used $\mathbb{E}(\xi_{(\ell)}) = \ell/(n+1)$, $\mathrm{Var}(\xi_{(\ell)}) \le C/n$ and Jensen's inequality. Combining (11) and (12) proves the lemma. □

By Lemma 2 and Lemma 3, we have the following theorem.

Theorem 3.1 (Mean square error of smooth hypergraphon)**.** Suppose the true parameter Θ admits the form (1) with $f \in \mathcal{H}(\alpha, L)$. Let $\hat{\Theta}$ be the estimator from (3). Then, there exists a positive constant $C > 0$ only depending on m and L such that

$$\mathbb{E}\left[\delta^2(f_{\hat{\Theta}}, f)\right] \leq C\left(n^{\frac{-2m\alpha}{m+2\alpha}} + \frac{\log n}{n^{m-1}} + \frac{1}{n^{\alpha}}\right).$$

Remark 3. The error rate is dominated by agnostic error $n^{-\alpha}$ for all combinations of $\{(m,\alpha)\}_{m\geq 2, \alpha\in(0,1]}$ except $(m,\alpha) = (2,1)$, when $\log n/n$ is the dominating error term. This agnostic error is unavoidable when considering mean square error because we estimate whole function values on $[0,1]^d$ from $\binom{n}{m} \asymp n^m$ sample size. My guess is sparsity parameter ρ_n can make different dominating error bounds depending on combinations of (m,α) because different powers of ρ_n will appear in three error terms.

3.2 Piecewise constant hypergraphon

Define $\mathcal{F}(k)$ the collection of k-piecewise constant hypergraphons such that for some $\mathcal{C} \in ([0,1]^k)^{\otimes m}$ and some $\phi\colon [0,1] \to [k]$,

$$f(x_1, \ldots, x_m) = \mathcal{C}_{\phi(x_1),\ldots,\phi(x_m)} \quad \text{for all } x_i \in [0,1], i = 1, \ldots m.$$

The estimation error associated to k-piecewise hypergraphon is bounded as follows.

Lemma 4 (Estimation error for k-piecewise constant hypergraphon)**.** Suppose the true parameter Θ admits the form (1) with $f \in \mathcal{F}(k)$. Let $\hat{\Theta}$ be the estimator from (3). Then

$$\frac{1}{n^m}\mathbb{E}\|\hat{\Theta} - \Theta\|_F^2 \leq C\left(\left(\frac{k}{n}\right)^m + \frac{\log k}{n^{m-1}}\right),$$

for some constant C only depending proportionally on L^2.

Proof. Expectation is with respect to Bernoulli distribution $\mathcal{A}$ and uniform distributions $\boldsymbol{\xi} := \{(\xi_{\omega_1}, \ldots, \xi_{\omega_m})\}_{\omega\in E}$. Notice that

$$\frac{1}{n^m}\mathbb{E}_{\mathcal{A},\boldsymbol{\xi}}\|\hat{\Theta} - \Theta\|_F^2 = \frac{1}{n^m}\mathbb{E}_{\boldsymbol{\xi}}\left[\mathbb{E}_{\mathcal{A}}\left[\|\hat{\Theta} - \Theta\|_F^2 \middle| \boldsymbol{\xi}\right]\right]$$

Similar to the proof of Therem 2.2, we have

$$\begin{aligned}
\mathbb{E}_{\mathcal{A}}\left[\|\hat{\Theta} - \Theta\|_F^2 \middle| \boldsymbol{\xi}\right] &\leq 4\mathbb{E}_{\mathcal{A}}\left[\left\langle \frac{\hat{\Theta} - \Theta}{\|\hat{\Theta} - \Theta\|_F}, \mathcal{A} - \Theta\right\rangle^2 \middle| \boldsymbol{\xi}\right] \\
&\leq 4\mathbb{E}_{\mathcal{A}}\left[\sup_{\Theta'\in\mathcal{P}_k}\sup_{\Theta''\in\mathcal{P}_k}\left\langle \frac{\text{cut}(\Theta') - \text{cut}(\Theta'')}{\|\text{cut}(\Theta') - \text{cut}(\Theta'')\|_F}, \mathcal{A} - \Theta\right\rangle^2 \middle| \boldsymbol{\xi}\right].
\end{aligned}$$

Notice the last inequality holds because $\Theta \in \text{cut}(\mathcal{P}_k)$. Therefore, we have the result following the proof of Lemma 2. $\square$

The egonostic error associated to k-piecewise hypergraphon is bounded as follows [I haven't proved this part yet].

Lemma 5 (Agnostic error for k-piecewise constant hypergraphon)**.** Consider the hypergraphon model (1). For all integers $k \leq n$, $f \in \mathcal{F}(k)$, we have

$$\mathbb{E}\left[\delta^2(f_{\Theta}, f)\right] \leq C\sqrt{\frac{k}{n}}$$

4 Technical lemmas

Lemma 6 (Sub-Gaussian probabilistic maximum bound)**.**

$$\mathbb{P}\left(\sup_{\boldsymbol{M}\in\mathcal{M}}\sup_{\mathcal{C}\in([-1,1]^k)^{\otimes m}}\left\langle\frac{\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))}{\|\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))\|_F},\mathcal{A}-\Theta\right\rangle\geq t\right)\leq\exp(-C_1t^2+C_2k^m+\log|\mathcal{M}|),$$

for some constant $C_1, C_2 > 0$.

Proof. Define a collection of normalized block tensors given a membership matrix as

$$\mathcal{S}(\boldsymbol{M},[-1,1])=\left\{\frac{\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))}{\|\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))\|_F}:\mathcal{C}\in([-1,1]^k)^{\otimes m}\right\}.$$

To ease of notation, we omit $[-1,1]$ in $S(\boldsymbol{M},[-1,1])$ here. Let $S'(\boldsymbol{M})$ be 1/2-net of $S(\boldsymbol{M})$ such that $|S'(\boldsymbol{M})|\leq\mathcal{N}(1/2,S(\boldsymbol{M}),\|\cdot\|_2)$ where $\mathcal{N}$ denotes the covering number. Notice that for any $\Theta_1\in\mathcal{S}(\boldsymbol{M})$, there exists $\Theta_2\in\mathcal{S}'(\boldsymbol{M})$ such that $\|\Theta_1-\Theta_2\|_F\leq 1/2$. Then,

$$\begin{aligned}|\langle\Theta_1,\mathcal{A}-\Theta\rangle|&\leq|\langle\Theta_1-\Theta_2,\mathcal{A}-\Theta\rangle|+|\langle\Theta_2,\mathcal{A}-\Theta\rangle|\\&=\|\Theta_1-\Theta_2\|_F\left|\left\langle\frac{\Theta_1-\Theta_2}{\|\Theta_1-\Theta_2\|_F},\mathcal{A}-\Theta\right\rangle\right|+|\langle\Theta_2,\mathcal{A}-\Theta\rangle|\\&\leq\frac{1}{2}\sup_{\Theta_1\in\mathcal{S}(\boldsymbol{M})}|\langle\Theta_1,\mathcal{A}-\Theta\rangle|+|\langle\Theta_2,\mathcal{A}-\Theta\rangle|.\end{aligned}$$

Taking sup and max on both side yields,

$$\sup_{\Theta_1\in\mathcal{S}(\boldsymbol{M})}|\langle\Theta_1,\mathcal{A}-\Theta\rangle|\leq 2\max_{\Theta_2\in\mathcal{S}'(\boldsymbol{M})}|\langle\Theta_2,\mathcal{A}-\Theta\rangle|.$$

Therefore, we have

$$\begin{aligned}\sup_{\boldsymbol{M}\in\mathcal{M}}\sup_{\mathcal{C}\in([-1,1]^k)^{\otimes m}}\left|\left\langle\frac{\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))}{\|\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))\|_F},\mathcal{A}-\Theta\right|\right\rangle&=\sup_{\boldsymbol{M}\in\mathcal{M}}\sup_{\Theta_1\in\mathcal{S}(\boldsymbol{M})}|\langle\Theta_1,\mathcal{A}-\Theta\rangle|\\&\leq\max_{\boldsymbol{M}\in\mathcal{M}}\max_{\Theta_2\in\mathcal{S}'(\boldsymbol{M})}|\langle\Theta_1,\mathcal{A}-\Theta\rangle|\end{aligned}\tag{13}$$

Using the maximum inequality over a finite set [Phillippe Rigollet, 2015, Theorem 1.14] and the fact $\mathcal{N}(1/2,S(\boldsymbol{M}),\|\cdot\|_F)\leq C^{k^m}$ for some constant $C>0$ completes the proof. ☐

Lemma 7 (Expected sub-exponential maximum bound)**.**

$$\mathbb{E}_{\mathcal{A}}\left[\sup_{\boldsymbol{M}\in\mathcal{M}}\sup_{\mathcal{C}\in([-1,1]^k)^{\otimes m}}\left\langle\frac{\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))}{\|\text{cut}(\Theta(\boldsymbol{M},\mathcal{C}))\|_F},\mathcal{A}-\Theta\right\rangle^2\middle|\boldsymbol{\xi}\right]\leq C\left(\log|\mathcal{M}|+k^m\right),$$

for some constant $C>0$.

Proof. Notice that for sub-Gausssian random variables $X_1,\ldots,X_N$ such that $X_i\sim\text{subG}(\sigma^2)$

$$\begin{aligned}\mathbb{E}\left[\max_{1\leq i\leq N}X_i^2\right]&=\frac{1}{s}\mathbb{E}\left[\log e^{s\max_{1\leq i\leq N}X_i^2}\right]\\&\leq\frac{1}{s}\log\mathbb{E}\left[e^{s\max_{1\leq i\leq N}X_i^2}\right]&&\text{[Jensen inequality]}\\&\leq\frac{1}{s}\log\sum_{1\leq i\leq N}\mathbb{E}\left[e^{sX_i^2}\right]\end{aligned}$$

$$
\begin{aligned}
&\le \frac{1}{s} \log \sum_{1\le i\le N} \left(1 + \sum_{k=2}^{\infty} \frac{s^k \mathbb{E} X_i^{2k}}{k!}\right) \\
&\le \frac{1}{s} \log \sum_{1\le i\le N} \left(1 + 2\sum_{k=2}^{\infty} (2s\sigma^2)^k\right) \quad \text{[Phillippe Rigollet, 2015, Lemma 1.4]} \\
&= \frac{1}{s} \log \sum_{1\le i\le N} \left(1 + 8s^2\sigma^4 \sum_{k=0}^{\infty} (2s\sigma^2)^k\right) \\
&\le \frac{1}{s} \log \sum_{1\le i\le N} \left(1 + 16 s^2\sigma^4\right) \qquad \text{taking } |s| = \frac{1}{4\sigma^2}
\end{aligned}
$$

Taking $s = 1/(4\sigma^2)$ yields,

$$
\mathbb{E}\left[\max_{1\le i\le N} X_i^2\right] \le 4\sigma^2(\log N + \log 2) \le 8\sigma^2 \log N. \tag{14}
$$

Combining (13) in the proof of Lemma 6 and (14) completes the proof. □

References

Chao Gao, Yu Lu, Harrison H Zhou, et al. Rate-optimal graphon estimation. *Annals of Statistics*, 43(6): 2624–2652, 2015.

Jan-Christian Hitter Phillippe Rigollet. High dimensional statistics. *Lecture notes for course 18S997*, 2015.