Hypergraphon estimation error

Chanwoo Lee
February 22, 2021

1 Theoretical guarantee of the estimation

We consider an undirected m-uniform hypergraph. The connectivity can be encoded by an adjacency tensor $\{\mathcal{A}_{i_1,\ldots,i_m}\}$ taking values in $(\{0,1\}^n)^{\otimes m}$. The model is $\mathcal{A}_{i_1,\ldots,i_m} = \mathcal{A}_{i_{\sigma(1)},\ldots,i_{\sigma(m)}} \sim \text{Bernoulli}(\Theta_{i_1,\ldots,i_m})$ for any permutation σ for $1 \leq i_l \leq n$, $l \in [m]$,where

$$\Theta_{i_1,\ldots,i_m} = f(\xi_{i_1},\ldots,\xi_{i_m}).$$

f: graphon.
Theta: discrete version of graphon

The sequence $\{\xi_i\}$ are random variables from Unif$[0,1]$. The function f assume to be symmetric such that $f(x_1,\ldots,x_m) = f(x_{\sigma(1)},\ldots,x_{\sigma(m)})$ for any permutation σ. Since f is symmetric, it is enough to consider the domain only $\mathcal{D} = \{\boldsymbol{x} = (x_1,\ldots,x_m) \in [0,1]^m : x_1 \geq \cdots \geq x_m\}$. Define the derivative operator by

Q: why do we need to introduce alpha-holder smoothness with alpha \in (0,\infty)?
can we just restrict ourselves to Lipschitz condition with alpha in (0,1)?

$$\nabla_{i_1,\ldots i_m} f(x_1,\ldots,x_m) = \frac{\partial^{i_1+\cdots+i_m}}{(\partial x_1)^{i_1}\cdots(\partial x_m)^{i_m}} f(x_1,\ldots,x_m),$$

and the Hölder norm is defined as

$$\|f\|_{\mathcal{H}_\alpha} = \max_{i_1+\cdots+i_m \leq \lfloor\alpha\rfloor} \sup_{\boldsymbol{x}\in\mathcal{D}} |\nabla_{i_1,\ldots,i_m} f(\boldsymbol{x})| + \max_{i_1+\cdots+i_m = \lfloor\alpha\rfloor} \sup_{\boldsymbol{x}\neq\boldsymbol{y}\in\mathcal{D}} \frac{|\nabla_{i_1,\ldots,i_m} f(\boldsymbol{x}) - \nabla_{i_1,\ldots,i_m} f(\boldsymbol{y})|}{(\|\boldsymbol{x}-\boldsymbol{y}\|_1)^{\alpha-\lfloor\alpha\rfloor}}.$$

The Hölder class is defined by w = (i_1,…,i_m)

$$\mathcal{H}_\alpha(M) = \{\|f\|_{\mathcal{H}_\alpha} \leq M : f \text{ is symmetric}\},$$

where $\alpha > 0$ is the smoothness parameter and $M > 0$ is the size of the class. Notice that a function $f \in \mathcal{H}_\alpha(M)$, satisfies the Lipschitz condition

x=(0,0)
y=(1,1)

alpha <= 1

$$|f(\boldsymbol{x}) - f(\mu)| \leq M(\|\boldsymbol{x}-\boldsymbol{y}\|_1)^\alpha,$$

Do we require alpha <= 1 here? Otherwise, the function f is a constant.

for any $\boldsymbol{x},\boldsymbol{y} \in \mathcal{D}$. We assume that the hypergraphon f belongs to the function class:

$$\mathcal{F}_\alpha(M) = \{0 \leq f \leq 1 : f \in \mathcal{H}_\alpha(M)\}.$$

For a given membership function $z\colon [n] \to [k]$, define the membership number function as $h\colon [k]^m \to [n]^k$ such that $h(a_1,\ldots,a_m) = (h_1,\ldots,h_m)$ where h_i is the number of i-th membership from $(a_1,\ldots,a_m) \in [k]^m$ for $i \in [m]$. Given a tensor $\Theta \in (\mathbb{R}^n)^{\otimes m}$, we define a block average on the set $z^{-1}(a_1)\times\cdots\times z^{-1}(a_m)$ for $a_i \in [k]$, $i \in [m]$ as

any way to simplify the notation?

$$\bar{\Theta}_{a_1,\ldots,a_m}(z) = \frac{1}{\cancel{\prod_{a\in\{a_1,\ldots,a_m\}}} |z^{-1}(a)||z^{-1}(a)-1|\cdots|z^{-1}(a)-h_a+1|} \sum_{\substack{(i_1,\ldots,i_m)\colon i_\ell\in z^{-1}(a_\ell),\ell\in[m] \\ \cancel{|\{i_1,\ldots,i_m\}|=m}}} \Theta_{i_1,\ldots,i_m}.$$

\bar Theta(\omega) = {1\over constant} sum_{\omega} Theta(\omega) 1(\omega and \omega' are in the same block)

We show that any hypergraphons in $\mathcal{F}_\alpha(M)$ can be approximated by the averaged block tensor.

Lemma 1. There exists $z^*\colon [n] \to [k]$, satisfying

(approximation error)

$$\frac{1}{n^m} \sum_{a_1,\ldots,a_m\in[k]} \sum_{\substack{(i_1,\ldots,i_m)\colon i_\ell\in(z^*)^{-1}(a_\ell),\ell\in[m] \\ \cancel{|\{i_1,\ldots,i_m\}|=m}}} (\Theta_{i_1,\ldots,i_m} - \bar{\Theta}_{a_1,\ldots,a_m}(z^*))^2 \leq CM^2\left(\frac{m^2}{k^2}\right)^\alpha.$$

\bar Theta: smooth version of Theta given block allocation.

Q: does Lemma 1 hold for any k?

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

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

Define $Z^*_{a_1,\dots,a_m} = \{(u_1, \dots, u_m) \colon z^*(u_i) = a_i \text{ for all } i \in [m]\}$. By the construction of z^* for $\xi_{i_\ell} \in [(a_{\ell-1} - 1)/k, a_\ell/k]$, when $|\{a_1, \dots a_m\}| = m$,

$$\begin{aligned}
|f(\xi_{i_1}, \dots, \xi_{i_m}) - \bar{\Theta}_{a_1,\dots,a_m}(z^*)| &= \left| f(\xi_{i_1}, \dots, \xi_{i_m}) - \frac{1}{\prod_{\ell=1}^m |(z^*)^{-1}(a_\ell)|} \sum_{(u_1,\dots,u_m)\in Z^*_{a_1\dots,a_m}} f(\xi_{u_1}, \dots, \xi_{u_m}) \right| \\
&\leq \frac{1}{\prod_{\ell=1}^m |(z^*)^{-1}(a_\ell)|} \sum_{(u_1,\dots,u_m)\in Z^*_{a_1\dots,a_m}} |f(\xi_{i_1}, \dots, \xi_{i_m}) - f(\xi_{u_1}, \dots, \xi_{u_m})| \\
&\leq \frac{1}{\prod_{\ell=1}^m |(z^*)^{-1}(a_\ell)|} \sum_{(u_1,\dots,u_m)\in Z^*_{a_1\dots,a_m}} M \|(\xi_{i_1}, \dots, \xi_{i_m}) - (\xi_{u_1}, \dots, \xi_{u_m})\|_1^\alpha \\
&\leq CM \left(\frac{m}{k}\right)^\alpha.
\end{aligned}$$

Similar results hold for the cases $|\{a_1, \dots, a_m\}| < m$. ◻

We estimate the hypergraphon $\Theta_{i_1,\dots,i_m} = f(\xi_{i_1}, \dots, \xi_{i_m})$ by

$$\hat{\Theta} = \operatorname*{arg\,min}_{\Theta \in \mathcal{P}_k} \|\mathcal{A} - \Theta\|_F^2, \tag{1}$$

where

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

Then we obtain the convergence rate for hypergraphon estimation with respect to the least square error.

Theorem 1.1. Let $\hat{\Theta}$ be the least square estimator from (1). Then, There exist two constants $C_1, C_2 > 0$ such that,

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

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

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

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

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

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

Next question:
1. current results are conditional on xi.
2. integrated error. that takes into account of both randomness in xi and y
E_{X, data involved in \hat } |\hat f(X) - f (X)|^2

Combining (2) 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

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

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

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

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 $t>0$.

$$\begin{aligned}\mathbb{P}\left(\|\hat{\Theta}-\Theta\|_F>t\right) &\le \mathbb{P}\left(\sup_{\boldsymbol{M},\boldsymbol{M}'\in\mathcal{M}}\sup_{\mathcal{C},\mathcal{C}'\in([0,1]^n)^{\otimes m}}\left|\left\langle\frac{\Theta(\boldsymbol{M},\mathcal{C})-\Theta(\boldsymbol{M}',\mathcal{C}')}{\|\Theta(\boldsymbol{M},\mathcal{C})-\Theta(\boldsymbol{M}',\mathcal{C}')\|_F},\mathcal{A}-\Theta\right\rangle\right|+\|\Theta-\Theta^*\|_F\ge\frac{t}{2}\right)\\ &\le \sum_{\mathcal{M},\mathcal{M}'\in\mathcal{M}}\mathbb{P}\left(\sup_{\mathcal{C},\mathcal{C}'\in([0,1]^n)^{\otimes m}}\left|\left\langle\frac{\Theta(\boldsymbol{M},\mathcal{C})-\Theta(\boldsymbol{M}',\mathcal{C}')}{\|\Theta(\boldsymbol{M},\mathcal{C})-\Theta(\boldsymbol{M}',\mathcal{C}')\|_F},\mathcal{A}-\Theta\right\rangle\right|+Cn^{m/2}M\left(\frac{m}{k}\right)^\alpha\ge\frac{t}{2}\right)\\ &\le |\mathcal{M}|^2C_1^{k^m}\exp\left(-C_2\left(t-n^{m/2}M\left(\frac{m}{k}\right)^\alpha\right)^2\right)\\ &= \exp\left(2n\log k+C_1k^m-C_2\left(t-n^{m/2}M\left(\frac{m}{k}\right)^\alpha\right)^2\right)\end{aligned}$$

For two universal constants $C_1,C_2>0$. The third line follows from Phillippe Rigollet [2015] and the fact that $\Theta=\Theta(\boldsymbol{M},\cdot)$ lies in a linear space of dimension k^m. Choosing $t=n^{m/2}M(m/k)^\alpha+C\sqrt{n\log k+k^m}$ yields

$$\frac{1}{n^m}\|\hat{\Theta}-\Theta\|_F\le C_1\left(\left(\frac{m}{k}\right)^{2\alpha}+\left(\frac{k}{n}\right)^m+\frac{\log k}{n^{m-1}}\right), \tag{3}$$

with probability at least $1-\exp\left(-C_2(n\log k+k^m)\right)$. Setting $k=\lceil n^{\frac{m}{m+2\alpha}}\rceil$ to balance (3), completes the theorem. $\square$

2 Discussion

Currently I am looking for the paper that guarantee the similar representation of exchangeable hypergraph. For exchangeable array such that $\mathcal{A}_{i_1\dots,i_m}=\mathcal{A}_{i_{\sigma(1)},\dots,i_{\sigma(m)}}$, It is known that there exists $f\colon[0,1]\times[0,1]^n\times$

$[0,1]^{\binom{n}{2}} \times \cdots \times [0,1]^{\binom{n}{n-1}} \times [0,1] \to [0,1]$ such that

$$\mathcal{A}_{i_1,\ldots,i_m} \sim \text{Bernoulli}(\Theta_{i_1,\ldots,i_m}), \quad \Theta_{i_1,\ldots,i_m} = f(\alpha, \xi_{i_1}, \ldots, \xi_{i_m}, \xi_{i_1 i_2}, \ldots, \xi_{i_1 i_2 \ldots i_m}).$$

[Austin et al., 2008]. I need to do more research for justification of modeling hypergraphon as

$$\Theta_{i_1,\ldots,i_m} = f(\xi_{i_1}, \ldots, \xi_{i_m}).$$

References

Tim Austin et al. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probability Surveys*, 5:80–145, 2008.

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