Hypergraphon estimation error 2

Chanwoo Lee
February 25, 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

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

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 smooth with a constant L.

Definition 1. A function $f \colon [0,1]^m \to [0,1]$ is a α-Hölder smooth with a constant L, denoted as $f \in \mathcal{H}(\alpha, L)$, if there exists a polynomial function $\mathcal{P}_{\lfloor \alpha \rfloor}(x)$ of degree $\lfloor \alpha \rfloor$ such that

$$|f(x) - \mathcal{P}_{\lfloor \alpha \rfloor}(x' - x)| \leq L \|x' - x\|_{\alpha - \lfloor \alpha \rfloor}^{\alpha - \lfloor \alpha \rfloor},$$

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

It follows from the standard embedding theorem for any $f \in \mathcal{H}(\alpha, L)$,

Can we define the function class of interest, H(alpha,L), using property(1) with alpha\in[0,1]?

$$|f(x) - f(y)| \leq M \|x - y\|_{\alpha \wedge 1}^{\alpha \wedge 1}, \tag{1}$$

where $M > 0$ is a global constant only depending on α and L. We only use the property (1) over the note for α-Hölder smoothness.

How does M depend on L?

2 Probability matrix estimation

For every integer k <=n, there exists ….

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

add name to Lemma 1

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

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 $\xi_{\omega_\ell} \in [(a_{\ell-1}-1)/k, a_\ell/k]$,

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

□

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

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

where

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

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

$$\mathrm{cut}(\Theta_\omega) = \begin{cases} \Theta_\omega & \text{if } \omega \in E, \\ 0 & \text{if } \omega \in [n]^m \setminus E. \end{cases} \tag{2}$$

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

Theorem 2.1 (hypergraphon model)**.** Let $\hat{\Theta}$ be the estimator from (2). Then, there exist two constants $C_1, C_2 > 0$ such that,

Add a remark.
Discuss which of the two terms dominates.
We will see different rate depending on (alpha, m).
Agree with intuition?

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

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

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

$$\|\hat{\Theta} - \Theta\|_F^2 \leq \underbrace{\|\hat{\Theta} - \Theta^*\|_F^2}_{\text{(i)}} + \underbrace{\|\Theta^* - \Theta\|_F^2}_{\text{(ii)}}. \tag{3}$$

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

$$\|\hat{\Theta} - \mathcal{A}\|_F^2 \leq \|\Theta^* - \mathcal{A}\|_F^2. \tag{4}$$

Combining (4) 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,

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

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 $t>0$.

$$\begin{aligned}&\mathbb{P}\left(\sup_{\boldsymbol{M},\boldsymbol{M}'\in\mathcal{M}}\sup_{\mathcal{C},\mathcal{C}'\in([0,1]^n)^{\otimes m}}\left|\left\langle\frac{\mathrm{cut}(\Theta(\boldsymbol{M},\mathcal{C}))-\mathrm{cut}(\Theta(\boldsymbol{M}',\mathcal{C}'))}{\|\mathrm{cut}(\Theta(\boldsymbol{M},\mathcal{C}))-\mathrm{cut}(\Theta(\boldsymbol{M}',\mathcal{C}'))\|_F},\mathcal{A}-\Theta\right\rangle\right|\geq t\right)\\ &\leq \sum_{\mathcal{M},\mathcal{M}'\in\mathcal{M}}\mathbb{P}\left(\sup_{\mathcal{C},\mathcal{C}'\in([0,1]^n)^{\otimes m}}\left|\left\langle\frac{\mathrm{cut}(\Theta(\boldsymbol{M},\mathcal{C}))-\mathrm{cut}(\Theta(\boldsymbol{M}',\mathcal{C}'))}{\|\mathrm{cut}(\Theta(\boldsymbol{M},\mathcal{C}))-\mathrm{cut}(\Theta(\boldsymbol{M}',\mathcal{C}'))\|_F},\mathcal{A}-\Theta\right\rangle\right|\cancel{\left(\frac{m}{k}\right)^{\alpha\wedge 1}}\geq t\right)\\ &\leq |\mathcal{M}|^2C_1^{k^m}\exp\left(-C_2t^2\right)\\ &= \exp\left(2n\log k+C_1k^m-C_2t^2\right),\end{aligned}$$

where $C_1,C_2>0$ are universal constants. The secibd 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 $t=C\sqrt{n\log k+k^m}$ yields

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

with probability at least $1-\exp\left(-C_2(n\log k+k^m)\right)$. Combinations of two error bounds in (3) and setting $k=\left\lceil n^{\frac{m}{m+2(\alpha\wedge 1)}}\right\rceil$, completes the theorem. □

Theorem 2.2 (stochastic block model)**.** Let $\hat{\Theta}$ be the estimator from (2). 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

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

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

Perform the similar discussion for Thm 2.1

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

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

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

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. □

3 Hypergraphon estimation

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

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

(i) directly follows from Thm 2.2

Currently, I only derived the upper bound of (ii).

Lemma 2. Suppose $f \in \mathcal{H}(\alpha, L)$. Then

Name this Lemma. (Agnostic error)

$$\mathbb{E}\left[\delta^2(f_\Theta, f)\right] \leq \frac{C}{n^{\alpha \wedge 1}},$$

C \prop L^2 m^2

for some constant C only depending on constants m and L.

For all future results:
Make the dependence on m explicit
—> Our goal is to investigate the impact of higher-order (m) compared to m=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 $\left(n^m - \binom{n}{m} n!\right)/n^m = C/n$. Hence it suffices to prove that

$$\mathbb{E}\left[\delta^2(f_{\Theta'}, f)\right] \le \frac{C}{n^{\alpha \wedge 1}}.$$

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 $x \in \prod_{\ell=1}^m [(\sigma(i_\ell) - 1)/n, \sigma(i_\ell)/n]$,

$$\begin{aligned} |f(x_1, \ldots, x_m) - f(\xi_1, \ldots, \xi_m)|^2 \le & 2\left| f(x_1, \ldots, x_m) - f\left(\frac{\sigma(i_1)}{n+1}, \ldots \frac{\sigma(i_m)}{n+1}\right)\right|^2 \\ & + 2\left| f\left(\frac{\sigma(i_1)}{n+1}, \ldots \frac{\sigma(i_m)}{n+1}\right) - f(\xi_{(\sigma(i_1))}, \ldots, \xi_{(\sigma(i_m))})\right|^2 \\ & + 2\left| f(\xi_{(\sigma(i_1))}, \ldots, \xi_{(\sigma(i_m))}) - f(\xi_{i_1}, \ldots, \xi_{i_m})\right|^2, \end{aligned} \tag{5}$$

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 (5) is 0 almost surely.

For the first summand in (5), 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 (1), we obtain

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

Perfect!
We've extended already >50% results in Klopp's et al.

Integrating and taking expectation on the first summand yields,

$$\mathbb{E}\left[\sum_{i_1, \ldots, i_m = 1}^{n} \int_{(\sigma(i_1)-1)/n}^{\sigma(i_1)/n} \cdots \int_{(\sigma(i_m)-1)/n}^{\sigma(i_m)/n} \left| f(x_1, \ldots, x_m) - f\left(\frac{\sigma(i_1)}{n+1}, \ldots \frac{\sigma(i_m)}{n+1}\right)\right|^2 dx_1 \cdots dx_m\right] \le m^2 M^2 \left(\frac{1}{n}\right)^{2(\alpha \wedge 1)}. \tag{6}$$

With (1), the second summand on (5) is bounded,

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

Integrating and taking expectation on the second summand yields,

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i_1,\ldots,i_m=1}^{n}\int_{(\sigma(i_1)-1)/n}^{\sigma(i_1)/n}\cdots\int_{(\sigma(i_m)-1)/n}^{\sigma(i_m)/n}\left|f\left(\frac{\sigma(i_1)}{n+1},\ldots\frac{\sigma(i_m)}{n+1}\right)-f(\xi_{(\sigma(i_1))},\ldots,\xi_{(\sigma(i_m))})\right|^2 dx_1\cdots dx_m\right] \\
&\leq \left[\sum_{i_1,\ldots,i_m=1}^{n}\int_{(\sigma(i_1)-1)/n}^{\sigma(i_1)/n}\cdots\int_{(\sigma(i_m)-1)/n}^{\sigma(i_m)/n} 2M^2\sum_{\ell=1}^{m}\left|\frac{\sigma(i_\ell)}{n+1}-\xi_{(\sigma(i_\ell))}\right|^{2(\alpha\wedge 1)} dx_1\cdots dx_m\right] \\
&= 2mM^2\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left|\frac{i}{n+1}-\xi_{(i))}\right|^{2(\alpha\wedge 1)}\right] \\
&\leq 2mM^2\max_{i=1,\ldots,n}\mathbb{E}\left[\left|\frac{i}{n+1}-\xi_{(i)}\right|^{2(\alpha\wedge 1)}\right] \\
&\leq 2mM^2\max_{i=1\ldots,n}\left[\mathrm{Var}(\xi_{(i)})\right]^{\alpha\wedge 1}\leq C\left(\frac{1}{n}\right)^{\alpha\wedge 1}, \qquad (7)
\end{aligned}
$$

where we have used $\mathbb{E}(\xi_{(\ell)})=\ell/(n+1)$, $\mathrm{Var}(\xi_{(\ell)})\leq C/n$ and Jensen's inequality. Combining (6) and (7) proves the lemma. □

References

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