Supplement to "Clustering with Statistical Error Control"

Michael Vogt
University of Bonn

Matthias Schmid
University of Bonn

In this supplement, we provide the proofs that are omitted in the paper. In particular, we derive Theorems 4.1–4.3 from Section 4. Throughout the supplement, we use the symbol C to denote a universal real constant which may take a different value on each occurrence.

Auxiliary results

In the proofs of Theorems 4.1–4.3, we frequently make use of the following uniform convergence result.

Lemma S.1. *Let $\mathcal{Z}_s = \{Z_{st} : 1 \le t \le T\}$ be sequences of real-valued random variables for $1 \le s \le S$ with the following properties: (i) for each s, the random variables in $\mathcal{Z}_s$ are independent of each other, and (ii) $\mathbb{E}[Z_{st}] = 0$ and $\mathbb{E}[|Z_{st}|^{\phi}] \le C < \infty$ for some $\phi > 2$ and $C > 0$ that depend neither on s nor on t. Suppose that $S = T^q$ with $0 \le q < \phi/2 - 1$. Then*

$$\mathbb{P}\Big(\max_{1 \le s \le S} \Big| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{st} \Big| > T^{\eta} \Big) = o(1),$$

where the constant $\eta > 0$ can be chosen as small as desired.

Proof of Lemma S.1. Define $\tau_{S,T} = (ST)^{1/\{(2+\delta)(q+1)\}}$ with some sufficiently small $\delta > 0$. In particular, let $\delta > 0$ be so small that $(2+\delta)(q+1) < \phi$. Moreover, set

$$\begin{aligned} Z_{st}^{\le} &= Z_{st} \mathbf{1}(|Z_{st}| \le \tau_{S,T}) - \mathbb{E}\big[Z_{st} \mathbf{1}(|Z_{st}| \le \tau_{S,T}) \big] \\ Z_{st}^{>} &= Z_{st} \mathbf{1}(|Z_{st}| > \tau_{S,T}) - \mathbb{E}\big[Z_{st} \mathbf{1}(|Z_{st}| > \tau_{S,T}) \big] \end{aligned}$$

and write

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{st} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{st}^{\le} + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{st}^{>}.$$

In what follows, we show that

$$\mathbb{P}\Big(\max_{1\le s\le S}\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_{st}^{>}\Big| > CT^{\eta}\Big) = o(1) \tag{S.1}$$

$$\mathbb{P}\Big(\max_{1\le s\le S}\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_{st}^{\le}\Big| > CT^{\eta}\Big) = o(1) \tag{S.2}$$

for any fixed constant $C > 0$. Combining (S.1) and (S.2) immediately yields the statement of Lemma S.1.

We start with the proof of (S.1): It holds that

$$\mathbb{P}\Big(\max_{1\le s\le S}\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_{st}^{>}\Big| > CT^{\eta}\Big) \le Q_1^{>} + Q_2^{>},$$

where

$$\begin{aligned}
Q_1^{>} &:= \sum_{s=1}^{S}\mathbb{P}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}|Z_{st}|\mathbf{1}(|Z_{st}|>\tau_{S,T}) > \frac{C}{2}T^{\eta}\Big) \\
&\le \sum_{s=1}^{S}\mathbb{P}\Big(|Z_{st}|>\tau_{S,T} \text{ for some } 1\le t\le T\Big) \\
&\le \sum_{s=1}^{S}\sum_{t=1}^{T}\mathbb{P}\big(|Z_{st}|>\tau_{S,T}\big) \le \sum_{s=1}^{S}\sum_{t=1}^{T}\mathbb{E}\Big[\frac{|Z_{st}|^{\phi}}{\tau_{S,T}^{\phi}}\Big] \\
&\le \frac{CST}{\tau_{S,T}^{\phi}} = o(1)
\end{aligned}$$

and

$$Q_2^{>} := \sum_{s=1}^{S}\mathbb{P}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{E}\big[|Z_{st}|\mathbf{1}(|Z_{st}|>\tau_{S,T})\big] > \frac{C}{2}T^{\eta}\Big) = 0$$

for S and T sufficiently large, since

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{E}\big[|Z_{st}|\mathbf{1}(|Z_{st}|>\tau_{S,T})\big] \\
&\le \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{E}\Big[\frac{|Z_{st}|^{\phi}}{\tau_{S,T}^{\phi-1}}\mathbf{1}(|Z_{st}|>\tau_{S,T})\Big] \\
&\le \frac{C\sqrt{T}}{\tau_{S,T}^{\phi-1}} = o(T^{\eta}).
\end{aligned}$$

This yields (S.1).

We next turn to the proof of (S.2): We apply the crude bound

$$\mathbb{P}\Big(\max_{1 \le s \le S} \Big| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{st}^{\le} \Big| > CT^{\eta} \Big) \le \sum_{s=1}^{S} \mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{st}^{\le} \Big| > CT^{\eta} \Big)$$

and show that for any $1 \le s \le S$,

$$\mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{st}^{\le} \Big| > CT^{\eta} \Big) \le C_0 T^{-\rho}, \tag{S.3}$$

where C_0 is a fixed constant and $\rho > 0$ can be chosen as large as desired by picking η slightly larger than $1/2 - 1/(2+\delta)$. Since $S = O(T^q)$, this immediately implies (S.2). To prove (S.3), we make use of the following facts:

(i) For a random variable Z and $\lambda > 0$, Markov's inequality says that

$$\mathbb{P}\big(\pm Z > \delta \big) \le \frac{\mathbb{E} \exp(\pm \lambda Z)}{\exp(\lambda \delta)}.$$

(ii) Since $|Z_{st}^{\le}/\sqrt{T}| \le 2\tau_{S,T}/\sqrt{T}$, it holds that $\lambda_{S,T}|Z_{st}^{\le}/\sqrt{T}| \le 1/2$, where we set $\lambda_{S,T} = \sqrt{T}/(4\tau_{S,T})$. As $\exp(x) \le 1 + x + x^2$ for $|x| \le 1/2$, this implies that

$$\mathbb{E}\Big[\exp\Big(\pm \lambda_{S,T} \frac{Z_{st}^{\le}}{\sqrt{T}} \Big) \Big] \le 1 + \frac{\lambda_{S,T}^2}{T} \mathbb{E}\big[(Z_{st}^{\le})^2 \big] \le \exp\Big(\frac{\lambda_{S,T}^2}{T} \mathbb{E}\big[(Z_{st}^{\le})^2 \big] \Big).$$

(iii) By definition of $\lambda_{S,T}$, it holds that

$$\lambda_{S,T} = \frac{\sqrt{T}}{4(ST)^{\frac{1}{(2+\delta)(q+1)}}} = \frac{\sqrt{T}}{4(T^{q+1})^{\frac{1}{(2+\delta)(q+1)}}} = \frac{T^{\frac{1}{2} - \frac{1}{2+\delta}}}{4}.$$

Using (i)–(iii) and writing $\mathbb{E}(Z_{st}^{\le})^2 \le C_Z < \infty$, we obtain that

$$\begin{aligned}
&\mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{st}^{\le} \Big| > CT^{\eta} \Big) \\
&\le \mathbb{P}\Big(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{st}^{\le} > CT^{\eta} \Big) + \mathbb{P}\Big(- \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{st}^{\le} > CT^{\eta} \Big) \\
&\le \exp\big(- \lambda_{S,T} CT^{\eta} \big) \Big\{ \mathbb{E}\Big[\exp\Big(\lambda_{S,T} \sum_{t=1}^{T} \frac{Z_{st}^{\le}}{\sqrt{T}} \Big) \Big] + \mathbb{E}\Big[\exp\Big(- \lambda_{S,T} \sum_{t=1}^{T} \frac{Z_{st}^{\le}}{\sqrt{T}} \Big) \Big] \Big\} \\
&= \exp\big(- \lambda_{S,T} CT^{\eta} \big) \Big\{ \prod_{t=1}^{T} \mathbb{E}\Big[\exp\Big(\lambda_{S,T} \frac{Z_{st}^{\le}}{\sqrt{T}} \Big) \Big] + \prod_{t=1}^{T} \mathbb{E}\Big[\exp\Big(- \lambda_{S,T} \frac{Z_{st}^{\le}}{\sqrt{T}} \Big) \Big] \Big\}
\end{aligned}$$

$$
\begin{aligned}
&\leq 2\exp\big(-\lambda_{S,T} C T^{\eta}\big)\prod_{t=1}^{T}\exp\Big(\frac{\lambda_{S,T}^2}{T}\mathbb{E}\big[(Z_{st}^{\leq})^2\big]\Big)\\
&= 2\exp\big(C_Z\lambda_{S,T}^2 - C\lambda_{S,T}T^{\eta}\big)\\
&= 2\exp\Big(\frac{C_Z}{16}\big(T^{\frac{1}{2}-\frac{1}{2+\delta}}\big)^2 - \frac{C}{4}T^{\frac{1}{2}-\frac{1}{2+\delta}}\,T^{\eta}\Big)\\
&\leq C_0 T^{-\rho},
\end{aligned}
$$

where $\rho > 0$ can be chosen arbitrarily large if we pick η slightly larger than $1/2 - 1/(2+\delta)$. $\square$

Proof of Theorem 4.1

We first prove that

$$\mathbb{P}\Big(\widehat{\mathcal{H}}^{[K_0]} \leq q(\alpha)\Big) = (1-\alpha) + o(1). \tag{S.4}$$

To do so, we derive a stochastic expansion of the individual statistics $\widehat{\Delta}_i^{[K_0]}$.

Lemma S.2. *It holds that*

$$\widehat{\Delta}_i^{[K_0]} = \Delta_i^{[K_0]} + R_i^{[K_0]},$$

where

$$\Delta_i^{[K_0]} = \frac{1}{\sqrt{p}}\sum_{j=1}^{p}\Big\{\frac{\varepsilon_{ij}^2}{\sigma^2} - 1\Big\}\Big/\kappa$$

and the remainder $R_i^{[K_0]}$ has the property that

$$\mathbb{P}\Big(\max_{1\leq i\leq n}\big|R_i^{[K_0]}\big| > p^{-\xi}\Big) = o(1) \tag{S.5}$$

for some $\xi > 0$.

The proof of Lemma S.2 as well as those of the subsequent Lemmas S.3–S.5 are postponed until the proof of Theorem 4.1 is complete. With the help of Lemma S.2, we can bound the probability of interest

$$P_\alpha := \mathbb{P}\Big(\widehat{\mathcal{H}}^{[K_0]} \leq q(\alpha)\Big) = \mathbb{P}\Big(\max_{1\leq i\leq n}\widehat{\Delta}_i^{[K_0]} \leq q(\alpha)\Big)$$

as follows: Since

$$\max_{1\leq i\leq n}\widehat{\Delta}_i^{[K_0]}\begin{cases}\leq \max_{1\leq i\leq n}\Delta_i^{[K_0]} + \max_{1\leq i\leq n}|R_i^{[K_0]}|\\ \geq \max_{1\leq i\leq n}\Delta_i^{[K_0]} - \max_{1\leq i\leq n}|R_i^{[K_0]}|,\end{cases}$$

it holds that

$$P_\alpha^{<} \le P_\alpha \le P_\alpha^{>},$$

where

$$\begin{aligned} P_\alpha^{<} &= \mathbb{P}\Big(\max_{1 \le i \le n} \Delta_i^{[K_0]} \le q(\alpha) - \max_{1 \le i \le n} |R_i^{[K_0]}| \Big) \\ P_\alpha^{>} &= \mathbb{P}\Big(\max_{1 \le i \le n} \Delta_i^{[K_0]} \le q(\alpha) + \max_{1 \le i \le n} |R_i^{[K_0]}| \Big). \end{aligned}$$

As the remainder $R_i^{[K_0]}$ has the property (S.5), we further obtain that

$$P_\alpha^{\ll} + o(1) \le P_\alpha \le P_\alpha^{\gg} + o(1), \tag{S.6}$$

where

$$\begin{aligned} P_\alpha^{\ll} &= \mathbb{P}\Big(\max_{1 \le i \le n} \Delta_i^{[K_0]} \le q(\alpha) - p^{-\xi} \Big) \\ P_\alpha^{\gg} &= \mathbb{P}\Big(\max_{1 \le i \le n} \Delta_i^{[K_0]} \le q(\alpha) + p^{-\xi} \Big). \end{aligned}$$

With the help of strong approximation theory, we can derive the following result on the asymptotic behaviour of the probabilities $P_\alpha^{\ll}$ and $P_\alpha^{\gg}$.

Lemma S.3. *It holds that*

$$\begin{aligned} P_\alpha^{\ll} &= (1-\alpha) + o(1) \\ P_\alpha^{\gg} &= (1-\alpha) + o(1). \end{aligned}$$

Together with (S.6), this immediately yields that $P_\alpha = (1-\alpha) + o(1)$, thus completing the proof of (S.4).

We next show that for any $K < K_0$,

$$\mathbb{P}\Big(\widehat{\mathcal{H}}^{[K]} \le q(\alpha) \Big) = o(1). \tag{S.7}$$

Consider a fixed $K < K_0$ and let $S \in \{\widehat{G}_k^{[K]} : 1 \le k \le K\}$ be any cluster with the following property:

$$\begin{aligned} &\#S \ge \underline{n} := \min\nolimits_{1 \le k \le K_0} \#G_k, \text{ and } S \text{ contains elements from at least two} \\ &\text{different classes } G_{k_1} \text{ and } G_{k_2}. \end{aligned} \tag{S.8}$$

It is not difficult to see that a cluster with the property (S.8) must always exist under our conditions. By $\mathscr{C} \subseteq \{\widehat{G}_k^{[K]} : 1 \le k \le K\}$, we denote the collection of clusters that have the property (S.8). With this notation at hand, we can derive the following stochastic expansion of the individual statistics $\widehat{\Delta}_i^{[K]}$.

Lemma S.4. *For any $i \in S$ and $S \in \mathscr{C}$, it holds that*

$$\widehat{\Delta}_i^{[K]} = \frac{1}{\kappa\sigma^2\sqrt{p}} \sum_{j=1}^{p} d_{ij}^2 + R_i^{[K]},$$

where $d_{ij} = \mu_{ij} - (\#S)^{-1} \sum_{i' \in S} \mu_{i'j}$ and the remainder $R_i^{[K]}$ has the property that

$$\mathbb{P}\Big(\max_{S \in \mathscr{C}} \max_{i \in S} \big| R_i^{[K]} \big| > p^{\frac{1}{2}-\xi} \Big) = o(1) \tag{S.9}$$

for some small $\xi > 0$.

Using (S.9) and the fact that

$$\max_{S \in \mathscr{C}} \max_{i \in S} \widehat{\Delta}_i^{[K]} \geq \max_{S \in \mathscr{C}} \max_{i \in S} \Big\{ \frac{1}{\kappa\sigma^2\sqrt{p}} \sum_{j=1}^{p} d_{ij}^2 \Big\} - \max_{S \in \mathscr{C}} \max_{i \in S} \big| R_i^{[K]} \big|,$$

we obtain that

$$\begin{aligned}
\mathbb{P}\Big(\widehat{\mathcal{H}}^{[K]} \leq q(\alpha) \Big) &= \mathbb{P}\Big(\max_{1 \leq i \leq n} \widehat{\Delta}_i^{[K]} \leq q(\alpha) \Big) \\
&\leq \mathbb{P}\Big(\max_{S \in \mathscr{C}} \max_{i \in S} \widehat{\Delta}_i^{[K]} \leq q(\alpha) \Big) \\
&\leq \mathbb{P}\Big(\max_{S \in \mathscr{C}} \max_{i \in S} \Big\{ \frac{1}{\kappa\sigma^2\sqrt{p}} \sum_{j=1}^{p} d_{ij}^2 \Big\} - \max_{S \in \mathscr{C}} \max_{i \in S} \big| R_i^{[K]} \big| \leq q(\alpha) \Big) \\
&\leq \mathbb{P}\Big(\max_{S \in \mathscr{C}} \max_{i \in S} \Big\{ \frac{1}{\kappa\sigma^2\sqrt{p}} \sum_{j=1}^{p} d_{ij}^2 \Big\} \leq q(\alpha) + p^{\frac{1}{2}-\xi} \Big) + o(1). \qquad \text{(S.10)}
\end{aligned}$$

The arguments from the proof of Lemma S.3, in particular (S.22), imply that $q(\alpha) \leq C\sqrt{\log n}$ for some fixed constant $C > 0$ and sufficiently large n. Moreover, we can prove the following result.

Lemma S.5. *It holds that*

$$\max_{S \in \mathscr{C}} \max_{i \in S} \Big\{ \frac{1}{\sqrt{p}} \sum_{j=1}^{p} d_{ij}^2 \Big\} \geq c\sqrt{p}$$

for some fixed constant $c > 0$.

Since $q(\alpha) \leq C\sqrt{\log n}$ and $\sqrt{\log n}/\sqrt{p} = o(1)$ by (C3), Lemma S.5 allows us to infer that

$$\mathbb{P}\Big(\max_{S \in \mathscr{C}} \max_{i \in S} \Big\{ \frac{1}{\kappa\sigma^2\sqrt{p}} \sum_{j=1}^{p} d_{ij}^2 \Big\} \leq q(\alpha) + p^{\frac{1}{2}-\xi} \Big) = o(1).$$

Together with (S.10), this yields that $\mathbb{P}(\widehat{\mathcal{H}}^{[K]} \leq q(\alpha)) = o(1)$. □

Proof of Lemma S.2. Let $n_k = \#G_k$ and write $\overline{\varepsilon}_i = p^{-1}\sum_{j=1}^p \varepsilon_{ij}$ along with $\overline{\mu}_i = p^{-1}\sum_{j=1}^p \mu_{ij}$. Since

$$\mathbb{P}\Big(\big\{\widehat{G}_k^{[K_0]} : 1 \le k \le K_0\big\} = \big\{G_k : 1 \le k \le K_0\big\}\Big) \to 1$$

by (3.1), we can ignore the estimation error in the clusters $\widehat{G}_k^{[K_0]}$ and replace them by the true classes G_k. For $i \in G_k$, we thus get

$$\widehat{\Delta}_i^{[K_0]} = \Delta_i^{[K_0]} + R_{i,A}^{[K_0]} + R_{i,B}^{[K_0]} - R_{i,C}^{[K_0]} + R_{i,D}^{[K_0]},$$

where

$$\begin{aligned}
R_{i,A}^{[K_0]} &= \Big(\frac{1}{\widehat{\kappa}} - \frac{1}{\kappa}\Big)\frac{1}{\sqrt{p}}\sum_{j=1}^p \Big\{\frac{\varepsilon_{ij}^2}{\sigma^2} - 1\Big\} \\
R_{i,B}^{[K_0]} &= \frac{1}{\widehat{\kappa}}\Big(\frac{1}{\widehat{\sigma}^2} - \frac{1}{\sigma^2}\Big)\frac{1}{\sqrt{p}}\sum_{j=1}^p \varepsilon_{ij}^2 \\
R_{i,C}^{[K_0]} &= \Big(\frac{2}{\widehat{\kappa}\widehat{\sigma}^2}\Big)\frac{1}{\sqrt{p}}\sum_{j=1}^p \varepsilon_{ij}\Big\{\overline{\varepsilon}_i + \frac{1}{n_k}\sum_{i' \in G_k}\big(\varepsilon_{i'j} - \overline{\varepsilon}_{i'}\big)\Big\} \\
R_{i,D}^{[K_0]} &= \Big(\frac{1}{\widehat{\kappa}\widehat{\sigma}^2}\Big)\frac{1}{\sqrt{p}}\sum_{j=1}^p \Big\{\overline{\varepsilon}_i + \frac{1}{n_k}\sum_{i' \in G_k}\big(\varepsilon_{i'j} - \overline{\varepsilon}_{i'}\big)\Big\}^2.
\end{aligned}$$

We now show that $\max_{i \in G_k}|R_{i,\ell}^{[K_0]}| = o_p(p^{-\xi})$ for any k and $\ell = A, \ldots, D$. This implies that $\max_{1 \le i \le n}|R_{i,\ell}^{[K_0]}| = \max_{1 \le k \le K_0}\max_{i \in G_k}|R_{i,\ell}^{[K_0]}| = o_p(p^{-\xi})$ for $\ell = A, \ldots, D$, which in turn yields the statement of Lemma S.2. Throughout the proof, we use the symbol $\eta > 0$ to denote a sufficiently small constant which results from applying Lemma S.1.

By assumption, $\widehat{\sigma}^2 = \sigma^2 + O_p(p^{-(1/2+\delta)})$ and $\widehat{\kappa} = \kappa + O_p(p^{-\delta})$ for some $\delta > 0$. Applying Lemma S.1 and choosing $\xi > 0$ such that $\xi < \delta - \eta$, we obtain that

$$\begin{aligned}
\max_{i \in G_k}\big|R_{i,A}^{[K_0]}\big| &\le \Big|\frac{1}{\widehat{\kappa}} - \frac{1}{\kappa}\Big| \max_{i \in G_k}\Big|\frac{1}{\sqrt{p}}\sum_{j=1}^p \Big\{\frac{\varepsilon_{ij}^2}{\sigma^2} - 1\Big\}\Big| \\
&= \Big|\frac{1}{\widehat{\kappa}} - \frac{1}{\kappa}\Big| O_p(p^{\eta}) = O_p(p^{-(\delta-\eta)}) = o_p(p^{-\xi})
\end{aligned}$$

and

$$\begin{aligned}
\max_{i \in G_k}\big|R_{i,B}^{[K_0]}\big| &\le \Big|\frac{1}{\widehat{\kappa}}\Big(\frac{1}{\widehat{\sigma}^2} - \frac{1}{\sigma^2}\Big)\Big| \Big\{\max_{i \in G_k}\Big|\frac{1}{\sqrt{p}}\sum_{j=1}^p \big(\varepsilon_{ij}^2 - \sigma^2\big)\Big| + \sigma^2\sqrt{p}\Big\} \\
&= \Big|\frac{1}{\widehat{\kappa}}\Big(\frac{1}{\widehat{\sigma}^2} - \frac{1}{\sigma^2}\Big)\Big| \big\{O_p(p^{\eta}) + \sigma^2\sqrt{p}\big\} = o_p(p^{-\xi}).
\end{aligned}$$

We next show that

$$\max_{i\in G_k}\big|R_{i,C}^{[K_0]}\big| = o_p\big(p^{-\frac{1}{4}}\big). \tag{S.11}$$

To do so, we work with the decomposition $R_{i,C}^{[K_0]} = \{2\widehat{\kappa}^{-1}\widehat{\sigma}^{-2}\}\{R_{i,C,1}^{[K_0]}+R_{i,C,2}^{[K_0]}-R_{i,C,3}^{[K_0]}\}$, where

$$\begin{aligned}
R_{i,C,1}^{[K_0]} &= \frac{1}{\sqrt{p}}\sum_{j=1}^{p}\varepsilon_{ij}\overline{\varepsilon}_i \\
R_{i,C,2}^{[K_0]} &= \frac{1}{\sqrt{p}}\sum_{j=1}^{p}\varepsilon_{ij}\Big(\frac{1}{n_k}\sum_{i'\in G_k}\varepsilon_{i'j}\Big) \\
R_{i,C,3}^{[K_0]} &= \Big(\frac{1}{\sqrt{p}}\sum_{j=1}^{p}\varepsilon_{ij}\Big)\Big(\frac{1}{n_k}\sum_{i'\in G_k}\overline{\varepsilon}_{i'}\Big).
\end{aligned}$$

With the help of Lemma S.1, we obtain that

$$\max_{i\in G_k}\big|R_{i,C,1}^{[K_0]}\big| \le \frac{1}{\sqrt{p}}\Big(\max_{i\in G_k}\Big|\frac{1}{\sqrt{p}}\sum_{j=1}^{p}\varepsilon_{ij}\Big|\Big)^2 = O_p\Big(\frac{p^{2\eta}}{\sqrt{p}}\Big). \tag{S.12}$$

Moreover,

$$\max_{i\in G_k}\big|R_{i,C,2}^{[K_0]}\big| = O_p\big(n_k^{-\frac{1}{4}}\big), \tag{S.13}$$

since

$$R_{i,C,2}^{[K_0]} = \frac{1}{n_k\sqrt{p}}\sum_{j=1}^{p}\big\{\varepsilon_{ij}^2-\sigma^2\big\} + \sigma^2\frac{\sqrt{p}}{n_k} + \frac{1}{n_k}\sum_{\substack{i'\in G_k\\ i'\neq i}}\frac{1}{\sqrt{p}}\sum_{j=1}^{p}\varepsilon_{ij}\varepsilon_{i'j},$$

$p \ll n_k$ and

$$\max_{i\in G_k}\Big|\frac{1}{n_k\sqrt{p}}\sum_{j=1}^{p}\big\{\varepsilon_{ij}^2-\sigma^2\big\}\Big| = O_p\Big(\frac{p^{\eta}}{n_k}\Big) \tag{S.14}$$

$$\max_{i\in G_k}\Big|\frac{1}{n_k}\sum_{\substack{i'\in G_k\\ i'\neq i}}\frac{1}{\sqrt{p}}\sum_{j=1}^{p}\varepsilon_{ij}\varepsilon_{i'j}\Big| = O_p\big(n_k^{-\frac{1}{4}}\big). \tag{S.15}$$

(S.14) is an immediate consequence of Lemma S.1. (S.15) follows upon observing that for any constant $C_0 > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{i\in G_k}\Big|\frac{1}{n_k}\sum_{\substack{i'\in G_k\\ i'\neq i}}\frac{1}{\sqrt{p}}\sum_{j=1}^{p}\varepsilon_{ij}\varepsilon_{i'j}\Big| > \frac{C_0}{n_k^{1/4}}\Big) \\
&\le \sum_{i\in G_k}\mathbb{P}\Big(\Big|\frac{1}{n_k}\sum_{\substack{i'\in G_k\\ i'\neq i}}\frac{1}{\sqrt{p}}\sum_{j=1}^{p}\varepsilon_{ij}\varepsilon_{i'j}\Big| > \frac{C_0}{n_k^{1/4}}\Big)
\end{aligned}$$

$$
\begin{aligned}
&\le \sum_{i\in G_k} \mathbb{E}\Big\{\frac{1}{n_k}\sum_{\substack{i'\in G_k\\ i'\ne i}} \frac{1}{\sqrt{p}}\sum_{j=1}^{p} \varepsilon_{ij}\varepsilon_{i'j}\Big\}^4 \Big/ \Big\{\frac{C_0}{n_k^{1/4}}\Big\}^4 \\
&\le \sum_{i\in G_k} \Big\{\frac{1}{n_k^4 p^2} \sum_{\substack{i'_1,\dots,i'_4\in G_k\\ i'_1,\dots,i'_4\ne i}} \sum_{j_1,\dots,j_4=1}^{p} \mathbb{E}\big[\varepsilon_{i\,j_1}\dots\varepsilon_{i\,j_4}\varepsilon_{i'_1 j_1}\dots\varepsilon_{i'_4 j_4}\big]\Big\}\Big/\Big\{\frac{C_0}{n_k^{1/4}}\Big\}^4 \\
&\le \frac{C}{C_0^4},
\end{aligned}
$$

the last inequality resulting from the fact that the mean $\mathbb{E}[\varepsilon_{i\,j_1}\dots\varepsilon_{i\,j_4}\varepsilon_{i'_1 j_1}\dots\varepsilon_{i'_4 j_4}]$ can only be non-zero if some of the index pairs (i'_ℓ, j_ℓ) for $\ell = 1,\dots,4$ are identical. Finally, with the help of Lemma S.1, we get that

$$
\max_{i\in G_k}\big|R^{[K_0]}_{i,C,3}\big| \le \Big|\frac{1}{n_k}\sum_{i'\in G_k}\overline{\varepsilon}_{i'}\Big| \max_{i\in G_k}\Big|\frac{1}{\sqrt{p}}\sum_{j=1}^{p}\varepsilon_{ij}\Big| = O_p\Big(\frac{p^\eta}{\sqrt{n_k p}}\Big). \qquad \text{(S.16)}
$$

Combining (S.12), (S.13) and (S.16), we arrive at the statement (S.11) on the remainder $R^{[K_0]}_{i,C}$.

We finally show that

$$
\max_{i\in G_k}\big|R^{[K_0]}_{i,D}\big| = O_p\Big(\frac{p^{2\eta}}{\sqrt{p}}\Big). \qquad \text{(S.17)}
$$

For the proof, we write $R^{[K_0]}_{i,D} = \{\widehat{\kappa}^{-1}\widehat{\sigma}^{-2}\}\{R^{[K_0]}_{i,D,1} + R^{[K_0]}_{i,D,2}\}$, where

$$
\begin{aligned}
R^{[K_0]}_{i,D,1} &= \frac{1}{\sqrt{p}}\Big(\frac{1}{\sqrt{p}}\sum_{j=1}^{p}\varepsilon_{ij}\Big)^2 \\
R^{[K_0]}_{i,D,2} &= \frac{1}{\sqrt{p}}\sum_{j=1}^{p}\Big\{\frac{1}{n_k}\sum_{i'\in G_k}\big(\varepsilon_{i'j} - \overline{\varepsilon}_{i'}\big)\Big\}^2.
\end{aligned}
$$

With the help of Lemma S.1, we obtain that

$$
\max_{i\in G_k}\big|R^{[K_0]}_{i,D,1}\big| = O_p\Big(\frac{p^{2\eta}}{\sqrt{p}}\Big). \qquad \text{(S.18)}
$$

Moreover, straightforward calculations yield that

$$
\max_{i\in G_k}\big|R^{[K_0]}_{i,D,2}\big| = O_p\Big(\frac{\sqrt{p}}{n_k}\Big). \qquad \text{(S.19)}
$$

(S.17) now follows upon combining (S.18) and (S.19). □

Proof of Lemma S.3. We make use of the following three results:

(R1) Let $\{W_i : 1 \le i \le n\}$ be independent random variables with a standard normal distribution and define $a_n = 1/\sqrt{2\log n}$ together with

$$b_n = \sqrt{2\log n} - \frac{\log\log n + \log(4\pi)}{2\sqrt{2\log n}}.$$

Then for any $w \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\Big(\max_{1\le i\le n} W_i \le a_n w + b_n \Big) = \exp(-\exp(-w)).$$

In particular, for $w(\alpha \pm \varepsilon) = -\log(-\log(1-\alpha\pm\varepsilon))$, we get

$$\lim_{n\to\infty} \mathbb{P}\Big(\max_{1\le i\le n} W_i \le a_n w(\alpha\pm\varepsilon) + b_n \Big) = 1-\alpha\pm\varepsilon.$$

The next result is known as Khintchine's Theorem.

(R2) Let F_n be distribution functions and G a non-degenerate distribution function. Moreover, let $\alpha_n > 0$ and $\beta_n \in \mathbb{R}$ be such that

$$F_n(\alpha_n x + \beta_n) \to G(x)$$

for any continuity point x of G. Then there are constants $\alpha_n' > 0$ and $\beta_n' \in \mathbb{R}$ as well as a non-degenerate distribution function G_* such that

$$F_n(\alpha_n' x + \beta_n') \to G_*(x)$$

at any continuity point x of G_* if and only if

$$\alpha_n^{-1}\alpha_n' \to \alpha_*, \quad \frac{\beta_n' - \beta_n}{\alpha_n} \to \beta_* \quad \text{and} \quad G_*(x) = G(\alpha_* x + \beta_*).$$

The final result exploits strong approximation theory and is a direct consequence of the so-called KMT Theorems; see Komlós et al. (1975, 1976):

(R3) Write

$$\Delta_i^{[K_0]} = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} X_{ij} \quad \text{with } X_{ij} = \Big\{ \frac{\varepsilon_{ij}^2}{\sigma^2} - 1 \Big\} \Big/ \kappa$$

and let F denote the distribution function of X_{ij}. It is possible to construct i.i.d. random variables $\{\widetilde{X}_{ij} : 1 \le i \le n, 1 \le j \le p\}$ with the distribution function F and independent standard normal random variables $\{Z_{ij} : 1 \le i \le$

n, $1 \leq j \leq p\}$ such that

$$\widetilde{\Delta}_i^{[K_0]} = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \widetilde{X}_{ij} \quad \text{and} \quad \Delta_i^* = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} Z_{ij}$$

have the following property:

$$\mathbb{P}\Big(\big| \widetilde{\Delta}_i^{[K_0]} - \Delta_i^* \big| > C p^{\frac{1}{2+\delta} - \frac{1}{2}} \Big) \leq p^{1 - \frac{\theta/2}{2+\delta}}$$

for some arbitrarily small but fixed $\delta > 0$ and some constant $C > 0$ that does not depend on i, p and n.

We now proceed as follows:

(i) We show that for any $w \in \mathbb{R}$,

$$\mathbb{P}\Big(\max_{1 \leq i \leq n} \Delta_i^{[K_0]} \leq a_n w + b_n \Big) \to \exp(-\exp(-w)). \tag{S.20}$$

This in particular implies that

$$\mathbb{P}\Big(\max_{1 \leq i \leq n} \Delta_i^{[K_0]} \leq w_n(\alpha \pm \varepsilon) \Big) \to 1 - \alpha \pm \varepsilon, \tag{S.21}$$

where $w_n(\alpha \pm \varepsilon) = a_n w(\alpha \pm \varepsilon) + b_n$ with a_n, b_n and $w(\alpha \pm \varepsilon)$ as defined in (R1). The proof of (S.20) is postponed until the arguments for Lemma S.3 are complete.

(ii) The statement (S.21) in particular holds in the special case that $\varepsilon_{ij} \sim N(0, \sigma^2)$. In this case, $q(\alpha)$ is the $(1-\alpha)$-quantile of $\max_{1 \leq i \leq n} \Delta_i^{[K_0]}$. Hence, we have

$$\begin{aligned}
&\mathbb{P}\Big(\max_{1 \leq i \leq n} \Delta_i^{[K_0]} \leq w_n(\alpha - \varepsilon) \Big) \to 1 - \alpha - \varepsilon \\
&\mathbb{P}\Big(\max_{1 \leq i \leq n} \Delta_i^{[K_0]} \leq q(\alpha) \Big) = 1 - \alpha \\
&\mathbb{P}\Big(\max_{1 \leq i \leq n} \Delta_i^{[K_0]} \leq w_n(\alpha + \varepsilon) \Big) \to 1 - \alpha + \varepsilon,
\end{aligned}$$

which implies that

$$w_n(\alpha - \varepsilon) \leq q(\alpha) \leq w_n(\alpha + \varepsilon) \tag{S.22}$$

for sufficiently large n.

(iii) Since $p^{-\xi}/a_n = p^{-\xi}\sqrt{2 \log n} = o(1)$ by (C3), we can use (S.20) together with (R2) to obtain that

$$\mathbb{P}\Big(\max_{1 \leq i \leq n} \Delta_i^{[K_0]} \leq w_n(\alpha \pm \varepsilon) \pm p^{-\xi} \Big) \to 1 - \alpha \pm \varepsilon. \tag{S.23}$$

As $w_n(\alpha-\varepsilon)-p^{-\xi} \leq q(\alpha)-p^{-\xi} \leq q(\alpha)+p^{-\xi} \leq w_n(\alpha+\varepsilon)+p^{-\xi}$ for sufficiently large n, it holds that

$$\begin{aligned}
P_{\alpha,\varepsilon}^{\ll} &:= \mathbb{P}\Big(\max_{1\leq i\leq n} \Delta_i^{[K_0]} \leq w_n(\alpha-\varepsilon) - p^{-\xi} \Big) \\
&\leq P_{\alpha}^{\ll} = \mathbb{P}\Big(\max_{1\leq i\leq n} \Delta_i^{[K_0]} \leq q(\alpha) - p^{-\xi} \Big) \\
&\leq P_{\alpha}^{\gg} = \mathbb{P}\Big(\max_{1\leq i\leq n} \Delta_i^{[K_0]} \leq q(\alpha) + p^{-\xi} \Big) \\
&\leq P_{\alpha,\varepsilon}^{\gg} := \mathbb{P}\Big(\max_{1\leq i\leq n} \Delta_i^{[K_0]} \leq w_n(\alpha+\varepsilon) + p^{-\xi} \Big)
\end{aligned}$$

for large n. Moreover, since $P_{\alpha,\varepsilon}^{\ll} \to 1-\alpha-\varepsilon$ and $P_{\alpha,\varepsilon}^{\gg} \to 1-\alpha+\varepsilon$ for any fixed $\varepsilon > 0$ by (S.23), we can conclude that $P_{\alpha}^{\ll} = (1-\alpha)+o(1)$ and $P_{\alpha}^{\gg} = (1-\alpha)+o(1)$, which is the statement of Lemma S.3.

It remains to prove (S.20): Using the notation from (R3) and the shorthand $w_n = a_n w + b_n$, we can write

$$\mathbb{P}\Big(\max_{1\leq i\leq n} \Delta_i^{[K_0]} \leq w_n \Big) = \mathbb{P}\Big(\max_{1\leq i\leq n} \widetilde{\Delta}_i^{[K_0]} \leq w_n \Big) = \prod_{i=1}^{n} \pi_i \tag{S.24}$$

with

$$\pi_i = \mathbb{P}\Big(\widetilde{\Delta}_i^{[K_0]} \leq w_n \Big).$$

The probabilities π_i can be decomposed into two parts as follows:

$$\pi_i = \mathbb{P}\Big(\Delta_i^* \leq w_n + \big\{ \Delta_i^* - \widetilde{\Delta}_i^{[K_0]} \big\} \Big) = \pi_i^{\leq} + \pi_i^{>},$$

where

$$\begin{aligned}
\pi_i^{\leq} &= \mathbb{P}\Big(\Delta_i^* \leq w_n + \big\{ \Delta_i^* - \widetilde{\Delta}_i^{[K_0]} \big\}, \big| \Delta_i^* - \widetilde{\Delta}_i^{[K_0]} \big| \leq C p^{\frac{1}{2+\delta}-\frac{1}{2}} \Big) \\
\pi_i^{>} &= \mathbb{P}\Big(\Delta_i^* \leq w_n + \big\{ \Delta_i^* - \widetilde{\Delta}_i^{[K_0]} \big\}, \big| \Delta_i^* - \widetilde{\Delta}_i^{[K_0]} \big| > C p^{\frac{1}{2+\delta}-\frac{1}{2}} \Big).
\end{aligned}$$

With the help of (R3) and the assumption that $n \ll p^{(\theta/4)-1}$, we can show that

$$\prod_{i=1}^{n} \pi_i = \prod_{i=1}^{n} \pi_i^{\leq} + R_n, \tag{S.25}$$

where R_n is a non-negative remainder term with

$$R_n \leq \sum_{i=1}^{n} \binom{n}{i} \Big(p^{1-\frac{\theta/2}{2+\delta}} \Big)^i = o(1).$$

Moreover, the probabilities $\pi_i^{\leq}$ can be bounded by

$$\pi_i^{\leq} \begin{cases} \leq \mathbb{P}\Big(\Delta_i^* \leq w_n + Cp^{\frac{1}{2+\delta}-\frac{1}{2}}\Big) \\ \geq \mathbb{P}\Big(\Delta_i^* \leq w_n - Cp^{\frac{1}{2+\delta}-\frac{1}{2}}\Big) - p^{1-\frac{\theta/2}{2+\delta}}, \end{cases}$$

the second line making use of (R3). From this, we obtain that

$$\prod_{i=1}^n \pi_i^{\leq} \begin{cases} \leq \overline{\Pi}_n \\ \geq \underline{\Pi}_n + o(1), \end{cases} \tag{S.26}$$

where

$$\begin{aligned} \overline{\Pi}_n &= \mathbb{P}\Big(\max_{1\leq i\leq n} \Delta_i^* \leq w_n + Cp^{\frac{1}{2+\delta}-\frac{1}{2}}\Big) \\ \underline{\Pi}_n &= \mathbb{P}\Big(\max_{1\leq i\leq n} \Delta_i^* \leq w_n - Cp^{\frac{1}{2+\delta}-\frac{1}{2}}\Big). \end{aligned}$$

By combining (S.24)–(S.26), we arrive at the intermediate result that

$$\underline{\Pi}_n + o(1) \leq \mathbb{P}\Big(\max_{1\leq i\leq n} \Delta_i^{[K_0]} \leq w_n\Big) \leq \overline{\Pi}_n + o(1). \tag{S.27}$$

Since $p^{\frac{1}{2+\delta}-\frac{1}{2}}/a_n = p^{\frac{1}{2+\delta}-\frac{1}{2}}\sqrt{2\log n} = o(1)$, we can use (R1) together with (R2) to show that

$$\overline{\Pi}_n \to \exp(-\exp(-w)) \qquad \text{and} \qquad \underline{\Pi}_n \to \exp(-\exp(-w)). \tag{S.28}$$

Plugging (S.28) into (S.27) immediately yields that

$$\mathbb{P}\Big(\max_{1\leq i\leq n} \Delta_i^{[K_0]} \leq w_n\Big) \to \exp(-\exp(-w)),$$

which completes the proof. □

Proof of Lemma S.4. We use the notation $n_S = \#S$ along with $\overline{\varepsilon}_i = p^{-1}\sum_{j=1}^p \varepsilon_{ij}$, $\overline{\mu}_i = p^{-1}\sum_{j=1}^p \mu_{ij}$ and $\overline{d}_i = p^{-1}\sum_{j=1}^p d_{ij}$. For any $i \in S$ and $S \in \mathscr{C}$, we can write

$$\widehat{\Delta}_i^{[K]} = \frac{1}{\kappa\sigma^2\sqrt{p}}\sum_{j=1}^p d_{ij}^2 + R_{i,A}^{[K]} + R_{i,B}^{[K]} + R_{i,C}^{[K]} + R_{i,D}^{[K]} - R_{i,E}^{[K]} + R_{i,F}^{[K]} + R_{i,G}^{[K]},$$

where

$$R_{i,A}^{[K]} = \Big(\frac{1}{\widehat{\kappa}\widehat{\sigma}^2} - \frac{1}{\kappa\sigma^2}\Big)\frac{1}{\sqrt{p}}\sum_{j=1}^p d_{ij}^2$$

$$R_{i,B}^{[K]} = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \Big\{ \frac{\varepsilon_{ij}^2}{\sigma^2} - 1 \Big\} \Big/ \kappa$$

$$R_{i,C}^{[K]} = \Big(\frac{1}{\widehat{\kappa}} - \frac{1}{\kappa} \Big) \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \Big\{ \frac{\varepsilon_{ij}^2}{\sigma^2} - 1 \Big\}$$

$$R_{i,D}^{[K]} = \frac{1}{\widehat{\kappa}} \Big(\frac{1}{\widehat{\sigma}^2} - \frac{1}{\sigma^2} \Big) \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \varepsilon_{ij}^2$$

$$R_{i,E}^{[K]} = \Big(\frac{2}{\widehat{\kappa}\widehat{\sigma}^2} \Big) \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \varepsilon_{ij} \Big\{ \overline{\varepsilon}_i + \frac{1}{n_S} \sum_{i' \in S} \big(\varepsilon_{i'j} - \overline{\varepsilon}_{i'} \big) \Big\}$$

$$R_{i,F}^{[K]} = \Big(\frac{1}{\widehat{\kappa}\widehat{\sigma}^2} \Big) \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \Big\{ \overline{\varepsilon}_i + \frac{1}{n_S} \sum_{i' \in S} \big(\varepsilon_{i'j} - \overline{\varepsilon}_{i'} \big) \Big\}^2$$

$$R_{i,G}^{[K]} = \Big(\frac{2}{\widehat{\kappa}\widehat{\sigma}^2} \Big) \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \Big\{ \varepsilon_{ij} - \overline{\varepsilon}_i - \frac{1}{n_S} \sum_{i' \in S} \big(\varepsilon_{i'j} - \overline{\varepsilon}_{i'} \big) \Big\} d_{ij}.$$

We now show that $\max_{S \in \mathscr{C}} \max_{i \in S} |R_{i,\ell}^{[K]}| = o_p(p^{1/2-\xi})$ for $\ell = A, \dots, G$. This immediately yields the statement of Lemma S.4. Throughout the proof, $\eta > 0$ denotes a sufficiently small constant that results from applying Lemma S.1.

With the help of Lemma S.1 and our assumptions on $\widehat{\sigma}^2$ and $\widehat{\kappa}$, it is straightforward to see that $\max_{S \in \mathscr{C}} \max_{i \in S} |R_{i,\ell}^{[K]}| \le \max_{1 \le i \le n} |R_{i,\ell}^{[K]}| = o_p(p^{1/2-\xi})$ for $\ell = A, B, C, D$ with some sufficiently small $\xi > 0$. We next show that

$$\max_{S \in \mathscr{C}} \max_{i \in S} \big| R_{i,E}^{[K]} \big| = O_p\big(p^{\eta}\big). \tag{S.29}$$

To do so, we write $R_{i,E}^{[K]} = \{2\widehat{\kappa}^{-1}\widehat{\sigma}^{-2}\}\{R_{i,E,1}^{[K]} + R_{i,E,2}^{[K]} - R_{i,E,3}^{[K]}\}$, where

$$R_{i,E,1}^{[K]} = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \varepsilon_{ij} \overline{\varepsilon}_i$$

$$R_{i,E,2}^{[K]} = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \varepsilon_{ij} \Big(\frac{1}{n_S} \sum_{i' \in S} \varepsilon_{i'j} \Big)$$

$$R_{i,E,3}^{[K]} = \Big(\frac{1}{\sqrt{p}} \sum_{j=1}^{p} \varepsilon_{ij} \Big) \Big(\frac{1}{n_S} \sum_{i' \in S} \overline{\varepsilon}_{i'} \Big).$$

Lemma S.1 yields that $\max_{S \in \mathscr{C}} \max_{i \in S} |R_{i,E,1}^{[K]}| \le \max_{1 \le i \le n} |R_{i,E,1}^{[K]}| = O_p(p^{2\eta}/\sqrt{p})$. Moreover, it holds that

$$\max_{S \in \mathscr{C}} \max_{i \in S} \big| R_{i,E,2}^{[K]} \big| = O_p\big(p^{\eta}\big),$$

since

$$R^{[K]}_{i,E,2} = \frac{1}{n_S\sqrt{p}} \sum_{j=1}^{p} \big\{ \varepsilon_{ij}^2 - \sigma^2 \big\} + \sigma^2 \frac{\sqrt{p}}{n_S} + \frac{1}{n_S} \sum_{\substack{i' \in S \\ i' \neq i}} \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \varepsilon_{ij}\varepsilon_{i'j}$$

and

$$\begin{aligned}
\max_{S \in \mathscr{C}} \max_{i \in S} \Big| \frac{1}{n_S\sqrt{p}} \sum_{j=1}^{p} \big\{ \varepsilon_{ij}^2 - \sigma^2 \big\} \Big| &\le \frac{1}{\underline{n}} \max_{1 \le i \le n} \Big| \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \big\{ \varepsilon_{ij}^2 - \sigma^2 \big\} \Big| = O_p\Big(\frac{p^\eta}{\underline{n}} \Big) \\
\max_{S \in \mathscr{C}} \max_{i \in S} \Big| \frac{1}{n_S} \sum_{\substack{i' \in S \\ i' \neq i}} \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \varepsilon_{ij}\varepsilon_{i'j} \Big| &\le \max_{1 \le i < i' \le n} \Big| \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \varepsilon_{ij}\varepsilon_{i'j} \Big| = O_p\big(p^\eta\big),
\end{aligned}$$

which follows upon applying Lemma S.1. Finally,

$$\max_{S \in \mathscr{C}} \max_{i \in S} \big| R^{[K]}_{i,E,3} \big| \le \frac{1}{\sqrt{p}} \Big\{ \max_{1 \le i \le n} \Big| \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \varepsilon_{ij} \Big| \Big\}^2 = O_p\Big(\frac{p^{2\eta}}{\sqrt{p}} \Big),$$

which can again be seen by applying Lemma S.1. Putting everything together, we arrive at (S.29). Similar arguments show that

$$\max_{S \in \mathscr{C}} \max_{i \in S} \big| R^{[K]}_{i,F} \big| = O_p\big(p^\eta\big) \tag{S.30}$$

as well.

To analyze the term $R^{[K]}_{i,G}$, we denote the signal vector of the group G_k by $\boldsymbol{m}_k = (m_{1,k}, \ldots, m_{p,k})^\top$ and write

$$\frac{1}{n_S} \sum_{i \in S} \mu_{ij} = \sum_{k=1}^{K_0} \lambda_{S,k} m_{j,k}$$

with $\lambda_{S,k} = \#(S \cap G_k)/n_S$. With this notation, we get

$$R^{[K]}_{i,G} = \{2\widehat{\kappa}^{-1}\widehat{\sigma}^{-2}\}\{R^{[K]}_{i,G,1} - R^{[K]}_{i,G,2} - R^{[K]}_{i,G,3} - R^{[K]}_{i,G,4} + R^{[K]}_{i,G,5}\},$$

where

$$\begin{aligned}
R^{[K]}_{i,G,1} &= \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \mu_{ij}\varepsilon_{ij} \\
R^{[K]}_{i,G,2} &= \sum_{k=1}^{K_0} \lambda_{S,k} \frac{1}{\sqrt{p}} \sum_{j=1}^{p} m_{j,k}\varepsilon_{ij} \\
R^{[K]}_{i,G,3} &= \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \overline{\varepsilon}_i d_{ij}
\end{aligned}$$

$$R^{[K]}_{i,G,4} = \frac{1}{n_S}\sum_{i'\in S}\frac{1}{\sqrt{p}}\sum_{j=1}^{p}\mu_{ij}\varepsilon_{i'j}$$

$$R^{[K]}_{i,G,5} = \frac{1}{n_S}\sum_{i'\in S}\sum_{k=1}^{K_0}\lambda_{S,k}\frac{1}{\sqrt{p}}\sum_{j=1}^{p}m_{j,k}\varepsilon_{i'j}.$$

With the help of Lemma S.1, it can be shown that $\max_{S\in\mathscr{C}}\max_{i\in S}\big|R^{[K]}_{i,G,\ell}\big| = O_p(p^\eta)$ for $\ell = 1,\ldots,5$. For example, it holds that

$$\max_{S\in\mathscr{C}}\max_{i\in S}\big|R^{[K]}_{i,G,4}\big| \le \max_{1\le i<i'\le n}\Big|\frac{1}{\sqrt{p}}\sum_{j=1}^{p}\mu_{ij}\varepsilon_{i'j}\Big| = O_p(p^\eta).$$

As a result, we obtain that

$$\max_{S\in\mathscr{C}}\max_{i\in S}\big|R^{[K]}_{i,G}\big| = O_p\big(p^\eta\big). \tag{S.31}$$

This completes the proof. □

Proof of Lemma S.5. Let $S\in\mathscr{C}$. In particular, suppose that $S\cap G_{k_1}\neq\emptyset$ and $S\cap G_{k_2}\neq\emptyset$ for some $k_1\neq k_2$. We show the following claim: there exists some $i\in S$ such that

$$\frac{1}{\sqrt{p}}\sum_{j=1}^{p}d_{ij}^2 \ge c\sqrt{p}, \tag{S.32}$$

where $c = (\sqrt{\delta_0}/2)^2$ with δ_0 defined in assumption (C2). From this, the statement of Lemma S.5 immediately follows.

For the proof of (S.32), we denote the Euclidean distance between vectors $v = (v_1,\ldots,v_p)^\top$ and $w = (w_1,\ldots,w_p)^\top$ by $d(v,w) = (\sum_{j=1}^{p}|v_j - w_j|^2)^{1/2}$. Moreover, as in Lemma S.4, we use the notation

$$\frac{1}{n_S}\sum_{i\in S}\mu_{ij} = \sum_{k=1}^{K_0}\lambda_{S,k}m_{j,k},$$

where $n_S = \#S$, $\lambda_{S,k} = \#(S\cap G_k)/n_S$ and $\boldsymbol{m}_k = (m_{1,k},\ldots,m_{p,k})^\top$ is the signal vector of the class G_k.

Take any $i\in S\cap G_{k_1}$. If

$$d\Big(\boldsymbol{\mu}_i,\sum_{k=1}^{K_0}\lambda_{S,k}\boldsymbol{m}_k\Big) = d\Big(\boldsymbol{m}_{k_1},\sum_{k=1}^{K_0}\lambda_{S,k}\boldsymbol{m}_k\Big) \ge \frac{\sqrt{\delta_0 p}}{2},$$

the proof is finished, as (S.32) is satisfied for i. Next consider the case that

$$d\Big(\boldsymbol{m}_{k_1}, \sum_{k=1}^{K_0} \lambda_{S,k} \boldsymbol{m}_k\Big) < \frac{\sqrt{\delta_0 p}}{2}.$$

By assumption (C2), it holds that $d(\boldsymbol{m}_k, \boldsymbol{m}_{k'}) \geq \sqrt{\delta_0 p}$ for $k \neq k'$. Hence, by the triangle inequality,

$$\begin{aligned}
\sqrt{\delta_0 p} &\leq d\big(\boldsymbol{m}_{k_1}, \boldsymbol{m}_{k_2}\big) \\
&\leq d\Big(\boldsymbol{m}_{k_1}, \sum_{k=1}^{K_0} \lambda_{S,k} \boldsymbol{m}_k\Big) + d\Big(\sum_{k=1}^{K_0} \lambda_{S,k} \boldsymbol{m}_k, \boldsymbol{m}_{k_2}\Big) \\
&< \frac{\sqrt{\delta_0 p}}{2} + d\Big(\sum_{k=1}^{K_0} \lambda_{S,k} \boldsymbol{m}_k, \boldsymbol{m}_{k_2}\Big),
\end{aligned}$$

implying that

$$d\Big(\sum_{k=1}^{K_0} \lambda_{S,k} \boldsymbol{m}_k, \boldsymbol{m}_{k_2}\Big) > \frac{\sqrt{\delta_0 p}}{2}.$$

This shows that the claim (S.32) is fulfilled for any $i' \in S \cap G_{k_2}$. □

Proof of Theorem 4.2

By Theorem 4.1,

$$\begin{aligned}
&\mathbb{P}\big(\widehat{K}_0 > K_0\big) \\
&= \mathbb{P}\Big(\widehat{\mathcal{H}}^{[K]} > q(\alpha) \text{ for all } K \leq K_0\Big) \\
&= \mathbb{P}\Big(\widehat{\mathcal{H}}^{[K_0]} > q(\alpha)\Big) - \mathbb{P}\Big(\widehat{\mathcal{H}}^{[K_0]} > q(\alpha), \widehat{\mathcal{H}}^{[K]} \leq q(\alpha) \text{ for some } K < K_0\Big) \\
&= \mathbb{P}\Big(\widehat{\mathcal{H}}^{[K_0]} > q(\alpha)\Big) + o(1) \\
&= \alpha + o(1)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}\big(\widehat{K}_0 < K_0\big) &= \mathbb{P}\Big(\widehat{\mathcal{H}}^{[K]} \leq q(\alpha) \text{ for some } K < K_0\Big) \\
&\leq \sum_{K=1}^{K_0-1} \mathbb{P}\Big(\widehat{\mathcal{H}}^{[K]} \leq q(\alpha)\Big) \\
&= o(1).
\end{aligned}$$

Moreover,

$$
\begin{aligned}
&\mathbb{P}\Big(\big\{\widehat{G}_k : 1 \le k \le \widehat{K}_0\big\} \ne \big\{G_k : 1 \le k \le K_0\big\}\Big) \\
&= \mathbb{P}\Big(\big\{\widehat{G}_k : 1 \le k \le \widehat{K}_0\big\} \ne \big\{G_k : 1 \le k \le K_0\big\}, \widehat{K}_0 = K_0\Big) \\
&\quad + \mathbb{P}\Big(\big\{\widehat{G}_k : 1 \le k \le \widehat{K}_0\big\} \ne \big\{G_k : 1 \le k \le K_0\big\}, \widehat{K}_0 \ne K_0\Big) \\
&= \alpha + o(1),
\end{aligned}
$$

since

$$
\begin{aligned}
&\mathbb{P}\Big(\big\{\widehat{G}_k : 1 \le k \le \widehat{K}_0\big\} \ne \big\{G_k : 1 \le k \le K_0\big\}, \widehat{K}_0 = K_0\Big) \\
&= \mathbb{P}\Big(\big\{\widehat{G}_k^{[K_0]} : 1 \le k \le K_0\big\} \ne \big\{G_k : 1 \le k \le K_0\big\}, \widehat{K}_0 = K_0\Big) \\
&\le \mathbb{P}\Big(\big\{\widehat{G}_k^{[K_0]} : 1 \le k \le K_0\big\} \ne \big\{G_k : 1 \le k \le K_0\big\}\Big) \\
&= o(1)
\end{aligned}
$$

by the consistency property (3.1) and

$$
\begin{aligned}
&\mathbb{P}\Big(\big\{\widehat{G}_k : 1 \le k \le \widehat{K}_0\big\} \ne \big\{G_k : 1 \le k \le K_0\big\}, \widehat{K}_0 \ne K_0\Big) \\
&= \mathbb{P}\big(\widehat{K}_0 \ne K_0\big) = \alpha + o(1).
\end{aligned}
$$

□

Proof of Theorem 4.3

With the help of Lemma S.1, we can show that

$$
\widehat{\rho}(i, i') = 2\sigma^2 + \frac{1}{p}\sum_{j=1}^{p}\big(\mu_{ij} - \mu_{i'j}\big)^2 + o_p(1) \tag{S.33}
$$

uniformly over i and i'. This together with (C2) allows us to prove the following claim:

> With probability tending to 1, the indices $i_1, \ldots, i_K$ belong to K different classes in the case that $K \le K_0$ and to K_0 different classes in the case that $K > K_0$. (S.34)

Now let $K = K_0$. With the help of (S.33) and (S.34), the starting values $\mathscr{C}_1^{[K_0]}, \ldots$ $\ldots, \mathscr{C}_{K_0}^{[K_0]}$ can be shown to have the property that

$$
\mathbb{P}\Big(\big\{\mathscr{C}_k^{[K_0]} : 1 \le k \le K_0\big\} = \big\{G_k : 1 \le k \le K_0\big\}\Big) \to 1. \tag{S.35}
$$

Together with Lemma S.1, (S.35) yields that

$$\widehat{\rho}_k^{(1)}(i) = \sigma^2 + \frac{1}{p}\sum_{j=1}^{p}\big(\mu_{ij} - m_{j,k}\big)^2 + o_p(1)$$

uniformly over i and k. Combined with (C2), this in turn implies that the k-means algorithm converges already after the first iteration step with probability tending to 1 and $\widehat{G}_k^{[K_0]}$ are consistent estimators of the classes G_k in the sense of (3.1). □

Proof of (3.16)

Suppose that (C1)–(C3) along with (3.15) are satisfied. As already noted in Section 3.4, the k-means estimators $\{\widehat{G}_k^A : 1 \le k \le K_{\max}\}$ can be shown to satisfy (3.14), that is,

$$\mathbb{P}\Big(\widehat{G}_k^A \subseteq G_{k'} \text{ for some } 1 \le k' \le K_0\Big) \to 1 \tag{S.36}$$

for any $k = 1, \dots, K_{\max}$. This can be proven by very similar arguments as the consistency property (3.1). We thus omit the details. Let E^A be the event that

$$\widehat{G}_k^A \subseteq G_{k'} \text{ for some } 1 \le k' \le K_0$$

holds for all clusters $\widehat{G}_k^A$ with $k = 1, \dots, K_{\max}$. E^A can be regarded as the event that the partition $\{\widehat{G}_k^A : 1 \le k \le K_{\max}\}$ is a refinement of the class structure $\{G_k : 1 \le k \le K_0\}$. By (S.36), the event E^A occurs with probability tending to 1.

Now consider the estimator

$$\widehat{\sigma}^2_{\text{RSS}} = \frac{1}{n\lfloor p/2\rfloor}\sum_{k=1}^{K_{\max}}\sum_{i\in\widehat{G}_k^A}\Big\|\widehat{\boldsymbol{Y}}_i^B - \frac{1}{\#\widehat{G}_k^A}\sum_{i'\in\widehat{G}_k^A}\widehat{\boldsymbol{Y}}_{i'}^B\Big\|^2.$$

Since the random variables $\widehat{\boldsymbol{Y}}_i^B$ are independent of the estimators $\widehat{G}_k^A$, it is not difficult to verify the following: for any $\delta > 0$, there exists a constant $C_\delta > 0$ (that does not depend on $\{\widehat{G}_k^A : 1 \le k \le K_{\max}\}$) such that on the event E^A,

$$\mathbb{P}\Big(\big|\widehat{\sigma}^2_{\text{RSS}} - \sigma^2\big| \ge \frac{C_\delta}{p}\,\Big|\,\big\{\widehat{G}_k^A : 1 \le k \le K_{\max}\big\}\Big) \le \delta.$$

From this, the first statement of (3.16) easily follows. The second statement can be obtained by similar arguments. □

References

Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV's, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **32** 111–131.

Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **34** 33–58.