Discussion of “Vintage factor analysis with Varimax performs statistical inference”∗

Rungang Han† and Anru R. Zhang‡

(May 23, 2022)

We wholeheartedly congratulate Drs. Rohe and Zeng for their insightful paper [2] on vintage factor analysis with Varimax rotation. Varimax rotation is a basic scheme to simplify the expression of a particular subspace and is included in build-in standard packages stats in R and PROC FACTOR statement in SAS. Drs. Rohe and Zeng nicely show that the principal component analysis with Varimax rotation actually performs statistical inference for the explainable factors.

Drs. Rohe and Zeng suggested leptokurtosis as a key identifiability condition for Varimax rotation; the number of factors often increases as the data dimension and sample size grow. It is thus natural to ask whether Varimax works with vanishing leptokurtosis and/or a growing number of factors. This note discusses when Varimax recovers the subspace rotation in such high-dimensional regimes. As a first step, we assume the factor matrix \( Z \in \mathbb{R}^{n \times k} \) includes a collection of i.i.d. centered random variables satisfying

\[
\mathbb{E} Z_{ij} = \mathbb{E} Z_{ij}^3 = 0, \quad \mathbb{E} Z_{ij}^2 = 1, \quad \mathbb{E} Z_{ij}^4 = \kappa > 3. \tag{1}
\]

We also assume \( Z_{ij} \)’s are sub-Gaussian such that \( \mathbb{E} \exp(\lambda Z_{ij}) \leq e^{c\lambda^2}, \forall \lambda \in \mathbb{R} \) for some constant \( c > 0 \). Let \( \hat{Z} \in \mathbb{R}^{p \times r} \) be the observed factors generated as

\[
\hat{Z} = ZR^*,
\]

where \( R^* \) is an unknown \( k \)-dimensional orthogonal matrix that represents the rotation to be recovered. Due to vanishing mean of \( Z \), we focus on the following centered Varimax:

\[
\hat{R} = \arg \max_{R \in O(k)} \sum_{i=1}^{n} \sum_{j=1}^{k} (\hat{Z}R^\top)_{ij}^4,
\]

∗Supported in part by NSF Grant CAREER-2203741 and NIH Grant R01 GM131399.
†Department of Statistical Science, Duke University
‡Departments of Biostatistics & Bioinformatics, Computer Science, Mathematics, and Statistical Science, Duke University
where $\mathcal{O}(k)$ is the set of $k$-by-$k$ orthogonal matrices. Consider the following error metric:

$$\text{dist}(R^*, \hat{R}) = \min_{P \in \mathcal{P}(k)} \| \hat{R} - P R^* \|_F = k^{-1/2} \min_{P \in \mathcal{P}(k)} \| \hat{R} R^\top - P \|_F,$$

where $\mathcal{P}(k) = \mathcal{O}(k) \cap \{0, \pm 1\}^{k^2}$ is the set of orthogonal matrices that allow for column reordering and sign changes as defined in Eqn. (12) in [2]. The following theorem characterizes the conditions of $n, k, \kappa$ under which Varimax works or fails.

**Theorem 1.** Let $\delta \in (0, 1/2]$ be any fixed value and $\kappa \leq C_0$ for some universal constant $C_0 > 3$.

- If $n \gtrsim \max \left\{ \frac{k \log n}{(\kappa - 3)^2}, \frac{k^2 \log^2 n}{\kappa - 3} \right\}$, then
  $$\limsup_{n \to \infty} \mathbb{P} \left( \text{dist}(R^*, \hat{R}) \geq \delta \right) = 0;$$

- if $k < n \lesssim k^2$, then
  $$\liminf_{n, k \to \infty} \mathbb{P} \left( \text{dist}(R^*, \hat{R}) \geq \delta \right) = 1.$$

If the kurtosis $\kappa > 3$ is a constant, Theorem 1 suggests that a necessary and nearly sufficient condition for consistent rotation recovery is $n \gtrsim k^2$ – this bound is tight up to $\log^2(n)$. Theorem 1 also shows when the leptokurtosis of factors is insignificant, i.e., $\kappa \to 3^+$, a sufficient sample size to ensure rotation recovery by Varimax is $\max\{k \log(n)/(\kappa - 3)^2, k^2 \log^2(n)/(\kappa - 3)\}$, while it is unclear if this bound is sharp. It is of future interest to investigate the tight condition that guarantees the consistency of Varimax.

**Proof of Theorem 1.** For any random matrix $Z$ and fixed $a \in \mathbb{R}^k$, define the following stochastic functions

$$v(Z; a) = \frac{1}{n} \sum_{i=1}^{n} \left( \sum_{j=1}^{k} a_j Z_{ij} \right)^4. \tag{2}$$

Note that the Varimax object function can be then written as $v(\hat{Z}; R^\top) = \sum_{j=1}^{k} v(\hat{Z}; R_j)$ where $R_j$, if the $j$th row of $R$. Our proof relies on the following lemmas.

**Lemma 1.** Suppose $A \in \mathcal{O}(k)$ and $\min_{P \in \mathcal{P}(k)} \| A - P \|_F > t$ for some $t > 0$. Then,

$$\sum_{i=1}^{k} \left( \sum_{j=1}^{k} A_{ij}^4 - 1 \right) \leq \frac{1}{16} t^2.$$
Lemma 2. Suppose $Z$ is a $n$-by-$k$ random matrix with i.i.d. mean-zero, unit-variance and sub-Gaussian entries. Then, with probability at least $1 - n^{-ck} - n^{-4}$,

$$\sup_{a: \|a\|_2 \leq 1} v(Z; a) - \mathbb{E}v(Z; a) \lesssim \sqrt{\frac{k \log n}{n}} + \frac{k^2 \log^2 n}{n}.$$  

Lemma 3. Suppose $X_1, \ldots, X_p$ are independent random variables with mean zero, variance one, and sub-Gaussian tails. Then there exists constants $C, c > 0$ such that

$$\mathbb{P} \left( \left| \sum_{i=1}^{p} (X_i^4 - \mathbb{E}X_i^4) \right| \geq C \sqrt{pt} + Ct^2 \right) \leq 2 \exp(-ct).$$

We start by proving the sufficient condition. Define

$$\delta := C_0 \sqrt{\frac{\kappa - 3}{\kappa - 3}} \max \left\{ \left( \frac{k^2 \log^2 n}{n} \right)^{1/2}, \left( \frac{k \log n}{n} \right)^{1/4} \right\},$$

for some positive constant $C_0$ to be specified later. It suffices to show that with high probability, for any $R \in \mathcal{O}(k)$ such that $\min_{P \in \mathcal{P}(k)} \|R^* R^\top - P\|_F > \delta k^{1/2}$, $R$ cannot be the solution of Varimax.

By Lemma 2, we know with probability at least $1 - o(1/n)$,

$$\sup_{a: \|a\|_2 \leq 1} v(Z; a) - \mathbb{E}v(Z; a) \leq C \left( \sqrt{\frac{k \log n}{n}} + \frac{k^2 \log^2 n}{n} \right) \leq c(\kappa - 3),$$

where $c$ is some small positive constant. In the meantime, Define $A = RR^\top = (a_1, a_2, \ldots, a_k)$, where $a_j$ is the $j$-th orthogonal column of $A$. Note that the Varimax loss function can be written as $v(\hat{Z}; R) = \sum_{j=1}^{k} v(Z; a_j)$. By calculating its expectation, we have:

$$\mathbb{E}v(\hat{Z}; R) = \mathbb{E} \sum_{j=1}^{k} v(Z; a_j) = 3k + (\kappa - 3) \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij}^4 \leq \kappa k \left( \sum_{j=1}^{k} a_{ij}^4 - 1 \right) \leq \kappa k - \frac{(\kappa - 3) \delta^2}{16} k.$$  

Meanwhile, note that

$$\mathbb{E}v(\hat{Z}, R^*) = \kappa k.$$  

3
Then we have
$$v(\hat{Z}; R^*) - v(\hat{Z}; R) = \left( E v(\hat{Z}; R^*) - E v(\hat{Z}; R) \right) + \left( v(\hat{Z}; R^*) - E(\hat{Z}; R^*) \right) - \left( v(\hat{Z}; R) - E(\hat{Z}; R) \right)$$

$$\geq \frac{(k - 3)\delta^2}{16} k - 2k \sup_{a; \|a\| \leq 1} |v(Z; a) - E v(Z; a)|.$$

$$\geq 0.$$ (3),(4)

Therefore, $R$ is not a solution of varimax. In other words, given $\hat{R} = \arg\max_{R \in O(k)} v(\hat{Z}, R)$, one must have $\min_{P \in P(k)} \left\| R^* R^T - P \right\|_F < \delta k^{1/2}$, which completes the proof of sufficient condition.

Next, we prove the necessary condition. Let $Z_1 \in \mathbb{R}^{[k/2] \times k}$ be the first $[k/2]$ rows of the factor matrix $Z$. Consider the QR-decomposition of $Z_1^T$: $Z_1^T = U_1 R_1$, where $U_1$ is a $k$-by-$[k/2]$ orthogonal matrix and $R_1$ is an $[k/2]$-by-$[k/2]$ upper triangular matrix. Let $U_{1\perp} \in \mathbb{R}^{k,k-[k/2]}$ be the perpendicular subspace of $U_1$ (such that $U_1^T U_{1\perp} = 0$), and denote $A = [U_1 U_{1\perp}]^T R^*$. Then,

$$\hat{Z} A^T_{[1:[k/2],:]} = Z_1 [U_1 U_{1\perp}] = [R_1^T O]$$

and it follows that

$$v(\hat{Z}, A) = \frac{1}{n} \sum_{i=1}^{[k/2]} \sum_{j=1}^{k} \left( (\hat{Z} A^T)_{ij} \right)^4$$

$$\geq \frac{1}{n} \left( \sum_{j=1}^{[k/2]} (R_1)_{jj}^4 + \sum_{i=k+1}^{n} \sum_{j=1}^{k} \left( (\hat{Z} A^T)_{ij} \right)^4 \right).$$ (7)

Now we establish the probabilistic bounds for $D_1$ and $D_2$.

- $D_1$. By random matrix theory [3], we know that with probability at least $1 - e^{-ck}$ that

$$\sigma_{[k/2]}(Z_1) \geq \sqrt{k} - \sqrt{\lfloor k/2 \rfloor} - c\sqrt{k} \geq \sqrt{0.1k}.$$ (8)

Note that $R_1$ shares the same non-zeros singular values with $X_1$. Thus it follows that

$$\prod_{j=1}^{[k/2]} (R_1)_{jj}^2 = |\det(R_1)|^2 = \prod_{j=1}^{[k/2]} \sigma_j^2(Z_1) \geq (0.1k)^{[k/2]}.$$
Then by Inequality of arithmetic and geometric means,

\[ D_1 = \sum_{j=1}^{\lfloor k/2 \rfloor} (R_1)_{jj}^4 \geq \lfloor k/2 \rfloor \left( \prod_{j=1}^{\lfloor k/2 \rfloor} (R_1)_{jj}^4 \right)^{1/\lfloor k/2 \rfloor} \geq 0.005k^3. \quad (9) \]

- \(D_2\). Denote \( B := R^* A^\top \) and we can rewrite

\[
D_2 = \sum_{i=k+1}^{n} \sum_{j=1}^{k} ((\tilde{Z} A^\top)_{ij})^4 = \sum_{i=k+1}^{n} \sum_{j=1}^{k} ((ZB)_{ij})^4 \\
\quad = \sum_{j=1}^{k} \sum_{i=k+1}^{n} \left( \sum_{l=1}^{k} B_{lj} Z_{il} \right)^4.
\]

Then, conditioning on \( A \) (or \( B \)), we have

\[
\mathbb{E}[D_2 | A] = \mathbb{E} \left[ \sum_{j=1}^{k} \sum_{i=k+1}^{n} \left( \sum_{l=1}^{k} B_{lj} Z_{il} \right)^4 \left| B \right. \right] \\
\quad = \sum_{j=1}^{k} (n - k) \mathbb{E} \left[ \left( \sum_{l=1}^{k} B_{lj} Z_{il} \right)^4 \left| B \right. \right] \\
\quad = (n - k) \left( 3k + (\kappa - 3) \sum_{l=1}^{k} \sum_{j=1}^{k} B_{lj}^4 \right) \\
\quad \geq 3(n - k)k.
\]

On the other hand, for each fixed \( j \in [k] \), \( \left\{ \sum_{l=1}^{k} B_{lj} Z_{il} \right\}_{i=k+1}^{n} \) are i.i.d. mean-0, variance-1 sub-Gaussian random variables (since \( B \) is independent of the last \( n - k \) rows of \( Z \)). By Lemma 3, we have

\[
\mathbb{P} \left( \sum_{i=k+1}^{n} \left( \sum_{l=1}^{k} B_{lj} Z_{il} \right)^4 - \mathbb{E} \left[ \sum_{i=k+1}^{n} \left( \sum_{l=1}^{k} B_{lj} Z_{il} \right)^4 \left| A \right. \right] < -C(\sqrt{nl} + t^2) \left| A \right. \right) \leq 2e^{-ct}.
\]

Taking \( t = c_0k \) for some small constant \( c_0 \) and applying union bound for each \( j \in [k] \), we know that

\[
\mathbb{P} \left( D_2 - \mathbb{E}[D_2 | A] < -c_0 \left( \sqrt{n k^3} + k^3 \right) \left| A \right. \right) \leq 2ke^{-ck}
\]
and it follows that
\[ P\left(D_2 < 3(n - k)k - c_0 \left(\sqrt{nk^3} + k^3\right) \right) \leq 2ke^{-ck}. \]

Integrate over the density of \(A\) and we obtain
\[ P\left(D_2 < 3(n - k)k - c_0 \left(\sqrt{nk^3} + k^3\right) \right) \leq 2ke^{-ck} \leq e^{-c'k}. \] (10)

In conclusion, we proved that with probability at least \(1 - e^{-ck}\),
\[ D_1 \geq 0.005k^3, \quad D_2 \geq 3(n - k)k - c_0 \left(\sqrt{nk^3} + k^3\right). \] (11)

Since \(n \leq ck^2\), by taking \(c_0\) to be sufficiently small, we have
\[ v(\hat{Z}, A) \geq \frac{1}{n} \left(0.004k^3 + 3nk\right). \] (12)

Next we show that for any \(R\) such that \(\text{dist}(\hat{R}, R^*) \leq \delta\), \(v(\hat{Z}, R)\) is upper bounded by the right-hand side of (12) with high probability. Note that
\[ v(\hat{Z}; R) = \mathbb{E}v(\hat{Z}; R^*) + (v(\hat{Z}, R^*) - \mathbb{E}v(\hat{Z}; R^*)) + (v(\hat{Z}, R) - v(\hat{Z}, R^*)) \]
\[ = \kappa k + (v(\hat{Z}; R^*) - \mathbb{E}v(\hat{Z}; R^*)) + \sup_{R: \text{dist}(R, R^*) \leq \delta} \left| v(\hat{Z}, R) - v(\hat{Z}, R^*) \right|. \] (13)

Following the same argument as (10), we know that with probability at least \(1 - e^{-ck}\) that
\[ v(\hat{Z}; R^*) - \mathbb{E}v(\hat{Z}, R^*) \leq \frac{c_0}{n} \left(\sqrt{nk^3} + k^3\right). \]

In the meantime,
\[ \sup_{R: \text{dist}(R, R^*) \leq \delta} \left| v(\hat{Z}, R) - v(\hat{Z}, R^*) \right| = \sup_{\tilde{R} \in \mathcal{O}(k): \|\tilde{R} - I\|_F \leq \delta \sqrt{k}} \left| v(Z; \tilde{R}) - v(Z; I) \right| \]
\[ \leq \sup_{\tilde{R} \in \mathcal{O}(k): \|\tilde{R} - I\|_F \leq \delta \sqrt{k}} \sum_{i=1}^{k} \left| v(Z; \tilde{R}_{[i,:]} - v(Z; e_i) \right|, \]
(14)

where \(e_i\) is the \(i\)-th canonical basis in \(\mathbb{R}^k\). Let \(\delta_i := \|\tilde{R}_{[i,:]} - e_i\|_2\). Then, by the proof of (20) in Lemma 2, for each \(i \in [k]\), we have
\[ \left| v(Z; \tilde{R}_{[i,:]} - v(Z; e_i) \right| \leq 6\delta_i \max_{j,l} Z_{jl}^4 \leq C\delta_i \log^2 n \]
(15)
hold with probability at least $1 - n^{-4}$. Combining (14) and (15), we obtain

\[
\sup_{R: \text{dist}(R, R^*) \leq \delta} \left| v(\hat{Z}, R) - v(\hat{Z}, R^*) \right| \leq C \log^2 n \sum_{i=1}^{k} \delta_i \leq C \log^2 n \left( \sum_{i=1}^{k} \delta_i^2 \right)^{1/2}
\]

\[
= C\sqrt{k} \log^2 n \| \tilde{R} - I \|_F \leq C\delta k \log^2 n.
\]

(16)

Thus, when $n \lesssim k^2 \log^2 n$, we have

\[
\sup_{R: \text{dist}(R, R^*) \leq \delta} v(\hat{Z}, R) \overset{\text{(13),(16)}}{\leq} \kappa k + \frac{c_0 (\sqrt{n}k^3 + k^3)}{n} + \delta k \log^2 n < \frac{0.004k^3 + 3nk}{n} \overset{\text{(12)}}{\leq} v(\hat{Z}, A).
\]

Here, the second inequality is obtained by choosing $c_0$ to be sufficiently small and $\delta = 1/2$. This suggests that the solution of Varimax function $\hat{R}$ must satisfy:

\[
\text{dist}(\hat{R}, R^*) \geq \delta = 1/2.
\]

\[\square\]

**Proof of Lemma 1.** Let $P^* = \arg \min_{P \in O(k)} \| A - P \|_F^2$. Without loss of generality, we can assume $P^* = I$ and $A_{11} \geq A_{22} \geq \ldots \geq A_{kk} \geq 0$. This suggests that

\[
t^2 < \sum_{i=1}^{k} \left( (1 - A_{ii})^2 + \sum_{j \neq i} A_{ij}^2 \right) = 2 \sum_{i=1}^{k} (1 - A_{ii}) \leq 2 \sum_{i=1}^{k} (1 - A_{ii}^2).
\]

(17)

It suffices to show that for each $i \in [k]$,

\[
\sum_{j=1}^{k} A_{ij}^4 - 1 \leq -\frac{1}{8} (1 - A_{ii}^2).
\]

Now consider the following two situations for each $i \in [k]$:

- $A_{ii}^2 \geq 1/16$.

\[
\sum_{j=1}^{k} A_{ij}^4 - 1 \leq A_{ii}^4 + \left( \sum_{j \neq i} A_{ij}^2 \right)^2 - 1
\]

\[
= A_{ii}^4 + (1 - A_{ii}^2)^2 - 1
\]

\[
= 2A_{ii}^2 (A_{ii}^2 - 1) \leq -\frac{1}{8} (1 - A_{jj}^2).
\]
• $A_{ii}^2 < 1/16$. We claim that $\max_j A_{ij}^2 < 3/4$. Suppose there exists $l \in [k]/i$ such that $A_{ii}^2 \geq 3/4$. Since $I = \arg \min_{P \in \mathcal{P}(k)} \|A - P\|_F^2$, we must have 
\[
\|A - I\|_F^2 \leq \|A - D^{(i,l)}\|_F^2, \tag{18}
\]
where $D^{(i,l)}$ is the permutation matrix such that $D^{(i,l)}_{ii} = \sgn(A_{ii})$, $D^{(i,l)}_{ij} = \sgn(A_{ij})$ and $D^{(i,l)}_{jj} = 1$ for $j \in [k]\{i,l\}$. This yields 
\[
(A_{ii} - 1)^2 + (A_{ii} - 1)^2 + A_{ii}^2 + A_{ii}^2 \leq (|A_{ii}| - 1)^2 + (|A_{ii}| - 1)^2 + A_{ii}^2 + A_{ii}^2,
\]
which is equivalent to 
\[
|A_{ii}| + |A_{ii}| \leq A_{ii} + A_{ii}.
\]
Then it follows that 
\[
A_{ii} \geq |A_{ii}| - A_{ii} \geq \frac{2\sqrt{3} - 1}{4}.
\]
Consequently, we get 
\[
1 \geq A_{ii}^2 + A_{ii}^2 \geq \frac{13 - 4\sqrt{3}}{16} + \frac{3}{4} > 1.
\]
This contradiction shows that we must have $\max_j A_{ij}^2 < 3/4$. Therefore,
\[
\sum_{j=1}^{k} A_{ij}^2 - 1 \leq \max_j A_{ij}^2 \sum_{j=1}^{k} A_{ij}^2 - 1 = -\frac{1}{4} \leq -\frac{1}{4}(1 - A_{ii}^2).
\]
The proof is finished by combining the two situations. \hfill \square

**Proof of Lemma 2.** We first construct an $\varepsilon$-net $\mathcal{N} = \{x^{(1)}, \ldots, x^{(N)}\} \subset \mathbb{S}^{k-1}$ such that for any $a \in \mathbb{S}^{k-1}$, there exists some $i_a \in [N]$ with $\|a - x^{(i_a)}\| \leq \varepsilon$. By [3], we can choose such an $\varepsilon$-net with $N \leq (3/\varepsilon)^k$. Then, by triangle inequality, 
\[
\sup_{a \in \mathbb{S}^{k-1}} |v(Z; a) - \mathbb{E}v(Z; a)| \\
\leq \sup_{a \in \mathbb{S}^{k-1}} \left( |\mathbb{E}v(Z; x^{(i_a)}) - \mathbb{E}v(Z; a)| + |v(Z; a) - v(Z; x^{(i_a)})| + |v(Z; x^{(i_a)}) - \mathbb{E}v(Z; x^{(i_a)})| \right) \\
\leq \sup_{a \in \mathbb{S}^{k-1}} \left( \underbrace{|\mathbb{E}v(Z; x^{(i_a)}) - \mathbb{E}v(Z; a)|}_{D_1} + \sup_{a \in \mathbb{S}^{k-1}} |v(Z; a) - v(Z; x^{(i_a)})| + \max_{i \in N} |v(Z; x^{(i)}) - \mathbb{E}v(Z; x^{(i)})| \right). \\
\]
We specify $\varepsilon = 1/n$ and $N \leq (3n)^k$. Next we obtain deterministic bounds for $D_1$ and $D_2$ and a probabilistic bound for $D_3$. 

8
\( \sup_{a \in S^{k-1}} |E v(Z; x^{(i_a)}) - E v(Z; a)| \). Note that for any fixed \( a \in S^{k-1} \), one can calculate that
\[
E v(Z; a) = E \left( \sum_{j=1}^{k} a_j X_{1j} \right)^4 = (\kappa - 3) \sum_{j=1}^{k} a_j^4 + 3.
\]
Therefore,
\[
|E v(Z; x^{(i_a)}) - E v(Z; a)| = \left( \kappa - 3 \right) \left| \sum_{j=1}^{k} \left( a_j^4 - (x_j^{(i_a)})^4 \right) \right|
\]
\[
\leq (\kappa - 3) \left( \max a_j^2 + (x_j^{(i_a)})^2 \right) \cdot \sum_{j=1}^{k} \left| a_j^2 - (x_j^{(i_a)})^2 \right| \quad (19)
\]
\[
\leq 2(\kappa - 3) \sum_{j=1}^{k} \left| a_j^2 - (x_j^{(i_a)})^2 \right|.
\]

Denote \( \delta = a - b \) for any \( b \in S^{k-1} \). Note that
\[
\sum_{j=1}^{k} |a_j - b_j| = \sum_{j=1}^{k} |a_j^2 + \delta_j^2 + 2b_j \delta_j - b_j^2| \leq \|\delta\|_2^2 + 2 \sum_{j=1}^{k} |b_j \delta_j| \leq \|\delta\|_2^2 + 2\|\delta\|_2,
\]
where we use Cauchy-Schwardz inequality and the fact \( \|b\| = 1 \) to obtain the last inequality. Now we replace \( b \) with \( x^{(i_a)} \), then \( D_1 \) can be further bounded as
\[
D_1 \leq 2(\kappa - 3) \sum_{j=1}^{k} \left| a_j^2 - (x_j^{(i_a)})^2 \right| \leq 2(\kappa - 3) \max_j |a_j + x_j^{(i_a)}| \cdot \sum_{j=1}^{k} \left| a_j - x_j^{(i_a)} \right|
\]
\[
\leq 12\varepsilon(\kappa - 3) \leq \frac{C}{n}. \quad (20)
\]
Now the proof is finished.\footnote{Discrete Mathematics, 2005.}

Note that \( \{X_{ij}^i\}_{i=1}^n \) are i.i.d. mean-zero, variance one, sub-Gaussian random variables, we have

\[
\Pr\left( v(Z; x(i)) - \mathbb{E}v(Z; x(i)) \right) \geq C \left( \sqrt{\frac{t}{n}} + \frac{t^2}{n} \right) \leq e^{-ct}.
\]

Taking \( t = Ck \log n \) and applying union bounds on \( x^{(i)} \) for each \( i \in \{N\} \), it follows that

\[
\Pr\left( D_3 \leq C \left( \sqrt{\frac{k \log n}{n}} + \frac{k^2 \log^2 n}{n} \right) \right) \geq 1 - e^{-ck \log n}.
\]

In the meantime, by the tail bound of sub-Gaussian Extreme values, we have

\[
\Pr\left( \max_{i,k} |X_{i,k}| \leq \sqrt{10\log(nk)} \right) \leq 1 - \sqrt{\frac{2}{\pi \log n}} n^{-4}.
\]

Combining (20), (19), (21) and (22), we know that with probability at least \( 1 - n^{-ck} - n^{-4} \) that

\[
\sup_{a \in \mathbb{S}^{k-1}} |v(Z; a) - \mathbb{E}v(Z; a)| \lesssim \sqrt{k \log n + \frac{k^2 \log^2 n}{n}}.
\]

Now the proof is finished. \( \square \)

\textbf{Proof of Lemma 3.} Note that \( \{X_{ij}^i\}_{i=1}^n \) are sub-Weibull(1/2) random variables and the result simply comes from the concentration developed by [1, Theorem 3.1]. \( \square \)
References

[1] Arun Kumar Kuchibhotla and Abhishek Chakrabortty. Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *arXiv preprint arXiv:1804.02605*, 2018. 10

[2] Karl Rohe and Muzhe Zeng. Vintage factor analysis with varimax performs statistical inference. *Journal of the Royal Statistical Association, Series B*, to appear, 2022. 1, 2

[3] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010. 4, 8