STAT 824 Lec 11 supplement

Some results based on the Lindeberg central limit theorem

Karl B. Gregory

Contents

1 The Lindeberg central limit theorem

We will make much use of the following central limit theorem.

Theorem 1 (Lindeberg central limit theorem)**.** *For each $n \geq 1$, let $\{U_{n1}, \dots, U_{nr_n}\}$ be a collection of independent random variables such that $\mathbb{E}U_{nj} = 0$ and $\operatorname{Var} U_{nj} < \infty$ for $j = 1, \dots, r_n$.*

$$\tilde{U}_{nj} = \frac{U_{nj}}{\sqrt{\sum_{k=1}^{r_n} \operatorname{Var} U_{nk}}}, \quad j = 1, \dots, r_n.$$

Then

$$\sum_{j=1}^{r_n} \tilde{U}_{nj} \to N(0,1) \text{ in distribution as } n \to \infty$$

if for every $\epsilon > 0$

$$\lim_{n\to\infty} \sum_{j=1}^{r_n} \mathbb{E}|\tilde{U}_{nj}|^2 \mathbb{1}(|\tilde{U}_{nj}| > \epsilon) = 0. \tag{1}$$

Remark 1. The sequence of collections of random variables $\{U_{n1}, \ldots, U_{nr_n}\}_{n\geq 1}$ introduced in the theorem is called a *triangular array*.

Remark 2. The variables $U_{n1}, \ldots, U_{nr_n}$ do not need to be identically distributed.

Remark 3. The condition in (1) is called the *Lindeberg condition*.

A proof of the Lindeberg central limit theorem appears in the Appendix. When stating corollaries to or applications of the Lindeberg central limit theorem, we may drop the somewhat cumbersome triangular array notation. For example, in the next result we introduce $X_1, \ldots, X_n$ and proceed as though with the triangular array $\{X_{n1}, \ldots, X_{nn}\}_{n\geq 1}$.

Corollary 1 (A simple central limit theorem)**.** *Let* $X_1, \ldots, X_n$ *be independent identically distributed random variables with* $\mathbb{E}X_1 = \mu$ *and* $\operatorname{Var} X_1 = \sigma^2 < \infty$ *and let* $\bar{X}_n = n^{-1}(X_1 + \cdots + X_n)$*. Then*

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma \to N(0, 1) \text{ in distribution as } n \to \infty.$$

Proof of Corollary 1. To analyze $X_1, \ldots, X_n$ for increasing n, introduce the triangular array $\{X_{n1}, \ldots, X_{nn}\}_{n\geq 1}$, with $\bar{X}_n = n^{-1}(X_{n1} + \cdots + X_{nn})$. Then define $U_{nj} = X_{nj} - \mu$ for $j = 1, \ldots, n$. Then $\sum_{j=1}^n \operatorname{Var} U_{nj} = n\sigma^2$ and

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma = \frac{\sum_{i=1}^n U_{ni}}{\sqrt{\sum_{j=1}^n \operatorname{Var} U_{nj}}}.$$

Thus by Theorem 1 it is sufficient to show that the random variables

$$\tilde{U}_{ni} = \frac{U_{ni}}{\sqrt{\sum_{j=1}^n \operatorname{Var} U_{nj}}} = \frac{X_{ni} - \mu}{\sqrt{n}\sigma}, \quad i = 1, \ldots, n, \quad n \geq 1,$$

satisfy the Lindeberg condition (1). For any $\epsilon > 0$ we have

$$\sum_{i=1}^n \mathbb{E}\left|\frac{X_{ni}-\mu}{\sqrt{n}\sigma}\right|^2 \mathbb{1}\left(\left|\frac{X_{ni}-\mu}{\sqrt{n}\sigma}\right| > \epsilon\right) = \frac{1}{\sigma^2}\mathbb{E}|X_{n1}-\mu|^2\mathbb{1}(|X_{n1}-\mu| > \epsilon\sigma\sqrt{n})$$
$$\to 0 \text{ as } n \to \infty$$

by the dominated convergence theorem, since $\mathbb{E}(X_{n1}-\mu)^2 < \infty$. □

Corollary 2 (Lindeberg CLT for linear combinations of iid rvs)**.** *For a seq. of iid rvs* $\xi_1, \xi_2, \ldots$ *with* $\mathbb{E}\xi_1 = 0$ *and* $\mathbb{E}\xi^2 = \sigma^2 < \infty$ *and a seq. of numbers* $a_1, a_2, \ldots$ *that satisfy*

$$\frac{\max_{1\le i\le n}|a_i|}{(\sum_{j=1}^n a_j^2)^{1/2}} \to 0 \qquad \textit{as } n \to \infty,$$

we have

$$\frac{\sum_{i=1}^n a_i \cdot \xi_i}{\sigma(\sum_{j=1}^n a_j^2)^{1/2}} \to N(0,1) \textit{ in dist. as } n \to \infty.$$

Proof of Corollary 2. For each $n \geq 1$ and $i = 1, \ldots, n$, let $U_{ni} = a_i\xi_i$, so that $\operatorname{Var} U_{ni} = a_i^2\sigma^2$. Accordingly set $\tilde{U}_{ni} = (\sum_{j=1}^n a_j\sigma^2)^{-1/2}a_i\xi_i$. Now we show that the triangular array $\{\tilde{U}_{ni},\ i = 1, \ldots, n\}$, $n \geq 1$ satisfies the Lindeberg condition. We have

$$\begin{aligned}
&\lim_{n\to\infty}\sum_{i=1}^n \mathbb{E}\left[\left|\frac{a_i\xi_i}{\sigma(\sum_{j=1}^n a_j^2)^{1/2}}\right|^2 \mathbb{1}\left(\left|\frac{a_i\xi_i}{\sigma(\sum_{j=1}^n a_j^2)^{1/2}}\right| > \epsilon\right)\right] \\
&\quad = \lim_{n\to\infty}\sum_{i=1}^n \frac{a_i^2}{\sigma^2\sum_{j=1}^n a_j^2}\mathbb{E}\xi_i^2\mathbb{1}\left(|a_i\xi_i| > \epsilon\sigma(\textstyle\sum_{j=1}^n a_j^2)^{1/2}\right) \\
&\quad \leq \lim_{n\to\infty}\sum_{i=1}^n \frac{a_i^2}{\sigma^2\sum_{j=1}^n a_j^2}\mathbb{E}\xi_1^2\mathbb{1}\left((\max_{1\le i\le n}|a_i|)|\xi_1| > \epsilon\sigma(\textstyle\sum_{j=1}^n a_j^2)^{1/2}\right) \\
&\quad = \frac{1}{\sigma^2}\lim_{n\to\infty}\mathbb{E}\xi_1^2\mathbb{1}\left(|\xi_1| > \epsilon\sigma\frac{(\sum_{j=1}^n a_j^2)^{1/2}}{\max_{1\le i\le n}|a_i|}\right) \\
&\quad = 0
\end{aligned}$$

by the dominated convergence theorem, since $\mathbb{E}\xi_1^2 < \infty$ and by the assumption on the sequence $a_1, a_2, \ldots$. □

2 Lindeberg CLT for regression

The Lindeberg central limit theorem is very useful for establishing the asymptotic Normality of linear regression coefficients, as these can be represented as the sum of differently-weighted residuals.

2.1 Simple linear regression with no intercept

Example 1 (Application of Lindeberg CLT to simple linear regression)**.** For each $n \geq 1$, let

$$Y_i = x_i\beta + \varepsilon_i, \quad i = 1, \dots, n,$$

where $\varepsilon_1, \dots, \varepsilon_n$ are independent identically distributed random variables such that $\mathbb{E}\varepsilon_1 = 0$ and $\operatorname{Var}\varepsilon_1 = \sigma^2 < \infty$ and $x_1, \dots, x_n$ are fixed constants, and let $\hat{\beta}_n = \sum_{i=1}^n x_i Y_i / \sum_{j=1}^n x_i^2$. Then if

$$\frac{\max_{1\leq i\leq n}|x_i|}{(\sum_{i=1}^n x_i^2)^{1/2}} \to 0 \text{ as } n \to \infty \tag{2}$$

we have

$$\sqrt{n}\sigma^{-1}(n^{-1}\textstyle\sum_{j=1}^n x_j^2)^{1/2}(\hat{\beta}_n - \beta) \to N(0,1) \text{ in distribution as } n \to \infty. \tag{3}$$

To show (3), we write

$$\sqrt{n}\sigma^{-1}(n^{-1}\textstyle\sum_{j=1}^n x_j^2)^{1/2}(\hat{\beta}_n - \beta) = \dfrac{\sum_{i=1}^n x_i\varepsilon_i}{\sigma(\sum_{j=1}^n x_j^2)^{1/2}}.$$

Then Corollary 2 gives that the condition (2) is sufficient for (3).

2.2 Multiple linear regression

The following is a multivariate extension of the Lindeberg central limit theorem.

Theorem 2 (Multivariate Lindeberg central limit theorem)**.** *For each* $n \geq 1$, *let* $U_{n1}, \dots, U_{nn} \in \mathbb{R}^d$ *be independent random vectors such that* $\mathbb{E}U_{ni} = 0$ *and* $\operatorname{Cov} U_{ni} < \infty$ *for* $i = 1, \dots, n$. *Let*

$$\tilde{U}_{ni} = \left(\textstyle\sum_{j=1}^n \operatorname{Cov} U_{nj}\right)^{-1/2} U_{ni}, \quad i = 1, \dots, n.$$

Then

$$\sum_{i=1}^{n} \tilde{U}_{ni} \to N(0, I_d) \text{ in distribution as } n \to \infty$$

if for every $\epsilon > 0$

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\|\tilde{U}_{ni}\|^2 \mathbb{1}(\|\tilde{U}_{ni}\| > \epsilon) = 0. \tag{4}$$

Remark 4. By $\operatorname{Cov} U_{ni} < \infty$, we mean that the largest eigenvalue of the covariance matrix $\operatorname{Cov} U_{ni}$ is finite. This means that for any vector $a \in \mathbb{R}$, $\operatorname{Var} a^T U_{ni} < \infty$; no linear combination of the variables in U_{ni} can result in an infinite-variance random variable.

Proof of Theorem 2. By the Cramer-Wold device, it is sufficient to show that for any $a \in \mathbb{R}^d$ such that $\|a\|_2 = 1$, we have

$$a^T \sum_{i=1}^{n} \tilde{U}_{ni} = \frac{\sum_{i=1}^{n} a^T \tilde{U}_{ni}}{\sqrt{\sum_{i=1}^{n} \operatorname{Var} a^T \tilde{U}_{ni}}} \to N(0, 1) \text{ in distribution as } n \to \infty. \tag{5}$$

We establish the convergence in distribution in (5) by showing that the random variables

$$a^T \tilde{U}_{ni}, \quad i = 1, \dots, n, \quad n \geq 1,$$

satisfy the univariate Lindeberg condition (1). Using the Cauchy-Schwarz inequality and the fact that a is a unit vector, for any $\epsilon > 0$ we have

$$\begin{aligned}\sum_{i=1}^{n} \mathbb{E}|a^T \tilde{U}_{ni}|^2 \mathbb{1}(|a^T \tilde{U}_{ni}| > \epsilon) &\leq \sum_{i=1}^{n} \mathbb{E}\|a\|_2^2 \|\tilde{U}_{ni}\|_2^2 \mathbb{1}(\|a\|_2 \|\tilde{U}_{ni}\|_2 > \epsilon) \\ &= \sum_{i=1}^{n} \mathbb{E}\|\tilde{U}_{ni}\|_2^2 \mathbb{1}(\|\tilde{U}_{ni}\|_2 > \epsilon),\end{aligned}$$

which goes to zero as $n \to \infty$ by assumption. □

The multivariate Lindeberg central limit theorem can be used to establish the joint asymptotic Normality of regression coefficients in multiple linear regression.

Example 2 (Application of multivariate Lindeberg CLT to linear regression)**.** For each $n \geq 1$, let

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent identically distributed random variables such that $\mathbb{E}\varepsilon_1 = 0$ and $\operatorname{Var} \varepsilon_1 = \sigma^2 < \infty$ and $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ are fixed vectors of dimension $d \leq n$, and let $\hat{\boldsymbol{\beta}}_n = (\mathbf{X}_n^T \mathbf{X}_n)^{-1} \mathbf{X}_n^T \mathbf{Y}_n$, where $\mathbf{X}_n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ and $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^T$. In addition let h_{ii}, $i = 1, \ldots, n$, be the diagonal entries of the matrix $\mathbf{X}_n (\mathbf{X}_n^T \mathbf{X}_n)^{-1} \mathbf{X}_n^T$. Then if

$$\max_{1 \leq i \leq n} h_{ii} \to 0 \text{ as } n \to \infty \tag{6}$$

we have

$$\sqrt{n} \sigma^{-1} (n^{-1} \mathbf{X}_n^T \mathbf{X}_n)^{1/2} (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \to N(0, I_d) \text{ in distribution as } n \to \infty. \tag{7}$$

To show (7), we set $U_i = \mathbf{x}_i \varepsilon_i$, which allows us to write

$$\sqrt{n} \sigma^{-1} (n^{-1} \mathbf{X}_n^T \mathbf{X}_n)^{1/2} (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = (\sigma^2 \mathbf{X}_n^T \mathbf{X}_n)^{-1/2} \sum_{i=1}^n \mathbf{x}_i \varepsilon_i = \left(\sum_{i=1}^n \operatorname{Cov} U_i \right)^{-1/2} \sum_{i=1}^n U_i.$$

Then by Theorem 2 it is sufficient to show that the random vectors

$$\tilde{U}_i = \left(\sum_{i=1}^n \operatorname{Cov} U_i \right)^{-1/2} U_i = (\sigma^2 \mathbf{X}_n^T \mathbf{X}_n)^{-1/2} \mathbf{x}_i \varepsilon_i, \quad i = 1, \ldots, n, \quad n \geq 1,$$

satisfy

$$\sum_{i=1}^n \mathbb{E} \|\tilde{U}_i\|_2^2 \mathbb{1}(\|\tilde{U}_i\|_2 > \delta) \to 0 \text{ as } n \to \infty$$

for every $\delta > 0$. Noting that $h_{ii} = \mathbf{x}_i^T (\mathbf{X}_n^T \mathbf{X}_n)^{-1} \mathbf{x}_i$ and

$$\sum_{i=1}^n h_{ii} = \operatorname{tr}(\mathbf{X}_n (\mathbf{X}_n^T \mathbf{X}_n)^{-1} \mathbf{X}_n^T) = \operatorname{tr}(\mathbf{X}_n^T \mathbf{X}_n (\mathbf{X}_n^T \mathbf{X}_n)^{-1}) = \operatorname{tr}(\mathbf{I}_d) = d,$$

we have, for each $\delta > 0$,

$$\sum_{i=1}^n \mathbb{E} \|\tilde{U}_i\|_2^2 \mathbb{1}(\|\tilde{U}_i\|_2 > \delta)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \mathbb{E}\|(\sigma^2 \mathbf{X}_n^T \mathbf{X}_n)^{-1/2} \mathbf{x}_i \varepsilon_i\|_2^2 \mathbb{1}(\|(\sigma^2 \mathbf{X}_n^T \mathbf{X}_n)^{-1/2} \mathbf{x}_i \varepsilon_i\|_2 > \delta) \\
&= \frac{1}{\sigma^2} \sum_{i=1}^{n} \mathbf{x}_i^T (\mathbf{X}_n^T \mathbf{X}_n)^{-1} \mathbf{x}_i \mathbb{E}\epsilon_i^2 \mathbb{1}(\mathbf{x}_i^T (\mathbf{X}_n^T \mathbf{X}_n)^{-1} \mathbf{x}_i \varepsilon_i^2 > \delta^2 \sigma^2) \\
&\leq \frac{d}{\sigma^2} \mathbb{E}\epsilon_1^2 \mathbb{1}\left((\max_{1 \leq i \leq n} h_{ii}) \varepsilon_1^2 > \delta^2 \sigma^2 \right) \\
&= \frac{d}{\sigma^2} \mathbb{E}\epsilon_1^2 \mathbb{1}\left(|\varepsilon_1| > \delta\sigma / \sqrt{\max_{1 \leq i \leq n} h_{ii}} \right) \\
&= 0
\end{aligned}
$$

by the dominated convergence theorem, since $\mathbb{E}\epsilon_1^2 < \infty$ and because of the assumption in (6). This gives the result.

2.3 Logistic regression

The multivariate Lindeberg CLT can be used to show the asymptotic Normality of the score function in Logistic regression.

Example 3 (Logistic regression)**.** Let $Y_1, \ldots, Y_n$ be independent random variables and $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ such that $Y_i \sim \text{Bernoulli}(\pi_i)$, where $\pi_i = 1/(1 + e^{-\eta_i})$, with $\eta_i = \mathbf{x}_i^T \boldsymbol{\theta}$. Then the log-likelihood function for estimating $\boldsymbol{\theta}$ based on $\{(Y_i, \mathbf{x}_i), i = 1, \ldots, n\}$ is given by

$$
\ell_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} [Y_i \log \pi_i + (1 - Y_i) \log(1 - \pi)].
$$

Setting $\mathbf{Y} = (Y_1, \ldots, Y_n)$, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$, $\mathbf{X}_n = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^T$, and $\mathbf{W}_n = \text{diag}(\pi_i(1 - \pi_i), i = 1, \ldots, n)$, the score and Hessian are given by

$$
\begin{aligned}
S_n(\boldsymbol{\theta}) &= \sum_{i=1}^{n} (Y_i - \pi_i) \mathbf{x}_i = \mathbf{X}_n^T (\mathbf{Y} - \boldsymbol{\pi}) \\
H_n(\boldsymbol{\theta}) &= -\sum_{i=1}^{n} \pi_i (1 - \pi_i) \mathbf{x}_i \mathbf{x}_i^T = -\mathbf{X}_n^T \mathbf{W}_n \mathbf{X}_n.
\end{aligned}
$$

The Lindeberg CLT gives

$$
(\mathbf{X}_n^T \mathbf{W}_n \mathbf{X}_n)^{-1/2} \mathbf{X}_n^T (\mathbf{Y} - \boldsymbol{\pi}) \to N(\mathbf{0}, \mathbf{I}_d) \tag{8}
$$

in distribution as $n \to \infty$ provided

$$\max_{1 \le i \le n} h_{ii}^W \to 0 \tag{9}$$

as $n \to \infty$, where h_{ii}^W is the ith diagonal entry of the matrix $\mathbf{X}_n(\mathbf{X}_n^T\mathbf{W}_n\mathbf{X}_n)^{-1}\mathbf{X}_n^T$. To show that (9) implies (8), set $U_i = (Y_i - \pi_i)\mathbf{x}_i$ and

$$\begin{aligned}
\tilde{U}_i &= \left(\textstyle\sum_{j=1}^n \operatorname{Cov} U_j\right)^{-1/2} U_i \\
&= \left(\textstyle\sum_{j=1}^n \pi_i(1-\pi_i)\mathbf{x}_i\mathbf{x}_i^T\right)^{-1/2} (Y_i - \pi_i)\mathbf{x}_i \\
&= (\mathbf{X}_n^T\mathbf{W}_n\mathbf{X}_n)^{-1}(Y_i - \pi_i)\mathbf{x}_i.
\end{aligned}$$

Then $\sum_{i=1}^n \tilde{U}_i = (\mathbf{X}_n^T\mathbf{W}_n\mathbf{X}_n)^{-1}\mathbf{X}_n^T(\mathbf{Y} - \boldsymbol{\pi})$. To establish (8) we must check whether $\tilde{U}_1, \ldots \tilde{U}_n$ satisfy the Lindeberg condition. For each $\varepsilon > 0$ we have

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{i=1}^n \mathbb{E}\|\tilde{U}_i\|_2^2 \mathbb{1}(\|\tilde{U}_i\|_2 > \varepsilon) \\
&\qquad = \lim_{n\to\infty} \sum_{i=1}^n \mathbb{E}\|(\mathbf{X}_n^T\mathbf{W}_n\mathbf{X}_n)^{-1}(Y_i-\pi_i)\mathbf{x}_i\|_2^2 \mathbb{1}(\|(\mathbf{X}_n^T\mathbf{W}_n\mathbf{X}_n)^{-1}(Y_i-\pi_i)\mathbf{x}_i\|_2 > \varepsilon) \\
&\qquad \le \lim_{n\to\infty} \sum_{i=1}^n \mathbf{x}_i^T(\mathbf{X}_n^T\mathbf{W}_n\mathbf{X}_n)^{-1}\mathbf{x}_i\pi_i(1-\pi_i)\mathbb{1}\left((\max_{1\le j\le n} \mathbf{x}_j^T(\mathbf{X}_n^T\mathbf{W}_n\mathbf{X}_n)^{-1}\mathbf{x}_j) > \varepsilon^2\right) \\
&\qquad = \operatorname{tr}(\mathbf{X}_n(\mathbf{X}_n^T\mathbf{W}_n\mathbf{X}_n)^{-1}\mathbf{X}_n^T\mathbf{W}_n) \lim_{n\to\infty}\left((\max_{1\le j\le n} h_{jj}^W) > \varepsilon^2\right) \\
&\qquad = d \lim_{n\to\infty}\left((\max_{1\le j\le n} h_{jj}^W) > \varepsilon^2\right) \\
&\qquad = 0,
\end{aligned}$$

by the assumption in (9).

3 Some bootstrap results based on the Lindeberg CLT

Theorem 3 (Consistency of the IID bootstrap for the sample mean)**.** *For each $n \ge 1$, let $X_{n1}, \ldots, X_{nn}$ be independent identically distributed random* *variables with* $\mathbb{E}X_{n1} = \mu$ *and* $\operatorname{Var} X_{n1} = \sigma^2 < \infty$ *and let* $\bar{X}_n = n^{-1}(X_{n1} + \cdots + X_{nn})$*. In addition, for each* $n \geq 1$*, conditional on* $X_{n1}, \ldots, X_{nn}$*, let* $X^*_{n1}, \ldots, X^*_{nn}$ *be independent random variables with distribution equal to the empirical distribution of* $X_{n1}, \ldots, X_{nn}$*. Then*

$$\sup_{x \in \mathbb{R}} \left| P_* \left(\sqrt{n} \frac{\bar{X}^*_n - \bar{X}_n}{\hat{\sigma}_n} \leq x \right) - P\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq x \right) \right| \to 0 \textit{ in probability as } n \to \infty,$$

where P_* *represents probability conditional on* $X_{n1}, \ldots, X_{nn}$*,* $n \geq 1$*,* $\bar{X}^*_n = n^{-1}(X^*_{n1} + \cdots + X^*_{nn})$*, and* $\hat{\sigma}^2_n = n^{-1} \sum_{i=1}^n (X_{ni} - \bar{X}_n)^2$*.*

Proof of Theorem 3. Since Corollary 1 gives

$$\sup_{x \in \mathbb{R}} \left| P\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq x \right) - \Phi(x) \right| \to 0 \text{ as } n \to \infty,$$

it is sufficient to show that

$$\sup_{x \in \mathbb{R}} \left| P_* \left(\sqrt{n} \frac{\bar{X}^*_n - \bar{X}_n}{\hat{\sigma}_n} \leq x \right) - \Phi(x) \right| \to 0 \text{ in probability as } n \to \infty. \quad (10)$$

Define $U^*_{ni} = X^*_{ni} - \bar{X}_n$. Then $\sum_{i=1}^n \operatorname{Var}_* U^*_{ni} = n\hat{\sigma}^2_n$ and

$$\sqrt{n}(\bar{X}^*_n - \bar{X}_n)/\hat{\sigma}_n = \frac{\sum_{i=1}^n U^*_{ni}}{\sqrt{\sum_{i=1}^n \operatorname{Var}_* U^*_{ni}}}.$$

Then by Theorem 1, (10) holds if for every $\epsilon > 0$ the random variables

$$\tilde{U}^*_{ni} = \frac{U^*_{ni}}{\sqrt{\sum_{j=1}^n \operatorname{Var}_* U^*_{nj}}} = \frac{X^*_{ni} - \bar{X}_n}{\sqrt{n}\hat{\sigma}_n}, \quad i = 1, \ldots, n, \quad n \geq 1,$$

satisfy

$$L_n(\epsilon) := \sum_{i=1}^n \mathbb{E}_* |\tilde{U}^*_{ni}|^2 \mathbb{1}(|\tilde{U}^*_{ni}| > \epsilon) \to 0 \text{ in probability as } n \to \infty. \quad (11)$$

We establish (11) by showing that for every $\delta > 0$, $P(L_n(\epsilon) > \delta) \to 0$ as $n \to \infty$, making use of the fact that $P(L_n(\epsilon) > \delta) \leq \delta^{-1}\mathbb{E}L_n(\epsilon)$, by Markov's inequality. Fix $\delta > 0$ and assume, without loss of generality, that $\mu = 0$ and $\sigma = 1$. Then

$$\delta^{-1}\mathbb{E}L_n(\epsilon) = \delta^{-1}\mathbb{E} \sum_{i=1}^n \mathbb{E}_* |\tilde{U}^*_{ni}|^2 \mathbb{1}(|\tilde{U}^*_{ni}| > \epsilon)$$

$$
\begin{aligned}
&= \delta^{-1}\mathbb{E}\sum_{i=1}^{n}\mathbb{E}_*\left|\frac{X_{ni}^* - \bar{X}_n}{\sqrt{n}\hat{\sigma}_n}\right|^2 \mathbb{1}\left(\left|\frac{X_{ni}^* - \bar{X}_n}{\sqrt{n}\hat{\sigma}_n}\right| > \epsilon\right) \\
&= \delta^{-1}\mathbb{E}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\left|\frac{X_{nj} - \bar{X}_n}{\sqrt{n}\hat{\sigma}_n}\right|^2 \mathbb{1}\left(\left|\frac{X_{nj} - \bar{X}_n}{\sqrt{n}\hat{\sigma}_n}\right| > \epsilon\right) \\
&= \delta^{-1}\mathbb{E}\left|\frac{X_{n1} - \bar{X}_n}{\hat{\sigma}_n}\right|^2 \mathbb{1}\left(\left|\frac{X_{n1} - \bar{X}_n}{\hat{\sigma}_n}\right| > \sqrt{n}\epsilon\right) \\
&\leq \delta^{-1}4\mathbb{E}\left[\left|\frac{X_{n1}}{\hat{\sigma}_n}\right|^2 \mathbb{1}\left(\left|\frac{X_{n1}}{\hat{\sigma}_n}\right| > \frac{\sqrt{n}\epsilon}{2}\right) + \left|\frac{\bar{X}_n}{\hat{\sigma}_n}\right|^2 \mathbb{1}\left(\left|\frac{\bar{X}_n}{\hat{\sigma}_n}\right| > \frac{\sqrt{n}\epsilon}{2}\right)\right] \\
&\leq \delta^{-1}4\mathbb{E}\left[\left|\frac{X_{n1}}{\hat{\sigma}_n}\right|^2 \mathbb{1}\left(\left|\frac{X_{n1}}{\hat{\sigma}_n}\right| > \frac{\sqrt{n}\epsilon}{2}\right) + \left|\frac{\bar{X}_n}{\hat{\sigma}_n}\right|^2\right] \\
&= \delta^{-1}4\mathbb{E}\left[\left|\frac{X_{n1}}{\hat{\sigma}_n}\right|^2 \mathbb{1}\left(\left|\frac{X_{n1}}{\hat{\sigma}_n}\right| > \frac{\sqrt{n}\epsilon}{2}\right)\mathbb{1}\left(\hat{\sigma}_n \geq \frac{1}{2}\right)\right. \\
&\qquad\left. + \left|\frac{X_{n1}}{\hat{\sigma}_n}\right|^2 \mathbb{1}\left(\left|\frac{X_{n1}}{\hat{\sigma}_n}\right| > \frac{\sqrt{n}\epsilon}{2}\right)\mathbb{1}\left(\hat{\sigma}_n < \frac{1}{2}\right) + \left|\frac{\bar{X}_n}{\hat{\sigma}_n}\right|^2\right] \\
&\leq \delta^{-1}8\mathbb{E}\left[|X_{n1}|^2 \mathbb{1}\left(|X_{n1}| > \frac{\sqrt{n}\epsilon}{4}\right)\mathbb{1}\left(\hat{\sigma}_n \geq \frac{1}{2}\right)\right. \\
&\qquad\left. + \left|\frac{X_{n1}}{\hat{\sigma}_n}\right|^2 \mathbb{1}\left(\hat{\sigma}_n < \frac{1}{2}\right) + \left|\frac{\bar{X}_n}{\hat{\sigma}_n}\right|^2\right],
\end{aligned}
$$

where the first inequality comes from the fact that for any two random variables U and V

$$
\mathbb{E}|U+V|^2\mathbb{1}(|U+V| > \gamma) \leq 4\mathbb{E}\left[|U|^2\mathbb{1}(U > \gamma/2) + |V|^2\mathbb{1}(V > \gamma/2)\right].
$$

Since $\hat{\sigma}_n \to 1$, $\mathbb{1}(\hat{\sigma}_n \geq 1/2) \to 1$ and $\mathbb{1}(\hat{\sigma}_n < 1/2) \to 0$ in probability, and the dominated convergence theorem gives that $\mathbb{E}|X_{n1}|^2\mathbb{1}(|X_{n1}| > \frac{\sqrt{n}\epsilon}{4}) \to 0$. Moreover $\mathbb{E}|X_{n1}/\hat{\sigma}_n|^2 < \infty$ and $\mathbb{E}|\bar{X}_n/\hat{\sigma}_n|^2 = O(n^{-1})$. This establishes (11), completing the proof. □

Theorem 4 (Consistency of residual bootstrap for linear regression)**.** *For each $n \geq 1$, let*

$$
Y_i = \mathbf{x}_i^T\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n,
$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent identically distributed random variables such that $\mathbb{E}\varepsilon_1 = 0$ and $\operatorname{Var}\varepsilon_1 = \sigma^2 < \infty$ and $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ are fixed vectors of dimension $d \leq n$, and let $\hat{\boldsymbol{\beta}}_n = (\mathbf{X}_n^T\mathbf{X}_n)^{-1}\mathbf{X}_n^T\mathbf{Y}_n$, where $\mathbf{X}_n =$

$(\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ *and* $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^T$. *Define* $\hat{\varepsilon}_i = Y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_n$ *for* $i = 1, \ldots, n$ *and, conditional on* $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$, *let* $\varepsilon_1^*, \ldots, \varepsilon_n^*$ *be independent random variables with distribution equal to the empirical distribution of* $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$. *Then let* $Y_i^* = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_n + \hat{\varepsilon}_i^*$, $i = 1, \ldots, n$, *and define* $\hat{\boldsymbol{\beta}}_n^* = (\mathbf{X}_n^T \mathbf{X}_n)^{-1} \mathbf{X}_n^T \mathbf{Y}_n^*$, *where* $\mathbf{Y}_n^* = (Y_1^*, \ldots, Y_n^*)^T$. *Finally, let* h_{ii}, $i = 1, \ldots, n$, *be the diagonal entries of* $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ *and assume*

$$\max_{1 \le i \le n} h_{ii} \to 0 \text{ as } n \to \infty. \tag{12}$$

Then

$$\sup_{A \in \mathcal{B}(\mathbb{R}^d)} \Big| P_* \left(\sqrt{n} \hat{\sigma}_n^{-1} (n^{-1} \mathbf{X}_n^T \mathbf{X}_n)^{1/2} (\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) \in A \right) - P \left(\sqrt{n} \sigma^{-1} (n^{-1} \mathbf{X}_n^T \mathbf{X}_n)^{1/2} (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \in A \right) \Big| \to 0 \text{ in probability as } n \to \infty,$$

where $\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_i^2$.

Proof of Theorem 4. From Example 2, we have

$$\sup_{A \in \mathcal{B}(\mathbb{R}^d)} \left| P \left(\sqrt{n} \sigma^{-1} (n^{-1} \mathbf{X}_n^T \mathbf{X}_n)^{1/2} (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \in A \right) - P(Z \in A) \right| \to 0 \text{ as } n \to \infty,$$

where Z has the $N(0, I_d)$ distribution. Therefore it is sufficient to show that

$$\sup_{A \in \mathcal{B}(\mathbb{R}^d)} \left| P_* \left(\sqrt{n} \hat{\sigma}_n^{-1} (n^{-1} \mathbf{X}_n^T \mathbf{X}_n)^{1/2} (\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) \in A \right) - P(Z \in A) \right| \to 0 \text{ in probability as } n \to \infty. \tag{13}$$

To show (13), we set $U_i^* = \mathbf{x}_i \varepsilon_i^*$, which allows us to write

$$\begin{aligned} \sqrt{n} \hat{\sigma}_n^{-1} (n^{-1} \mathbf{X}_n^T \mathbf{X}_n)^{1/2} (\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) &= (\hat{\sigma}_n^2 \mathbf{X}_n^T \mathbf{X}_n)^{-1/2} \sum_{i=1}^n \mathbf{x}_i \varepsilon_i^* \\ &= \left(\sum_{i=1}^n \operatorname{Cov} U_i^* \right)^{-1/2} \sum_{i=1}^n U_i^*. \end{aligned}$$

Then, by Theorem 2, it is sufficient to show that the random vectors

$$\tilde{U}_i^* = \left(\sum_{i=1}^n \operatorname{Cov} U_i^* \right)^{-1/2} U_i^* = (\hat{\sigma}_n^2 \mathbf{X}_n^T \mathbf{X}_n)^{-1/2} \mathbf{x}_i \varepsilon_i^*, \quad i = 1, \ldots, n, \quad n \ge 1,$$

satisfy

$$L_n(\delta) := \sum_{i=1}^{n} \mathbb{E}_* \|\tilde{U}_i^*\|_2^2 \mathbb{1}(\|\tilde{U}_i^*\|_2 > \delta) \to 0 \text{ in probability as } n \to \infty \tag{14}$$

for every $\delta > 0$. We establish (14) by showing that for every $\delta > 0$ and every $\eta > 0$, $P(L_n(\delta) > \eta) \to 0$ as $n \to \infty$, making use of the fact that $P(L_n(\delta) > \eta) \le \eta^{-1}\mathbb{E}L_n(\delta)$, by Markov's inequality. Fixing $\eta > 0$ and $\delta > 0$ we have

$$\begin{aligned}
\eta^{-1}\mathbb{E}L_n(\epsilon) &= \eta^{-1}\mathbb{E}\sum_{i=1}^{n} \mathbb{E}_* \|\tilde{U}_i^*\|_2^2 \mathbb{1}(\|\tilde{U}_i^*\|_2 > \delta) \\
&= \eta^{-1}\mathbb{E}\sum_{i=1}^{n} \mathbb{E}_* \|(\hat{\sigma}_n^2 \mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\varepsilon_i^*\|_2^2 \mathbb{1}(\|(\hat{\sigma}_n^2 \mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\varepsilon_i^*\|_2 > \delta) \\
&= \eta^{-1}\mathbb{E}\sum_{i=1}^{n} \frac{1}{n}\sum_{j=1}^{n} \|(\hat{\sigma}_n^2 \mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\hat{\varepsilon}_j\|_2^2 \mathbb{1}(\|(\hat{\sigma}_n^2 \mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\hat{\varepsilon}_j\|_2 > \delta) \\
&\le \eta^{-1}4\mathbb{E}\sum_{i=1}^{n} \frac{1}{n}\sum_{j=1}^{n} \|(\hat{\sigma}_n^2 \mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\varepsilon_j\|_2^2 \mathbb{1}(\|(\hat{\sigma}_n^2 \mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\varepsilon_j\|_2 > \delta/2) \\
&\quad + \eta^{-1}4\mathbb{E}\sum_{i=1}^{n} \frac{1}{n}\sum_{j=1}^{n} \|(\hat{\sigma}_n^2 \mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\mathbf{x}_j^T(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})\|_2^2 \\
&=: A_n + B_n, \text{ say,}
\end{aligned}$$

where the inequality comes from using the fact that $\hat{\varepsilon}_j = \varepsilon_j + \mathbf{x}_j^T(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ along with the fact that for any two random variables U and V

$$\mathbb{E}|U+V|^2\mathbb{1}(|U+V| > \gamma) \le 4\mathbb{E}\left[|U|^2\mathbb{1}(U > \gamma/2) + |V|^2\right].$$

To show $A_n \to 0$ and $B_n \to 0$ in probability, we will need to use the consistency result

$$\hat{\sigma}_n^2 \to \sigma^2 \text{ in probability as } n \to \infty, \tag{15}$$

which we will show as the last part of this proof. Now, assuming without loss of generality that $\sigma = 1$ and separating the cases $\hat{\sigma}_n > 1/2$ and $\hat{\sigma}_n \le 1/2$, we have

$$A_n = \eta^{-1}4\sum_{i=1}^{n} \mathbb{E}\|(\hat{\sigma}_n^2 \mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\varepsilon_1\|_2^2 \mathbb{1}(\|(\hat{\sigma}_n^2 \mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\varepsilon_1\|_2 > \delta/2)$$

$$
\begin{aligned}
&\leq \eta^{-1}16\sum_{i=1}^{n}\mathbb{E}\|(\mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\varepsilon_1\|_2^2\mathbb{1}(\|(\mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\varepsilon_1\|_2 > \delta/4)\mathbb{1}(\hat{\sigma}_n > 1/2)\\
&\qquad + \eta^{-1}4\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\|(\hat{\sigma}_n^2\mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\varepsilon_j\|_2^2\mathbb{1}(\hat{\sigma}_n \leq 1/2)\\
&\leq \eta^{-1}16\sum_{i=1}^{n}\mathbb{E}\|(\mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\varepsilon_1\|_2^2\mathbb{1}(\|(\mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\varepsilon_1\|_2 > \delta/4)\\
&\qquad + \eta^{-1}4\cdot d\cdot\mathbb{E}\left[\hat{\sigma}_n^{-2}n^{-1}\sum_{j=1}^{n}\varepsilon_j^2\mathbb{1}(\hat{\sigma}_n \leq 1/2)\right]\\
&\to 0 \text{ as } n\to\infty,
\end{aligned}
$$

where the first term goes to zero by the arguments in Example 2 and the second term goes to zero since $P(\hat{\sigma}_n \leq 1/2) \to 0$ and $\mathbb{E}\hat{\sigma}_n^{-2}n^{-1}\sum_{i=1}^{n}\varepsilon_j^2 \to 1$ as $n\to\infty$. In addition

$$
\begin{aligned}
B_n &= \eta^{-1}4\mathbb{E}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\|(\hat{\sigma}_n^2\mathbf{X}_n^T\mathbf{X}_n)^{-1/2}\mathbf{x}_i\mathbf{x}_j^T(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})\|_2^2\\
&= \eta^{-1}4\mathbb{E}\hat{\sigma}_n^{-2}\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\mathbf{x}_i^T(\mathbf{X}_n^T\mathbf{X}_n)^{-1}\mathbf{x}_i(\mathbf{x}_j^T(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}))^2\\
&= \eta^{-1}4\cdot d\cdot\mathbb{E}\hat{\sigma}_n^{-2}\frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j^T(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}))^2\\
&\leq \eta^{-1}4\cdot d\cdot\mathbb{E}\hat{\sigma}_n^{-2}\frac{1}{n}\|\mathbf{X}_n(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})\|_2^2\\
&\to 0 \text{ as } n\to\infty
\end{aligned}
$$

since $\mathbb{E}\hat{\sigma}_n^{-2}\|\mathbf{X}_n(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})\|_2^2 \to d$ (noting that $\|\mathbf{X}_n(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})\|_2^2$ converges to a χ_d^2 distribution by (7)).

We now prove the consistency result in (15). We have

$$
\begin{aligned}
\hat{\sigma}_n^2 &= \frac{1}{n}\hat{\varepsilon}_i^2\\
&= \frac{1}{n}\sum_{i=1}^{n}[\varepsilon_i - \mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})]^2\\
&= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - 2\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}) + \frac{1}{n}\sum_{i=1}^{n}|\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})|^2
\end{aligned}
$$

$$= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - \frac{1}{n}\|\mathbf{X}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})\|_2^2,$$

where the first term converges in probability to σ^2 by the WLLN. Since $\|\mathbf{X}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})\|_2^2$ converges to a χ_d^2 distribution, the second term converges in probability to zero, giving the result. □

Appendix

We now give a proof of the Lindeberg central limit theorem, which is essentially reproduced from [1].

Proof of Theorem 1. Define the notation

$$\sigma_{ni}^2 := \operatorname{Var} U_{ni}, \quad i = 1, \ldots, n, \quad n \geq 1,$$

and without loss of generality, let $\sum_{i=1}^n \sigma_{ni}^2 = 1$ for $n \geq 1$ (we can always divide each U_{ni} by σ_{ni}). Then, for some sequence $\varepsilon_n \to 0$, we have

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}|U_{ni}|^2 \mathbb{1}(|U_{ni}| > \varepsilon_n) = 0. \tag{16}$$

It is sufficient to show that the characteristic function of $\sum_{i=1}^n U_{ni}$ converges to that of the standard Normal distribution. Letting ϕ_{ni} represent the characteristic function of U_{ni}, $i = 1, \ldots, n$, $n \geq 1$, for any $t \in \mathbb{R}$, we have

$$\begin{aligned}
&\left|\mathbb{E}\exp\left(\iota t \sum_{j=1}^{n} U_{ni}\right) - \exp\left(-\frac{t^2}{2}\right)\right| \\
&\quad \leq \left|\prod_{i=1}^{n}\phi_{ni}(t) - \prod_{i=1}^{n}\left(1 - \frac{t^2\sigma_{ni}^2}{2}\right)\right| + \left|\prod_{i=1}^{n}\left(1 - \frac{t^2\sigma_{ni}^2}{2}\right) - \prod_{i=1}^{n}\exp\left(-\frac{t^2\sigma_{ni}^2}{2}\right)\right| \\
&\quad \leq \sum_{i=1}^{n}\left|\phi_{ni}(t) - \left(1 - \frac{t^2\sigma_{ni}^2}{2}\right)\right| + \sum_{i=1}^{n}\left|\exp\left(-\frac{t^2\sigma_{ni}^2}{2}\right) - \left(1 - \frac{t^2\sigma_{ni}^2}{2}\right)\right| \\
&\quad = A_n + B_n, \text{ say,}
\end{aligned}$$

where the second inequality comes from Lemma 11.1.3 of [1]. We show that A_n and B_n go to zero as $n \to \infty$. Since $|\exp(\iota x) - (1 + \iota x + (\iota x)^2/2)| \leq$

$\min\{|x|^3/3!, |x|^2\}$ for all $x \in \mathbb{R}$, for all $t \in \mathbb{R}$ we have

$$
\begin{aligned}
A_n &:= \sum_{i=1}^{n} \left| \phi_{ni}(t) - \left(1 - \frac{t^2\sigma_{ni}^2}{2}\right) \right| \\
&= \sum_{i=1}^{n} \left| \mathbb{E}\exp(\iota t U_{ni}) - \left(1 + \mathbb{E}\iota t U_{ni} + \frac{(\iota t)^2}{2!}\mathbb{E}U_{ni}^2\right) \right| \\
&\leq \sum_{i=1}^{n} \mathbb{E}\min\left\{\frac{|tU_{ni}|^3}{3!}, |tU_{ni}|^2\right\} \\
&\leq \sum_{i=1}^{n} \mathbb{E}|tU_{ni}|^3 \mathbb{1}(|U_{ni} \leq \varepsilon_n|) + \sum_{i=1}^{n} \mathbb{E}|tU_{ni}|^2 \mathbb{1}(|U_{ni}| > \varepsilon_n) \\
&\leq t^3\varepsilon_n \sum_{i=1}^{n} \mathbb{E}U_{ni}^2 + t^2 \sum_{i=1}^{n} \mathbb{E}|U_{ni}|^2 \mathbb{1}(|U_{ni}| > \varepsilon_n) \\
&\to 0 \text{ as } n \to \infty,
\end{aligned}
$$

since $\sum_{i=1}^n \mathbb{E}U_{ni}^2 = 1$ and $\varepsilon_n \to 0$ and by (16). Now, since $|e^x - 1 - x| \leq x^2 e^{|x|}$ for all $x \in \mathbb{R}$ (see pg. 347 of [1]), we may write

$$
\begin{aligned}
B_n &:= \sum_{i=1}^{n} \left| 1 - \frac{t^2\sigma_{ni}^2}{2} - \exp\left(-\frac{t^2\sigma_{ni}^2}{2}\right) \right| \\
&\leq \sum_{i=1}^{n} \left(\frac{t^2\sigma_{ni}^2}{2}\right) \exp\left(\frac{t^2\sigma_{ni}^2}{2}\right) \\
&\leq \frac{t^4}{4} \left(\max_{1\leq i\leq n} \sigma_{ni}^2\right) \exp\left[\frac{t^2}{2}\left(\max_{1\leq i\leq n} \sigma_{ni}^2\right)\right] \sum_{i=1}^{n} \sigma_{n,i}^2 \\
&\leq t^4 \left(\max_{1\leq i\leq n} \sigma_{ni}^2\right) \exp\left[t^2\left(\max_{1\leq i\leq n} \sigma_{ni}^2\right)\right].
\end{aligned}
$$

Lastly, we have

$$
\begin{aligned}
\max_{1\leq i\leq n} \sigma_{ni}^2 &= \max_{1\leq i\leq n} \mathbb{E}U_{ni}^2 \\
&= \max_{1\leq i\leq n} \mathbb{E}\left[|U_{ni}|^2 \mathbb{1}(|U_{ni}| \leq \varepsilon_n) + |U_{n,i}|^2 \mathbb{1}(|U_{ni}| > \varepsilon_n)\right] \\
&\leq \varepsilon_n^2 + \sum_{i=1}^{n} \mathbb{E}|U_{ni}|^2 \mathbb{1}(|U_{ni}| > \varepsilon_n)
\end{aligned}
$$

$$\to 0 \text{ as } n \to \infty,$$

by (16). This completes the proof. $\square$

References

[1] Krishna B Athreya and Soumendra N Lahiri. *Measure theory and probability theory.* Springer Science & Business Media, 2006.