# Asymptotics for Asymmetric Weighted U-Statistics: Central Limit Theorem and Bootstrap under Data Heterogeneity

Fang Han* and Tianchen Qian†

September 22, 2016

**Abstract**

Theoretical analysis of weighted U-statistics is long-time known to be complicated. Motivated from challenges on studying a new correlation measurement being popularized in evaluating online ranking algorithms' performance, this paper founds new theory for nondegenerate weighted U-statistics. Without any commonly adopted assumption, we establish the central limit theorem, and verify Efron's bootstrap and a new resampling procedure's inference validity. Specifically, our theory allows kernels and weights asymmetric and data points not identically distributed, which are all new issues that historically have not been addressed. For achieving strict generalization, for example, we have to carefully control the order of the "degenerate" term in U-statistics which are no longer degenerate under the empirical measure for non-i.i.d. data. Our theory applies to the motivating task, giving the region at which solid statistical inference can be made.

**Keywords:** weighted U-statistics, nondegeneracy, bootstrap inference, data heterogeneity, rank correlation, average-precision correlation.

## 1 Introduction

This paper studies asymptotics for the following nondegenerate weighted U-statistic of degree $m$:

$$U_n = \frac{(n-m)!}{n!} \sum_{\substack{1 \le i_1, i_2, \ldots, i_m \le n: \\ i_j \ne i_k \text{ if } j \ne k}} a_n(i_1, \ldots, i_m) h_n(X_{i_1}, \ldots, X_{i_m}). \tag{1.1}$$

Here we assume $X_1, \ldots, X_n$ are independent but not necessarily identically distributed random variables, taking values in a measurable space $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ (Korolyuk and Borovskich, 2013). The weight function $a_n(\cdot)$ and kernel function $h_n(\cdot)$ are both possibly asymmetric, and they are both allowed to be sample size dependent.

*Department of Statistics, University of Washington, Seattle, WA 98195, USA; e-mail: fanghan@uw.edu

†Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA; e-mail: tqian2@jhu.edu

Our study on asymmetric weighted U-statistics is motivated from the following new correlation measurement popularized in the information retrieval area (Yilmaz et al., 2008). It is formulated as a weighted U-statistic of asymmetric kernels and weights:

$$\tau^{\mathrm{AP}} := \frac{2}{n-1}\sum_{i=2}^{n}\frac{\sum_{j=1}^{i-1}\mathbb{1}(X_j > X_i)}{i-1} - 1. \tag{1.2}$$

Here $\mathbb{1}(\cdot)$ represents the indicator function and $X_1,\dots,X_n$ are specified to be real-valued. For this specific example, $X_1,\dots,X_n$ correspond to the scores the ranking machine gives for each online page, aligned by the rankings of human labels. The data points $X_1,\dots,X_n$ are usually modeled by a location-scale model, and are usually non-i.i.d.. The statistic in (1.2), named average-precision (AP) correlation, aims to evaluate the performance of any given online ranking algorithm by calculating a reweighted rank correlation measurement between the algorithm's rankings, while "giving more weights to the errors at high rankings". For the AP correlation, it is desirable to derive confidence intervals for solid inference.

Obviously, $\tau^{\mathrm{AP}}$ is an extension to the Kendall's tau statistic:

$$\tau^{\mathrm{Ken}} := \frac{2}{n(n-1)}\sum_{i\neq j}\Big\{\mathbb{1}(X_i > X_j)\mathbb{1}(i<j) + \mathbb{1}(X_i < X_j)\mathbb{1}(i>j)\Big\} - 1. \tag{1.3}$$

Compared to $\tau^{\mathrm{Ken}}$, the analysis of $\tau^{\mathrm{AP}}$ is much more involved, but naturally falls into the application regime of our theory.

The analysis of unweighted U-statistics (i.e., $a_n(\cdot)\equiv 1$) has a long history. There has been a vast literature on evaluating their asymptotic behaviors since the seminal paper of Hoeffding (1948). Specifically, regarding the simple independent and identically distributed (i.i.d.) setting, inference results have been summarized in Lee (1990), Serfling (2009), and Korolyuk and Borovskich (2013). For extensions, Lee (1990) showed the asymptotic normality under a Lyapunov-type non-i.i.d. condition. Yoshihara (1976) and Dehling and Wendler (2010) derived central limit theorem and (block) bootstrap inference validity for stationary weakly dependent time series. Csörgő and Nasari (2013) proved the $m$-out-of-$n$ bootstrap inference validity.

Weighted U-statistic is comparably less touched in the literature. Here, under the i.i.d. setting, Shapiro and Hubert (1979) and O'Neil and Redner (1993) conducted asymptotic analysis for weighted U-statistics of degree two. Major (1994) and Rifi and Utzet (2000) made extensions to weighted U-statistics of degree $m\geq 2$, with focus on the degenerate cases. Hsing and Wu (2004) relaxed the independence assumption, showing the asymptotic normality for a wide range of stationary stochastic processes. Recently, Zhou (2014) generalized the results in Hsing and Wu (2004), proving central and noncentral limit theorems for a class of nonstationary time series.

Despite the above substantial advances, two sets of assumptions are commonly required in the analysis of (weighted) U-statistics. First, the kernels and weights are required to be symmetric. Allowing both of them to be asymmetric, though requiring much more involved combinatorial analysis, is necessary for building rigorous inference for statistics like $\tau^{\mathrm{AP}}$[1]. Secondly, i.i.d. or sta-

[1] Of note, when a U-statistic has either kernel or weight function symmetric, it could be easily rewritten as U-

tionary assumption is commonly posed, especially for proving Efron's bootstrap inference validity. A notable exception is Zhou (2014), who established central limit theorem for nonstationary time series. However, bootstrap inference is not discussed, and the regularity conditions therein are too strong to include statistics like $\tau^{\mathrm{AP}}$.

Motivated from our study on the AP correlation, this manuscript aims to fill these gaps. In particular, we build unified theory for analyzing nondegenerate weighted U-statistics, namely, establishing sufficient conditions for their asymptotic normality and bootstrap inference validity. Both Efron's bootstrap and a new resampling procedure stemmed from Politis and Romano (1994) and Bickel et al. (1997) are considered. For this, we waive the above two sets of assumptions, allowing researchers to analyze statistics like $\tau^{\mathrm{AP}}$ in practical settings.

### 1.1 Other related work

Our results are very related to bootstrap inference under data heterogeneity. In Liu (1988), Regina Liu pioneered the study on Efron's bootstrap inference validity for non-i.i.d. models. Her results showed that bootstrap is robust to these specific non-i.i.d. settings with common locations (means). However, bootstrap is very sensitive to mean differences. The inference validity is captured by a function of $\{\mu_i := EX_i\}_{i=1}^n$, which she called "heterogeneity factors" (Liu, 1988; Liu and Singh, 1995). For example, for the sample mean, at the worse case, the distance between the largest and smallest means needs to shrink to zero as $n \to \infty$ for bootstrap consistency. Mammen (2012) summarized the existing results, providing necessary and sufficient conditions of bootstrap validity for the sample-mean-type statistics under non-i.i.d. settings.

Politis and Romano's subsampling (Politis et al., 1999) and many other resampling schemes (Bickel et al., 1997) are appealing alternatives to Efron's bootstrap. They are designed to correct the bootstrap inference inconsistency problem in many different settings, where the data could be, for example, dependent or heavy-tailed. In this paper, we examine a new resampling procedure's inference validity for weighted U-statistics.

### 1.2 Notation

Let $\mathbb{R}$ be the set of real numbers, and $\mathbb{Z}$ be the set of integers. For a positive integer $n$, we write $[n] = \{a \in \mathbb{Z} : 1 \le a \le n\}$. For any set $\mathcal{A}$, let $\mathrm{card}(\mathcal{A})$ represent the cardinality of $\mathcal{A}$. Let $\stackrel{d}{\to}$ denote "convergence in distribution", and $\stackrel{P}{\to}$ denote "convergence in probability". Let "a.s." be the abbreviation of "almost surely". Let $\Phi(t)$ be the cumulative distribution function of the standard Gaussian. For two positive integers $m < n$, define

$$\binom{n}{m} = \frac{n!}{(n-m)!m!},$$

where $n!$ represents the factorial of $n$. Let $C$ be a generic absolute positive constant, whose actual value may vary at different locations. For any two real sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$, or

statistic of both weight and kernel functions symmetric. However, this type of argument does not apply to the U-statistic of both weight and kernel functions asymmetric, like the AP correlation.

equivalently $b_n \gtrsim a_n$, if there exists an absolute constant $C$ such that $|a_n| \leq C|b_n|$ for all sufficiently large $n$. We write $a_n \asymp b_n$ if both $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ hold. We write $a_n \gtrnsim b_n$, or equivalently $b_n \lnsim a_n$, if $a_n \gtrsim b_n$ holds, but $a_n \lesssim b_n$ does not. We write $a_n = O(b_n)$ if $a_n \lesssim b_n$, and $a_n = o(b_n)$ if $a_n = O(b_n)$ and $b_n \neq O(a_n)$. We write $a_n = O_P(b_n)$ or $a_n = o_P(b_n)$ if $a_n = O(b_n)$ or $a_n = o_P(b_n)$ holds stochastically.

### 1.3 Structure of the paper

The rest of the paper is organized as follows. In Section 2 we provide the unified theory for asymmetric weighted U-statistic, proving central limit theorem, bootstrap, and a new resampling procedure's inference validity under data non-i.i.d. settings. In Section 3, we apply the developed theory to study Kendall's tau in (1.3) and AP correlation in (1.2). Section 4 contains the main proofs, with more relegated to the supplementary appendix.

## 2 Main results

Throughout the paper, we focus on the following triangular array setting: Assume we have $n$ independent random variables $\{X_{n,i}\}, n \geq 1, 1 \leq i \leq n$. Each $X_{n,i}$ follows the distribution $P_{n,i}$. $\{P_{n,i}, i \in [n]\}$ are not necessarily equal to each other. When $n$ increases, $P_{n,i}$ could possibly change. For notational simplicity, in the sequel we drop $n$ in the subscripts of $X_{n,i}$ and $P_{n,i}$ when no confusion could be made. All probabilities and expectations are in the outer measure sense.

We are focused on the following weighted U-statistic of degree $m$, with weight function $a(\cdot): \mathbb{Z}^m \to \mathbb{R}$ and kernel $h(\cdot): \mathcal{X}^m \to \mathbb{R}$:

$$U_n = U_n(X_1, \ldots, X_n) = \frac{(n-m)!}{n!} \sum_{I_n^m} a_n(i_1, \ldots, i_m) h_n(X_{i_1}, \ldots, X_{i_m}). \tag{2.1}$$

Here the summation is over all possible $m$ elements in $[n]$ without overlap:

$$I_n^m := \{1 \leq i_1, i_2, \ldots, i_m \leq n : i_j \neq i_k \text{ if } j \neq k\}.$$

Such $U_n$ is usually referred to as a weighted U-statistic in the literature (Serfling, 2009). We do not assume symmetry of $a_n(\cdot)$ or $h_n(\cdot)$ in their arguments. For notation simplicity, in the sequel we omit the subscript $n$ in $a_n(\cdot)$ and $h_n(\cdot)$.

Let's define

$$\theta(i_1, \ldots, i_m) := E\{h(X_{i_1}, \ldots, X_{i_m})\} = \int h(y_1, \ldots, y_m) dP_{i_1}(y_1) \ldots dP_{i_m}(y_m) \tag{2.2}$$

to be the population mean of $h(X_{i_1}, \ldots, X_{i_m})$. For any $l \in [m]$, define $\pi_l(\cdot;\cdot)$ to be a function that takes two arguments (a scalar and a vector of length $m-1$), and returns a vector of length $m$ by inserting the first argument into the $l$-th position of the second argument. Formally, we define

$$\pi_l(y; y_1, y_2, \ldots, y_{m-1}) := (y_1, \ldots, y_{l-1}, y, y_l, \ldots, y_{m-1}).$$

We further define

$$a^{(l)}(i;i_1,i_2,\dots,i_{m-1}) := a\{\pi_l(i;i_1,i_2,\dots,i_{m-1})\},$$
$$h^{(l)}(x;x_1,\dots,x_{m-1}) := h\{\pi_l(x;x_1,\dots,x_{m-1})\},$$
$$\theta^{(l)}(i;i_1,i_2,\dots,i_{m-1}) := \theta\{\pi_l(i;i_1,i_2,\dots,i_{m-1})\}.$$

Define the first order expansion of $h(\cdot)$ for each $X_i$, regarding the specific sequence $X_{i_1},\dots,X_{i_{m-1}}$, to be:

$$h_{1,i;i_1,\dots,i_{m-1}}(x) := \sum_{l=1}^{m} a^{(l)}(i;i_1,\dots,i_{m-1})\big\{f^{(l)}_{i_1,\dots,i_{m-1}}(x) - \theta^{(l)}(i;i_1,\dots,i_{m-1})\big\},$$

where

$$\begin{aligned} f^{(l)}_{i_1,\dots,i_{m-1}}(x) &:= E_{i_1,\dots,i_{m-1}}\{h^{(l)}(x;Y_1,\dots,Y_{m-1})\} \\ &= \int h^{(l)}(x;y_1,\dots,y_{m-1})dP_{i_1}(y_1)\dots dP_{i_{m-1}}(y_{m-1}). \end{aligned} \tag{2.3}$$

Define the first order expansion of $h(\cdot)$ for $X_i$ to be

$$h_{1,i}(x) := \frac{(n-m)!}{(n-1)!}\sum_{I^{m-1}_{n-1}(-i)} h_{1,i;i_1,\dots,i_{m-1}}(x), \tag{2.4}$$

where the summation is over

$$I^{m-1}_{n-1}(-i) := \Big\{1 \le i_1,\dots,i_{m-1} \le n : i_j \ne i_k \text{ if } j \ne k, \text{ and } i_j \ne i \text{ for all } j \in [m-1]\Big\}.$$

For $l \in [m]$, we write $(i_1,\dots,i_m)\backslash i_l := (i_1,\dots,i_{l-1},i_{l+1},\dots,i_m)$, and define

$$h_{2;i_1,\dots,i_m}(x_1,\dots,x_m) := h(x_1,\dots,x_m) - \sum_{l=1}^{m} f^{(l)}_{(i_1,\dots,i_m)\backslash i_l}(x_l) + (m-1)\theta(i_1,\dots,i_m), \tag{2.5}$$

where by (2.3) we have

$$f^{(l)}_{(i_1,\dots,i_m)\backslash i_l}(x) = \int h(y_1,\dots,y_{l-1},x,y_{l+1},\dots,y_m)dP_{i_1}(y_1)\dots dP_{i_{l-1}}(y_{l-1})dP_{i_{l+1}}(y_{l+1})\dots dP_{i_m}(y_m).$$

Before presenting the main theorem, we have to introduce more notation on the weight function $a(\cdot)$. For $K,q \in \mathbb{Z}$ with $K \ge 2$ and $0 \le q \le m$, let $(I^m_n)^{\otimes K}_{\ge q}$ be the collection of all $K$-dimensional index vectors from $I^m_n$ that share at least $q$ common indices:

$$(I^m_n)^{\otimes K}_{\ge q} := \Big\{(i^{(1)}_1,\dots,i^{(1)}_m) \in I^m_n,\dots,(i^{(K)}_1,\dots,i^{(K)}_m) \in I^m_n : \operatorname{card}\Big(\bigcap_{k=1}^{K}\{i^{(k)}_1,\dots,i^{(k)}_m\}\Big) \ge q\Big\},$$

and $(I^m_n)^{\otimes K}_{=q}$ be the collection of all $K$-dimensional index vectors from $I^m_n$ that share exactly $q$ indices in common:

$$(I^m_n)^{\otimes K}_{=q} = \Big\{(i^{(1)}_1,\dots,i^{(1)}_m) \in I^m_n,\dots,(i^{(K)}_1,\dots,i^{(K)}_m) \in I^m_n : \operatorname{card}\Big(\bigcap_{k=1}^{K}\{i^{(k)}_1,\dots,i^{(k)}_m\}\Big) = q\Big\}.$$

With fixed $K,q,m$, it is easy to observe $\text{card}\{(I_n^m)_{\geq q}^{\otimes K}\} \asymp \text{card}\{(I_n^m)_{=q}^{\otimes K}\}$ as $n \to \infty$, and

$$\text{card}\{(I_n^m)_{=q}^{\otimes K}\} \asymp \binom{n}{q}\binom{n-q}{m-q}\cdots\binom{n-(K-1)m-q}{m-q} \asymp n^{q+K(m-q)}.$$

In particular, we have $\text{card}\{(I_n^m)_{\geq 2}^{\otimes 2}\} \asymp n^{2m-2}$, $\text{card}\{(I_n^m)_{\geq 1}^{\otimes 2}\} \asymp n^{2m-1}$, and $\text{card}\{(I_n^m)_{\geq 1}^{\otimes 3}\} \asymp n^{3m-2}$. Define the average weight, $A_{K,q}(n)$, as

$$A_{K,q}(n) := \frac{1}{\text{card}\{(I_n^m)_{\geq q}^{\otimes K}\}} \sum_{(I_n^m)_{\geq q}^{\otimes K}} \left| a(i_1^{(1)},\ldots,i_m^{(1)}) \cdots a(i_1^{(K)},\ldots,i_m^{(K)}) \right|. \tag{2.6}$$

The following theorem gives sufficient conditions on the weights and distributions of $\{X_i\}$ for guaranteeing $U_n$ to be asymptotically normal.

**Theorem 2.1** (Sufficient condition for asymptotic normality of $U_n$)**.** For each $n$, assume there exists a positive constant $M(n) > 0$ only depending on $n$ such that

$$\sup_{(i_1,\ldots,i_m)\in I_n^m} E\{h(X_{i_1},\ldots,X_{i_m})^4\} \leq M(n). \tag{2.7}$$

Define $V(n) = \text{Var}\{n^{-1}\sum_{i=1}^n h_{1,i}(X_i)\}$ with $h_{1,i}(\cdot)$ defined in (2.4). Assume the following conditions hold:

$$n^{-2}V(n)^{-1}A_{2,2}(n)M(n)^{1/2} \to 0, \tag{2.8}$$

$$n^{-2}V(n)^{-3/2}A_{3,1}(n)M(n)^{3/4} \to 0. \tag{2.9}$$

Then we have

$$\text{Var}(U_n)/V(n) \to 1, \tag{2.10}$$

and

$$\text{Var}(U_n)^{-1/2}\{U_n - E(U_n)\} \xrightarrow{d} N(0,1). \tag{2.11}$$

The proof of Theorem 2.1 is very involved. However, the first step of the proof, which establishes a von-Mises-expansion type result, is simple yet inspiring. Of note, for symmetric weighted U-statistics under i.i.d. settings, an analogous theorem has been (inexplicitly) stated in Shapiro and Hubert (1979).

**Lemma 2.2** (Hoeffding's decomposition)**.** With $h_{1,i}(\cdot)$ and $h_{2;i_1,\ldots,i_m}(\cdot)$ defined in (2.4) and (2.5), we have

$$U_n - E(U_n) = \frac{1}{n}\sum_{i=1}^n h_{1,i}(X_i) + U_n(a,h_2), \tag{2.12}$$

where

$$U_n(a,h_2) := \frac{(n-m)!}{n!}\sum_{I_n^m} a(i_1,\ldots,i_m)h_{2;i_1,\ldots,i_m}(X_{i_1},\ldots,X_{i_m}), \tag{2.13}$$

and for any $i,k \in [n]$ and $(i_1,\ldots,i_m) \in I_n^m$,

$$E\{h_{1,i}(X_i)\} = 0, \tag{2.14}$$

$$E\{h_{2;i_1,\ldots,i_m}(X_{i_1},\ldots,X_{i_m}) \mid X_k\} = 0 \ \text{ a.s..} \tag{2.15}$$

For putting Theorem 2.1 appropriately in the literature, let's first give a brief review on the most relevant existing results. The first proof of asymptotic normality for (unweighted) nondegenerate U-statistics was given in Hoeffding (1948). Grams and Serfling (1973) studied general unweighted U-statistics of degree $m \geq 2$ and bounded their central moments . The techniques therein also play a central role in our analysis. Shapiro and Hubert (1979) and O'Neil and Redner (1993) analyzed the asymptotic behavior of weighted U-statistics of degree 2. They assumed weight function $a(\cdot)$ symmetric. The above results all assume data i.i.d.-ness. For unweighted U-statistics, Lee (1990) outlined an extension to non-i.i.d. data.

Theorem 2.1 is stronger than the results in the literature, allowing $a(\cdot)$ and $h(\cdot)$ asymmetric, and the $X_i$'s non-i.i.d.. By examining the proof, one can also easily check that, when the corresponding symmetry, boundedness, or i.i.d. assumptions are made, our results can reduce to the ones in Hoeffding (1948), Shapiro and Hubert (1979), O'Neil and Redner (1993), and Lee (1990).

**Remark 2.3.** Condition (2.8) is added to enforce domination of $n^{-1}\sum_{i=1}^n h_{1,i}(X_i)$ over $U_n(a,h_2)$ in (2.12). Condition (2.9) evolves from the Lyapunov condition with $\delta = 1$, which is readily weaken to the condition of a smaller $0 < \delta < 1$ or the Lindeberg-Feller condition. Condition (2.7) is made and could be weakened based on the same argument. For presentation clearness, we choose the current conditions.

Inferring the distribution of $U_n$ or approximating $\mathrm{Var}(U_n)$ is usually challenging in practice. Resampling procedures are hence recommended. The rest of this section gives asymptotic results for Efron's bootstrap (Efron, 1979) and a new resampling procedure for approximating $\mathrm{Var}(U_n)$.

Due to the heterogeneity in $P_i$, it is well known that bootstrap could possibly no longer be consistent (Liu, 1988). However, it is still possible to recover bootstrap consistency by restricting the heterogeneity degree. But before that, let's first provide a theoretically interesting theorem. It states that, under very mild conditions, bootstrapped mean from the set $\{h_{1,i}(X_i) : 1 \leq i \leq n\}$ approximates the distribution of $n^{-1}\sum_{i=1}^n h_{1,i}(X_i)$ consistently. This is consistent to the discovery in Liu (1988) by noting that $E\{h_{1,i}(X_i)\} = 0$ no matter how different $\{P_i\}_{i=1}^n$ are.

**Theorem 2.4** (Sufficient condition for bootstrapping main term to work)**.** Denote

$$\sigma_n^2 := \mathrm{Var}(U_n). \tag{2.16}$$

Consider the term $n^{-1}\sum_{i=1}^n h_{1,i}(X_i)$ with $h_{1,i}(X_i)$ defined in (2.4) and its bootstrapped version $n^{-1}\sum_{i=1}^n \{h_{1,i}(X_i)\}^*$, where conditional on $X_1,\ldots,X_n$ the $\{h_{1,i}(X_i)\}^*$'s are i.i.d. draws from the empirical distribution of $\{h_{1,j}(X_j) : 1 \leq j \leq n\}$. Assume (2.7) and (2.9) hold. In addition, assume for every $\epsilon > 0$, we have

$$\sup_{1\le i\le n} P\Big\{\Big|\frac{h_{1,i}(X_i)}{n\sigma_n}\Big| \ge \epsilon\Big\} \to 0, \tag{2.17}$$

$$\sum_{i=1}^{n}\Big[E\Big\{\frac{h_{1,i}(X_i)}{n\sigma_n}\mathbb{1}\Big(\Big|\frac{h_{1,i}(X_i)}{n\sigma_n}\Big| \le \epsilon\Big)\Big\}\Big]^2 \to 0. \tag{2.18}$$

Then

$$\sup_{t\in\mathbb{R}}\Big|P^*\Big\{\sum_{i=1}^{n}\frac{\{h_{1,i}(X_i)\}^*}{n\sigma_n} - \sum_{i=1}^{n}\frac{h_{1,i}(X_i)}{n\sigma_n} \le t\Big\} - P\Big\{\sum_{i=1}^{n}\frac{h_{1,i}(X_i)}{n\sigma_n} \le t\Big\}\Big| \xrightarrow{P} 0, \tag{2.19}$$

where $P^*$ denotes the conditional probability given $X_1,\ldots,X_n$. If further (2.8) holds, then

$$\sup_{t\in\mathbb{R}}\Big|P^*\Big\{\sum_{i=1}^{n}\frac{\{h_{1,i}(X_i)\}^*}{n\sigma_n} - \sum_{i=1}^{n}\frac{h_{1,i}(X_i)}{n\sigma_n} \le t\Big\} - P\Big\{\mathrm{Var}(U_n)^{-1/2}\{U_n - E(U_n)\} \le t\Big\}\Big| \xrightarrow{P} 0. \tag{2.20}$$

**Remark 2.5.** Equations (2.17) and (2.18) are rather mild constraints. As we will show in Corollary 3.1, usually they can be directly deduced from the asymptotic normality of $U_n$. However, unless we know much about $X_i$, the form of $h_{1,i}(\cdot)$ is unknown.

We now focus on bootstrapping the original U-statistic for estimating $\mathrm{Var}(U_n)$. The following theorem shows that Efron's bootstrap still gives consistent variance estimate for $U_n$ under some additional conditions on data heterogeneity. Although the bootstrap inference validity for U-statistics under i.i.d. assumptions has been established (check, for example, Korolyuk and Borovskich (2013)), the corresponding one for non-i.i.d. settings, even for the simplest unweighted U-statistics, is still absent in the literature. Our paper fills this gap.

**Theorem 2.6** (Sufficient condition for consistent bootstrap variance estimation)**.** Given $X_1,\ldots,X_n$, let $X_1^*,\ldots,X_n^*$ denote the bootstrapped sample, which are i.i.d. draws from the empirical distribution of $X_1,\ldots,X_n$. Define the bootstrapped U-statistic

$$U_n^* = \frac{(n-m)!}{n!}\sum_{I_n^m} a(i_1,\ldots,i_m)h(X_{i_1}^*,\ldots,X_{i_m}^*).$$

Assume all conditions in Theorem 2.1 are satisfied. Also assume the following conditions hold:

(i) Bounded second moment of von-Mises type kernel:

$$\limsup_{n\to\infty}\max_{1\le i_1,\ldots,i_m\le n} E\{h(X_{i_1},\ldots,X_{i_m})^2\} < \infty. \tag{2.21}$$

(ii) Control of heterogeneity in the distributions of $X_i$:

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big\{\frac{h_{1,i}(X_j)}{n\sigma_n}\Big\}^2 \xrightarrow{P} 1, \tag{2.22}$$

$$\frac{1}{n^2}\sum_{i=1}^{n}\Big\{\sum_{j=1}^{n}\frac{h_{1,i}(X_j)}{n\sigma_n}\Big\}^2 \xrightarrow{P} 0, \tag{2.23}$$

and

$$n^{-1}\sigma_n^{-2}A_{2,1}(n)\{M_1(n)^2+M_2(n)+n^{-1}\}\to 0, \tag{2.24}$$

where

$$M_1(n)=\max_{(I_n^m)^{\otimes 2}_{\geq 0}}|\theta(i_1,\ldots,i_m)-\theta(j_1,\ldots,j_m)|, \tag{2.25}$$

$$M_2(n)=\max_{1\leq p,q\leq m}\max_{\substack{\boldsymbol{r},\boldsymbol{s}\in(I_n^m)^{\otimes 2}_{=1}\\ \boldsymbol{r}\cap\boldsymbol{s}=r_p=s_q}}\max_{\substack{\boldsymbol{k}\in I_n^m\\ \boldsymbol{k}\cap\boldsymbol{s}=k_p=s_q}}\Big|E[E\{h(X_{r_1},\ldots,X_{r_m})h(X_{s_1}\ldots X_{s_m})\mid X_{k_p}\}]$$

$$-E[E\{h(X_{k_1},\ldots,X_{k_m})h(X_{s_1},\ldots,X_{s_m})\mid X_{k_p}\}]\Big|. \tag{2.26}$$

Here we define $\boldsymbol{r}:=(r_1,\ldots,r_m)$, and similarly for $\boldsymbol{s},\boldsymbol{k}$.

Then we have

$$\left|\mathrm{Var}^*(\sigma_n^{-1}U_n^*)-\mathrm{Var}(\sigma_n^{-1}U_n)\right|\xrightarrow{P}0, \tag{2.27}$$

where the operator $\mathrm{Var}^*(\cdot)$ denotes the conditional variance given $X_1,\ldots,X_n$.

The detailed proof of Theorem 2.6 is very involved and highly combinatorial. We defer it to Section 4. Of note, in the theorem, (2.21) comes from Bickel and Freedman (1981), ensuring that the bootstrapped U-statistic won't explode. Equations (2.22) and (2.23) ensure that the conditional variance of $n^{-1}\sum_{i=1}^n h_{1,i}(X_i^*)$ approximates $\mathrm{Var}(U_n)$. Equation (2.24) ensures that $U_n^*(a,h_2)$ is negligible compared to $n^{-1}\sum_{i=1}^n h_{1,i}(X_i^*)$.

**Remark 2.7.** Although $U_n(a,h_2)$ in the decomposition (2.12) is degenerate and hence negligible under the conditions of Theorem 2.1, its bootstrapped version $U_n^*(a,h_2)$ is not necessarily degenerate. This makes $U_n^*(a,h_2)$ not necessarily negligible compared to the bootstrapped version of the main term, $n^{-1}\sum_{i=1}^n h_{1,i}(X_i^*)$. The reason for the blown-up $U_n^*(a,h_2)$ is the heterogeneity in $\{P_i,i\in[n]\}$. Therefore, bootstrap may fail without careful control on both the main term and the remainder $U_n^*(a,h_2)$. We developed delicate analysis to bound $U_n^*(a,h_2)$ and showed that it is negligible under the constraint (2.24).

**Remark 2.8.** Condition (2.24) puts homogeneity conditions mainly on the means. This is consistent to Theorem 2.4 and the discoveries in Liu (1988), who showed that bootstrap is most sensitive to mean differences. To illustrate, assume $a(\cdot)\equiv 1$ and the kernel $h(\cdot)$ to be a bounded function. Assume the assumptions in Theorem 2.1 hold, so that we have asymptotic normality of $U_n$. Equation (2.9) requires $\sigma_n^2\gtrsim n^{-4/3}$. Therefore, for (2.24) to hold, it is necessary that $M_1(n)^2\lesssim n^{-1/3}$ and $M_2(n)\lesssim n^{-1/3}$. The space to improve our requirements, if existing, is relatively small. This is by noting that, even for the simplest sample-mean-type statistics, for most cases, Liu (1988) required the mean differences shrink to zero as $n\to\infty$ for bootstrap consistency.

An immediate implication of Theorem 2.6 proves the validity of bootstrapping weighted U-statistics for i.i.d. data.

**Corollary 2.1.** Assume that $X_1,\ldots,X_n$ are i.i.d., and that (2.7), (2.8), (2.9), and (2.21) hold. In addition, assume $n^{-2}\sigma_n^{-2}A_{2,1}(n)\to 0$. Then (2.22), (2.23), and (2.24) hold, and we have

$$\left|\mathrm{Var}^*(\sigma_n^{-1}U_n^*)-\mathrm{Var}(\sigma_n^{-1}U_n)\right|\xrightarrow{P}0.$$

**Remark 2.9.** The assumption $n^{-2}\sigma_n^{-2}A_{2,1}(n)\to 0$ is mild. Actually it follows immediately from (2.8) if we have $A_{2,1}(n)\lesssim A_{2,2}(n)$. It is reasonable to expect $A_{2,1}(n)$ and $A_{2,2}(n)$ to be of similar order because of their definitions in (2.6). Indeed, for the two applications in Section 3, we have $A_{2,1}(n)\asymp A_{2,2}(n)$ for $U_n^{\mathrm{Ken}}$ and $A_{2,1}(n)\lesssim A_{2,2}(n)\lesssim A_{2,1}(n)\log n$ for $U_n^{\mathrm{AP}}$.

In many cases, although the data are in general non-i.i.d., they possess some locally stationary property. For example, consider the following nonparametric regression model. Assume $X_i\sim N(\mu_i,1)$ with $\mu_i=g_n(i/n)$ for $i=1,\ldots,n$. If the function $g_n(\cdot)$ is smooth enough (e.g., $\epsilon(n)$-Lipschitz), then, although $\max|g_n(1)-g_n(0)|$ could increase to infinity, the subsample $\{X_i,X_{i+1},\ldots,X_{i+b-1}\}$, for each $i\in 1,\ldots,n-b+1$, can be approximately i.i.d..

Adopting this thinking, we consider the following revised resampling procedure whose idea comes from Politis and Romano (1994) and Bickel et al. (1997), but is tailored for non-i.i.d. data. In detail, for $m<b\to\infty$, we consider the following statistic:

$$V_n^*=\frac{1}{h_n(n-b+1)}\sum_{i=1}^{n-b+1}\mathrm{Var}^*(U_{b,i}^*),\quad \text{where}\quad U_{b,i}^*:=\frac{(b-m)!}{b!}\sum_{I_b^m}a(i_1,\ldots,i_m)h(X_{i_1,b,i}^*,\ldots,X_{i_m,b,i}^*),$$

and for each $i\in[n-b+1]$, $X_{i_1,b,i}^*,\ldots,X_{i_m,b,i}^*$ are independently drawn from the empirical distribution of $\{X_i,\ldots,X_{i+b-1}\}$ with replacement. The tuning parameter $h_n$ regulates the scale.

The following theorem verifies the new resampling procedure's inference consistency for $V_n^*$, showing that the procedure tends to give conservative variance estimate under non-i.i.d. settings. It also shows that the inference is more tractable compared to Efron's bootstrap when we have more prior information on the heterogeneity degree, reflected in the consistency rate of $U_n$ and the choice of $h_n$. We also refer the readers to Remark 3.4 and discussions therein for the order of $\sigma_n$ in a specific example.

**Theorem 2.10.** Assume that all conditions in Theorem 2.6 hold for each "moving block" $\{X_i,\ldots,X_{i+b-1}\}$ of $i\in[n-b+1]$ as $n,b\to\infty$. Assume $\mathrm{Var}(U_b(X_i,\ldots,X_{i+b-1}))=\sigma_b^2(1+o(1))$ for any $i\in[n-b+1]$, and $\sigma_b^2/\sigma_n^2=\zeta_{n,b}\cdot(1+o(1))$ for some $\zeta_{n,b}>0$. We then have

$$\sigma_n^{-2}V_n^*-\mathrm{Var}(\sigma_n^{-1}U_n)=\frac{\zeta_{n,b}}{h_n}\cdot(1+o_P(1))-1.$$

The proof is a simple consequence of the Rao-Blackwell theorem combined with the proofs in Theorem 2.6. The details are hence omitted.

# 3 Application

This section studies our motivating statistics, the Kendall's tau (denoted as $\tau^{\text{Ken}}$) (Kendall, 1938) and average-precision (AP) correlation (denoted as $\tau^{\text{AP}}$) (Yilmaz et al., 2008):

$$\tau^{\text{Ken}} = \frac{2}{n(n-1)} \sum_{i \neq j} \{\mathbb{1}(X_i > X_j)\mathbb{1}(i < j) + \mathbb{1}(X_j > X_i)\mathbb{1}(j < i)\} - 1,$$

$$\tau^{\text{AP}} = \frac{2}{n-1} \sum_{i=2}^{n} \frac{\sum_{j=1}^{i-1} \mathbb{1}(X_j > X_i)}{i-1} - 1.$$

Without loss of generality, we focus on the transformed versions of these two statistics:

$$U_n^{\text{Ken}} = \frac{\tau^{\text{Ken}} + 1}{4} = \frac{1}{n(n-1)} \sum_{i \neq j} \mathbb{1}(j < i)\mathbb{1}(X_j > X_i),$$

$$U_n^{\text{AP}} := \frac{\tau^{\text{AP}} + 1}{2} = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{n\mathbb{1}(j < i)}{i-1} \mathbb{1}(X_j > X_i).$$

We assume $\{P_i, i \in [n]\}$ to be absolutely continuous with regard to the Lebesgue measure. Obviously, both $U_n^{\text{Ken}}$ and $U_n^{\text{AP}}$ enjoy the distribution-free property (Kendall and Stuart, 1973) when the data are i.i.d..

## 3.1 Asymptotic theory

Note that the statistics $U_n^{\text{Ken}}$ and $U_n^{\text{AP}}$ have the same kernel $h(x,y) = \mathbb{1}(y > x)$. Using the definition in (2.2), we have $\theta(i,j) = E\{h(X_i, X_j)\} = P(X_j > X_i)$. The forms of $h_{1,i}(\cdot)$ and $h_{2;i,j}(\cdot)$ for $U_n^{\text{Ken}}$ and $U_n^{\text{AP}}$ are then summarized in the following two lemmas.

**Lemma 3.1** (Hoeffding's decomposition of $U_n^{\text{Ken}}$)**.** We have

$$U_n^{\text{Ken}} - E(U_n^{\text{Ken}}) = \frac{1}{n} \sum_{i=1}^{n} h_{1,i}^{\text{Ken}}(X_i) + \frac{1}{n(n-1)} \sum_{i \neq j} \mathbb{1}(j < i) h_{2;i,j}^{\text{Ken}}(X_i, X_j),$$

where

$$h_{1,i}^{\text{Ken}}(x) = \frac{1}{n-1} \sum_{j=1}^{n} \{\mathbb{1}(j < i) - \mathbb{1}(j > i)\}\{P(X_j > x) - \theta(i,j)\} \tag{3.1}$$

and

$$h_{2;i,j}^{\text{Ken}}(x,y) = \mathbb{1}(y > x) - P(X_j > x) - P(y > X_i) + \theta(i,j).$$

**Lemma 3.2** (Hoeffding's Decomposition of $U_n^{AP}$)**.** We have

$$U_n^{\text{AP}} - E(U_n^{\text{AP}}) = \frac{1}{n} \sum_{i=1}^{n} h_{1,i}^{\text{AP}}(X_i) + \frac{1}{n(n-1)} \sum_{i \neq j} \frac{n\mathbb{1}(j < i)}{i-1} h_{2;i,j}^{\text{AP}}(X_i, X_j),$$

where

$$h_{1,i}^{\text{AP}}(x) = \frac{1}{n-1} \sum_{j=1}^{n} \left\{ \frac{n\mathbb{1}(j < i)}{i-1} - \frac{n\mathbb{1}(j > i)}{j-1} \right\} \{P(X_j > x) - \theta(i,j)\}, \tag{3.2}$$

and

$$h^{\mathrm{AP}}_{2;i,j}(x,y) = \mathbb{1}(y > x) - P(X_j > x) - P(y > X_i) + \theta(i,j).$$

In (3.2), by convention, we have $0/0 := 0$.

The next theorem characterizes sufficient distributional conditions for $U_n^{\mathrm{Ken}}$ and $U_n^{\mathrm{AP}}$ to be asymptotically normal, allowing for data non-i.i.d.-ness.

**Theorem 3.3** (Sufficient conditions for asymptotic normality of $U_n^{\mathrm{Ken}}$ and $U_n^{\mathrm{AP}}$)**.** Assume a sequence $\{\delta_n \in (0,1)\}_{n=1}^{\infty}$ and a sequence $\{p_n \in (0,1)\}_{n=1}^{\infty}$ such that for any sufficiently large $n$ and for each $i \in [n]$, one of the following two conditions holds:

(i) $P\{P(X_j > X_i \mid X_i) - P(X_j > X_i) \in [\delta_n, 1], \forall j \in [n]\backslash\{i\}\} \geq p_n$;

(ii) $P\{P(X_j > X_i \mid X_i) - P(X_j > X_i) \in [-1, -\delta_n], \forall j \in [n]\backslash\{i\}\} \geq p_n$.

In addition, if

$$\delta_n^3 p_n \gtrsim n^{-1/3}, \tag{3.3}$$

then $U_n^{\mathrm{Ken}}$ is asymptotically normal,

$$\mathrm{Var}(U_n^{\mathrm{Ken}})^{-1/2}\{U_n^{\mathrm{Ken}} - E(U_n^{\mathrm{Ken}})\} \xrightarrow{d} N(0,1).$$

If we have

$$\delta_n^3 p_n \gtrsim n^{-1/3}(\log n)^2, \tag{3.4}$$

then $U_n^{\mathrm{AP}}$ is asymptotically normal,

$$\mathrm{Var}(U_n^{\mathrm{AP}})^{-1/2}\{U_n^{\mathrm{AP}} - E(U_n^{\mathrm{AP}})\} \xrightarrow{d} N(0,1).$$

The proof of Theorem 3.3 exploits Theorem 2.1. A key step in the proof is to bound $V(n) := n^{-2}\sum_i \mathrm{Var}\{h_{1,i}(X_i)\}$ from below. The magnitude of $\mathrm{Var}\{h_{1,i}(X_i)\}$ varies greatly with different $i$, making it a challenging task to bound the entire summation. To tackle this, we break $V(n)$ into summations over multiple subsets of $[n]$. Within each of these summations, the magnitude of $\mathrm{Var}\{h_{1,i}(X_i)\}$ is stable. Then we develop bounds on the summations for $i$ with large $\mathrm{Var}\{h_{1,i}(X_i)\}$. The detailed proof is put in Section 4.

The sequences $\{\delta_n\}$ and $\{p_n\}$ in Conditions (i) and (ii) of Theorem 3.3 characterize the heterogeneity degree among the $P_i$'s. If all $P_i$'s are identical, it is easy to check that there exist absolute constants $\delta_n$ and $p_n$ not depending on $n$ such that Condition (i) or (ii) holds. Equations (3.3) and (3.4) allow $\delta_n$ and $p_n$ to decay to zero as $n \to \infty$. The legitimate decaying rate of $\delta_n^3 p_n$ depends on the average weight of each of the two statistics. The conditions for asymptotic normality of $U_n^{\mathrm{AP}}$ (3.4) are slightly stronger than that for $U_n^{\mathrm{Ken}}$ (3.3), because for $U_n^{\mathrm{AP}}$ the weight is more asymmetric.

**Remark 3.4.** In the literature about Kendall's tau, the classical result gives root-$n$ convergence rate (Sen, 1968). Theorem 3.3 gives a more general result regarding the convergence rate due to the non-i.i.d.-ness of $\{X_1, \ldots, X_n\}$. In the proof of Theorem 3.3, we show that the $\mathrm{Var}(U_n^{\mathrm{Ken}}) \gtrsim n^{-1}\delta_n^3 p_n$.

As we vary the distribution of $X_i$ from i.i.d. to the more heterogeneous ones, $\delta_n^3 p_n$ changes from $O(1)$ to $O(n^{-1/3+\epsilon})$ for some small $\epsilon > 0$. Therefore, the upper bound on the order of $\mathrm{Var}(U_n^{\mathrm{Ken}})^{-1/2}$ can vary from $n^{1/2}$ to $n^{2/3-\epsilon/2}$.

Motivated by the studies in Yilmaz et al. (2008), in the sequel we consider the following specific location-scale model. In particular, given two sets of real values $\mu_i$ with $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$ and $\sigma_1^2, \ldots, \sigma_n^2 > 0$, let's consider absolute continuous (with respect to Lebesgue measure) probability distribution $P_i$ with mean $\mu_i$ and variance $\sigma_i^2$ for $i \in [n]$. Assume $X_1, \ldots, X_n$ are independent draws from $P_1, \ldots, P_n$. The following theorem characterizes the explicit sufficient conditions on $\{(\mu_i, \sigma_i), i \in [n]\}$ for Kendall's tau and AP correlation to be asymptotically normal.

**Theorem 3.5** (Sufficient condition for asymptotic normality of $U_n^{\mathrm{Ken}}$ and $U_n^{\mathrm{AP}}$ under two tail conditions)**.** For each $i \in [n]$, assume $X_i$ follows distribution $P_i$ with mean $\mu_i$ and variance $\sigma_i^2$. Define

$$r_{ij} := (\mu_i - \mu_j)/\sigma_i, \quad R_n := \max_{1 \leq i \neq j \leq n} |r_{ij}|, \quad \rho_{ij} := \sigma_i/\sigma_j, \text{ and } \quad \rho_n := \max_{1 \leq i \neq j \leq n} \rho_{ij}.$$

For $n, i, j$ such that $1 \leq i \neq j \leq n$, define

$$F_j^c(t) = P\Big(\frac{X_j - \mu_j}{\sigma_j} > t\Big) \quad \text{and} \quad F_{ji}^c(t) = P\Big\{\frac{X_j - X_i - (\mu_j - \mu_i)}{(\sigma_i^2 + \sigma_j^2)^{1/2}} > t\Big\}. \tag{3.5}$$

Then the following results hold.

(i) Assume there exist absolute constants $c_1, c_2 > 0$, $b_1 > b_2 > 0$, and $t_0 > 0$, such that for any $n, i, j$ with $1 \leq i \neq j \leq n$ and for any $t \geq t_0$,

$$c_1 t^{-b_1} \leq F_j^c(t) \leq c_2 t^{-b_2}, \tag{3.6}$$

$$c_1 t^{-b_1} \leq F_{ji}^c(t) \leq c_2 t^{-b_2}. \tag{3.7}$$

Then the sufficient condition for asymptotic normality of $U_n^{\mathrm{Ken}}$ is

$$R_n^{(3b_1 b_2 + b_1^2)/b_2} \rho_n^{b_1} \lesssim n^{1/3}, \tag{3.8}$$

and the sufficient condition for asymptotic normality of $U_n^{\mathrm{AP}}$ is

$$R_n^{(3b_1 b_2 + b_1^2)/b_2} \rho_n^{b_1} \lesssim n^{1/3} (\log n)^{-2}. \tag{3.9}$$

(ii) Assume there exist absolute constants $c_1, c_2 > 0$, $b_1 > b_2 > 0$, and $t_0 > 0$, such that for any $n, i, j$ with $1 \leq i \neq j \leq n$ and for any $t \geq t_0$,

$$c_1 \exp(-b_1 t^\lambda) \leq F_j^c(t) \leq c_2 \exp(-b_2 t^\lambda), \tag{3.10}$$

$$c_1 \exp(-b_1 t^\lambda) \leq F_{ji}^c(t) \leq c_2 \exp(-b_2 t^\lambda). \tag{3.11}$$

Then the sufficient condition for asymptotic normality of $U_n^{\mathrm{Ken}}$ is

$$3b_1 R_n^\lambda + b_1 (R_n + K_3 \rho_n + K_4 \rho_n R_n)^\lambda \lesssim \frac{1}{3} \log n, \tag{3.12}$$

and the sufficient condition for asymptotic normality of $U_n^{\mathrm{AP}}$ is

$$3b_1 R_n^\lambda + b_1 (R_n + K_3 \rho_n + K_4 \rho_n R_n)^\lambda \lesssim \frac{1}{3} \log n - 2 \log\log n, \tag{3.13}$$

where

$$K_3 := t_0 + \Big(-\frac{1}{b_2}\log\frac{c_1}{2c_2} + \frac{b_1}{b_2}t_0^\lambda\Big)^{1/\lambda} + \xi(\lambda^{-1})\Big(-\frac{1}{b_2}\log\frac{c_1}{2c_2}\Big)^{1/\lambda},$$
$$K_4 := \xi(\lambda^{-1})\Big(\frac{b_1}{b_2}\Big)^{1/\lambda}, \tag{3.14}$$

and $\xi(p) := \mathbb{1}(p \le 1) + 2^{p-1}\mathbb{1}(p > 1)$.

**Remark 3.6.** It is worth noting that distributions satisfying (3.6) in Theorem 3.5(i) are commonly referred to as"heavy-tailed" distributions, whereas distributions satisfying (3.10) in Theorem 3.5(ii) are considered to be "light-tailed" (Mikosch, 1999; Resnick, 2007).

We compare Condition (3.8) in (i) and Condition (3.12) in (ii) for $U_n^{\text{Ken}}$. Assume $\sigma_i = 1$ for all $i \in [n]$. In this case, we have $\rho_n = 1$, and $R_n = \max_{1\le i\neq j\le n}|\mu_i - \mu_j|$ becomes the spread of the means. Equation (3.8) becomes

$$R_n \lesssim n^{\frac{b_2}{3(3b_1b_2+b_1^2)}}. \tag{3.15}$$

Equation (3.12) becomes

$$3b_1R_n^\lambda + b_1(R_n + K_3 + K_4R_n)^\lambda \lesssim \frac{1}{3}\log n. \tag{3.16}$$

Lemma A.14 yields $(R_n + K_3 + K_4R_n)^\lambda \le \xi(\lambda)(1+K_4)^\lambda R_n^\lambda + \xi(\lambda)K_3^\lambda$. So for (3.16) to hold, it suffices to have $\xi(\lambda)(1+K_4)^\lambda R_n^\lambda + 3b_1R_n^\lambda \lesssim (\log n)/3$. Rearranging terms, we obtain a sufficient condition for (3.16) to hold:

$$R_n \lesssim \Big[\frac{\log n}{3b_1\{3+\xi(\lambda)(1+K_4)^\lambda\}}\Big]^{1/\lambda}. \tag{3.17}$$

For heavy-tailed distributions in (i), (3.15) implies that the spread of means should not grow faster than a polynomial of $n$. For light-tailed distributions in (ii), (3.17) implies that the spread of means should not grow faster than the logarithm of $n$ (up to some constant scaling factor). Of note, under both tail conditions, $R_n$ is allowed to increase to infinity at proper rates.

**Example 3.1.** A special distribution satisfying the conditions in Theorem 3.5(ii) is the Gaussian. Again, consider $U_n^{\text{Ken}}$ and assume $\sigma_i = 1$ for all $i \in [n]$. Note in this case $F_j^c(\cdot)$ is the survival function for Gaussian with variance 1, whereas $F_{ji}^c(\cdot)$ is for Gaussian with variance 2. Let $\lambda = 2$, $b_1 = 1/2+\epsilon$, $b_2 = 1/4-\epsilon$ for arbitrarily small $\epsilon > 0$, and $c_1, c_2, t_0$ be properly chosen constants (whose value does not affect the rate in (3.17)). Equations (3.6) and (3.7) are satisfied due to Lemma A.16. It then follows from (3.17) that

$$R_n \lesssim \Big(\frac{2\log n}{27+12\sqrt{2}}\Big)^{1/2}$$

is sufficient for $U_n^{\text{Ken}}$ to be asymptotically normal.

**Remark 3.7.** We comment on a modified version of Theorem 3.5(i), with a condition alternative to (3.6) (A similar modification applies to Theorem 3.5(ii)). In detail, define $F_j(t) = P\{(X_j - \mu_j)/\sigma_j \le$

$t\}$ to be the standardized cumulative distribution function that is complement to the survival function $F_j^c(t)$. The conclusion in Theorem 3.5(i) still holds if we replace the condition (3.6) by

$$c_1 t^{-b_1} \le F_j(-t) \le c_2 t^{-b_2}. \tag{3.18}$$

For comparison, (3.6) regulates the upper-tail behavior of $X_j$, whereas (3.18) regulates the lower-tail of $X_j$. Technically speaking, the proof of Theorem 3.5(i) examines Condition (ii) in Theorem 3.5, whereas the alternative version examines Condition (i) in Theorem 3.5. Note that (3.7) is required in both versions, which regulates both the upper- and lower-tail behaviors of $X_j - X_i$.

The following three corollaries give asymptotic results for bootstrapping $U_n^{\rm Ken}$ and $U_n^{\rm AP}$. The first of them states that bootstrapping the main term is very insensitive to data non-i.i.d.-ness. This is as expected by the results in Liu (1988).

**Corollary 3.1** (Bootstrap of main term works for $U_n^{\rm Ken}$ and $U_n^{\rm AP}$)**.** If (3.3) holds, we have that (2.17) and (2.18) hold for $h_{1,i}^{\rm Ken}$. If (3.4) holds, we have that (2.17) and (2.18) hold for $h_{1,i}^{\rm AP}$.

As has been shown in Section 2, bootstrapping the whole U-statistic requires much stronger assumptions for guaranteeing its consistency. The following two corollaries provide sufficient conditions for bootstrap inference validity of the two considered U-statistics.

**Corollary 3.2** (Sufficient condition for consistent bootstrap variance estimation of $U_n^{\rm Ken}$)**.** Assume (3.3) holds. Assume there exist $\theta > 0$ and an absolute constant $C > 0$ such that for all $(i,j) \in I_n^2$,

$$|P(X_i > X_j) - \theta| \le C n^{-1/6}. \tag{3.19}$$

In addition, assume there exist $\eta^2 > 0$ and an absolute constant $C > 0$ such that for all $i \in [n]$ and all $1 \le j,k \le n$ such that $j \ne i$ and $k \ne i$,

$$|E\{P(X_j > X_i \mid X_i) P(X_k > X_i \mid X_i)\} - \eta^2| \le C n^{-1/3}. \tag{3.20}$$

Assume $\eta^2 \ne \theta^2$. Then we have

$$|{\rm Var}^*(\sigma_n^{-1} U_n^{\rm Ken *}) - {\rm Var}(\sigma_n^{-1} U_n^{\rm Ken})| \xrightarrow{P} 0.$$

**Corollary 3.3** (Sufficient condition for consistent bootstrap variance estimation of $U_n^{\rm AP}$)**.** Assume (3.4) holds. Assume there exist $\theta > 0$ and an absolute constant $C > 0$ such that for all $(i,j) \in I_n^2$,

$$|P(X_i > X_j) - \theta| \le C n^{-1/6} \log n. \tag{3.21}$$

In addition, assume there exist $\eta^2 > 0$ and an absolute constant $C > 0$ such that for all $1 \le i \le n$ and all $1 \le j,k \le n$ such that $j \ne i$ and $k \ne i$,

$$|E\{P(X_j > X_i \mid X_i) P(X_k > X_i \mid X_i)\} - \eta^2| \le C n^{-1/3} (\log n)^2. \tag{3.22}$$

Assume $\eta^2 \ne \theta^2$. Then we have

$$|{\rm Var}^*(\sigma_n^{-1} U_n^{\rm AP *}) - {\rm Var}(\sigma_n^{-1} U_n^{\rm AP})| \xrightarrow{P} 0.$$

In the proof of Corollaries 3.2 and 3.3, for verifying (2.22), we exploit the weak law of large numbers for independent but not identically distributed variables. For verifying (2.23), we break

Table 1: Testing for normality of $U_n^{\mathrm{Ken}}$ and $U_n^{\mathrm{AP}}$ with $n=100$ under different values of $R_n := \max|\theta_i - \theta_j|$. This table gives the $p$-values under three normality tests: CvM, L, and SF. CvM stands for Cramer-von Mises test. L stands for Lilliefors test. SF stands for Shapiro-Francia test. $p$-values of the three tests are calculated using R package "Rnortest". For each $R_n$, $\{\theta_i : 1 \le i \le n\}$ are equally spaced between $R_n$ and 0, and we simulate $X_i \sim N(\theta_i, 1)$.

| $R_n$ | | 0 | 10 | 30 | 50 |
|---|---|---|---|---|---|
| $U_n^{\mathrm{Ken}}$ | CvM | 0.17 | 0.28 | 0.002 | $<0.001$ |
| | L | 0.15 | 0.10 | $<0.001$ | $<0.001$ |
| | SF | 0.60 | 0.025 | 0.003 | $<0.001$ |
| $U_n^{\mathrm{AP}}$ | CvM | 0.25 | 0.14 | $<0.001$ | $<0.001$ |
| | L | 0.23 | 0.32 | $<0.001$ | $<0.001$ |
| | SF | 0.69 | 0.003 | $<0.001$ | $<0.001$ |

the left-hand side into the sum of an unweighted U-statistic and a negligible term, and apply the law of large numbers for unweighted U-statistics. The detailed proof is very lengthy, and is relegated to Section 4.

**Remark 3.8.** The condition $\eta^2 \neq \theta^2$ in Corollaries 3.2 and 3.3 is mild. Under the i.i.d. case, it essentially requires that the $X_i$'s are not degenerate random variables. To see this, let $\theta := P(X_1 > X_2)$ and $\eta^2 := E\{P(X_1 > X_2 \mid X_1)^2\}$. Since the $X_i$'s are i.i.d., it follows that

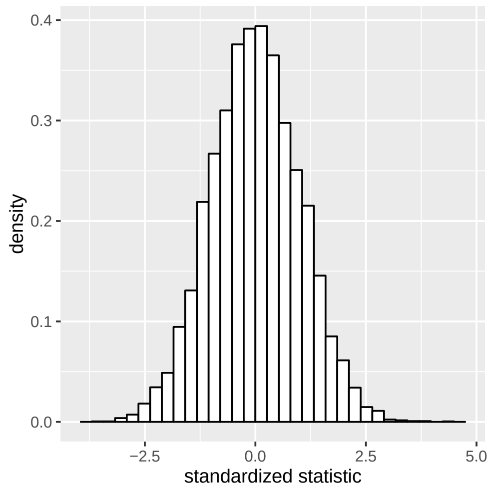
$$|P(X_i > X_j) - \theta| = 0 \quad \text{and} \quad |E\{P(X_j > X_i \mid X_i) P(X_k > X_i \mid X_i)\} - \eta^2| = 0.$$

Jensen's inequality implies that $\eta^2 \geq \theta^2$, with equality only if $X_i$ is a degenerate random variable.

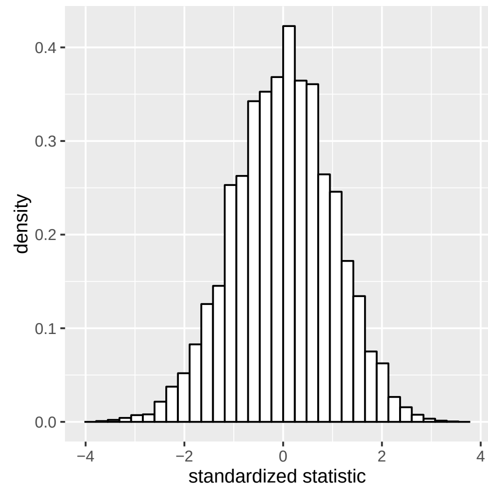
### 3.2 Numerical experiments

In this section, we evaluate the developed theory and examine the finite sample behavior of Kendall's tau and AP correlation via synthetic data analysis. Both central limit theorem and bootstrap inference validity are checked under different data heterogeneity degree.

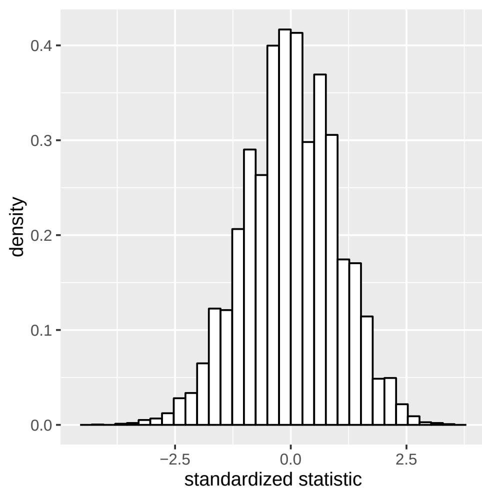
First, we examine the validity of central limit theorem for Kendall's tau and AP correlation. For this, each time, we generate the data sequence $X_1, \ldots, X_n$ with $X_i \sim N(\theta_i, 1)$ for $i \in [n]$. Here the sample size $n$ is picked to be 100, and the means $\{\theta_i, i \in [n]\}$ are assigned equally spaced between $R_n$ and 0, with $R_n = \max|\theta_i - \theta_j|$ representing the heterogeneity degree, taking values from 0 to 50. We repeat the simulation for 5,000 times, and use three goodness-of-fit tests to examine the normality of the considered statistics: Cramer-von Mises test (CvM), Lilliefors test (L), and Shapiro-Francia test (SF). All three tests are implemented in the R package "Rnortest", and we refer the readers to Thode (2002) for detailed descriptions on these tests.



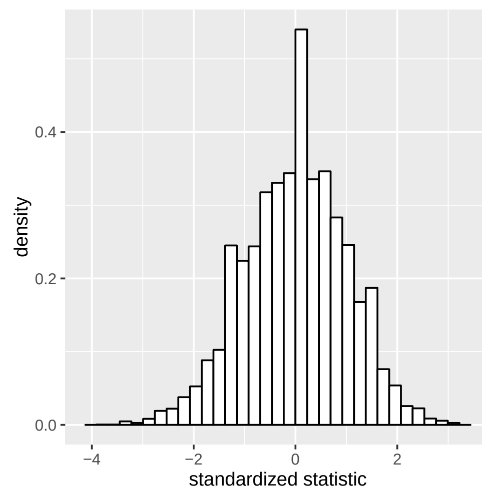

(a) $R_n = 0$




(b) $R_n = 10$




(c) $R_n = 30$

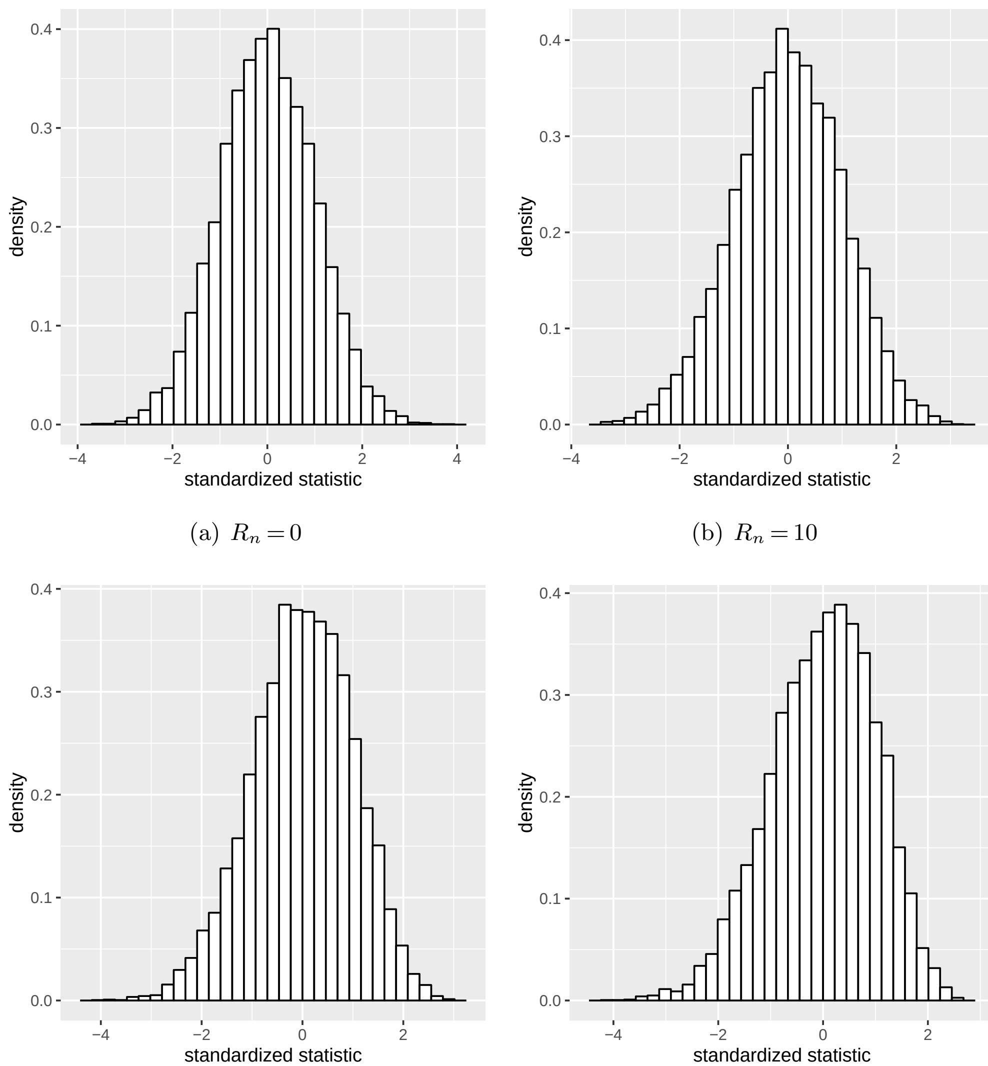



(d) $R_n = 50$

Figure 1: Illustration of the distribution of standardized $U_n^{\text{Ken}}$ with $n = 100$ under different values of $R_n := \max|\theta_i - \theta_j|$. For each $R_n$, $\{\theta_i : 1 \le i \le n\}$ are equally spaced between $R_n$ and 0, and we simulate $X_i \sim N(\theta_i, 1)$. Each histogram is based on 5,000 simulations.

Table 1 illustrates the $p$-values of three tests for normality. For both $U_n^{\text{Ken}}$ and $U_n^{\text{AP}}$, normality is plausible for $R_n$ as large as 10, where two of the three tests fail to reject at significance level 0.1. Figures 1 and 2 give the histograms of simulated statistics (standardized) under the above settings, which also support the results in Table 1. More simulations show that, for fixed $R_n$ as $n$ becomes larger, the two considered statistics are closer to the normal.

In the second simulation study, we examine the bootstrap variance estimation consistency of the following three approaches: (i) bootstrapping the main term of the U-statistic (as in Theorem 2.4);




(a) $R_n = 0$

(b) $R_n = 10$

(c) $R_n = 30$

(d) $R_n = 50$

Figure 2: Illustration of the distribution of standardized $U_n^{\mathrm{AP}}$ with $n = 100$ under different values of $R_n := \max|\theta_i - \theta_j|$. For each $R_n$, $\{\theta_i : 1 \leq i \leq n\}$ are equally spaced between $R_n$ and 0, and we simulate $X_i \sim N(\theta_i, 1)$. Each histogram is based on 5,000 simulations.

Table 2: Three bootstrap variance estimates for $U_n^{\text{Ken}}$ and $U_n^{\text{AP}}$ with $n = 1{,}000$ under different values of $R_n := \max|\theta_i - \theta_j|$. The three considered estimates are: (i) variance estimated by bootstrapping the main term $\{h_{1,i}(X_i)\}$, as in Theorem 2.4; (ii) variance estimated by bootstrapping the original U-statistic, as in Theorem 2.6; (iii) variance estimated by the "moving-block" bootstrap, as in Theorem 2.10 by picking $h_n = n/b$ and $b = 200$. The number listed is the averaged bootstrap variance estimates, scaled by multiplying by $\sigma_n^{-2}$, the inverse of the variance of the U-statistic. For each $R_n$, $\{\theta_i : 1 \le i \le n\}$ are equally spaced between $R_n$ and 0, and we simulate $X_i \sim N(\theta_i, 1)$. Each value is based on 5,000 simulations, and the number of bootstrap replicates within each simulation is 1,000 in (i) and (ii), 100 for each block in (iii).

| | $R_n$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|
| $U_n^{\text{Ken}}$ | main term | 1.029 | 1.032 | 1.031 | 1.030 |
| | original U-stat | 1.032 | 1.155 | 1.557 | 2.331 |
| | moving-block | 1.045 | 1.171 | 1.580 | 2.366 |
| $U_n^{\text{AP}}$ | main term | 1.033 | 1.031 | 1.026 | 1.019 |
| | original U-stat | 1.043 | 1.116 | 1.328 | 1.653 |
| | moving-block | 1.093 | 1.169 | 1.392 | 1.733 |

(ii) bootstrapping the original U-statistic (as in Theorem 2.6); (iii) the new resampling strategy, termed as "moving-block" bootstrap (as in Theorem 2.10 by picking $h_n = n/b$ and $b = 200$). For this, we simulate $X_i \sim N(\theta_i, 1)$ for $1 \le i \le n$. We set the sample size $n$ to be 1,000. The means $\{\theta_i, i \in [n]\}$ are equally spaced between $R_n$ and 0, and the degree of heterogeneity $R_n$ is set to be 0, 1, 2, and 3. We set the number of bootstrap replicates within each simulation to be 1,000 in bootstrap approaches (i) and (ii), 100 for each block in bootstrap approach (iii). We repeat the simulation for 5,000 times.

Table 2 shows the averaged bootstrap variance estimates using the three approaches. The value is scaled by multiplying $\sigma_n^{-2}$, inverse of the variance of the U-statistic. Consistent bootstrap variance estimates would concentrate around 1.

The observations are three-fold. First, we observe that bootstrapping the main term gives consistent variance estimate for all the $R_n$ we considered. This is as expected due to Corollary 3.1 and the asymptotic normality of $U_n^{\text{Ken}}$ and $U_n^{\text{AP}}$. Secondly, bootstrapping the original U-statistic gives consistent variance estimate when $R_n = 0$, i.e., when the $X_i$'s are i.i.d.. It becomes more conservative as the data sequence becomes more heterogeneous. Similar phenomena occur to the moving-block bootstrap. Lastly, by comparing the results derived from the two simulation studies, it is easy to observe that the central limit theorem for our considered statistics holds under much weaker homogeneity conditions than the resampling procedures. This is as expected.

## 4 Proofs

This section contains the proof of main results. More proofs and technical lemmas are provided in the supplementary appendix.

### 4.1 Proof of Theorem 2.1

*Proof.* By Lemma 2.2, we have

$$\left\{\frac{\mathrm{Var}(U_n)}{V(n)}\right\}^{1/2}\frac{U_n-E(U_n)}{\mathrm{Var}(U_n)^{1/2}}=\frac{n^{-1}\sum_{i=1}^n h_{1,i}(X_i)}{V(n)^{1/2}}+\frac{U_n(a,h_2)}{V(n)^{1/2}}. \tag{4.1}$$

For proving Theorem 2.1, by Slutsky's theorem it suffices to establish the following results:

$$V(n)^{-1/2}n^{-1}\sum_{i=1}^n h_{1,i}(X_i)\xrightarrow{d}N(0,1), \tag{4.2}$$

$$V(n)^{-1/2}U_n(a,h_2)\xrightarrow{P}0, \tag{4.3}$$

and

$$\mathrm{Var}(U_n)/V(n)\to 1. \tag{4.4}$$

First we show (4.2) using Lyapunov's Central Limit Theorem (Lemma A.9). The following lemma gives bound on $\sum_{i=1}^n E|h_{1,i}(X_i)|^3$.

**Lemma 4.1.** For $A_{3,1}(n)$ defined in (2.6) and $M(n)$ defined in (2.7), we have

$$\sum_{i=1}^n E|h_{1,i}(X_i)|^3\le CnA_{3,1}(n)M(n)^{3/4},$$

where $C$ is some absolute constant.

By Lemma 4.1 and the fact that $E\{h_{1,i}(X_i)\}=0$, we deduce

$$\sum_{i=1}^n E|h_{1,i}(X_i)-E\{h_{1,i}(X_i)\}|^3\le CnA_{3,1}(n)M(n)^{3/4}. \tag{4.5}$$

Since $V(n):=n^{-2}\sum_{i=1}^n\mathrm{Var}\{h_{1,i}(X_i)\}$, it follows from (4.5) and (2.9) that

$$\frac{\sum_{i=1}^n E|h_{1,i}(X_i)-E\{h_{1,i}(X_i)\}|^3}{\left[\sum_{i=1}^n E|h_{1,i}(X_i)-E\{h_{1,i}(X_i)\}|^2\right]^{3/2}}\le\frac{CnA_{3,1}(n)M(n)^{3/4}}{n^3V(n)^{3/2}}\to 0. \tag{4.6}$$

Equation (4.6) and Lemma A.9 with $\delta=1$ yield (4.2).

Next we show (4.3). To simplify notation, let $\boldsymbol{i}$ denote the index vector $(i_1,\dots,i_m)$ and $X_{\boldsymbol{i}}$ denote $(X_{i_1},\dots,X_{i_m})$. Consider two index vectors $\boldsymbol{i},\boldsymbol{j}$ from $I_n^m$. If $\boldsymbol{i}\cap\boldsymbol{j}=\emptyset$, by independence of the $X_i$'s we have $\mathrm{Cov}\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{i}}),h_{2;\boldsymbol{j}}(X_{\boldsymbol{j}})\}=0$. If $\boldsymbol{i}\cap\boldsymbol{j}=i_p=j_q$ for some $p,q\in[n]$ (i.e., the two vectors only share one common index), Lemma A.7 and (2.15) imply that

$$\mathrm{Cov}\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{i}}),h_{2;\boldsymbol{j}}(X_{\boldsymbol{j}})\}=\mathrm{Cov}[E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{i}})\mid X_{i_p}\},E\{h_{2;\boldsymbol{j}}(X_{\boldsymbol{j}})\mid X_{j_q}\}]=0.$$

Therefore, we have

$$\operatorname{Var}\{U_n(a,h_2)\} = \Big\{\frac{(n-m)!}{n!}\Big\}^2 \sum_{\boldsymbol{i},\boldsymbol{j}\in(I_n^m)^{\otimes 2}_{\geq 2}} a(\boldsymbol{i})a(\boldsymbol{j})\operatorname{Cov}\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{i}}),h_{2;\boldsymbol{j}}(X_{\boldsymbol{j}})\}. \tag{4.7}$$

By Lemma A.6(i) and Cauchy-Schwarz inequality, the right-hand side of (4.7) is bounded by $Cn^{-2}A_{2,2}(n)M(n)^{1/2}$ for some absolute constant $C$, where $A_{2,2}(n)$ is defined in (2.6). This combined with (2.8) yields that

$$V(n)^{-1}\operatorname{Var}\{U_n(a,h_2)\} \leq CV(n)^{-1}n^{-2}A_{2,2}(n)M(n) \to 0. \tag{4.8}$$

Equation (4.3) follows from (4.8) and Lemma A.8.

Lastly, we establish (4.4). Taking variance on both sides of (4.1) gives

$$\frac{\operatorname{Var}(U_n)}{V(n)} = 1 + \frac{\operatorname{Var}\{U_n(a,h_2)\}}{V(n)} + \operatorname{Cov}\Big\{\frac{\sum_{i=1}^n h_{1,i}(X_i)}{nV(n)^{1/2}}, \frac{U_n(a,h_2)}{V(n)^{1/2}}\Big\}. \tag{4.9}$$

By Cauchy-Schwarz inequality and (4.8), we have

$$\Big|\operatorname{Cov}\Big\{\frac{\sum_{i=1}^n h_{1,i}(X_i)}{nV(n)^{1/2}}, \frac{U_n(a,h_2)}{V(n)^{1/2}}\Big\}\Big| \leq \operatorname{Var}\Big\{\frac{\sum_{i=1}^n h_{1,i}(X_i)}{nV(n)^{1/2}}\Big\}^{1/2} \frac{\operatorname{Var}\{U_n(a,h_2)\}^{1/2}}{V(n)^{1/2}} \to 0. \tag{4.10}$$

Equations (4.8), (4.9), and (4.10) imply that

$$\operatorname{Var}(U_n)/V(n) \to 1.$$

This completes the proof. □

## 4.2 Proof of Theorem 2.6

*Proof.* By the definition of $\sigma_n^2$, we have $\operatorname{Var}(\sigma_n^{-1}U_n) = 1$. For proving Theorem 2.6 it suffices to show that

$$\operatorname{Var}^*(\sigma_n^{-1}U_n^*) \xrightarrow{P} 1. \tag{4.11}$$

In Lemma 2.2, replacing $X_i$ by $X_i^*$ yields

$$U_n^* - E(U_n) = \frac{1}{n}\sum_{i=1}^n h_{1,i}(X_i^*) + U_n^*(a,h_2), \tag{4.12}$$

where

$$U_n^*(a,h_2) := \frac{(n-m)!}{n!}\sum_{I_n^m} a(i_1,\ldots,i_m)h_{2;i_1,\ldots,i_m}(X_{i_1}^*,\ldots,X_{i_m}^*). \tag{4.13}$$

Multiplying $\sigma_n^{-1}$ and then taking $\operatorname{Var}^*$ on both sides of (4.12) yields

$$\operatorname{Var}^*(\sigma_n^{-1}U_n^*) = \operatorname{Var}^*\Big\{\sum_{i=1}^n \frac{h_{1,i}(X_i^*)}{n\sigma_n}\Big\} + \operatorname{Var}^*\Big\{\frac{U_n^*(a,h_2)}{\sigma_n}\Big\} + \operatorname{Cov}^*\Big\{\sum_{i=1}^n \frac{h_{1,i}(X_i^*)}{n\sigma_n}, \frac{U_n^*(a,h_2)}{\sigma_n}\Big\}, \tag{4.14}$$

where $\mathrm{Cov}^*(\cdot)$ denotes the covariance operator on the empirical measure. By (4.14) and Slutsky's theorem, for proving (4.11) it suffices to show the following:

$$\mathrm{Var}^*\Big\{\sum_{i=1}^n \frac{h_{1,i}(X_i^*)}{n\sigma_n}\Big\} \xrightarrow{P} 1, \tag{4.15}$$

$$\mathrm{Var}^*\Big\{\frac{U_n^*(a,h_2)}{\sigma_n}\Big\} \xrightarrow{P} 0, \tag{4.16}$$

and

$$\mathrm{Cov}^*\Big\{\sum_{i=1}^n \frac{h_{1,i}(X_i^*)}{n\sigma_n}, \frac{U_n^*(a,h_2)}{\sigma_n}\Big\} \xrightarrow{P} 0, \tag{4.17}$$

First we prove (4.15). Since conditional on $X_1,\ldots,X_n$ the $X_i^*$'s are i.i.d. draws from the empirical distribution of $X_1,\ldots,X_n$, we have

$$E^*\Big[\Big\{\sum_{i=1}^n \frac{h_{1,i}(X_i^*)}{n\sigma_n}\Big\}^2\Big] = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\Big\{\frac{h_{1,i}(X_j)}{n\sigma_n}\Big\}^2 + \frac{1}{n^2}\sum_{i_1\neq i_2}\sum_{j_1=1}^n\sum_{j_2=1}^n \frac{h_{1,i_1}(X_{j_1})}{n\sigma_n}\frac{h_{1,i_2}(X_{j_2})}{n\sigma_n}, \tag{4.18}$$

and

$$\begin{aligned}\Big[E^*\Big\{\sum_{i=1}^n \frac{h_{1,i}(X_i^*)}{n\sigma_n}\Big\}\Big]^2 &= \frac{1}{n^2}\Big\{\sum_{i_1=1}^n\sum_{j_1=1}^n \frac{h_{1,i_1}(X_{j_1})}{n\sigma_n}\Big\}\Big\{\sum_{i_2=1}^n\sum_{j_2=1}^n \frac{h_{1,i_2}(X_{j_2})}{n\sigma_n}\Big\}\\ &= \frac{1}{n^2}\sum_{i=1}^n\Big\{\sum_{j_1=1}^n \frac{h_{1,i}(X_{j_1})}{n\sigma_n}\Big\}\Big\{\sum_{j_2=1}^n \frac{h_{1,i}(X_{j_2})}{n\sigma_n}\Big\} + \frac{1}{n^2}\sum_{i_1\neq i_2}\sum_{j_1=1}^n\sum_{j_2=1}^n \frac{h_{1,i_1}(X_{j_1})}{n\sigma_n}\frac{h_{1,i_2}(X_{j_2})}{n\sigma_n}.\end{aligned} \tag{4.19}$$

Equations (4.18) and (4.19) yield

$$\begin{aligned}\mathrm{Var}^*\Big\{\sum_{i=1}^n \frac{h_{1,i}(X_i^*)}{n\sigma_n}\Big\} &= E^*\Big[\Big\{\sum_{i=1}^n \frac{h_{1,i}(X_i^*)}{n\sigma_n}\Big\}^2\Big] - \Big[E^*\Big\{\sum_{i=1}^n \frac{h_{1,i}(X_i^*)}{n\sigma_n}\Big\}\Big]^2\\ &= \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\Big\{\frac{h_{1,i}(X_j)}{n\sigma_n}\Big\}^2 - \frac{1}{n^2}\sum_{i=1}^n\Big\{\sum_{j=1}^n \frac{h_{1,i}(X_j)}{n\sigma_n}\Big\}^2.\end{aligned} \tag{4.20}$$

Equation (4.15) follows from (4.20), (2.22), (2.23), and Slutsky's theorem.

The following lemma establishes (4.16).

**Lemma 4.2.** Under conditions of Theorem 2.6, we have $\mathrm{Var}^*\{U_n^*(a,h_2)/\sigma_n\} \xrightarrow{P} 0$, where $U_n^*(a,h_2)$ is defined in (4.13).

Equation (4.17) follows from (4.15), (4.16), and Cauchy-Schwarz inequality. This completes the proof. $\square$

## 4.3 Proof of Theorem 3.3

*Proof.* Here we present the proof for $U_n^{\mathrm{AP}}$. The proof for $U_n^{\mathrm{Ken}}$ is similar and can be found in the appendix.

By Theorem 2.1, for proving asymptotic normality of $U_n^{\mathrm{AP}}$, it suffices to show that (2.7), (2.8), and (2.9) hold under the assumption of Theorem 3.1. Equation (2.7) holds trivially with $M(n)=1$ due to boundedness of the kernel function $h(\cdot)$. In the following, we establish (2.8) and (2.9) by calculating the orders of $A_{2,2}(n)$, $A_{3,1}(n)$, and $V(n)$.

First we derive upper bound on $A_{2,2}(n)$ and $A_{3,1}(n)$. We will repeatedly use Lemma A.13 to bound the partial sum of harmonic series. By the definition of $A_{2,2}(n)$ in (2.6), we have

$$A_{2,2}(n):=\frac{1}{n^2}\sum_{(I_n^2)_{\geq 2}^{\otimes 2}}|a(i_1,j_1)a(i_2,j_2)|=\frac{1}{n^2}\sum_{(i,j)\in I_n^2}\big|a(i,j)^2+a(i,j)a(j,i)\big|. \tag{4.21}$$

Since $a(i,j)=n(i-1)^{-1}\mathbb{1}(j<i)$, we have $a(i,j)a(j,i)=0$ and $a(i,i)=0$. It then follows from (4.21) that

$$A_{2,2}(n)=\frac{1}{n^2}\sum_{i=2}^{n}\sum_{j=1}^{i-1}(\frac{n}{i-1})^2=\sum_{i=2}^{n}\frac{1}{i-1}\leq 1+\log(n-1). \tag{4.22}$$

By the definition of $A_{3,1}(n)$ in (2.6), we have

$$\begin{aligned}A_{3,1}(n)=\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j_1,j_2,j_3=1}^{n}\Big\{&|a(i,j_1)a(i,j_2)a(i,j_3)|+3|a(i,j_1)a(i,j_2)a(j_3,i)|\\&+3|a(i,j_1)a(j_2,i)a(j_3,i)|+|a(j_1,i)a(j_2,i)a(j_3,i)|\Big\}.\end{aligned} \tag{4.23}$$

The term $|a(i,j_1)a(i,j_2)a(i,j_3)|$ is nonzero only if $j_1,j_2,j_3<i$, so the corresponding summation in (4.23) equals

$$\frac{1}{n^4}\sum_{i=2}^{n}\sum_{j_1,j_2,j_3=1}^{i-1}\frac{n}{i-1}\cdot\frac{n}{i-1}\cdot\frac{n}{i-1}\leq\frac{1}{n}\sum_{i=1}^{n-1}(i-1)^3(\frac{1}{i-1})^3=\frac{n-1}{n}. \tag{4.24}$$

The term $|a(i,j_1)a(i,j_2)a(j_3,i)|$ is nonzero only if $j_1,j_2<i<j_3$, so the corresponding summation in (4.23) equals

$$\frac{3}{n^4}\sum_{i=2}^{n-1}\sum_{j_1,j_2=1}^{i-1}\sum_{j_3=i+1}^{n}(\frac{n}{i-1})^2(\frac{n}{j_3-1})=\frac{3}{n}\sum_{i=2}^{n-1}\sum_{j_3=i+1}^{n}\frac{1}{j_3-1}\leq\frac{3}{n}\sum_{i=2}^{n-1}\log\frac{n-1}{i-1}\leq 3\log n. \tag{4.25}$$

The term $|a(i,j_1)a(j_2,i)a(j_3,i)|$ is nonzero only if $j_1<i<j_2,j_3$, so the corresponding summation in (4.23) equals

$$\frac{3}{n^4}\sum_{i=2}^{n-1}\sum_{j_1=1}^{i-1}\sum_{j_2,j_3=i+1}^{n}(\frac{n}{i-1})(\frac{n}{j_2-1})(\frac{n}{j_3-1})\leq\frac{3}{n}\sum_{i=2}^{n-1}\left(\log\frac{n-1}{i-1}\right)^2\leq 3(\log n)^2. \tag{4.26}$$

The term $|a(j_1,i)a(j_2,i)a(j_3,i)|$ is nonzero only if $j_1,j_2,j_3>i$, so the corresponding summation in (4.23) equals

$$\frac{1}{n^4}\sum_{i=1}^{n-1}\sum_{j_1,j_2,j_3=i+1}^{n}\frac{n}{j_1-1}\cdot\frac{n}{j_2-1}\cdot\frac{n}{j_3-1}\leq\frac{1}{n}\sum_{i=1}^{n-1}\left(\log\frac{n-1}{i-1}\right)^3\leq(\log n)^3. \tag{4.27}$$

By (4.24)-(4.27), it follows from (4.23) that

$$A_{3,1}(n)\leq C(\log n)^3. \tag{4.28}$$

Next we establish lower bound on $V(n):=n^{-2}\sum_{i=1}^n \mathrm{Var}\{h_{1,i}^{\mathrm{AP}}(X_i)\}$. The following lemma gives lower bound on $|h_{1,i}^{\mathrm{AP}}(X_i)|$.

**Lemma 4.3.** Consider a fixed $i$ with $2\le i\le n$. If $\delta_n/2\ge \log\{(n-1)/(i-1)\}$, either Condition (i) or Condition (ii) in Theorem 3.3 implies

$$P\{|h_{1,i}^{\mathrm{AP}}(X_i)|\ge \delta_n/2\}\ge p_n. \tag{4.29}$$

If $\delta_n\log(n/i)\ge 2$, either Condition (i) or Condition (ii) in Theorem 3.3 implies

$$P\{|h_{1,i}^{\mathrm{AP}}(X_i)|\ge 1\}\ge p_n. \tag{4.30}$$

If $i\ge 1+(n-1)\exp(-\delta_n/2)$, we have $\delta_n/2\ge\log\{(n-1)/(i-1)\}$. Lemma 4.3 implies that (4.29) holds. By Chebyshev's inequality we deduce

$$\mathrm{Var}\{h_{1,i}^{\mathrm{AP}}(X_i)\}\ge \frac{1}{4}\delta_n^2 p_n. \tag{4.31}$$

If $2\le i\le n\exp(-2/\delta_n)$, we have $\delta_n\log(n/i)\ge 2$. Lemma 4.3 implies that (4.30) holds. By Chebyshev's inequality we deduce

$$\mathrm{Var}\{h_{1,i}^{\mathrm{AP}}(X_i)\}\ge p_n. \tag{4.32}$$

By (4.31) and (4.32), we have

$$\begin{aligned}
\sum_{i=1}^n \mathrm{Var}\{h_{1,i}^{\mathrm{AP}}(X_i)\} &\ge \sum_{i=2}^{\lfloor n\exp(-\frac{2}{\delta_n})\rfloor} p_n + \sum_{i=\lfloor 1+(n-1)\exp(-\frac{\delta_n}{2})\rfloor+1}^{n} \frac{1}{4}\delta_n^2 p_n \\
&\ge \Big\{n\exp\Big(-\frac{2}{\delta_n}\Big)-2\Big\}p_n+\frac{1}{4}\Big\{n-(n-1)\exp\Big(-\frac{\delta_n}{2}\Big)-1\Big\}\delta_n^2p_n \\
&= n\exp\Big(-\frac{2}{\delta_n}\Big)p_n+\frac{n\delta_n^2p_n}{4}\Big\{1-\exp\Big(-\frac{\delta_n}{2}\Big)\Big\}+\frac{\delta_n^2p_n}{4}\Big\{\exp\Big(-\frac{\delta_n}{2}\Big)-1\Big\}-2p_n. 
\end{aligned} \tag{4.33}$$

By (3.4) we have

$$n\delta_n^2p_n\Big\{1-\exp\Big(-\frac{\delta_n}{2}\Big)\Big\}\asymp n\delta_n^3p_n\gtrsim n^{2/3}(\log n)^2. \tag{4.34}$$

Note that

$$n\exp\Big(-\frac{2}{\delta_n}\Big)p_n\ge 0 \quad\text{and}\quad \frac{\delta_n^2p_n}{4}\Big\{\exp\Big(-\frac{\delta_n}{2}\Big)-1\Big\}-2p_n=O(1). \tag{4.35}$$

Combining (4.33) with (4.34) and (4.35) gives

$$\sum_{i=1}^n \mathrm{Var}\{h_{1,i}^{\mathrm{AP}}(X_i)\}\gtrsim n^{2/3}(\log n)^2.$$

This implies

$$V(n):=\frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\{h_{1,i}^{\mathrm{AP}}(X_i)\}\gtrsim n^{-4/3}(\log n)^2. \tag{4.36}$$

Equations (4.22), (4.28), and (4.36) yield (2.8) and (2.9). The asymptotic normality of $U_n^{\mathrm{AP}}$ then follows from Theorem 2.1. This completes the proof for Part I.

$\square$

## 4.4 Proof of Theorem 3.5

*Proof.* Define

$$f_{ij}(x) := P(X_j > x) - P(X_j > X_i), \tag{4.37}$$

and

$$z_i = z_i(x) := (x - \mu_i)/\sigma_i. \tag{4.38}$$

Using the definitions of $F_j^c$ and $F_{ji}^c$ in (3.5), we have

$$\begin{aligned} f_{ij}(x) &= P\Big\{\frac{X_j - \mu_j}{\sigma_j} > \frac{(x-\mu_i)+(\mu_i-\mu_j)}{\sigma_i}\cdot\frac{\sigma_i}{\sigma_j}\Big\} - P\Big\{\frac{X_j - X_i - (\mu_j-\mu_i)}{(\sigma_i^2+\sigma_j^2)^{1/2}} > \frac{\mu_i-\mu_j}{\sigma_i}\cdot\frac{\sigma_i}{(\sigma_i^2+\sigma_j^2)^{1/2}}\Big\} \\ &= F_j^c\{\rho_{ij}(z_i + r_{ij})\} - F_{ji}^c\{r_{ij}(1+\rho_{ij}^{-2})^{-1/2}\}. \end{aligned} \tag{4.39}$$

For proving Theorem 3.5, it suffices to show the existence of $\delta_n$ and $p_n$ satisfying the conditions in Theorem 3.3. Because the proofs for $U_n^{\mathrm{Ken}}$ and $U_n^{\mathrm{AP}}$ are almost identical, we give detailed proof for $U_n^{\mathrm{Ken}}$ and comment on the proof for $U_n^{\mathrm{AP}}$ at the end. We divide the proof for $U_n^{\mathrm{Ken}}$ into two parts. In Part I we construct such $\delta_n$ and $p_n$ under conditions (3.6), (3.7), and (3.8). In Part II we construct such $\delta_n$ and $p_n$ under conditions (3.10), (3.11), and (3.12).

**Part I:** Assume (3.6), (3.7), and (3.8) hold. The following lemma gives bound on $f_{ij}(x)$.

**Lemma 4.4.** Define

$$K_1 = t_0 + \Big(t_0^{-b_1}\frac{c_1}{2c_2}\Big)^{-1/b_2} \quad \text{and} \quad K_2 = \Big(\frac{c_1}{2c_2}\Big)^{-1/b_2}.$$

Consider a fixed $i \in [n]$. If $x$ satisfies

$$z_i(x) \geq R_n + K_1\rho_n + K_2\rho_n R_n^{b_1/b_2}, \tag{4.40}$$

then for all $j \in [n]\backslash\{i\}$ we have

$$f_{ij}(x) \leq -\min\Big\{\frac{c_1}{2}R_n^{-b_1}, \frac{c_1}{2}t_0^{-b_1}, \frac{1}{2}\Big\}. \tag{4.41}$$

Define $\delta_n := \min\{\frac{c_1}{2}R_n^{-b_1}, \frac{c_1}{2}t_0^{-b_1}, \frac{1}{2}\}.$, $Z_i := (X_i - \mu_i)/\sigma_i$, and

$$p_n := P\{Z_i \geq R_n + K_1\rho_n + K_2\rho_n R_n^{b_1/b_2}\}. \tag{4.42}$$

Lemma 4.4 yields that

$$P\{f_{ij}(x) \leq -\delta_n, \forall j \in [n]\backslash\{i\}\} \geq p_n.$$

Since $\rho_n \geq 1$, by the definition of $K_1$ we have

$$R_n + K_1\rho_n + K_2\rho_n R_n^{b_1/b_2} \geq K_1\rho_n \geq t_0. \tag{4.43}$$

Combining (4.42), (4.43) and (3.6) yields

$$p_n \geq c_1(R_n + K_1\rho_n + K_2\rho_n R_n^{b_1/b_2})^{-b_1}.$$

Thus by dropping constants we obtain

$$\delta_n^3 p_n \gtrsim (R_n+\rho_n+\rho_n R_n^{b_1/b_2})^{-b_1}\min(R_n^{-3b_1},1). \tag{4.44}$$

In the following we show that (4.44) and (3.8) imply

$$\delta_n^3 p_n \gtrsim n^{-1/3}. \tag{4.45}$$

If $\limsup_{n\to\infty} R_n=\infty$, the fact that $\rho_n\geq 1$ and $b_1>b_2>0$ yields

$$(R_n+\rho_n+\rho_n R_n^{b_1/b_2})^{-b_1}\asymp \rho_n^{-b_1}R_n^{-b_1^2/b_2} \tag{4.46}$$

and

$$\min(R_n^{-3b_1},1)\asymp R_n^{-3b_1}. \tag{4.47}$$

Equation (4.44) together with (4.46) and (4.47) gives

$$\delta_n^3 p_n \gtrsim \rho_n^{-b_1}R_n^{-b_1^2/b_2}R_n^{-3b_1}. \tag{4.48}$$

By (4.48) and (3.8), we deduce (4.45). If $\limsup_{n\to\infty} R_n<\infty$, by (4.44) we have

$$\delta_n^3 p_n \gtrsim \rho_n^{-b_1}. \tag{4.49}$$

Equation (3.8) implies

$$\rho_n^{-b_1}\gtrsim n^{-1/3}. \tag{4.50}$$

Combining (4.49) and (4.50) yields (4.45). Therefore, the asymptotic normality of $U_n^{\mathrm{Ken}}$ follows from Theorem 3.3.

This completes the proof of Part I for $U_n^{\mathrm{Ken}}$. For $U_n^{\mathrm{AP}}$ the proof is almost the same, except that (3.8) is replaced by (3.9), and the right-hand side of (4.45) and (4.50) is replaced by $n^{-1/3}(\log n)^2$.

**Part II:** Assume (3.9), (3.10), and (3.11) hold. The following lemma gives bound on $f_{ij}(x)$.

**Lemma 4.5.** For a fixed $i\in[n]$, assume that

$$z_i\geq R_n+K_3\rho_n+K_4\rho_n R_n, \tag{4.51}$$

where $K_3,K_4$ are defind in (3.14). Then for all $j\in[n]\backslash\{i\}$ we have

$$f_{ij}(x)\leq -\min\left\{\frac{c_1}{2}\exp(-b_1R_n^\lambda),\frac{c_1}{2}\exp(-b_1t_0^\lambda),\frac{1}{2}\right\}. \tag{4.52}$$

Define $\delta_n=\min\left\{\frac{c_1}{2}\exp(-b_1R_n^\lambda),\frac{c_1}{2}\exp(-b_1t_0^\lambda),\frac{1}{2}\right\}$, $Z_i=(X_i-\mu_i)/\sigma_i$, and

$$p_n=P\{Z_i\geq R_n+K_3\rho_n+K_4\rho_nR_n\}. \tag{4.53}$$

Lemma 4.5 yields that

$$P\{f_{ij}(x)\leq -\delta_n, \forall j\in[n]\backslash\{i\}\}\geq p_n.$$

Since $\rho_n\geq 1$, by the definition of $K_3$, we have

$$R_n+K_3\rho_n+K_4\rho_nR_n\geq K_3\rho_n\geq t_0. \tag{4.54}$$

Combining (4.53), (4.54), and (3.10) yields

$$p_n \geq c_1 \exp\{-b_1(R_n + K_3\rho_n + K_4\rho_n R_n)^\lambda\}.$$

Thus by dropping constants we obtain

$$\begin{aligned}\delta_n^3 p_n &\gtrsim \exp\{-b_1(R_n + K_3\rho_n + K_4\rho_n R_n)^\lambda\}\min\{\exp(-3b_1 R_n^\lambda), 1\}.\\ &\asymp \min\left[\exp\{-3b_1 R_n^\lambda - b_1(R_n + K_3\rho_n + K_4\rho_n R_n)^\lambda\}, \exp\{-b_1(R_n + K_3\rho_n + K_4\rho_n R_n)^\lambda\}\right] \quad (4.55)\end{aligned}$$

With an argument similar to (4.46)-(4.50), it follows from (4.55) and (3.12) that

$$\delta_n^3 p_n \gtrapprox n^{-1/3}. \tag{4.56}$$

This completes the proof of Part II for $U_n^{\mathrm{Ken}}$. For $U_n^{\mathrm{AP}}$ the proof is almost the same, except that (3.12) is replaced by (3.13), and the right-hand side of (4.56) is replaced by $n^{-1/3}(\log n)^2$. $\square$

## 4.5 Proof of Corollary 3.3

*Proof.* By Theorem 2.6, for proving Corollary 3.3, it suffices to show that (2.21), (2.22), (2.23), and (2.24) hold. For $U_n^{\mathrm{AP}}$ we have $|h(x,y)| \leq 1$ for any $x, y$. This implies (2.21).

Now we establish (2.24). For $U_n^{\mathrm{AP}}$, we have $a(i,j) = \mathbb{1}(j<i)n/(i-1)$. It follows that $a(i,j)a(j,i) = 0$ and $a(i,i) = 0$. By the definition in (2.6), we have

$$\begin{aligned}A_{2,1}(n) &= n^{-3} \sum_{(i,j)\in I_n^2} \sum_{k=1, k\neq i}^{n} \{|a(i,j)a(i,k)| + |a(i,j)a(k,i)|\}.\\ &= n^{-3}\Big\{\sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{i-1} \frac{n}{i-1}\cdot\frac{n}{i-1} + \sum_{i=2}^{n-1}\sum_{j=1}^{i-1}\sum_{k=i+1}^{n} \frac{n}{i-1}\cdot\frac{n}{k-1}\Big\}. \end{aligned}\tag{4.57}$$

By algebra, we have

$$\sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{i-1} \frac{n}{i-1}\cdot\frac{n}{i-1} = n^2(n-1), \tag{4.58}$$

and

$$\sum_{i=2}^{n-1}\sum_{j=1}^{i-1}\sum_{k=i+1}^{n} \frac{n}{i-1}\cdot\frac{n}{k-1} = n^2\sum_{k=3}^{n}\sum_{i=2}^{k-1}\frac{1}{k-1} = n^2\sum_{k=3}^{n}\frac{k-2}{k-1} = O(n^3). \tag{4.59}$$

Combining (4.57) with (4.58) and (4.59) yields

$$A_{2,1}(n) = O(1). \tag{4.60}$$

By (2.25) and (3.21) we have

$$M_1(n) \lesssim n^{-1/6}\log n. \tag{4.61}$$

By (2.26) and (3.22) we have

$$M_2(n) \lesssim n^{-1/3}(\log n)^2. \tag{4.62}$$

Equation (3.4) implies (4.36) by Theorem 3.3. Combining (4.36) and (2.10) yields

$$\sigma_n^2 \gtrsim n^{-4/3}(\log n)^2. \tag{4.63}$$

Equation (2.24) follows from (4.60), (4.61), (4.62), and (4.63).

Next we establish (2.22). The following lemma gives useful bounds.

**Lemma 4.6.** Under the assumptions of Corollary 3.3, we have

$$\sum_{i=1}^{n} E\{h_{1,i}^{\mathrm{AP}}(X_j)^2\} = \frac{n^2}{n-1}(\eta^2-\theta^2) + O(n^{5/6}\log n), \tag{4.64}$$

and

$$\sum_{i=1}^{n} E\{h_{1,i}^{\mathrm{AP}}(X_i)^2\} = \frac{n^2}{n-1}(\eta^2-\theta^2) + O(n^{5/6}\log n). \tag{4.65}$$

By (2.10) we have

$$\sigma_n^2 = n^{-2}\sum_{i=1}^{n} E\{h_{1,i}^{\mathrm{AP}}(X_i)^2\}\{1+o(1)\}. \tag{4.66}$$

Using (4.66) and (4.65) we obtain

$$n^2\sigma_n^2 = \{1+o(1)\}\Big\{\frac{n^2}{n-1}(\eta^2-\theta^2) + O(n^{5/6}\log n)\Big\}. \tag{4.67}$$

Note that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E\Big[\Big\{\frac{h_{1,i}^{\mathrm{AP}}(X_j)}{n\sigma_n}\Big\}^2\Big] = \frac{n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} E\{h_{1,i}^{\mathrm{AP}}(X_j)^2\}}{n^2\sigma_n^2}. \tag{4.68}$$

Combining (4.68) with (4.64), (4.67), and the fact that $\eta^2 \neq \theta^2$ yields

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E\Big[\Big\{\frac{h_{1,i}^{\mathrm{AP}}(X_j)}{n\sigma_n}\Big\}^2\Big] = \frac{n^2(n-1)^{-1}(\eta^2-\theta^2) + O(n^{5/6}\log n)}{\{1+o(1)\}\Big\{n^2(n-1)^{-1}(\eta^2-\theta^2) + O(n^{5/6}\log n)\Big\}} \to 1. \tag{4.69}$$

By (3.2) we have $|h_{1,i}^{\mathrm{AP}}(x)| \le 1+\varphi(n-1)-\varphi(i-1)$ for all $x$. This combined with Lemma A.13 yields

$$|h_{1,i}^{\mathrm{AP}}(x)| \le 1+\log\frac{n}{i} \le 1+\log n. \tag{4.70}$$

It then follows from (4.70) that

$$\Big|\sum_{i=1}^{n}\Big\{\frac{h_{1,i}^{\mathrm{AP}}(x)}{n\sigma_n}\Big\}^2\Big| = \Big|\frac{\sum_{i=1}^{n} h_{1,i}^{\mathrm{AP}}(x)^2}{n^2\sigma_n^2}\Big| \le \frac{(1+\log n)^2}{n\sigma_n^2}. \tag{4.71}$$

Equations (4.67) and (4.71) imply that

$$\sum_{j=1}^{n}\mathrm{Var}\Big[\sum_{i=1}^{n}\Big\{\frac{h_{1,i}^{\mathrm{AP}}(X_j)}{n\sigma_n}\Big\}^2\Big] \le \sum_{j=1}^{n}\frac{(1+\log n)^4}{n^2\sigma_n^4} = O\{n(\log n)^4\} = o(n^2).$$

It then follows from Lemma A.12 that

$$\frac{1}{n}\sum_{j=1}^{n}\Big[\sum_{i=1}^{n}\Big\{\frac{h_{1,i}^{\mathrm{AP}}(X_j)}{n\sigma_n}\Big\}^2 - E\Big\{\sum_{i=1}^{n}\Big(\frac{h_{1,i}^{\mathrm{AP}}(X_j)}{n\sigma_n}\Big)^2\Big\}\Big] \xrightarrow{P} 0. \tag{4.72}$$

Equation (2.22) follows from (4.69) and (4.72).

Lastly, we prove (2.23). By algebra we have

$$\frac{1}{n^2}\sum_{i=1}^{n}\Big\{\sum_{j=1}^{n}\frac{h_{1,i}^{\mathrm{AP}}(X_j)}{n\sigma_n}\Big\}^2=\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\Big\{\frac{h_{1,i}^{\mathrm{AP}}(X_j)}{n\sigma_n}\Big\}^2+\frac{1}{n^2}\sum_{j_1\neq j_2}\sum_{i=1}^{n}\frac{h_{1,i}^{\mathrm{AP}}(X_{j_1})h_{1,i}^{\mathrm{AP}}(X_{j_2})}{n^2\sigma_n^2}. \tag{4.73}$$

By (2.22) we have

$$\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\Big\{\frac{h_{1,i}^{\mathrm{AP}}(X_j)}{n\sigma_n}\Big\}^2\xrightarrow{P}0. \tag{4.74}$$

The second term on the right-hand side of (4.73) is $(n-1)/n$ times a U-statistic with symmetric kernel $g(x,y)=n^{-2}\sigma_n^{-2}\sum_{i=1}^{n}h_{1,i}^{\mathrm{AP}}(x)h_{1,i}^{\mathrm{AP}}(y)$. By (4.70) and (3.21) we have

$$E\{h_{1,i}^{\mathrm{AP}}(X_j)\}=\frac{1}{n-1}\sum_{k=1}^{n}\Big\{\frac{n\mathbb{1}(j<i)}{i-1}-\frac{n\mathbb{1}(j>i)}{j-1}\Big\}O(n^{-1/6}\log n).$$

This combined with Lemma A.13 yields

$$E\{h_{1,i}^{\mathrm{AP}}(X_j)\}=O\{n^{-1/6}(\log n)^2\}. \tag{4.75}$$

It follows from (4.75) and (4.67) that

$$E\{g(X_{j_1},X_{j_2})\}=n^{-2}\sigma_n^{-2}\sum_{i=1}^{n}E\{h_{1,i}^{\mathrm{AP}}(X_{j_1})\}E\{h_{1,i}^{\mathrm{AP}}(X_{j_2})\}\to 0. \tag{4.76}$$

By (4.76) and Theorem 1 in Lee (1990, Section 3.7.2), we deduce

$$\frac{1}{n^2}\sum_{j_1\neq j_2}\sum_{i=1}^{n}\frac{h_{1,i}^{\mathrm{AP}}(X_{j_1})h_{1,i}^{\mathrm{AP}}(X_{j_2})}{n^2\sigma_n^2}\xrightarrow{P}0. \tag{4.77}$$

Equation (2.23) follows from (4.73), (4.74), and (4.77).

This completes the proof. □

## References


Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9(6):1196–1217.

Bickel, P. J., Götze, F., and van Zwet, W. R. (1997). Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistica Sinica*, 7:1–31.

Csörgő, M. and Nasari, M. M. (2013). Asymptotics of randomly weighted U-and V-statistics: Application to bootstrap. *Journal of Multivariate Analysis*, 121:176–192.

Dehling, H. and Wendler, M. (2010). Central limit theorem and the bootstrap for U-statistics of strongly mixing data. *Journal of Multivariate Analysis*, 101(1):126–137.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Grams, W. F. and Serfling, R. J. (1973). Convergence rates for U-statistics and related statistics. *The Annals of Statistics*, 1(1):153–160.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325.

Hsing, T. and Wu, W. B. (2004). On weighted U-statistics for stationary processes. *The Annals of Probability*, 32(2):1600–1631.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2):81–93.

Kendall, M. G. and Stuart, A. (1973). *The Advanced Theory of Statistics*, volume 2. Charles Griffin, 3 edition.

Korolyuk, V. S. and Borovskich, Y. V. (2013). *Theory of U-Statistics*, volume 273. Springer.

Lee, J. (1990). *U-Statistics: Theory and Practice*. CRC Press.

Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *The Annals of Statistics*, 16(4):1696–1708.

Liu, R. Y. and Singh, K. (1995). Using i.i.d. bootstrap inference for general non-i.i.d. models. *Journal of Statistical Planning and Inference*, 43(1):67–75.

Major, P. (1994). Asymptotic distributions for weighted U-statistics. *The Annals of Probability*, 21(2):1514–1535.

Mammen, E. (2012). *When Does Bootstrap Work? Asymptotic Results and Simulations*, volume 77. Springer.

Mikosch, T. (1999). Regular variation, subexponentiality and their applications in probability theory. Technical report, Eindhoven University of Technology.

O'Neil, K. A. and Redner, R. A. (1993). Asymptotic distributions of weighted U-statistics of degree 2. *The Annals of Probability*, 21(2):1159–1169.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, 22(4):2031–2050.

Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling*. Springer.

Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer.

Rifi, M. and Utzet, F. (2000). On the asymptotic behavior of weighted U-statistics. *Journal of Theoretical Probability*, 13(1):141–167.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63(324):1379–1389.

Serfling, R. J. (2009). *Approximation Theorems of Mathematical Statistics*, volume 162. John Wiley and Sons.

Shapiro, C. P. and Hubert, L. (1979). Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions. *The Annals of Statistics*, 7(4):788–794.

Thode, H. C. (2002). *Testing for Normality*. CRC Press.

Yilmaz, E., Aslam, J. A., and Robertson, S. (2008). A new rank correlation coefficient for information retrieval. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 587–594. ACM.

Yoshihara, K.-i. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. *Probability Theory and Related Fields*, 35(3):237–252.

Zhou, Z. (2014). Inference of weighted V-statistics for nonstationary time series and its applications. *The Annals of Statistics*, 42(1):87–114.

# Supplement to “Asymptotics for Asymmetric Weighted U-Statistics: Central Limit Theorem and Bootstrap under Data Heterogeneity”

Fang Han[*] and Tianchen Qian[†]

September 22, 2016

## A Appendix

This supplementary appendix contains the rest of technical proofs.

### A.1 Proof of the rest of main results

#### A.1.1 Proof of Lemma 2.2

*Proof.* We have

$$U_n - E(U_n) = \sum_{i=1}^n \{E(U_n \mid X_i) - E(U_n)\} + \Big[U_n - E(U_n) - \sum_{i=1}^n \{E(U_n \mid X_i) - E(U_n)\}\Big].$$

For proving Lemma 2.2, it suffices to show

$$\sum_{i=1}^n \{E(U_n \mid X_i) - E(U_n)\} = \frac{1}{n}\sum_{i=1}^n h_{1,i}(X_i) \tag{A.1}$$

and

$$U_n - E(U_n) - \sum_{i=1}^n \{E(U_n \mid X_i) - E(U_n)\} = \frac{(n-m)!}{n!}\sum_{I_n^m} a(i_1,\ldots,i_m) h_{2;i_1,\ldots,i_m}(X_{i_1},\ldots,X_{i_m}), \tag{A.2}$$

where $h_{1,i}(\cdot)$ and $h_{2;i_1,\ldots,i_m}(\cdot)$ are defined in (2.14) and (2.15), respectively.

First we establish (A.1). We have

$$E(U_n \mid X_i) - E(U_n) = \frac{(n-m)!}{n!}\sum_{I_n^m} a(i_1,\ldots,i_m)\Big[E\{h(X_{i_1},\ldots,X_{i_m}) \mid X_i\} - \theta(i_1,\ldots,i_m)\Big]. \tag{A.3}$$

Consider a fixed $i \in [n]$ and fixed $(i_1,\ldots,i_m) \in I_n^m$. If $i \notin \{i_1,\ldots,i_m\}$,

$$E\{h(X_{i_1},\ldots,X_{i_m}) \mid X_i\} - \theta(i_1,\ldots,i_m) = 0 \ \text{ a.s..}$$

[*]Department of Statistics, University of Washington, Seattle, WA 98195, USA; e-mail: `fanghan@uw.edu`
[†]Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA; e-mail: `tqian2@jhu.edu`

It follows that

$$
\begin{aligned}
&\sum_{I_n^m} a(i_1,\ldots,i_m)\Big[E\{h(X_{i_1},\ldots,X_{i_m}) \mid X_i\} - \theta(i_1,\ldots,i_m)\Big] \\
&= \sum_{I_{n-1}^{m-1}(-i)} a(i,i_1,\ldots,i_{m-1})\Big[E\{h(X_i,X_{i_1},\ldots,X_{i_{m-1}}) \mid X_i\} - \theta(i,i_1,\ldots,i_{m-1})\Big] \\
&\quad + \sum_{I_{n-1}^{m-1}(-i)} a(i_1,i,i_2,\ldots,i_{m-1})\Big[E\{h(X_{i_1},X_i,X_{i_2},\ldots,X_{i_{m-1}}) \mid X_i\} - \theta(i_1,i,i_2,\ldots,i_{m-1})\Big] + \cdots \\
&\quad + \sum_{I_{n-1}^{m-1}(-i)} a(i_1,\ldots,i_{m-1},i)\Big[E\{h(X_{i_1},\ldots,X_{i_{m-1}},X_i) \mid X_i\} - \theta(i_1,\ldots,i_{m-1},i)\Big] \\
&= \sum_{I_{n-1}^{m-1}(-i)} \sum_{l=1}^{m} a^{(l)}(i;i_1,\ldots,i_{m-1})\Big[E\{h^{(l)}(X_i;X_{i_1},\ldots,X_{i_{m-1}}) \mid X_i\} - \theta^{(l)}(i;i_1,\ldots,i_{m-1})\Big]. \qquad \text{(A.4)}
\end{aligned}
$$

By the definition of $h_{1,i}(\cdot)$, (A.4) equals $\{(n-1)!/(n-m)!\}h_{1,i}(X_i)$. Combining this with (A.3) yields (A.1).

Next we establish (A.2). The following lemma shows that $\sum_{i=1}^n\{E(U_n \mid X_i) - E(U_n)\}$ is a U-statistic.

**Lemma A.1.** We have

$$
\begin{aligned}
&\sum_{l=1}^{m}\sum_{i=1}^{n}\sum_{I_{n-1}^{m-1}(-i)} a^{(l)}(i;i_1,\ldots,i_{m-1})\Big[E\{h^{(l)}(X_i;X_{i_1},\ldots,X_{i_{m-1}}) \mid X_i\} - \theta^{(l)}(i;i_1,\ldots,i_{m-1})\Big] \\
&= \sum_{I_n^m} a(i_1,\ldots,i_m)\Big[\sum_{j=1}^{m} E\{h(X_{i_1},\ldots,X_{i_m}) \mid X_{i_j}\} - m\theta(i_1,\ldots,i_m)\Big]. \qquad \text{(A.5)}
\end{aligned}
$$

Using Lemma A.1, it follows from (A.4) that

$$
\sum_{i=1}^{n}\{E(U_n \mid X_i) - E(U_n)\} = \sum_{I_n^m} a(i_1,\ldots,i_m)\Big[\sum_{j=1}^{m} E\{h(X_{i_1},\ldots,X_{i_m}) \mid X_{i_j}\} - m\theta(i_1,\ldots,i_m)\Big].
$$

By the definition of $h_{2;i_1,\ldots,i_m}(\cdot)$, we deduce that (A.2) holds.

Equations (2.14) and (2.15) follow immediately from the definitions in (2.4) and (2.5). This completes the proof. $\square$

### A.1.2 Proof of Theorem 2.4

*Proof.* In Lemma A.11, let $Y_{n,i} = \sigma_n^{-1}h_{1,i}(X_i)$, $g_n$ be the identity function, $t_n = 0$ and $\sigma_n^2 = 1$. By the definition of $\widehat{T}_n$ we have $\widehat{T}_n = n^{-1}\sum_{i=1}^n \sigma_n^{-1}h_{1,i}(X_i)$. Equation (2.17) implies (A.187). (2.18) implies (A.188). Equations (4.2), (2.10) and Slutsky's theorem imply that for any $t \in \mathbb{R}$,

$$
P\left\{\widehat{T}_n - t_n \le t\right\} - \Phi(t) \to 0.
$$

By Lemma A.10 the above convergence is uniform in $t \in \mathbb{R}$. This yields (A.189). Therefore, all conditions in Lemma A.11 hold, which implies

$$\sup_{t\in\mathbb{R}}\left|P^*\left\{\sum_{i=1}^{n}\frac{\{h_{1,i}(X_i)\}^*}{n\sigma_n}-\sum_{i=1}^{n}\frac{h_{1,i}(X_i)}{n\sigma_n}\le t\right\}-P\left\{\sum_{i=1}^{n}\frac{h_{1,i}(X_i)}{n\sigma_n}\le t\right\}\right|\xrightarrow{P}0.$$

This proves (2.19). Equation (2.20) follows immediately from Theorem 2.1. □

### A.1.3 Proof of Corollary 2.1

*Proof.* For proving Corollary 2.1, by Theorem 2.6, it suffices to show (2.22), (2.23), and (2.24) when the $X_i$'s are i.i.d..

First we show (2.22). Equations (2.7), (2.8), and (2.9) imply (2.10) according to Theorem 2.1. By the i.i.d.-ness of the $X_i$'s we have $E\{h_{1,i}(X_j)\}=E\{h_{1,i}(X_i)\}=0$ and $E\{h_{1,i}(X_j)^2\}=E\{h_{1,i}(X_i)^2\}$. It follows from (2.10) that for any $j\in[n]$,

$$E\Big[\sum_{i=1}^{n}\Big\{\frac{h_{1,i}(X_j)}{n\sigma_n}\Big\}^2\Big]=\sum_{i=1}^{n}\frac{E\{h_{1,i}(X_j)^2\}}{n^2\sigma_n^2}=\frac{\sum_{i=1}^{n}\mathrm{Var}\{h_{1,i}(X_i)\}}{n^2\sigma_n^2}\to 1. \tag{A.6}$$

By the weak law of large numbers for i.i.d. random variables, we have

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\Big\{\frac{h_{1,i}(X_j)}{n\sigma_n}\Big\}^2-E\Big[\sum_{i=1}^{n}\Big\{\frac{h_{1,i}(X_j)}{n\sigma_n}\Big\}^2\Big]\xrightarrow{P}0. \tag{A.7}$$

Equations (A.6), (A.7), and Slutsky's theorem yield (2.22).

Next we prove (2.23). By algebra we have

$$\frac{1}{n^2}\sum_{i=1}^{n}\Big\{\sum_{j=1}^{n}\frac{h_{1,i}(X_j)}{n\sigma_n}\Big\}^2=\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\Big\{\frac{h_{1,i}(X_j)}{n\sigma_n}\Big\}^2+\frac{1}{n^2}\sum_{j_1\neq j_2}\sum_{i=1}^{n}\frac{h_{1,i}(X_{j_1})h_{1,i}(X_{j_2})}{n^2\sigma_n^2}. \tag{A.8}$$

Equation (2.22) implies that the first term on the right-hand side of (A.8) converges to 0 in probability. The second term on the right-hand side of (A.8) equals $(n-1)/n$ times a U-statistic with symmetric kernel $g(x,y)=n^{-2}\sigma_n^{-2}\sum_{i=1}^{n}h_{1,i}(x)h_{1,i}(y)$. By the triangle inequality, Jensen's inequality, and the i.i.d.-ness of the $X_i$'s, we deduce

$$E|g(X_1,X_2)|\le\sum_{i=1}^{n}E\Big|\frac{h_{1,i}(X_1)}{n\sigma_n}\Big|\;E\Big|\frac{h_{1,i}(X_2)}{n\sigma_n}\Big|\le\sum_{i=1}^{n}E\Big\{\Big(\frac{h_{1,i}(X_i)}{n\sigma_n}\Big)^2\Big\}=1. \tag{A.9}$$

The i.i.d.-ness of the $X_i$'s and the fact that $E\{h_{1,i}(X_i)\}=0$ yield

$$E\{g(X_1,X_2)\}=n^{-2}\sigma_n^{-2}\sum_{i=1}^{n}E\{h_{1,i}(X_1)\}E\{h_{1,i}(X_2)\}=0. \tag{A.10}$$

By (A.9) and (A.10), it follows from the weak law of large numbers for U-statistics of i.i.d. variables (Serfling, 2009, Theorem 5.4 A) that the second term on the right-hand side of (A.8) converges to 0 in probability. Therefore, by Slutsky's theorem, the left-hand side of (A.8) converges to 0 in probability, which establishes (2.23).

Lastly, we establish (2.24). By the definition of $\theta(\cdot)$ in (2.2), we have $\theta(i_1,\ldots,i_m)-\theta(j_1,\ldots,j_m)=0$ for any $(i_1,\ldots,i_m)$ and $(j_1,\ldots,j_m)$ in $I_n^m$. This implies that $M_1(n)=0$. For any $p,q\in[m]$ and

$\boldsymbol{r},\boldsymbol{s},\boldsymbol{k}\in I_n^m$ such that $\boldsymbol{r}\cap\boldsymbol{s}=\boldsymbol{k}\cap\boldsymbol{s}=r_p=s_q=k_p$, by the i.i.d.-ness of the $X_i$'s, we have

$$\begin{aligned}&E\Big[E\Big\{h(X_{r_1},\dots,X_{r_m})h(X_{s_1}\dots X_{s_m})\mid X_{k_p}\Big\}\Big]\\&\quad=E\Big[E\Big\{h(X_1,\dots,X_m)h(X_{m+1},\dots,X_{m+q-1},X_p,X_{m+q},\dots,X_{2m-1})\mid X_p\Big\}\Big],\end{aligned}\tag{A.11}$$

and

$$\begin{aligned}&E\Big[E\Big\{h(X_{k_1},\dots,X_{k_m})h(X_{s_1},\dots,X_{s_m})\mid X_{k_p}\Big\}\Big]\\&\quad=E\Big[E\Big\{h(X_1,\dots,X_m)h(X_{m+1},\dots,X_{m+q-1},X_p,X_{m+q},\dots,X_{2m-1})\mid X_p\Big\}\Big].\end{aligned}\tag{A.12}$$

Equations (A.11) and (A.12) imply that $M_2(n)=0$. Therefore, (2.24) follows from the fact that $M_1(n)=M_2(n)=0$ and the assumption that $n^{-2}\sigma_n^{-2}A_{2,1}(n)\to 0$. □

### A.1.4 Proof of Lemma 3.1

*Proof.* For $U_n^{\mathrm{Ken}}$ we have $a(i,j)=\mathbb{1}(j<i)$ and $h(X_i,X_j)=\mathbb{1}(X_j>X_i)$. Using definitions in (2.2) and (2.3), we have $f_i^{(1)}(x)=E\{h(x,X_i)\}=P(X_i>x)$, $f_i^{(2)}(x)=E\{h(X_i,x)\}=1-P(X_i>x)$, and $\theta(i,j)=1-\theta(j,i)$. By Lemma 2.2 we obtain

$$\begin{aligned}h_{1,i}^{\mathrm{Ken}}(x)&=\frac{1}{n-1}\sum_{\substack{j=1\\j\neq i}}^{n}a(i,j)\{f_j^{(1)}(x)-\theta(i,j)\}+a(j,i)\{f_j^{(2)}(x)-\theta(j,i)\}\\&=\frac{1}{n-1}\sum_{j=1}^{n}\{\mathbb{1}(j<i)-\mathbb{1}(j>i)\}\{P(X_j>x)-\theta(i,j)\},\end{aligned}$$

and

$$h_{2;i,j}^{\mathrm{Ken}}(x,y)=h(x,y)-f_j^{(1)}(x)-f_i^{(2)}(y)+\theta(i,j)=\mathbb{1}(y>x)-P(X_j>x)-P(y>X_i)+\theta(i,j).$$

This completes the proof. □

### A.1.5 Proof of Lemma 3.2

*Proof.* For $U_n^{\mathrm{AP}}$, we have $a(i,j)=n(i-1)^{-1}\mathbb{1}(j<i)$ and $h(X_i,X_j)=\mathbb{1}(X_j>X_i)$. The form of $f_i^{(1)}(x)$ and $f_i^{(2)}(x)$ is the same as in the proof of Lemma 3.1. By Lemma 2.2 we obtain

$$\begin{aligned}h_{1,i}^{\mathrm{AP}}(x)&=\frac{1}{n-1}\sum_{\substack{j=1\\j\neq i}}^{n}a(i,j)\{f_j^{(1)}(x)-\theta(i,j)\}+a(j,i)\{f_j^{(2)}(x)-\theta(j,i)\}\\&=\frac{1}{n-1}\sum_{j=1}^{n}\left\{\frac{n\mathbb{1}(j<i)}{i-1}-\frac{n\mathbb{1}(j>i)}{j-1}\right\}\{P(X_j>x)-\theta(i,j)\},\end{aligned}$$

and

$$h_{2;i,j}^{AP}(x,y)=h(x,y)-f_j^{(1)}(x)-f_i^{(2)}(y)+\theta(i,j)=\mathbb{1}(y>x)-P(X_j>x)-P(y>X_i)+\theta(i,j).$$

This completes the proof. □

### A.1.6 Proof of Theorem 3.3, $U_n^{\mathrm{Ken}}$ part

*Proof.* Proof for $U_n^{\mathrm{Ken}}$ follows the same logic as the proof for $U_n^{\mathrm{AP}}$. In the following we calculate the orders of $A_{2,2}(n)$, $A_{3,1}(n)$, and $V(n)$ for $U_n^{\mathrm{Ken}}$.

Since $a(i,j)=\mathbb{1}(j<i)$ for $U_n^{\mathrm{Ken}}$, we have $a(i,j)a(j,i)=0$ and $a(i,i)=0$. It then follows from (4.21) that

$$A_{2,2}(n)=\frac{1}{n^2}\sum_{i=2}^{n}\sum_{j=1}^{i-1}1=O(1). \tag{A.13}$$

By (4.23), following the same argument as in (4.24)-(4.27) we deduce

$$A_{3,1}(n)=O(1). \tag{A.14}$$

Next we establish lower bound on $V(n):=n^{-2}\sum_{i=1}^{n}\mathrm{Var}\{h_{1,i}^{\mathrm{Ken}}(X_i)\}$. The following lemma gives lower bound on $|h_{1,i}^{\mathrm{Ken}}(X_i)|$.

**Lemma A.2.** Consider a fixed $i\in[n]$. If $n-i\le(i-1)\delta_n/2$, either Condition (i) or Condition (ii) in Theorem 3.3 implies

$$P\Big\{|h_{1,i}^{\mathrm{Ken}}(X_i)|\ge\frac{i-1}{n-1}\frac{\delta_n}{2}\Big\}\ge p_n. \tag{A.15}$$

If $i-1\le(n-i)\delta_n/2$, either Condition (i) or Condition (ii) in Theorem 3.3 implies

$$P\Big\{|h_{1,i}^{\mathrm{Ken}}(X_i)|\ge\frac{n-i}{n-1}\frac{\delta_n}{2}\Big\}\ge p_n. \tag{A.16}$$

If $i\ge(2n-\delta_n)/(2+\delta_n)$, we have $n-i\le(i-1)\delta_n/2$ and $(i-1)/(n-1)\ge 2/(\delta_n+2)$. Lemma A.2 implies that (A.15) holds. By Chebyshev's inequality we deduce

$$\mathrm{Var}\{h_{1,i}^{\mathrm{Ken}}(X_i)\}\ge\{\frac{i-1}{n-1}\frac{\delta_n}{2}\}^2p_n\ge\frac{4}{(2+\delta_n)^2}\delta_n^2p_n. \tag{A.17}$$

If $i\le(n\delta_n+2)/(2+\delta_n)$, we have $i-1\le(n-i)\delta_n/2$ and $(n-i)/(n-1)\ge 2/(2+\delta_n)$. Lemma A.2 implies that (A.16) holds. By Chebyshev's inequality we deduce

$$\mathrm{Var}\{h_{1,i}^{\mathrm{Ken}}(X_i)\}\ge\{\frac{n-i}{n-1}\frac{\delta_n}{2}\}^2p_n\ge\frac{1}{(2+\delta_n)^2}\delta_n^2p_n. \tag{A.18}$$

By (A.17) and (A.18), we have

$$\sum_{i=1}^{n}\mathrm{Var}\{h_{1,i}^{\mathrm{Ken}}(X_i)\}\ge\sum_{i=1}^{\lfloor(n\delta_n+2)/(2+\delta_n)\rfloor}\frac{1}{(2+\delta_n)^2}\delta_n^2p_n+\sum_{i=\lfloor(2n-\delta_n)/(2+\delta_n)\rfloor+1}^{n}\frac{4}{(2+\delta_n)^2}\delta_n^2p_n. \tag{A.19}$$

Note that

$$\sum_{i=1}^{\lfloor(n\delta_n+2)/(2+\delta_n)\rfloor}\frac{1}{(2+\delta_n)^2}\delta_n^2p_n=\frac{(n\delta_n+2)/(2+\delta_n)-1}{(2+\delta_n)^2}\delta_n^2p_n\asymp n\delta_n^3p_n. \tag{A.20}$$

Combining (A.19) and (A.20) yields

$$\sum_{i=1}^{n}\mathrm{Var}\{h_{1,i}^{\mathrm{Ken}}(X_i)\}\gtrsim n\delta_n^3p_n. \tag{A.21}$$

It follows from (3.3) and (A.21) that

$$V(n) := \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\{h_{1,i}^{\mathrm{Ken}}(X_i)\} \gtrsim n^{-4/3}. \tag{A.22}$$

Equations (A.13), (A.14), and (A.22) yield (2.8) and (2.9). The asymptotic normality of $U_n^{\mathrm{Ken}}$ then follows from Theorem 2.1. This completes the proof for Part II.

□

### A.1.7 Proof of Corollary 3.1

*Proof.* We divide the proof into two parts. In Part I, we show that (2.17) and (2.18) hold for $h_{1,i}^{\mathrm{Ken}}$. In Part II, we show that (2.17) and (2.18) hold for $h_{1,i}^{\mathrm{AP}}$.

**Part I.** By (3.3) and Theorem 3.3, we have that (2.10) holds. This combined with (A.22) gives

$$n\sigma_n \gtrsim n^{1/3}, \tag{A.23}$$

where $\sigma_n^2 := \mathrm{Var}(U_n^{\mathrm{Ken}})$. By (3.1), we have $|h_{1,i}^{\mathrm{Ken}}(x)| \le 1$ for any $x$. It then follows from Markov's inequality that for any $\epsilon > 0$,

$$P\Big\{\Big|\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}\Big| \ge \epsilon\Big\} \le \frac{E|h_{1,i}^{\mathrm{Ken}}(X_i)|}{\epsilon n\sigma_n} \le \frac{1}{\epsilon n\sigma_n}. \tag{A.24}$$

Taking $\sup_{1\le i\le n}$ on both sides of (A.24), we deduce (2.17) from (A.23).

By (2.14) we have

$$E\Big\{\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}1\Big(\Big|\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}\Big| \le \epsilon\Big)\Big\} = -E\Big\{\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}1\Big(\Big|\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}\Big| > \epsilon\Big)\Big\}. \tag{A.25}$$

Cauchy-Schwarz inequality gives

$$\Big|E\Big\{\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}1\Big(\Big|\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}\Big| > \epsilon\Big)\Big\}\Big| \le \Big[E\Big\{\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}\Big\}^2\Big]^{1/2} P\Big(\Big|\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}\Big| > \epsilon\Big)^{1/2}. \tag{A.26}$$

Combining (A.25) and (A.26) yields

$$\Big[E\Big\{\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}1\Big(\Big|\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}\Big| \le \epsilon\Big)\Big\}\Big]^2 \le E\Big\{\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}\Big\}^2 P\Big(\Big|\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}\Big| > \epsilon\Big). \tag{A.27}$$

Taking summation over $1 \le i \le n$ on both sides of (A.27), it follows from (A.24) that

$$\sum_{i=1}^{n}\Big[E\Big\{\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}1\Big(\Big|\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}\Big| \le \epsilon\Big)\Big\}\Big]^2 \le \frac{1}{\epsilon n\sigma_n}\sum_{i=1}^{n}E\Big\{\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}\Big\}^2. \tag{A.28}$$

By (2.14) and (2.10) we obtain

$$\sum_{i=1}^{n}E\Big\{\frac{h_{1,i}^{\mathrm{Ken}}(X_i)}{n\sigma_n}\Big\}^2 = \sigma_n^{-2}n^{-2}\sum_{i=1}^{n}\mathrm{Var}\{h_{1,i}^{\mathrm{Ken}}(X_i)\} \to 1. \tag{A.29}$$

Equation (2.18) then follows from (A.23), (A.28), and (A.29).

**Part II.** By (3.4) and Theorem 3.3, we have that (2.10) hold. This combined with (4.36) gives

$$n\sigma_n \gtrsim n^{1/3}\log n, \tag{A.30}$$

where $\sigma_n^2 := \mathrm{Var}(U_n^{\mathrm{AP}})$. By (3.2) and the fact that $|\{\mathbb{1}(j<i)-\mathbb{1}(j>i)\}\{P(X_j>x)-\theta(i,j)\}|\le 1$, we obtain

$$|h_{1,i}^{\mathrm{AP}}(x)| \le \frac{n}{n-1}\Big(\sum_{j=1}^{i-1}\frac{1}{i-1}+\sum_{j=i+1}^{n}\frac{1}{j-1}\Big). \tag{A.31}$$

It follows from (A.31) and Lemma A.13 that

$$|h_{1,i}^{\mathrm{AP}}(x)| \le \frac{n}{n-1}\{1+1+\log(n-1)\} \le 4+2\log n. \tag{A.32}$$

By Markov's inequality and (A.32) we have for any $\epsilon>0$,

$$P\Big\{\Big|\frac{h_{1,i}^{\mathrm{AP}}(X_i)}{n\sigma_n}\Big| \ge \epsilon\Big\} \le \frac{E|h_{1,i}^{\mathrm{AP}}(X_i)|}{\epsilon n\sigma_n} \le \frac{4+2\log n}{\epsilon n \sigma_n}. \tag{A.33}$$

Taking $\sup_{1\le i\le n}$ on both sides of (A.33), we deduce (2.17) from (A.30).

Equations (A.25), (A.26) and (A.27) hold for $h_{1,i}^{\mathrm{AP}}$ as well. Taking summation over $1\le i\le n$ on both sides of (A.27), it follows from (A.33) that

$$\sum_{i=1}^{n}\Big[E\Big\{\frac{h_{1,i}^{\mathrm{AP}}(X_i)}{n\sigma_n}1\Big(\Big|\frac{h_{1,i}^{\mathrm{AP}}(X_i)}{n\sigma_n}\Big|\le\epsilon\Big)\Big\}\Big]^2 \le \frac{4+2\log n}{\epsilon n\sigma_n}\sum_{i=1}^{n}E\Big\{\frac{h_{1,i}^{\mathrm{AP}}(X_i)}{n\sigma_n}\Big\}^2. \tag{A.34}$$

By (2.14) and (2.10) we obtain

$$\sum_{i=1}^{n}E\Big\{\frac{h_{1,i}^{\mathrm{AP}}(X_i)}{n\sigma_n}\Big\}^2 = \sigma_n^{-2}n^{-2}\sum_{i=1}^{n}\mathrm{Var}\{h_{1,i}^{\mathrm{AP}}(X_i)\}\to 1. \tag{A.35}$$

Equation (2.18) then follows from (A.30), (A.34), and (A.35).

This completes the proof. □

### A.1.8 Proof of Corollary 3.2

*Proof.* By Theorem 2.6, for proving Corollary 3.2, it suffices to show that (2.21), (2.22), (2.23), and (2.24) hold. For $U_n^{\mathrm{Ken}}$ we have $|h(x,y)|\le 1$ for any $x,y$. This implies (2.21).

Now we establish (2.24). For $U_n^{\mathrm{Ken}}$, we have $|a(i,j)| = |\mathbb{1}(j<i)|\le 1$. By the definition in (2.6), we have

$$A_{2,1}(n) = n^{-3}\sum_{(I_n^2)_{\ge 1}^{\otimes 2}}|a(i_1,j_1)a(i_2,j_2)| = O(1). \tag{A.36}$$

By (2.25) and (3.19) we have

$$M_1(n)\lesssim n^{-1/6}. \tag{A.37}$$

By (2.26) and (3.20) we have

$$M_2(n)\lesssim n^{-1/3}. \tag{A.38}$$

Equation (3.3) implies (A.22) by Theorem 3.3. Combining (A.22) and (2.10) yields

$$\sigma_n^2 \gtrsim n^{-4/3}. \tag{A.39}$$

Equation (2.24) follows from (A.36), (A.37), (A.38), and (A.39).

Next we establish (2.22). The following lemma gives bounds on $\sum_{i=1}^n E\{h_{1,i}^{\mathrm{Ken}}(X_j)^2\}$ and $\sum_{i=1}^n E\{h_{1,i}^{\mathrm{Ken}}(X_i)^2\}$.

**Lemma A.3.** Under the assumptions of Corollary 3.2, we have

$$\sum_{i=1}^n E\{h_{1,i}^{\mathrm{Ken}}(X_j)^2\} = \frac{n(n+1)}{3(n-1)}(\eta^2-\theta^2)+O(n^{5/6}), \tag{A.40}$$

and

$$\sum_{i=1}^n E\{h_{1,i}^{\mathrm{Ken}}(X_i)^2\} = \frac{n(n+1)}{3(n-1)}(\eta^2-\theta^2)+O(n^{5/6}). \tag{A.41}$$

By (2.10) we have

$$\sigma_n^2 = n^{-2}\sum_{i=1}^n E\{h_{1,i}^{\mathrm{Ken}}(X_i)^2\}\{1+o(1)\}. \tag{A.42}$$

Using (A.41) and (A.42) we obtain

$$n^2\sigma_n^2 = \{1+o(1)\}\Big\{\frac{n(n+1)}{3(n-1)}(\eta^2-\theta^2)+O(n^{5/6})\Big\}. \tag{A.43}$$

Note that

$$\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n E\Big[\Big\{\frac{h_{1,i}^{\mathrm{Ken}}(X_j)}{n\sigma_n}\Big\}^2\Big] = \frac{n^{-1}\sum_{i=1}^n\sum_{j=1}^n E\{h_{1,i}^{\mathrm{Ken}}(X_j)^2\}}{n^2\sigma_n^2}. \tag{A.44}$$

Combining (A.44) with (A.40), (A.43), and the fact that $\eta^2 \neq \theta^2$ yields

$$\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n E\Big[\Big\{\frac{h_{1,i}^{\mathrm{Ken}}(X_j)}{n\sigma_n}\Big\}^2\Big] = \frac{3^{-1}(n-1)^{-1}n(n+1)(\eta^2-\theta^2)+O(n^{5/6})}{\{1+o(1)\}\Big\{3^{-1}(n-1)^{-1}n(n+1)(\eta^2-\theta^2)+O(n^{5/6})\Big\}} \to 1. \tag{A.45}$$

By (3.1) we have $|h_{1,i}(x)| \le 1$. Therefore, for any $x$

$$\Big|\sum_{i=1}^n\Big\{\frac{h_{1,i}(x)}{n\sigma_n}\Big\}^2\Big| = \Big|\frac{\sum_{i=1}^n h_{1,i}(x)^2}{n^2\sigma_n^2}\Big| \le \frac{1}{n\sigma_n^2}. \tag{A.46}$$

Equations (A.43) and (A.46) imply that

$$\sum_{j=1}^n \mathrm{Var}\Big[\sum_{i=1}^n\Big\{\frac{h_{1,i}(X_j)}{n\sigma_n}\Big\}^2\Big] \le \sum_{j=1}^n \frac{1}{n^2\sigma_n^4} = O(n) = o(n^2).$$

It then follows from Lemma A.12 that

$$\frac{1}{n}\sum_{j=1}^n\Big[\sum_{i=1}^n\Big\{\frac{h_{1,i}(X_j)}{n\sigma_n}\Big\}^2 - E\Big\{\sum_{i=1}^n\Big(\frac{h_{1,i}(X_j)}{n\sigma_n}\Big)^2\Big\}\Big] \xrightarrow{P} 0. \tag{A.47}$$

Equation (2.22) follows from (A.45) and (A.47).

Lastly, we prove (2.23). By algebra we have

$$\frac{1}{n^2}\sum_{i=1}^n\Big\{\sum_{j=1}^n\frac{h_{1,i}^{\mathrm{Ken}}(X_j)}{n\sigma_n}\Big\}^2 = \frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n\Big\{\frac{h_{1,i}^{\mathrm{Ken}}(X_j)}{n\sigma_n}\Big\}^2 + \frac{1}{n^2}\sum_{j_1\neq j_2}\sum_{i=1}^n\frac{h_{1,i}^{\mathrm{Ken}}(X_{j_1})h_{1,i}^{\mathrm{Ken}}(X_{j_2})}{n^2\sigma_n^2}. \tag{A.48}$$

By (2.22) we have

$$\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\left\{\frac{h_{1,i}^{\mathrm{Ken}}(X_j)}{n\sigma_n}\right\}^2 \xrightarrow{P} 0. \tag{A.49}$$

The second term on the right-hand side of (A.48) is $(n-1)/n$ times a U-statistic with symmetric kernel $g(x,y) = n^{-2}\sigma_n^{-2}\sum_{i=1}^{n} h_{1,i}^{\mathrm{Ken}}(x)h_{1,i}^{\mathrm{Ken}}(y)$. By (3.1) and (3.19) we have

$$E\{h_{1,i}^{\mathrm{Ken}}(X_j)\} = \frac{1}{n-1}\sum_{k=1}^{n}\mathrm{sgn}(i-k)\{P(X_k > X_j) - P(X_k > X_i)\} = O(n^{-1/6}). \tag{A.50}$$

It follows from (A.50) and (A.43) that

$$E\{g(X_{j_1},X_{j_2})\} = n^{-2}\sigma_n^{-2}\sum_{i=1}^{n} E\{h_{1,i}^{\mathrm{Ken}}(X_{j_1})\}E\{h_{1,i}^{\mathrm{Ken}}(X_{j_2})\} \to 0. \tag{A.51}$$

By (A.51) and the weak law of large numbers for U-statistics with independent but not identically distributed variables (Lee, 1990, Theorem 1, Section 3.7.2), we deduce

$$\frac{1}{n^2}\sum_{j_1\neq j_2}\sum_{i=1}^{n}\frac{h_{1,i}^{\mathrm{Ken}}(X_{j_1})h_{1,i}^{\mathrm{Ken}}(X_{j_2})}{n^2\sigma_n^2} \xrightarrow{P} 0. \tag{A.52}$$

Equation (2.23) follows from (A.48), (A.49), and (A.52).

This completes the proof. □

## A.2 Proofs of the supporting lemmas

### A.2.1 Proof of Lemma A.1

*Proof.* We prove Lemma A.1 by showing that for each $(i_1^*,\ldots,i_m^*) \in I_n^m$, the coefficients of $a(i_1^*,\ldots,i_m^*)$ on both sides of (A.5) are equal. In the following we fix $(i_1^*,\ldots,i_m^*) \in I_n^m$.

For the left-hand side of (A.5), we enumerate the combinations in

$$\{l,i,(i_1,\ldots,i_{m-1}) : l \in [m], i \in [n], (i_1,\ldots,i_{m-1}) \in I_{n-1}^{m-1}(-i)\}$$

such that $a^{(l)}(i;i_1,\ldots,i_{m-1}) = a(i_1^*,\ldots,i_m^*)$, as follows:

$$\begin{aligned}
l = 1, i = i_1^*, (i_1,\ldots,i_{m-1}) &= (i_1^*,\ldots,i_m^*)\backslash i_1^*;\\
&\vdots\\
l = j, i = i_j^*, (i_1,\ldots,i_{m-1}) &= (i_1^*,\ldots,i_m^*)\backslash i_j^*;\\
&\vdots\\
l = m, i = i_m^*, (i_1,\ldots,i_{m-1}) &= (i_1^*,\ldots,i_m^*)\backslash i_m^*.
\end{aligned} \tag{A.53}$$

When $l = j, i = i_j^*, (i_1,\ldots,i_{m-1}) = (i_1^*,\ldots,i_m^*)\backslash i_j^*$,

$$E\{h^{(l)}(X_i;X_{i_1},\ldots,X_{i_{m-1}}) \mid X_i\} - \theta^{(l)}(i;i_1,\ldots,i_{m-1}) = E\{h(X_{i_1^*},\ldots,X_{i_m^*}) \mid X_{i_j^*}\} - \theta(i_1^*,\ldots,i_m^*).$$

So the coefficient of $a(i_1^*,\dots,i_m^*)$ on the left-hand side of (A.5) is

$$\sum_{j=1}^{m}\Big[E\{h(X_{i_1^*},\dots,X_{i_m^*}\mid X_{i_j^*}\}-\theta(i_1^*,\dots,i_m^*)\Big].$$

This equals the coefficient of $a(i_1^*,\dots,i_m^*)$ on the right-hand side of (A.5). This completes the proof. □

### A.2.2 Proof of Lemma 4.1

*Proof.* To simplify notation, define $\boldsymbol{i}=(i_1,\dots,i_m)$, $X_{\boldsymbol{i}}=(X_{i_1},\dots,X_{i_m})$, and $\boldsymbol{i}_{-m}=(i_1,\dots,i_{m-1})$. By definition of $h_{1,i}(X_i)$ in (2.4) we have

$$\sum_{i=1}^{n}E|h_{1,i}(X_i)|^3=\sum_{i=1}^{n}\Big\{\frac{(n-m)!}{(n-1)!}\Big\}^3E\Big|\sum_{I_{n-1}^{m-1}(-i)}\sum_{l=1}^{m}a^{(l)}(i;\boldsymbol{i}_{-m})\Big\{f_{\boldsymbol{i}_{-m}}^{(l)}(X_i)-\theta^{(l)}(i;\boldsymbol{i}_{-m})\Big\}\Big|^3. \tag{A.54}$$

Define

$$T_{\boldsymbol{i}_{-m}}^{(l_1)}(X_i)=f_{\boldsymbol{i}_{-m}}^{(l_1)}(X_i)-\theta^{(l_1)}(i;\boldsymbol{i}_{-m}),$$

and define $T_{\boldsymbol{j}_{-m}}^{(l_2)}(X_i)$ and $T_{\boldsymbol{k}_{-m}}^{(l_3)}(X_i)$ similarly. The right-hand side of (A.54) equals

$$\Big\{\frac{(n-m)!}{(n-1)!}\Big\}^3\sum\Big|a^{(l_1)}(i;\boldsymbol{i}_{-m})a^{(l_2)}(i;\boldsymbol{j}_{-m})a^{(l_3)}(i;\boldsymbol{k}_{-m})\Big|E\Big|T_{\boldsymbol{i}_{-m}}^{(l_1)}(X_i)T_{\boldsymbol{j}_{-m}}^{(l_2)}(X_i)T_{\boldsymbol{k}_{-m}}^{(l_3)}(X_i)\Big|, \tag{A.55}$$

where the summation is over $i\in[n]$, $l_1,l_2,l_3\in[m]$, and $\boldsymbol{i}_{-m},\boldsymbol{j}_{-m},\boldsymbol{k}_{-m}\in I_{n-1}^{m-1}(-i)$. By Cauchy-Schwarz inequality and Lemma A.6(ii), we have

$$\begin{aligned}E\Big|T_{\boldsymbol{i}_{-m}}^{(l_1)}(X_i)T_{\boldsymbol{j}_{-m}}^{(l_2)}(X_i)T_{\boldsymbol{k}_{-m}}^{(l_3)}(X_i)\Big|&\le\Big[E\{T_{\boldsymbol{i}_{-m}}^{(l_1)}(X_i)^2T_{\boldsymbol{j}_{-m}}^{(l_2)}(X_i)^2\}\Big]^{1/2}\Big[E\{T_{\boldsymbol{k}_{-m}}^{(l_3)}(X_i)^2\}\Big]^{1/2}\\&\le\Big[E\{T_{\boldsymbol{i}_{-m}}^{(l_1)}(X_i)^4\}\Big]^{1/4}\Big[E\{T_{\boldsymbol{j}_{-m}}^{(l_2)}(X_i)^4\}\Big]^{1/4}\Big[E\{T_{\boldsymbol{k}_{-m}}^{(l_3)}(X_i)^4\}\Big]^{1/4}\le CM(n)^{3/4}.\end{aligned} \tag{A.56}$$

By the definition of $A_{3,1}(n)$ in (2.6) and algebra, we have

$$\sum|a^{(l_1)}(i;\boldsymbol{i}_{-m})a^{(l_2)}(i;\boldsymbol{j}_{-m})a^{(l_3)}(i;\boldsymbol{k}_{-m})|\le C\sum_{(I_n^m)_{\ge1}^{\otimes3}}|a(\boldsymbol{i})a(\boldsymbol{j})a(\boldsymbol{k})|=Cn^{3m-2}A_{3,1}(n), \tag{A.57}$$

where the summation in the leftmost part of (A.57) is over $i\in[n]$, $l_1,l_2,l_3\in[m]$, and $\boldsymbol{i}_{-m},\boldsymbol{j}_{-m},\boldsymbol{k}_{-m}\in I_{n-1}^{m-1}(-i)$. By (A.54), (A.55), (A.56), and (A.57), we deduce

$$\sum_{i=1}^{n}E|h_{1,i}(X_i)|^3\le CnA_{3,1}(n)M(n)^{3/4}. \tag{A.58}$$

This completes the proof. □

### A.2.3 Proof of Lemma 4.2

*Proof.* Define $\boldsymbol{i} := (i_1, \ldots, i_m)$ and $X_{\boldsymbol{i}} := (X_{i_1}, \ldots, X_{i_m})$. By (4.13) we have

$$E^*\{\sigma_n^{-2} U_n^*(h_2)^2\} = \sigma_n^{-2} \Big\{ \frac{(n-m)!}{n!} \Big\}^2 \sum_{(I_n^m)^{\otimes 2}_{\geq 0}} a(\boldsymbol{i}) a(\boldsymbol{j}) E^*\{h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}}) h_{2;\boldsymbol{j}}(X^*_{\boldsymbol{j}})\} \tag{A.59}$$

$$\text{and} \quad [E^*\{\sigma_n U_n^*(h_2)\}]^2 = \sigma_n^{-2} \Big\{ \frac{(n-m)!}{n!} \Big\}^2 \sum_{(I_n^m)^{\otimes 2}_{\geq 0}} a(\boldsymbol{i}) a(\boldsymbol{j}) E^*\{h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}})\} E^*\{h_{2;\boldsymbol{j}}(X^*_{\boldsymbol{j}})\}. \tag{A.60}$$

Define

$$g(\boldsymbol{i},\boldsymbol{j}) := a(\boldsymbol{i}) a(\boldsymbol{j}) \Big[ E^*\{h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}}) h_{2;\boldsymbol{j}}(X^*_{\boldsymbol{j}})\} - E^*\{h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}})\} E^*\{h_{2;\boldsymbol{j}}(X^*_{\boldsymbol{j}})\} \Big]. \tag{A.61}$$

It follows from (A.59) and (A.60) that

$$\mathrm{Var}^*\{\sigma_n^{-1} U_n^*(h_2)\} = \sigma_n^{-2} \Big\{ \frac{(n-m)!}{n!} \Big\}^2 \sum_{(\boldsymbol{i},\boldsymbol{j}) \in (I_n^m)^{\otimes 2}_{\geq 0}} g(\boldsymbol{i},\boldsymbol{j}). \tag{A.62}$$

The following proof consists of two steps. In the first step, we establish

$$\mathrm{Var}^*\{\sigma_n^{-1} U_n^*(h_2)\} = \sigma_n^{-2} \Big\{ \frac{(n-m)!}{n!} \Big\}^2 \sum_{(\boldsymbol{i},\boldsymbol{j}) \in (I_n^m)^{\otimes 2}_{=1}} g(\boldsymbol{i},\boldsymbol{j}) + o_P(1). \tag{A.63}$$

In the second step, we show that

$$\sigma_n^{-2} \Big\{ \frac{(n-m)!}{n!} \Big\}^2 \sum_{(\boldsymbol{i},\boldsymbol{j}) \in (I_n^m)^{\otimes 2}_{=1}} g(\boldsymbol{i},\boldsymbol{j}) \xrightarrow{P} 0. \tag{A.64}$$

Lemma 4.2 then follows from (A.63), (A.64), and Slutsky's theorem.

**Step I.** If $(\boldsymbol{i},\boldsymbol{j}) \in (I_n^m)^{\otimes 2}_{=0}$, we have $E^*\{h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}}) h_{2;\boldsymbol{j}}(X^*_{\boldsymbol{j}})\} = E^*\{h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}})\} E^*\{h_{2;\boldsymbol{j}}(X^*_{\boldsymbol{j}})\}$ a.s.. This implies

$$\sum_{(\boldsymbol{i},\boldsymbol{j}) \in (I_n^m)^{\otimes 2}_{=0}} g(\boldsymbol{i},\boldsymbol{j}) = 0 \text{ a.s..} \tag{A.65}$$

For any $(\boldsymbol{i},\boldsymbol{j}) \in (I_n^m)^{\otimes 2}_{\geq 2}$, by the law of iterated expectation, Cauchy-Schwarz inequality, and triangular inequality we have

$$E\Big| E^*\{h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}}) h_{2;\boldsymbol{j}}(X^*_{\boldsymbol{j}})\} \Big| \leq E\Big\{ |h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}})||h_{2;\boldsymbol{j}}(X^*_{\boldsymbol{j}})| \Big\} \leq \Big[ E\{h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}})^2\} E\{h_{2;\boldsymbol{j}}(X^*_{\boldsymbol{j}})^2\} \Big]^{\frac{1}{2}}. \tag{A.66}$$

Similarly, by Jensen's inequality and triangular inequality we have

$$\begin{aligned} & E\Big| E^*\{h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}})\} E^*\{h_{2;\boldsymbol{j}}(X^*_{\boldsymbol{j}})\} \Big| \leq E\Big\{ E^*|h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}})| E^*|h_{2;\boldsymbol{j}}(X^*_{\boldsymbol{j}})| \Big\} \\ \leq & \Big[ E\{E^*|h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}})|\}^2 E\{E^*|h_{2;\boldsymbol{j}}(X^*_{\boldsymbol{j}})|\}^2 \Big]^{\frac{1}{2}} \leq \Big[ E\{h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}})^2\} E\{h_{2;\boldsymbol{j}}(X^*_{\boldsymbol{j}})^2\} \Big]^{\frac{1}{2}}. \end{aligned} \tag{A.67}$$

Using the law of iterated expectation, we deduce

$$E\{h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}})^2\} = E[E^*\{h_{2;\boldsymbol{i}}(X^*_{\boldsymbol{i}})^2\}] = n^{-m} \sum_{1 \leq j_1, \ldots, j_m \leq n} E\{h_{2;\boldsymbol{i}}(X_{j_1}, \ldots, X_{j_m})^2\}. \tag{A.68}$$

By Lemma A.6(iii) and (2.21), there exists an absolute constant $C > 0$ such that for any $n$, for any

$\boldsymbol{i}\in I_n^m$, and for any $1\le j_1,\ldots,j_m\le n$,

$$E\{h_{2;\boldsymbol{i}}(X_{j_1},\ldots,X_{j_m})^2\}\le C. \tag{A.69}$$

Combining (A.68) and (A.69) yields that $E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{i}}^*)^2\}\le C$. It then follows from (A.66) and (A.67) that

$$E\Big|E^*\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{i}}^*)h_{2;\boldsymbol{j}}(X_{\boldsymbol{j}}^*)\}\Big|\le C, \tag{A.70}$$

$$\text{and}\qquad E\Big|E^*\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{i}}^*)\}E^*\{h_{2;\boldsymbol{j}}(X_{\boldsymbol{j}}^*)\}\Big|\le C. \tag{A.71}$$

Equations (A.61), (A.70), and (A.71) imply that

$$E\Big|\sum_{(I_n^m)^{\otimes 2}_{\ge 2}}g(\boldsymbol{i},\boldsymbol{j})\Big|\le 2C\sum_{(I_n^m)^{\otimes 2}_{\ge 2}}|a(\boldsymbol{i})a(\boldsymbol{j})|=2Cn^{2m-2}A_{2,2}(n). \tag{A.72}$$

By (2.8), (2.10), and (A.72), we deduce

$$\sigma_n^{-2}\Big\{\frac{(n-m)!}{n!}\Big\}^2E\Big|\sum_{(I_n^m)^{\otimes 2}_{\ge 2}}g(\boldsymbol{i},\boldsymbol{j})\Big|\to 0.$$

It then follows from Markov's inequality that

$$\sigma_n^{-2}\Big\{\frac{(n-m)!}{n!}\Big\}^2\sum_{(I_n^m)^{\otimes 2}_{\ge 2}}g(\boldsymbol{i},\boldsymbol{j})\xrightarrow{P}0. \tag{A.73}$$

Combining (A.62), (A.65), and (A.73) yields (A.63). This concludes **Step I.**

**Step II.** Consider a fixed $(\boldsymbol{i},\boldsymbol{j})\in(I_n^m)^{\otimes 2}_{=1}$. Without loss of generality assume $\boldsymbol{i}\cap\boldsymbol{j}=\{i_p\}=\{j_q\}$ for some $1\le p,q\le m$. By the i.i.d.-ness of $X_i^*$'s given $X_1,\ldots,X_n$, we have

$$E[E^*\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{i}}^*)h_{2;\boldsymbol{j}}(X_{\boldsymbol{j}}^*)\}]=n^{-(2m-1)}\sum_{\substack{\boldsymbol{r},\boldsymbol{s}\in[n]^{2m}\\ r_p=s_q}}E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\} \tag{A.74}$$

and

$$E[E^*\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{i}}^*)\}E^*\{h_{2;\boldsymbol{j}}(X_{\boldsymbol{j}}^*)\}]=n^{-2m}\sum_{\boldsymbol{r},\boldsymbol{s}\in[n]^{2m}}E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\}. \tag{A.75}$$

The number of pairs $(\boldsymbol{r},\boldsymbol{s})$ in $\{(\boldsymbol{r},\boldsymbol{s})\in[n]^{2m}:r_p=s_q\}$ satisfying any of the following three statements is of order $O(n^{2m-2})$: (1) $\boldsymbol{r}$ or $\boldsymbol{s}$ has duplicate indices (i.e., $\boldsymbol{r}\notin I_n^m$ or $\boldsymbol{s}\notin I_n^m$); (2) $\boldsymbol{i}\cap\boldsymbol{r}\ne\emptyset$; or (3) $\boldsymbol{j}\cap\boldsymbol{s}\ne\emptyset$. It then follows from (A.70) that

$$\sum_{\substack{\boldsymbol{r},\boldsymbol{s}\in[n]^{2m}\\ r_p=s_q}}E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\}=\sum_{\substack{(\boldsymbol{r},\boldsymbol{s})\in(I_n^m)^{\otimes 2}_{=1}\\ r_p=s_q,\boldsymbol{i}\cap\boldsymbol{r}=\emptyset=\boldsymbol{j}\cap\boldsymbol{s}}}E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\}+O(n^{2m-2}). \tag{A.76}$$

The following lemma gives bound on the right-hand side of (A.76).

**Lemma A.4.** For any $(\boldsymbol{i},\boldsymbol{j})\in(I_n^m)^{\otimes 2}_{=1}$, under the assumptions of Theorem 2.6, there exists a constant

$C$ such that

$$\Big|\sum_{\substack{(\boldsymbol{r},\boldsymbol{s})\in(I_n^m)_{=1}^{\otimes 2}\\ r_p=s_q,\boldsymbol{i}\cap\boldsymbol{r}=\emptyset=\boldsymbol{j}\cap\boldsymbol{s}}} E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\}\Big| \le Cn^{2m-1}\{M_1(n)^2+M_2(n)\}. \tag{A.77}$$

It follows from (A.74), (A.76) and Lemma A.4 that

$$|E[E^*\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{i}}^*)h_{2;\boldsymbol{j}}(X_{\boldsymbol{j}}^*)\}]| \le C\{M_1(n)^2+M_2(n)+n^{-1}\}. \tag{A.78}$$

Using an argument similar to (A.76), we have

$$\sum_{\boldsymbol{r},\boldsymbol{s}\in\{1,\ldots,n\}^{2m}} E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\} = \sum_{\substack{(\boldsymbol{r},\boldsymbol{s})\in(I_n^m)_{=0}^{\otimes 2}\\ \boldsymbol{i}\cap\boldsymbol{r}=\emptyset=\boldsymbol{j}\cap\boldsymbol{s}}} E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\}+O(n^{2m-1}). \tag{A.79}$$

The following lemma gives bound on the right-hand side of (A.79).

**Lemma A.5.** For any $(\boldsymbol{i},\boldsymbol{j})\in(I_n^m)_{=1}^{\otimes 2}$, under the assumptions of Theorem 2.6, there exists a constant $C$ such that

$$\Big|\sum_{\substack{(\boldsymbol{r},\boldsymbol{s})\in(I_n^m)_{=0}^{\otimes 2}\\ \boldsymbol{i}\cap\boldsymbol{r}=\emptyset=\boldsymbol{j}\cap\boldsymbol{s}}} E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\}\Big| \le Cn^{2m}M_1(n)^2. \tag{A.80}$$

It follows from (A.75), (A.79) and Lemma A.5 that

$$|E[E^*\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{i}}^*)\}E^*\{h_{2;\boldsymbol{j}}(X_{\boldsymbol{j}}^*)\}]| \le CM_1(n)^2. \tag{A.81}$$

Combining (A.61) with (A.78) and (A.81) yields that, for any $(\boldsymbol{i},\boldsymbol{j})\in(I_n^m)_{=1}^{\otimes 2}$,

$$|g(\boldsymbol{i},\boldsymbol{j})| \le C|a(\boldsymbol{i})a(\boldsymbol{j})|\{M_1(n)^2+M_2(n)+n^{-1}\}.$$

Therefore, by the definition of $A_{2,1}(n)$ in (2.6), we have

$$\Big|\sum_{(I_n^m)_{=1}^{\otimes 2}} g(\boldsymbol{i},\boldsymbol{j})\Big| \le Cn^{2m-1}A_{2,1}(n)\{M_1(n)^2+M_2(n)+n^{-1}\}. \tag{A.82}$$

Equation (A.64) follows from (A.82) and (2.24). This concludes **Step II**.

The proof is thus finished. □

### A.2.4 Proof of Lemma 4.3

*Proof.* Define

$$f_{ij}(x) := P(X_j > x) - \theta(i,j), \quad S_i^{(1)}(x) := \sum_{j=1}^{i-1}\frac{n}{i-1}f_{ij}(x), \quad \text{and} \quad S_i^{(2)}(x) := \sum_{j=i+1}^{n}\frac{n}{j-1}f_{ij}(x). \tag{A.83}$$

By (3.2) we have $h_{1,i}^{\mathrm{AP}}(X_i) = \{S_i^{(1)}(X_i) - S_i^{(2)}(X_i)\}/(n-1)$ for any $i\in[n]$. In the following we use Lemma A.13 repeatedly to bound $\varphi(n) := \sum_{k=1}^n k^{-1}$.

First, we show that (4.29) and (4.30) hold under Condition (i) of Theorem 3.3. Using $f_{ij}(\cdot)$ notation, Condition (i) becomes

$$P\{\delta_n \le f_{ij}(X_i) \le 1, \forall j \in [n]\backslash\{i\}\} \ge p_n. \tag{A.84}$$

If $\delta_n \le f_{ij}(x) \le 1, \forall j \in [n]\backslash\{i\}$, we have

$$n\delta_n = \sum_{j=1}^{i-1} \frac{n}{i-1}\delta_n \le S_i^{(1)}(x) \le \sum_{j=1}^{i-1} \frac{n}{i-1} = n, \tag{A.85}$$

and

$$S_i^{(2)}(x) \ge \sum_{j=i+1}^{n} \frac{n}{j-1}\delta_n = n\delta_n\{\varphi(n-1)-\varphi(i-1)\} \ge n\delta_n \log\frac{n}{i}, \tag{A.86}$$

$$S_i^{(2)}(x) \le \sum_{j=i+1}^{n} \frac{n}{j-1} = n\{\varphi(n-1)-\varphi(i-1)\} \le n\log\frac{n-1}{i-1}. \tag{A.87}$$

Using (A.85), (A.86), and (A.87), it follows from (A.84) that

$$P\left\{n\delta_n \le S_i^{(1)}(X_i) \le n, n\delta_n\log\frac{n}{i} \le S_i^{(2)}(X_i) \le n\log\frac{n-1}{i-1}\right\} \ge p_n. \tag{A.88}$$

If $\log\{(n-1)/(i-1)\} \le \delta_n/2$, the monotonicity property of probability measure gives

$$\begin{aligned} &P\{h_{1,i}(X_i) \ge \delta_n/2\} \ge P\left\{\frac{S_i^{(1)}(X_i)}{n-1} \ge \frac{n}{n-1}\delta_n, \frac{S_i^{(2)}(X_i)}{n-1} \le \frac{n}{n-1}\log\frac{n-1}{i-1}\right\} \\ =&P\left\{S_i^{(1)} \ge n\delta_n, S_i^{(2)} \le n\log\frac{n-1}{i-1}\right\} \ge P\left\{n\delta_n \le S_i^{(1)} \le n, n\delta_n\log\frac{n}{i} \le S_i^{(2)} \le n\log\frac{n-1}{i-1}\right\}. \end{aligned} \tag{A.89}$$

Note that

$$P\{|h_{1,i}(X_i)| \ge \delta_n/2\} \ge P\{h_{1,i}(X_i) \ge \delta_n/2\}. \tag{A.90}$$

Equation (4.29) follows from (A.88), (A.89), and (A.90). If $\delta_n\log(n/i) \ge 2$, the monotonicity property of probability measure gives

$$\begin{aligned} &P\{h_{1,i}(X_i) \le -1\} \ge P\left\{\frac{S_i^{(1)}(X_i)}{n-1} \le \frac{n}{n-1}, \frac{S_i^{(2)}(X_i)}{n-1} \ge \frac{n}{n-1}\delta_n\log\frac{n}{i}\right\} \\ =&P\left\{S_i^{(1)} \le n, S_i^{(2)} \ge n\delta_n\log\frac{n}{i}\right\} \ge P\left\{n\delta_n \le S_i^{(1)} \le n, n\delta_n\log\frac{n}{i} \le S_i^{(2)} \le n\log\frac{n-1}{i-1}\right\}. \end{aligned} \tag{A.91}$$

Note that

$$P\{|h_{1,i}(X_i)| \ge 1\} \ge P\{h_{1,i}(X_i) \le -1\}. \tag{A.92}$$

Equation (4.30) follows from (A.88), (A.91), and (A.92).

Secondly, we show that (4.29) and (4.30) hold under Condition (ii) of Theorem 3.3. Using $f_{ij}(\cdot)$ notation, Condition (ii) becomes

$$P\{-1 \le f_{ij}(X_i) \le -\delta_n, \forall j \in [n]\backslash\{i\}\} \ge p_n. \tag{A.93}$$

By an argument similar to (A.85)-(A.87), if $-1 \le f_{ij}(x) \le -\delta_n, \forall j \in [n]\backslash\{i\}$ we have

$$-n \le S_i^{(1)}(x) \le -n\delta_n, \quad \text{and} \quad -n\log\frac{n-1}{i-1} \le S_i^{(2)}(x) \le -n\delta_n\log\frac{n}{i}. \tag{A.94}$$

By (A.94), Condition (ii) in Theorem 3.3 implies that

$$P\Big\{-n \le S_i^{(1)}(X_i) \le -n\delta_n, -n\log\frac{n-1}{i-1} \le S_i^{(2)}(X_i) \le -n\delta_n \log\frac{n}{i}\Big\} \ge p_n. \tag{A.95}$$

If $\log\{(n-1)/(i-1)\} \le \delta_n/2$, the monotonicity property of probability measure gives

$$\begin{aligned}
&P\{h_{1,i}(X_i) \le -\delta_n/2\} \ge P\Big\{\frac{S_i^{(1)}(X_i)}{n-1} \le -\frac{n}{n-1}\delta_n, \frac{S_i^{(2)}(X_i)}{n-1} \ge -\frac{n}{n-1}\log\frac{n-1}{i-1}\Big\} \\
=&P\Big\{S_i^{(1)} \le -n\delta_n, S_i^{(2)} \ge -n\log\frac{n-1}{i-1}\Big\} \\
\ge&P\Big\{-n \le S_i^{(1)} \le -n\delta_n, -n\log\frac{n-1}{i-1} \le S_i^{(2)} \le -n\delta_n\log\frac{n}{i}\Big\}.
\end{aligned} \tag{A.96}$$

Note that

$$P\{|h_{1,i}(X_i)| \ge \delta_n/2\} \ge P\{h_{1,i}(X_i) \le -\delta_n/2\}. \tag{A.97}$$

Equation (4.29) follows from (A.95), (A.96), and (A.97). If $\delta_n \log(n/i) \ge 2$, the monotonicity property of probability measure gives

$$\begin{aligned}
&P\{h_{1,i}(X_i) \ge 1\} \ge P\Big\{\frac{S_i^{(1)}(X_i)}{n-1} \ge -\frac{n}{n-1}, \frac{S_i^{(2)}(X_i)}{n-1} \le -\frac{n}{n-1}\delta_n\log\frac{n}{i}\Big\} \\
=&P\Big\{S_i^{(1)} \ge -n, S_i^{(2)} \le -n\delta_n\log\frac{n}{i}\Big\} \\
\ge&P\Big\{-n \le S_i^{(1)} \le -n\delta_n, -n\log\frac{n-1}{i-1} \le S_i^{(2)} \le -n\delta_n\log\frac{n}{i}\Big\}.
\end{aligned} \tag{A.98}$$

Note that

$$P\{|h_{1,i}(X_i)| \ge 1\} \ge P\{h_{1,i}(X_i) \ge 1\}. \tag{A.99}$$

Equation (4.30) follows from (A.95), (A.98), and (A.99).

This completes the proof. □

### A.2.5 Proof of Lemma A.2

*Proof.* Define

$$f_{ij}(x) := P(X_j > x) - \theta(i,j), \quad S_i^{(1)}(x) := \sum_{j=1}^{i-1} f_{ij}(x), \quad \text{and} \quad S_i^{(2)}(x) := \sum_{j=i+1}^{n} f_{ij}(x). \tag{A.100}$$

By (3.1) we have $h_{1,i}^{\mathrm{AP}}(X_i) = \{S_i^{(1)}(X_i) - S_i^{(2)}(X_i)\}/(n-1)$ for $2 \le i \le n$.

First, we show that (A.15) and (A.16) hold under Condition (i) of Theorem 3.3. Using $f_{ij}(\cdot)$ notation, Condition (i) becomes

$$P\{\delta_n \le f_{ij}(X_i) \le 1, \forall j \in [n]\backslash\{i\}\} \ge p_n. \tag{A.101}$$

If $\delta_n \le f_{ij}(x) \le 1, \forall j \in [n]\backslash\{i\}$, we have

$$(i-1)\delta_n = \sum_{j=1}^{i-1}\delta_n \le S_i^{(1)}(x) \le \sum_{j=1}^{i-1} 1 = i-1, \tag{A.102}$$

and

$$(n-i)\delta_n = \sum_{j=i+1}^{n} \delta_n \leq S_i^{(2)}(x) \leq \sum_{j=i+1}^{n} 1 = n-i. \tag{A.103}$$

Using (A.102) and (A.103), it follows from (A.101) that

$$P\Big\{(i-1)\delta_n \leq S_i^{(1)}(X_i) \leq i-1, (n-i)\delta_n \leq S_i^{(2)}(X_i) \leq n-i\Big\} \geq p_n. \tag{A.104}$$

If $n-i \leq (i-1)\delta_n/2$, the monotonicity property of probability measure gives

$$\begin{aligned}
&P\Big\{h_{1,i}(X_i) \geq \frac{i-1}{n-1}\frac{\delta_n}{2}\Big\} \geq P\Big\{\frac{S_i^{(1)}(X_i)}{n-1} \geq \frac{i-1}{n-1}\delta_n, \frac{S_i^{(2)}(X_i)}{n-1} \leq \frac{n-i}{n-1}\Big\} \\
=& P\Big\{S_i^{(1)} \geq (i-1)\delta_n, S_i^{(2)} \leq n-i\Big\} \\
\geq& P\Big\{(i-1)\delta_n \leq S_i^{(1)} \leq i-1, (n-i)\delta_n \leq S_i^{(2)} \leq n-i\Big\}. \qquad (A.105)
\end{aligned}$$

Note that

$$P\Big\{|h_{1,i}(X_i)| \geq \frac{i-1}{n-1}\frac{\delta_n}{2}\Big\} \geq P\Big\{h_{1,i}(X_i) \geq \frac{i-1}{n-1}\frac{\delta_n}{2}\Big\}. \tag{A.106}$$

Equation (A.15) follows from (A.104), (A.105), and (A.106). If $i-1 \leq (n-i)\delta_n/2$, the monotonicity property of probability measure gives

$$\begin{aligned}
&P\{h_{1,i}(X_i) \leq -\frac{n-i}{n-1}\frac{\delta_n}{2}\} \geq P\Big\{\frac{S_i^{(1)}(X_i)}{n-1} \leq \frac{i-1}{n-1}, \frac{S_i^{(2)}(X_i)}{n-1} \geq \frac{n-i}{n-1}\delta_n\Big\} \\
=& P\Big\{S_i^{(1)} \leq i-1, S_i^{(2)} \geq (n-i)\delta_n\Big\} \\
\geq& P\Big\{(i-1)\delta_n \leq S_i^{(1)} \leq i-1, (n-i)\delta_n \leq S_i^{(2)} \leq n-i\Big\}. \qquad (A.107)
\end{aligned}$$

Note that

$$P\Big\{|h_{1,i}(X_i)| \geq \frac{n-i}{n-1}\frac{\delta_n}{2}\Big\} \geq P\Big\{h_{1,i}(X_i) \leq -\frac{n-i}{n-1}\frac{\delta_n}{2}\Big\}. \tag{A.108}$$

Equation (A.16) follows from (A.104), (A.107), and (A.108).

Secondly, we show that (A.15) and (A.16) hold under Condition (ii) of Theorem 3.3. Using $f_{ij}(\cdot)$ notation, Condition (ii) becomes

$$P\{-1 \leq f_{ij}(X_i) \leq -\delta_n, \forall j \in [n] \backslash \{i\}\} \geq p_n. \tag{A.109}$$

If $-1 \leq f_{ij}(x) \leq -\delta_n, \forall j \in [n]\backslash\{i\}$, we have

$$-(i-1) = -\sum_{j=1}^{i-1} 1 \leq S_i^{(1)}(x) \leq -\sum_{j=1}^{i-1} \delta_n = -(i-1)\delta_n, \tag{A.110}$$

and

$$-(n-i) = -\sum_{j=i+1}^{n} 1 \leq S_i^{(2)}(x) \leq -\sum_{j=i+1}^{n} \delta_n = -(n-i)\delta_n. \tag{A.111}$$

Using (A.110) and (A.111), it follows from (A.101) that

$$P\Big\{-(i-1) \leq S_i^{(1)}(X_i) \leq -(i-1)\delta_n, -(n-i) \leq S_i^{(2)}(X_i) \leq -(n-i)\delta_n\Big\} \geq p_n. \tag{A.112}$$

If $n-i \le (i-1)\delta_n/2$, the monotonicity property of probability measure gives

$$\begin{aligned}
&P\Big\{h_{1,i}(X_i) \le -\frac{i-1}{n-1}\frac{\delta_n}{2}\Big\} \ge P\Big\{\frac{S_i^{(1)}(X_i)}{n-1} \le -\frac{i-1}{n-1}\delta_n, \frac{S_i^{(2)}(X_i)}{n-1} \ge -\frac{n-i}{n-1}\Big\}\\
=&P\Big\{S_i^{(1)} \le -(i-1)\delta_n, S_i^{(2)} \ge -(n-i)\Big\}\\
\ge&P\Big\{-(i-1) \le S_i^{(1)} \le -(i-1)\delta_n, -(n-i) \le S_i^{(2)} \le -(n-i)\delta_n\Big\}. \qquad \text{(A.113)}
\end{aligned}$$

Note that

$$P\Big\{|h_{1,i}(X_i)| \ge \frac{i-1}{n-1}\frac{\delta_n}{2}\Big\} \ge P\Big\{h_{1,i}(X_i) \le -\frac{i-1}{n-1}\frac{\delta_n}{2}\Big\}. \qquad \text{(A.114)}$$

Equation (A.15) follows from (A.112), (A.113), and (A.114). If $i-1 \le (n-i)\delta_n/2$, the monotonicity property of probability measure gives

$$\begin{aligned}
&P\{h_{1,i}(X_i) \ge \frac{n-i}{n-1}\frac{\delta_n}{2}\} \ge P\Big\{\frac{S_i^{(1)}(X_i)}{n-1} \ge -\frac{i-1}{n-1}, \frac{S_i^{(2)}(X_i)}{n-1} \le -\frac{n-i}{n-1}\delta_n\Big\}\\
=&P\Big\{S_i^{(1)} \ge -(i-1), S_i^{(2)} \le -(n-i)\delta_n\Big\}\\
\ge&P\Big\{-(i-1) \le S_i^{(1)} \le -(i-1)\delta_n, -(n-i) \le S_i^{(2)} \le -(n-i)\delta_n\Big\}. \qquad \text{(A.115)}
\end{aligned}$$

Note that

$$P\Big\{|h_{1,i}(X_i)| \ge \frac{n-i}{n-1}\frac{\delta_n}{2}\Big\} \ge P\Big\{h_{1,i}(X_i) \ge \frac{n-i}{n-1}\frac{\delta_n}{2}\Big\}. \qquad \text{(A.116)}$$

Equation (A.16) follows from (A.112), (A.115), and (A.116).

This completes the proof. □

### A.2.6 Proof of Lemma 4.4

*Proof.* As in the statement of Lemma 4.4, we consider a fixed $i \in [n]$. For any $j \in [n]\backslash\{i\}$, we have $\rho_{ij}^{-1} \le \rho_n$ and $-r_{ij} \le R_n$. This combined with (4.40) implies that $z_i \ge \rho_{ij}^{-1}t_0 - r_{ij}$, or equivalently

$$\rho_{ij}(z_i + r_{ij}) \ge t_0. \qquad \text{(A.117)}$$

Equations (A.117) and (3.6) imply that

$$F_j^c\{\rho_{ij}(z_i+r_{ij})\} \le c_2\{\rho_{ij}(z_i+r_{ij})\}^{-b_2}. \qquad \text{(A.118)}$$

Define

$$\delta_n := \min\Big\{\frac{c_1}{2}R_n^{-b_1}, \frac{c_1}{2}t_0^{-b_1}, \frac{1}{2}\Big\}.$$

This implies that $\delta_n \in (0,1)$ and

$$-\frac{\delta_n}{c_2} + \frac{c_1}{c_2}t_0^{-b_1} \ge \frac{c_1}{2c_2}t_0^{-b_1}, \qquad \text{(A.119)}$$

$$\text{and} \quad -\frac{\delta_n}{c_2} + \frac{c_1}{c_2}R_n^{-b_1} \ge \frac{c_1}{2c_2}R_n^{-b_1}. \qquad \text{(A.120)}$$

For an arbitrary $j \in [n]\backslash\{i\}$, either $r_{ij}(1+\rho_{ij}^{-2})^{-1/2} \le t_0$ or $r_{ij}(1+\rho_{ij}^{-2})^{-1/2} > t_0$ holds. In the following we show $f_{ij}(x) \le -\delta_n$ for all $j \in [n]\backslash\{i\}$ under these two mutually exclusive and collectively exhaustive cases.

**Case 1**: Assume that for a fixed $j$ we have

$$r_{ij}(1+\rho_{ij}^{-2})^{-1/2} \leq t_0. \tag{A.121}$$

By the monotonicity of $F_{ji}^c(\cdot)$ we have

$$F_{ji}^c\{r_{ij}(1+\rho_{ij}^{-2})^{-1/2}\} \geq F_{ji}^c(t_0). \tag{A.122}$$

By (3.7) we have

$$F_{ji}^c(t_0) \geq c_1 t_0^{-b_1}. \tag{A.123}$$

Combining (A.122) and (A.123) yields

$$F_{ji}^c\{r_{ij}(1+\rho_{ij}^{-2})^{-1/2}\} \geq c_1 t_0^{-b_1}. \tag{A.124}$$

Combining (4.39), (A.118), and (A.124) gives

$$f_{ij}(x) \leq c_2\{\rho_{ij}(z_i+r_{ij})\}^{-b_2} - c_1 t_0^{-b_1}. \tag{A.125}$$

Equation (A.119) implies

$$\Big(-\frac{\delta_n}{c_2}+\frac{c_1}{c_2}t_0^{-b_1}\Big)^{-1/b_2} \leq \Big(t_0^{-b_1}\frac{c_1}{2c_2}\Big)^{-1/b_2}. \tag{A.126}$$

Noting that $t_0>0$ and $R_n \geq -r_{ij}$, (4.40) implies

$$z_i \geq -r_{ij} + \Big(t_0^{-b_1}\frac{c_1}{2c_2}\Big)^{-1/b_2}\rho_n, \tag{A.127}$$

Combining (A.126) and (A.127) gives

$$\rho_{ij}(z_i+r_{ij}) \geq \Big(-\frac{\delta_n}{c_2}+\frac{c_1}{c_2}t_0^{-b_1}\Big)^{-1/b_2}. \tag{A.128}$$

Therefore, by (A.125) and (A.128) we deduce

$$f_{ij}(x) \leq -\delta_n + c_1 t_0^{-b_1} - c_1 t_0^{-b_1} = -\delta_n.$$

**Case 2**: Assume that for a fixed $j$ we have

$$r_{ij}(1+\rho_{ij}^{-2})^{-1/2} > t_0. \tag{A.129}$$

By (3.7) we have

$$F_{ji}^c\{r_{ij}(1+\rho_{ij}^{-2})^{-1/2}\} \geq c_1\{r_{ij}(1+\rho_{ij}^{-2})^{-1/2}\}^{-b_1}. \tag{A.130}$$

Combining (4.39), (A.118), and (A.130) gives

$$f_{ij}(x) \leq c_2\{\rho_{ij}(z_i+r_{ij})\}^{-b_2} - c_1\{r_{ij}(1+\rho_{ij}^{-2})^{-1/2}\}^{-b_1}. \tag{A.131}$$

Equation (A.120) implies

$$\Big(-\frac{\delta_n}{c_2}+\frac{c_1}{c_2}R_n^{-b_1}\Big)^{-1/b_2} \leq \Big(\frac{c_1}{2c_2}R_n^{-b_1}\Big)^{-1/b_2}. \tag{A.132}$$

Noting that $t_0 > 0$ and $R_n \geq -r_{ij}$, (4.40) implies

$$z_i \geq -r_{ij} + \rho_{ij}^{-1} \left( \frac{c_1}{2c_2} \right)^{-1/b_2} R_n^{b_1/b_2}. \tag{A.133}$$

Combining (A.132) and (A.133) gives

$$\rho_{ij}(z_i + r_{ij}) \geq \left( -\frac{\delta_n}{c_2} + \frac{c_1}{c_2} R_n^{-b_1} \right)^{-1/b_2}. \tag{A.134}$$

Equation (A.134) implies

$$c_2\{\rho_{ij}(z_i + r_{ij})\}^{-b_2} \leq -\delta_n + c_1 R_n^{-b_1}. \tag{A.135}$$

Since $r_{ij} \leq R_n$ and $(1+\rho_{ij}^{-2})^{-1/2} \leq 1$, we have

$$c_1\{r_{ij}(1+\rho_{ij}^{-2})^{-1/2}\}^{-b_1} \geq c_1 R_n^{-b_1}. \tag{A.136}$$

Therefore, by (A.131), (A.135), and (A.136) we deduce

$$f_{ij}(x) \leq -\delta_n + c_1 R_n^{-b_1} - c_1 R_n^{-b_1} = -\delta_n.$$

This completes the proof. $\square$

### A.2.7 Proof of Lemma 4.5

*Proof.* For any $j \in [n] \setminus \{i\}$, we have $\rho_{ij}^{-1} \leq \rho_n$ and $-r_{ij} \leq R_n$. This combined with (4.51) implies that $z_i \geq \rho_{ij}^{-1} t_0 - r_{ij}$, or equivalently

$$\rho_{ij}(z_i + r_{ij}) \geq t_0. \tag{A.137}$$

Equations (A.137) and (3.10) imply that

$$F_j^c\{\rho_{ij}(z_i + r_{ij})\} \leq c_2 \exp[-b_2\{\rho_{ij}(z_i + r_{ij})\}^{\lambda}]. \tag{A.138}$$

Define

$$\delta_n := \min\left\{ \frac{c_1}{2} \exp(-b_1 R_n^{\lambda}), \frac{c_1}{2} \exp(-b_1 t_0^{\lambda}), \frac{1}{2} \right\}.$$

This implies that $\delta_n \in (0,1)$ and that

$$-\frac{\delta_n}{c_2} + \frac{c_1}{c_2} \exp(-b_1 t_0^{\lambda}) \geq \frac{c_1}{2c_2} \exp(-b_1 t_0^{\lambda}) \tag{A.139}$$

$$\text{and} \quad -\frac{\delta_n}{c_2} + \frac{c_1}{c_2} \exp(-b_1 R_n^{\lambda}) \geq \frac{c_1}{2c_2} \exp(-b_1 R_n^{\lambda}). \tag{A.140}$$

In the following we show $f_{ij}(x) \leq -\delta_n$ for all $j \in [n] \setminus \{i\}$ under these two mutually exclusive and collectively exhaustive cases.

**Case 1**: Assume that for a fixed $j$ we have

$$r_{ij}(1+\rho_{ij}^{-2})^{-1/2} \leq t_0. \tag{A.141}$$

By the monotonicity of $F_{ji}^c(\cdot)$ we have

$$F_{ji}^c\{r_{ij}(1+\rho_{ij}^{-2})^{-1/2}\} \geq F_{ji}^c(t_0). \tag{A.142}$$

By (3.11) we have

$$F_{ji}^c(t_0) \geq c_1 \exp(-b_1 t_0^\lambda). \tag{A.143}$$

Combining (A.142) and (A.143) yields

$$F_{ji}^c\{r_{ij}(1+\rho_{ij}^{-2})^{-1/2}\} \geq c_1 \exp(-b_1 t_0^\lambda). \tag{A.144}$$

Combining (4.39), (A.138), and (A.144) gives

$$f_{ij}(x) \leq c_2 \exp[-b_2\{\rho_{ij}(z_i + r_{ij})\}^\lambda] - c_1 \exp(-b_1 t_0^\lambda). \tag{A.145}$$

Equation (A.139) implies

$$-\frac{1}{b_2}\log\{-\frac{\delta_n}{c_2} + \frac{c_1}{c_2}\exp(-b_1 t_0^\lambda)\} \leq -\frac{1}{b_2}\log\frac{c_1}{2c_2} + \frac{b_1}{b_2}t_0^\lambda. \tag{A.146}$$

Noting that $t_0 > 0$ and $R_n \geq -r_{ij}$, (4.51) implies

$$z_i \geq -r_{ij} + \rho_{ij}^{-1}K_3 \geq -r_{ij} + \rho_{ij}^{-1}\Big(-\frac{1}{b_2}\log\frac{c_1}{2c_2} + \frac{b_1}{b_2}t_0^\lambda\Big)^{1/\lambda}. \tag{A.147}$$

Combining (A.146) and (A.147) gives

$$\rho_{ij}(z_i + r_{ij}) \geq \Big[-\frac{1}{b_2}\log\Big\{-\frac{\delta_n}{c_2} + \frac{c_1}{c_2}\exp(-b_1 t_0^\lambda)\Big\}\Big]^{1/\lambda}. \tag{A.148}$$

Therefore, by (A.145) and (A.148) we deduce

$$f_{ij}(x) \leq -\delta_n + c_1\exp(-b_1 t_0^\lambda) - c_1\exp(-b_1 t_0^\lambda) = -\delta_n.$$

**Case 2**: Assume that for a fixed $j$ we have

$$r_{ij}(1+\rho_{ij}^{-2})^{-1/2} > t_0. \tag{A.149}$$

By (3.11) we have

$$F_{ji}^c\{r_{ij}(1+\rho_{ij}^{-2})^{-1/2}\} \geq c_1 \exp\Big[-b_1\Big\{r_{ij}\Big(1+\rho_{ij}^{-2}\Big)^{-1/2}\Big\}^\lambda\Big]. \tag{A.150}$$

Combining (4.39), (A.138), and (A.150) gives

$$f_{ij}(x) \leq c_2\exp[-b_2\{\rho_{ij}(z_i + r_{ij})\}^\lambda] - c_1\exp[-b_1\{r_{ij}(1+\rho_{ij}^{-2})^{-1/2}\}^\lambda]. \tag{A.151}$$

Equation (A.140) implies

$$-\frac{1}{b_2}\log\{-\frac{\delta_n}{c_2} + \frac{c_1}{c_2}\exp(-b_1 R_n^\lambda)\} \leq -\frac{1}{b_2}\log\frac{c_1}{2c_2} + \frac{b_1}{b_2}R_n^\lambda. \tag{A.152}$$

Equation (4.51) implies

$$z_i \geq R_n + \rho_n \xi(\lambda^{-1})\Big\{\Big(-\frac{1}{b_2}\log\frac{c_1}{2c_2}\Big)^{1/\lambda} + \Big(\frac{b_1}{b_2}R_n^\lambda\Big)^{1/\lambda}\Big\}. \tag{A.153}$$

It follows from (A.153) and Lemma A.14 that

$$z_i \geq R_n + \rho_n\Big(-\frac{1}{b_2}\log\frac{c_1}{2c_2} + \frac{b_1}{b_2}R_n^\lambda\Big)^{1/\lambda}. \tag{A.154}$$

Noting that $t_0>0$ and $R_n\geq -r_{ij}$, (A.154) implies

$$z_i\geq -r_{ij}+\rho_{ij}^{-1}\Big(-\frac{1}{b_2}\log\frac{c_1}{2c_2}+\frac{b_1}{b_2}R_n^\lambda\Big)^{1/\lambda}. \tag{A.155}$$

Combining (A.152) and (A.155) gives

$$\rho_{ij}(z_i+r_{ij})\geq\Big[-\frac{1}{b_2}\log\{-\frac{\delta_n}{c_2}+\frac{c_1}{c_2}\exp(-b_1R_n^\lambda)\}\Big]^{1/\lambda}. \tag{A.156}$$

Equation (A.156) implies

$$c_2\exp[-b_2\{\rho_{ij}(z_i+r_{ij})\}^\lambda]\leq -\delta_n+c_1\exp(-b_1R_n^\lambda). \tag{A.157}$$

Since $r_{ij}\leq R_n$ and $(1+\rho_{ij}^{-2})^{-1/2}\leq 1$, we have

$$c_1\exp\Big[-b_1\{r_{ij}(1+\rho_{ij}^{-2})^{-1/2}\}^\lambda\Big]\geq c_1\exp(-b_1R_n^\lambda). \tag{A.158}$$

Therefore, by (A.151), (A.157), and (A.158) we deduce

$$f_{ij}(x)\leq -\delta_n+c_1\exp(-b_1R_n^\lambda)-c_1\exp(-b_1R_n^\lambda)=-\delta_n.$$

This completes the proof. □

### A.2.8 Proof of Lemma A.3

*Proof.* Consider an arbitrary vector $(l_1,\dots,l_n)$, with each $l_i\in[n]$. Define the sign function $\mathrm{sgn}(x):=\mathbb{1}(x>0)-\mathbb{1}(x<0)$. It follows from (3.1) that

$$\begin{aligned}\sum_{i=1}^n E\{h_{1,i}^{\mathrm{Ken}}(X_{l_i})^2\}&=\frac{1}{(n-1)^2}\sum_{i=1}^n E\Big[\sum_{k=1}^n\mathrm{sgn}(i-k)\{P(X_k>X_{l_i}\mid X_{l_i})-P(X_k>X_i)\}\Big]^2\\&=T_1-2T_2+T_3,\end{aligned} \tag{A.159}$$

where

$$T_1=\frac{1}{(n-1)^2}\sum_{i=1}^n\sum_{k_1,k_2=1}^n\mathrm{sgn}(i-k_1)\mathrm{sgn}(i-k_2)E[P(X_{k_1}>X_{l_i}\mid X_{l_i})P(X_{k_2}>X_{l_i}\mid X_{l_i})], \tag{A.160}$$

$$T_2=\frac{1}{(n-1)^2}\sum_{i=1}^n\sum_{k_1,k_2=1}^n\mathrm{sgn}(i-k_1)\mathrm{sgn}(i-k_2)P(X_{k_1}>X_j)P(X_{k_2}>X_i), \tag{A.161}$$

$$T_3=\frac{1}{(n-1)^2}\sum_{i=1}^n\sum_{k_1,k_2=1}^n\mathrm{sgn}(i-k_1)\mathrm{sgn}(i-k_2)P(X_{k_1}>X_i)P(X_{k_2}>X_i). \tag{A.162}$$

We have

$$\sum_{i=1}^n\sum_{k_1,k_2=1}^n\mathrm{sgn}(i-k_1)\mathrm{sgn}(i-k_2)=\sum_{i=1}^n(2i-n-1)^2=\frac{1}{3}n(n-1)(n+1). \tag{A.163}$$

It follows from (3.20), (A.160), and (A.163) that

$$T_1=\frac{n(n-1)(n+1)}{3(n-1)^2}\{\eta^2+O(n^{-1/3})\}=\frac{n(n+1)}{3(n-1)}\eta^2+O(n^{2/3}). \tag{A.164}$$

It follows from (3.19), (A.161), (A.162), and (A.163) that

$$T_2 = \frac{n(n-1)(n+1)}{3(n-1)^2}\{\theta + O(n^{-1/6})\}^2 = \frac{n(n+1)}{3(n-1)}\theta^2 + O(n^{5/6}), \tag{A.165}$$

$$T_3 = \frac{n(n-1)(n+1)}{3(n-1)^2}\{\theta + O(n^{-1/6})\}^2 = \frac{n(n+1)}{3(n-1)}\theta^2 + O(n^{5/6}). \tag{A.166}$$

Combining (A.159) with (A.164), (A.165), and (A.166) yields

$$\sum_{i=1}^{n} E\left\{h_{1,i}^{\mathrm{Ken}}(X_{l_i})^2\right\} = T_1 - 2T_2 + T_3 = \frac{n(n+1)}{3(n-1)}(\eta^2 - \theta^2) + O(n^{5/6}). \tag{A.167}$$

In (A.167), letting $(l_1,\ldots,l_n) = (j,j,\ldots,j)$ yields (A.40), and letting $(l_1,\ldots,l_n) = (j,j,\ldots,j)$ yields (A.41).

This completes the proof. □

### A.2.9 Proof of Lemma 4.6

*Proof.* Consider an arbitrary vector $(l_1,\ldots,l_n)$, with each $l_i \in [n]$. It follows from (3.2) that

$$\begin{aligned}\sum_{i=1}^{n} E\{h_{1,i}^{\mathrm{AP}}(X_{l_i})^2\} &= \frac{1}{(n-1)^2}\sum_{i=1}^{n} E\Big[\sum_{k=1}^{n}\Big\{\frac{n\mathbb{1}(k<i)}{i-1} - \frac{n\mathbb{1}(k>i)}{k-1}\Big\}\{P(X_k > X_{l_i} \mid X_{l_i}) - P(X_k > X_i)\}\Big]^2 \\ &= T_1 - 2T_2 + T_3,\end{aligned} \tag{A.168}$$

where

$$T_1 = \frac{n^2}{(n-1)^2}\sum_{i=1}^{n}\sum_{k_1,k_2=1}^{n}\gamma(i,k_1,k_2)E[P(X_{k_1} > X_{l_i} \mid X_{l_i})P(X_{k_2} > X_{l_i} \mid X_{l_i})], \tag{A.169}$$

$$T_2 = \frac{n^2}{(n-1)^2}\sum_{i=1}^{n}\sum_{k_1,k_2=1}^{n}\gamma(i,k_1,k_2)P(X_{k_1} > X_j)P(X_{k_2} > X_i), \tag{A.170}$$

$$T_3 = \frac{n^2}{(n-1)^2}\sum_{i=1}^{n}\sum_{k_1,k_2=1}^{n}\gamma(i,k_1,k_2)P(X_{k_1} > X_i)P(X_{k_2} > X_i), \tag{A.171}$$

and

$$\gamma(i,k_1,k_2) := \Big\{\frac{\mathbb{1}(k_1<i)}{i-1} - \frac{\mathbb{1}(k_1>i)}{k_1-1}\Big\}\Big\{\frac{\mathbb{1}(k_2<i)}{i-1} - \frac{\mathbb{1}(k_2>i)}{k_2-1}\Big\}.$$

By Lemma A.15 and Lemma A.13 we have

$$\sum_{i=1}^{n}\sum_{k_1,k_2=1}^{n}\gamma(i,k_1,k_2) = (n-1) + \varphi(n-1) = (n-1) + O(\log n). \tag{A.172}$$

It follows from (3.22), (A.169), and (A.172) that

$$T_1 = \frac{n^2\{(n-1) + O(\log n)\}}{(n-1)^2}\{\eta^2 + O(n^{-1/3}(\log n)^2)\} = \frac{n^2}{n-1}\eta^2 + O\{n^{2/3}(\log n)^2\}. \tag{A.173}$$

It follows from (3.21), (A.170), (A.171), and (A.172) that

$$T_2 = \frac{n^2\{(n-1)+O(\log n)\}}{(n-1)^2}\{\theta+O(n^{-1/6}\log n)\}^2 = \frac{n^2}{n-1}\theta^2 + O(n^{5/6}\log n), \tag{A.174}$$

$$T_3 = \frac{n^2\{(n-1)+O(\log n)\}}{(n-1)^2}\{\theta+O(n^{-1/6}\log n)\}^2 = \frac{n^2}{n-1}\theta^2 + O(n^{5/6}\log n). \tag{A.175}$$

Combining (A.168) with (A.173), (A.174), and (A.175) yields

$$\sum_{i=1}^{n} E\{h_{1,i}^{\mathrm{AP}}(X_{l_i})^2\} = T_1 - 2T_2 + T_3 = \frac{n^2}{n-1}(\eta^2-\theta^2) + O(n^{5/6}\log n). \tag{A.176}$$

In (A.176), letting $(l_1,\dots,l_n)=(j,j,\dots,j)$ yields (A.40), and letting $(l_1,\dots,l_n)=(j,j,\dots,j)$ yields (A.41). This completes the proof. □

### A.2.10 Proof of Lemma A.4

*Proof.* For a fixed $(\boldsymbol{i},\boldsymbol{j}) \in (I_n^m)^{\otimes 2}_{=1}$, consider any $(\boldsymbol{r},\boldsymbol{s}) \in (I_n^m)^{\otimes 2}_{=1}$ with $r_p = s_q$ and $\boldsymbol{i}\cap\boldsymbol{r} = \emptyset = \boldsymbol{j}\cap\boldsymbol{s}$. By the law of iterated expectation and the independence of $X_i$'s we have

$$E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\} = E[E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}}) \mid X_{r_p}\}E\{h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}}) \mid X_{s_q}\}]. \tag{A.177}$$

For $\boldsymbol{i} = (i_1,\dots,i_m)$ and $l \in [m]$, define

$$\boldsymbol{i}\backslash i_l := (i_1,\dots,i_{l-1},i_{l+1},\dots,i_m).$$

Using the definition of $h_{2;\boldsymbol{i}}(\cdot)$ in (2.5) we have

$$\begin{aligned} E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}}) \mid X_{r_p}\} &= E\{h(X_{\boldsymbol{r}}) \mid X_{r_p}\} \\ &\quad - \sum_{l=1}^{m} E[E_{\boldsymbol{i}\backslash i_l}\{h^{(l)}(X_{r_l};Y_1,\dots,Y_{m-1}) \mid X_{r_l}\} \mid X_{r_p}] + (m-1)\theta(\boldsymbol{i}). \end{aligned} \tag{A.178}$$

By the independence of the $X_i$'s we have

$$\begin{aligned} &\sum_{l=1}^{m} E[E_{\boldsymbol{i}\backslash i_l}\{h^{(l)}(X_{r_l};Y_1,\dots,Y_{m-1}) \mid X_{r_l}\} \mid X_{r_p}]. \\ &\qquad = \sum_{\substack{l=1 \\ l\neq p}}^{m} \theta^{(l)}(r_l;\boldsymbol{i}\backslash i_l) + E_{\boldsymbol{i}\backslash i_p}\{h^{(l)}(X_{r_p};Y_1,\dots,Y_{m-1}) \mid X_{r_p}\} \end{aligned} \tag{A.179}$$

Using (A.178) and (A.179) we obtain

$$\begin{aligned} E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}}) \mid X_{r_p}\} &= E\{h(X_{\boldsymbol{r}}) \mid X_{r_p}\} - \sum_{\substack{l=1 \\ l\neq p}}^{m} \theta^{(l)}(r_l;\boldsymbol{i}\backslash i_l) \\ &\qquad - E_{\boldsymbol{i}\backslash i_p}\{h^{(l)}(X_{r_p};Y_1,\dots,Y_{m-1}) \mid X_{r_p}\} + (m-1)\theta(\boldsymbol{i}). \end{aligned} \tag{A.180}$$

We introduce some notation:

$$
\begin{aligned}
&\boldsymbol{i}\backslash i_l \oplus k := (i_1,\dots,i_{l-1},k,i_{l+1},\dots,i_m),\\
&X_{\boldsymbol{i}\backslash i_l\oplus k} := (X_{i_1},\dots,X_{i_{l-1}},X_k,X_{i_{l+1}},\dots,X_{i_m}),\\
&\theta(\boldsymbol{i}\,|\,i_l) := E\{h(X_{\boldsymbol{i}})\,|\,X_{i_l}\}.
\end{aligned}
$$

Using the new notation, (A.180) becomes

$$
E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})\,|\,X_{r_p}\} = \theta(\boldsymbol{r}\,|\,r_p) - \sum_{\substack{l=1\\ l\neq p}}^{m}\theta(\boldsymbol{i}\backslash i_l\oplus r_l) - \theta(\boldsymbol{i}\backslash i_p\oplus r_p\,|\,r_p) + (m-1)\theta(\boldsymbol{i}). \tag{A.181}
$$

Similarly, we have

$$
E\{h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\,|\,X_{s_q}\} = \theta(\boldsymbol{s}\,|\,s_q) - \sum_{\substack{l=1\\ l\neq q}}^{m}\theta(\boldsymbol{j}\backslash j_l\oplus s_l) - \theta(\boldsymbol{j}\backslash j_q\oplus s_q\,|\,s_q) + (m-1)\theta(\boldsymbol{j}). \tag{A.182}
$$

By algebra and the law of iterated expectation, we derive from (A.181) and (A.182) that

$$
E[E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})\,|\,X_{r_p}\}E\{h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\,|\,X_{s_q}\}] = T_1+T_2+T_3+T_4+T_5, \tag{A.183}
$$

where

$$
\begin{aligned}
T_1 &= E\{\theta(\boldsymbol{r}\,|\,r_p)\theta(\boldsymbol{s}\,|\,s_q)\} - E\{\theta(\boldsymbol{r}\,|\,r_p)\theta(\boldsymbol{j}\backslash j_q\oplus s_q\,|\,s_q)\}\\
&\quad - E\{\theta(\boldsymbol{i}\backslash i_p\oplus r_p\,|\,r_p)\theta(\boldsymbol{s}\,|\,s_q)\} + E\{\theta(\boldsymbol{i}\backslash i_p\oplus r_p\,|\,r_p)\theta(\boldsymbol{j}\backslash j_q\oplus s_q\,|\,s_q)\},\\
T_2 &= (m-1)\theta(\boldsymbol{r})\theta(\boldsymbol{j}) - \theta(\boldsymbol{r})\sum_{l\neq q}\theta(\boldsymbol{j}\backslash j_l\oplus s_l) + (m-1)\theta(\boldsymbol{i})\theta(\boldsymbol{s}) - \theta(\boldsymbol{s})\sum_{l\neq p}\theta(\boldsymbol{i}\backslash i_l\oplus r_l),\\
T_3 &= \Big\{\sum_{l=1}^{m}\theta(\boldsymbol{i}\backslash i_l\oplus r_l) - m\theta(\boldsymbol{i})\Big\}\Big\{\sum_{l=1}^{m}\theta(\boldsymbol{j}\backslash j_l\oplus s_l) - m\theta(\boldsymbol{j})\Big\},\\
T_4 &= \theta(\boldsymbol{i})\sum_{l=1}^{m}\theta(\boldsymbol{j}\backslash j_l\oplus s_l) + \theta(\boldsymbol{j})\sum_{l=1}^{m}\theta(\boldsymbol{i}\backslash i_l\oplus r_l) - 2m\theta(\boldsymbol{i})\theta(\boldsymbol{j}),\\
T_5 &= \theta(\boldsymbol{i})\theta(\boldsymbol{j}) - \theta(\boldsymbol{i}\backslash i_p\oplus r_p)\theta(\boldsymbol{j}\backslash j_q\oplus s_q).
\end{aligned}
$$

By the definitions of $M_1(n)$ and $M_2(n)$ in (2.25) and (2.26), we have $|T_1|\le 2M_2(n)$, $|T_2|\le CM_1(n)$, $|T_3|\le CM_1(n)^2$, $|T_4|\le CM_1(n)$, and $|T_5|\le CM_1(n)$. Therefore, it follows from (A.183) that

$$
|E[E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})\,|\,X_{r_p}\}E\{h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\,|\,X_{s_q}\}]| \le C\{M_1(n)^2 + M_2(n)\}.
$$

This yields (A.77). The proof is thus finished. □

### A.2.11 Proof of Lemma A.5

*Proof.* For a fixed $(\boldsymbol{i},\boldsymbol{j})\in(I_n^m)^{\otimes 2}_{=1}$, consider any $(\boldsymbol{r},\boldsymbol{s})\in(I_n^m)^{\otimes 2}_{=0}$ such that $\boldsymbol{i}\cap\boldsymbol{r}=\emptyset=\boldsymbol{j}\cap\boldsymbol{s}$. By independence of the $X_i$'s we have

$$
E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\} = E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})\}E\{h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\}. \tag{A.184}
$$

By the definition of $h_{2;\boldsymbol{i}}(\cdot)$ in (2.5), we have

$$E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})\} = E\{h(X_{\boldsymbol{r}})\} - \sum_{l=1}^{m} E[E_{\boldsymbol{i}\setminus i_l}\{h^{(l)}(X_{r_l};Y_1,\ldots,Y_{m-1}) \mid X_{r_l}\}] + (m-1)\theta(\boldsymbol{i})$$
$$= \theta(\boldsymbol{r}) - \sum_{l=1}^{m} \theta^{(l)}(r_l;\boldsymbol{i}\setminus i_l) + (m-1)\theta(\boldsymbol{i}).$$

It then follows from the definition of $M_1(n)$ in (2.25) that

$$|E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}})\}| \le m M_1(n). \tag{A.185}$$

Combining (A.184) and (A.185) yields that

$$|E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{r}}) h_{2;\boldsymbol{j}}(X_{\boldsymbol{s}})\}| \le m^2 M_1(n).$$

This implies (A.80). The proof is thus finished. □

### A.3 Auxiliary lemmas

**Lemma A.6.** There exists a constant $c_m$ which only depends on $m$, such that the following results hold.

(i) For any $n$ and any $(i_1,\ldots,i_m) \in I_n^m$,
$$E\{h_{2;i_1,\ldots,i_m}(X_{i_1},\ldots,X_{i_m})^2\} \le c_m E\{h(X_{i_1},\ldots,X_{i_m})^2\}.$$

(ii) For any $n$, any $i \in [n]$, any $(i_1,\ldots,i_{m-1}) \in I_{n-1}^{m-1}(-i)$, and any $l \in [m]$,
$$E[\{f^{(l)}_{i_1,\ldots,i_{m-1}}(X_i) - \theta^{(l)}(i;i_1,\ldots,i_{m-1})\}^4] \le c_m E\{h^{(l)}(X_i;X_{i_1},\ldots,X_{i_{m-1}})^4\}.$$

(iii) For any $n$, any $(i_1,\ldots,i_m) \in I_n^m$, and any $j_1,\ldots,j_m \in [n]$,
$$E\{h_{2;i_1,\ldots,i_m}(X_{j_1},\ldots,X_{j_m})^2\} \le c_m \sup_{1\le k_1,\ldots,k_m \le n} E\{h(X_{k_1},\ldots,X_{k_m})^2\}.$$

**Lemma A.7.** Consider three random variables $X,Y,Z$. Assume $Y$ is independent of $Z$ conditional on $X$. Then for two measurable functions $f,g:\mathbb{R}^2 \to \mathbb{R}$, we have

$$\mathrm{Cov}\{f(X,Y),g(X,Z)\} = \mathrm{Cov}\Big[E\{f(X,Y) \mid X\}, E\{g(X,Z) \mid X\}\Big].$$

**Lemma A.8.** Consider a sequence of random variables $X_1,X_2,\ldots,$ with $E(X_n)=0$ for all $X_n$. If $\mathrm{Var}(X_n) \to 0$, then $X_n \xrightarrow{P} 0$.

**Lemma A.9** (Lyapunov's central limit theorem)**.** Let $X_1,X_2,\ldots$ be a sequence of independent random variables and let $S_n = n^{-1}\sum_{i=1}^n X_i$. If there exists $\delta > 0$ such that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n E|X_i - E(X_i)|^{2+\delta}}{\left\{\sum_{i=1}^n E|X_i - E(X_i)|^2\right\}^{\frac{2+\delta}{2}}} = 0, \tag{A.186}$$

then

$$\mathrm{Var}(S_n)^{-1/2}\{S_n - E(S_n)\} \xrightarrow{d} N(0,1).$$

**Lemma A.10** (Lehmann, 1999, Theorem 2.6.1)**.** If a sequence of cumulative distribution functions $H_n$ tends to a continuous cdf $H$, then $H_n(x)$ converges to $H(x)$ uniformly in $x$.

**Lemma A.11** (Mammen, 2012, Theorem 2.2)**.** Consider a sequence $Y_{n,1},\dots,Y_{n,n}$ of independent random variables with distribution $P_{n,i}$. For a function $g_n$ define $\widehat{T}_n = n^{-1}\sum_{i=1}^n g_n(Y_{n,i})$. Consider a bootstrap sample $Y^*_{n,1},\dots,Y^*_{n,n}$ and define $\widehat{T}^*_n = n^{-1}\sum_{i=1}^n g_n(Y^*_{n,i})$. Then for every sequence $t_n$ the following assertions are equivalent:

(i) There exists $\sigma_n$ such that for every $\epsilon > 0$

$$\sup_{1\le i\le n} P\Big\{\Big|\frac{g_n(Y_{n,i})-t_n}{n\sigma_n}\Big| \ge \epsilon\Big\} \to 0, \tag{A.187}$$

$$\sum_{i=1}^n \Big(E\Big[\frac{g_n(Y_{n,i})-t_n}{n\sigma_n} 1\Big\{\Big|\frac{g_n(Y_{n,i})-t_n}{n\sigma_n}\Big| \le \epsilon\Big\}\Big]\Big)^2 \to 0, \tag{A.188}$$

$$\sup_{t\in\mathbb{R}} |P(\widehat{T}_n - t_n \le t) - \Phi(t)| \to 0. \tag{A.189}$$

(ii) Bootstrap works:

$$\sup_{t\in\mathbb{R}} |P(\widehat{T}^*_n - \widehat{T}_n \le t \mid Y_{n,1},\dots,Y_{n,n}) - P(\widehat{T}_n - t_n \le t)| \xrightarrow{P} 0.$$

**Lemma A.12** (Serfling, 2009, Theorem 1.8 C)**.** Let $X_1, X_2, \dots$ be uncorrelated with means $\mu_1, \mu_2, \dots$ and variances $\sigma_1^2, \sigma_2^2, \dots$. If $\sum_{i=1}^n \sigma_i^2 = o(n^{-2})$, $n \to \infty$, then

$$\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mu_i \xrightarrow{P} 0.$$

**Lemma A.13** (Bound on the partial sum of harmonic series)**.** Denote $\varphi(n) = \sum_{k=1}^n k^{-1}$. Then for any two integers $m, n$ such that $1 \le m \le n$,

$$\log\frac{n+1}{m+1} \le \varphi(n) - \varphi(m) \le \log\frac{n}{m}, \tag{A.190}$$

$$\log(n+1) \le \varphi(n) \le 1 + \log n. \tag{A.191}$$

**Lemma A.14.** For any two positive real numbers $a, b$ and real number $p > 0$, we have

$$(a+b)^p \le \xi(p)(a^p + b^p),$$

where

$$\xi(p) = \begin{cases} 2^{p-1} & \text{if } p \ge 1, \\ 1 & \text{if } 0 < p < 1. \end{cases}$$

**Lemma A.15.** We have

$$\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n \Big\{\frac{\mathbb{1}(j<i)}{i-1} - \frac{\mathbb{1}(j>i)}{j-1}\Big\}\Big\{\frac{\mathbb{1}(k<i)}{i-1} - \frac{\mathbb{1}(k>i)}{k-1}\Big\} = (n-1) + \varphi(n-1), \tag{A.192}$$

where we define $0/0 := 0$ and $\varphi(n) := \sum_{k=1}^n k^{-1}$.

**Lemma A.16.** Define $\Phi^c(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-\frac{t^2}{2}) dt$ to be the complement distribution function for the standard Gaussian. We have the following bounds for $\Phi^c(x)$:

$$\begin{aligned} \frac{1}{\sqrt{2\pi}}\left(\frac{1}{x} - \frac{1}{x^3}\right)\exp(-\frac{x^2}{2}) &\leq \Phi^c(x) \leq \quad \frac{1}{\sqrt{2\pi}}\frac{1}{x}\exp(-\frac{x^2}{2}), &&\text{if } x > 0, \\ 1 + \frac{1}{\sqrt{2\pi}}\frac{1}{x}\exp(-\frac{x^2}{2}) &\leq \Phi^c(x) \leq \quad 1 + \frac{1}{\sqrt{2\pi}}\left(\frac{1}{x} - \frac{1}{x^3}\right)\exp(-\frac{x^2}{2}), &&\text{if } x < 0. \end{aligned}$$

## A.4 Proof of auxiliary lemmas

### A.4.1 Proof of Lemma A.6

*Proof.* Define $\boldsymbol{i} = (i_1, \ldots, i_m)$, $X_{\boldsymbol{i}} = (X_{i_1}, \ldots, X_{i_m})$, and $\boldsymbol{i}_{-m} = (i_1, \ldots, i_{m-1})$.

(i) By the definition of $h_{2;\boldsymbol{i}}(\cdot)$ in (2.5) we have

$$E\{h_{2;}(X_{\boldsymbol{i}})^2\} \leq 2^{m+2}\Big[E\{h(X_{\boldsymbol{i}})^2\} + \sum_{l=1}^m E\{f^{(l)}_{\boldsymbol{i}\setminus i_l}(X_{i_l})^2\} + (m-1)^2\theta^2(\boldsymbol{i})\Big]. \tag{A.193}$$

Jensen's inequality and the law of iterated expectation yield

$$E\{f^{(l)}_{\boldsymbol{i}\setminus i_l}(X_{i_l})^2\} = E_{i_l}[E_{\boldsymbol{i}\setminus i_l}\{h^{(l)}(X_{i_l}; Y_1, \ldots Y_{m-1}) \mid X_{i_l}\}^2] \leq E\{h(X_{\boldsymbol{i}})^2\} \tag{A.194}$$

and

$$\theta^2(\boldsymbol{i}) \leq E\{h(X_{\boldsymbol{i}})^2\}. \tag{A.195}$$

Equations (A.193), (A.194), and (A.195) imply

$$E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{i}})^2\} \leq 2^{m+2}\{1 + m + (m-1)^2\}E\{h(X_{\boldsymbol{i}})^2\}.$$

This proves (i).

(ii) We have

$$E[\{f^{(l)}_{\boldsymbol{i}_{-m}}(X_i) - \theta^{(l)}(i; \boldsymbol{i}_{-m})\}^4] \leq 2^4[E\{f^{(l)}_{\boldsymbol{i}_{-m}}(X_i)^4\} + \theta^{(l)}(i; \boldsymbol{i}_{-m})^4] \tag{A.196}$$

By the definition of $f^{(l)}_{\boldsymbol{i}_{-m}}(\cdot)$ in (2.3) and Jensen's inequality we have

$$E\{f^{(l)}_{\boldsymbol{i}_{-m}}(X_i)^4\} \leq E_i[E_{\boldsymbol{i}_{-m}}\{h^{(l)}(X_i; Y_1, \ldots Y_{m-1})^4 \mid X_i\}] = E\{h^{(l)}(X_i; X_{\boldsymbol{i}_{-m}})^4\}. \tag{A.197}$$

Jensen's inequality also implies that

$$\theta^{(l)}(i; \boldsymbol{i}_{-m})^4 = \{Eh^{(l)}(X_i; X_{\boldsymbol{i}_{-m}})\}^4 \leq E\{h^{(l)}(X_i; X_{\boldsymbol{i}_{-m}})^4\}. \tag{A.198}$$

Combining (A.196) with (A.197) and (A.198) yields

$$E[\{f^{(l)}_{\boldsymbol{i}_{-m}}(X_i) - \theta^{(l)}(i; \boldsymbol{i}_{-m})\}^4] \leq 2^4 E\{h^{(l)}(X_i; X_{\boldsymbol{i}_{-m}})^4\}.$$

This proves (ii).

(iii) Consider $\boldsymbol{j} := (j_1, \ldots, j_m)$ with each $j_l \in [m]$. Define $X_{\boldsymbol{j}} := (X_{j_1}, \ldots, X_{j_m})$. By the definition of

$h_{2;\boldsymbol{i}}(\cdot)$ in (2.5) we have

$$E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{j}})^2\} \le 2^{m+2}\Big[E\{h(X_{\boldsymbol{j}})^2\} + \sum_{l=1}^{m} E\{f^{(l)}_{\boldsymbol{i}\setminus i_l}(X_{j_l})^2\} + (m-1)^2\theta^2(\boldsymbol{i})\Big]. \tag{A.199}$$

By the definition of $f^{(l)}_{\boldsymbol{i}_{-m}}(\cdot)$ in (2.3) and Jensen's inequality we have

$$E\{f^{(l)}_{\boldsymbol{i}_{-m}}(X_{j_l})^2\} \le E_{j_l}[E_{\boldsymbol{i}_{-m}}\{h^{(l)}(X_{j_l};Y_1,\ldots Y_{m-1})^2 \mid X_{j_l}\}] = E\{h^{(l)}(X_{j_l};X_{\boldsymbol{i}_{-m}})^2\}. \tag{A.200}$$

Combining (A.199), (A.200), and (A.195) yields

$$\begin{aligned} E\{h_{2;\boldsymbol{i}}(X_{\boldsymbol{j}})^2\} &\le 2^{m+2}\Big[E\{h(X_{\boldsymbol{j}})^2\} + \sum_{l=1}^{m} E\{h^{(l)}(X_{j_l};X_{\boldsymbol{i}_{-m}})^2\} + (m-1)^2 E\{h(X_{\boldsymbol{i}})^2\}\Big] \\ &\le 2^{m+2}m^2 \sup_{1\le k_1,\ldots,k_m\le n} E\{h(X_{k_1},\ldots,X_{k_m})^2\}. \end{aligned}$$

This proves (iii).

The proof is thus finished. □

### A.4.2 Proof of Lemma A.7

*Proof.* Define $\theta(f) = E\{f(X,Y)\}$ and $\theta(g) = E\{g(X,Z)\}$. By the law of iterated expectation we have

$$\begin{aligned} \mathrm{Cov}\{f(X,Y),g(X,Z)\} &= E[\{f(X,Y)-\theta(f)\}\{g(X,Z)-\theta(g)\}] \\ &= E(E[\{f(X,Y)-\theta(f)\}\{g(X,Z)-\theta(g)\} \mid X]) \end{aligned} \tag{A.201}$$

By independence between $Y$ and $Z$ conditional on $X$, we have

$$\begin{aligned} E(E[\{f(X,Y)-\theta(f)\}\{g(X,Z)-\theta(g)\} \mid X]) &= E([E\{f(X,Y) \mid X\}-\theta(f)][E\{g(X,Z) \mid X\}-\theta(g)]) \\ &= \mathrm{Cov}[E\{f(X,Y) \mid X\}, E\{g(X,Z) \mid X\}]. \end{aligned} \tag{A.202}$$

Lemma A.7 follows from (A.201) and (A.202). □

### A.4.3 Proof of Lemma A.13

*Proof.* We have $\varphi(n) - \varphi(m) = \sum_{k=m+1}^{n} k^{-1}$. By integral bound, we have

$$\log\frac{n+1}{m+1} = \int_{m+1}^{n+1} \frac{1}{x}dx \le \sum_{k=m+1}^{n} \frac{1}{k} \le \int_{m}^{n} \frac{1}{x}dx = \log\frac{n}{m},$$

which yields (A.190). We also have

$$\log(n+1) \le \int_{1}^{n+1} \frac{1}{x}dx \le \sum_{k=1}^{n} \frac{1}{k} \le 1 + \int_{1}^{n} \frac{1}{x}dx \le 1 + \log n,$$

which yields (A.191). The proof is thus finished. □

### A.4.4 Proof of Lemma A.15

*Proof.* By algebra we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left\{\frac{\mathbb{1}(j<i)}{i-1}-\frac{\mathbb{1}(j>i)}{j-1}\right\}\left\{\frac{\mathbb{1}(k<i)}{i-1}-\frac{\mathbb{1}(k>i)}{k_2-1}\right\}=T_1-T_2-T_3+T_4, \tag{A.203}$$

where

$$T_1=\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\mathbb{1}(j<i)}{i-1}\cdot\frac{\mathbb{1}(k<i)}{i-1},\quad T_2=\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\mathbb{1}(j<i)}{i-1}\cdot\frac{\mathbb{1}(k>i)}{k-1},$$
$$T_3=\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\mathbb{1}(j>i)}{j-1}\cdot\frac{\mathbb{1}(k<i)}{i-1},\quad T_4=\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\mathbb{1}(j>i)}{j-1}\cdot\frac{\mathbb{1}(k>i)}{k-1}.$$

For $T_1$ we have

$$T_1=\sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{i-1}\frac{1}{(i-1)^2}=n-1. \tag{A.204}$$

For $T_2$ we have

$$T_2=\sum_{i=2}^{n-1}\sum_{j=1}^{i-1}\sum_{k=i+1}^{n}\frac{1}{i-1}\cdot\frac{1}{k-1}=\sum_{k=3}^{n}\sum_{i=2}^{k-1}\frac{1}{k-1}=\sum_{k=3}^{n}\left(1-\frac{1}{k-1}\right)=(n-1)-\varphi(n-1). \tag{A.205}$$

By symmetry $T_2=T_3$, so

$$T_3=(n-1)-\varphi(n-1). \tag{A.206}$$

For $T_4$ we have

$$T_4=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=i+1}^{n}\frac{1}{j-1}\cdot\frac{1}{k-1}=\sum_{j=2}^{n}\sum_{k=j+1}^{n}\sum_{i=1}^{j-1}\frac{1}{j-1}\cdot\frac{1}{k-1}+\sum_{j=2}^{n}\sum_{k=2}^{j}\sum_{i=1}^{k-1}\frac{1}{j-1}\cdot\frac{1}{k-1}. \tag{A.207}$$

Note that

$$\sum_{j=2}^{n}\sum_{k=j+1}^{n}\sum_{i=1}^{j-1}\frac{1}{j-1}\cdot\frac{1}{k-1}=\sum_{j=2}^{n}\sum_{k=j+1}^{n}\frac{1}{k-1}=\sum_{k=3}^{n}\sum_{j=2}^{k-1}\frac{1}{k-1}=(n-1)-\varphi(n-1) \tag{A.208}$$

and
$$\sum_{j=2}^{n}\sum_{k=2}^{j}\sum_{i=1}^{k-1}\frac{1}{j-1}\cdot\frac{1}{k-1}=\sum_{j=2}^{n}\sum_{k=2}^{j}\frac{1}{j-1}=\sum_{j=2}^{n}1=n-1. \tag{A.209}$$

Combining (A.207) with (A.208) and (A.209) yields

$$T_4=2(n-1)-\varphi(n-1). \tag{A.210}$$

Equation (A.192) follows from (A.203), (A.204), (A.205), and (A.210).

This completes the proof. □

## References


Lee, J. (1990). *U-Statistics: Theory and Practice.* CRC Press.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer Science and Business Media.

Mammen, E. (2012). *When Does Bootstrap Work? Asymptotic Results and Simulations*, volume 77. Springer.

Serfling, R. J. (2009). *Approximation Theorems of Mathematical Statistics*, volume 162. John Wiley and Sons.