Kernel-Based Adaptive Randomization Toward Balance in Continuous and Discrete Covariates

Fei Jiang
Department of Statistics and Actuarial Science, University of Hong Kong
Yanyuan Ma
Department of Statistics, Pennsylvania State University
Guosheng Yin
Department of Statistics and Actuarial Science, University of Hong Kong

May 2, 2017

Abstract

Covariate balance among different treatment arms is critical in clinical trials, as confounding effects can be effectively eliminated when patients in different arms are alike. To balance the prognostic factors across different arms, we propose a new dynamic scheme for patient allocation. Our approach does not require discretizing continuous covariates to multiple categories, and can handle both continuous and discrete covariates naturally. This is achieved through devising a statistical measure to characterize the similarity between a new patient and all the existing patients in the trial. Under the similarity weighting scheme, we develop a covariate-adaptive biased coin design and establish its theoretical properties, as well as improving the original Pocock–Simon design. We conduct extensive simulation studies to examine the design operating characteristics and illustrate our method with a real data example. The new approach is demonstrated to be superior to other existing methods in terms of performance.

Keywords: Biased coin design; Clinical trial; Covariate–adaptive randomization; Covariate balance; Pocock and Simon design; Similarity measure; Stratification.

1 Introduction

Peter Hall was one of the most influential and prolific researchers in modern statistics. His contributions are broad and cover many important areas. From interactions with him, the authors have been greatly influenced by his statistical thinking, especially in how to use "smoothing" methods to increase modeling flexibility and reduce estimation error. One of the nonparametric devices, called kernel smoothing, is widely used in density estimation and nonparametric regression. In density estimation, Hall (1981) derived the law of the iterated logarithm for the kernel estimator, discussed the choice on the order of kernels (Hall & Marron 1988), and addressed the issues on constructing confidence intervals (Hall 1992). In nonparametric regression, Hall (1984) investigated the asymptotic properties of the kernel regression estimator. A series of his follow-up works focused on the confidence intervals and confidence bands for kernel estimators, which include Hall & Marron (1988), Hall (1992), Hall (1993), and Hall & Horowitz (2013). Motivated by kernel estimation, we propose a kernel-based covariate-adaptive randomization design. In addition, we apply the martingale convergence theorem in Hall & Heyde (1980) extensively in deriving the asymptotic properties of the proposed design, which reinforces Peter Hall's impact, especially in the area of sequential analysis.

The primary goals of randomized clinical trials are to differentiate the treatment effects efficiently as well as to treat patients effectively. If the treatment effects of different drugs can be quickly discriminated, then patients outside of the trial would benefit from the more effective therapy sooner. To achieve this goal, allocation of patients is random to balance out both known and unknown prognostic factors that may affect the response of interest, and the numbers of patients should also be balanced across different treatment arms to achieve high statistical power. For discrete covariates, various approaches have been developed for patient allocation to achieve covariate balancing (Hu & Hu 2012). These include the biased coin covariate-adaptive randomization design (Wei 1978, Antognini & Giovagnoli 2004), which is an extension of the biased coin design (Efron 1971) for balancing the sample size, and the Pocock–Simon design which is based on a minimization method for sequential treatment assignment (Taves 1974, Pocock & Simon 1975). Despite their popularity, the main drawbacks of the these designs are that continuous covariates must be categorized into several groups, while clinical trials often collect a large number of continuous covariates and different ways of categorization may lead to different imbalanced structures. In addition, breaking down continuous covariates into sub-categories often changes the nature of the covariates and makes distributional balance unattainable (Ma & Hu 2013). If the sub-categories are not appropriately defined, it can even lead to error and loss of efficiency in the randomization procedure (Stigsby & Taves 2010).

Such a problem has arisen in many clinical trials, which is illustrated with an AIDS Clinical Trials Group study (Campbell et al. 2012). To evaluate several antiretroviral regimens in diverse populations, patients in the A5175 trial were randomly assigned to the antiretroviral therapies with efavirenz plus lamivudine-zidovudine (arm 1) and atazanavir, didanosine-EC plus emtricitabine (arm 2). The study endpoint was the CD4 count at week 96. The baseline covariate CD4 cell count at screening was found to be strongly associated with the endpoint with a p-value less than 2×10^{-16} in a simple linear regression analysis. To balance the CD4 cell count at screening, there was a controversy on the choice of the cutoffs, either the clinically meaningful low CD4 count 200 or the sample average 169. In a simulated clinical trial study, we compared the performances of using these two cutoffs under the same covariate–adaptive procedure. The resulting absolute mean difference between the two groups was 71.76 for the cutoff 200, and 43.73 for the cutoff 169 with corresponding p-values of 0.002 and 0.06 for the two sample t-test of the mean differences. This suggested that a slight variation in the cutoff may lead to substantially different allocation results. To handle continuous covariates, Frane (1998) proposed to calculate the p-value for the mean difference of each covariate, presuming that a new patient is assigned to each treatment group. Using the minimal p-value as a representation of the imbalance of assigning a new patient to a specific treatment, the new patient is then assigned to the treatment with the largest minimal p-value. Stigsby & Taves (2010) considered the rank-sum based covariate adaptive procedure, and Su (2011) discussed a method using quantiles of the covariate differences. Ma & Hu (2013) proposed a randomization procedure by defining the imbalance of the covariates through kernel density estimators, which summarize all the information in the covariate distributions.

To improve the overall balance among both continuous and discrete covariates, we develop a kernel-based adaptive randomization framework that can simultaneously handle a large number of continuous covariates in a single step. In particular, we define a similarity measure between each incoming patient and all the existing patients, and then allocate the new patient with the largest probability to the arm that has the least overall similarity to the new patient. Through weighing each observation by taking into account his/her similarity with the new patient, the proposed method handles both discrete and continuous covariates in a natural way and further broadens the traditional counting from integer values to all nonnegative values.

The rest of the paper is organized as follows. Section 2 describes our covariate–adaptive randomization procedure via introducing the similarity measure and modifying the biased coin design. In Section 3, we cast the Pocock–Simon design in our new framework so as to accommodate continuous covariates. We carry out simulation studies and a real data example to illustrate the performance of the new designs in Section 4. Section 5 concludes with some remarks. Theoretical results are delineated in the Appendix and the corresponding technical proofs are presented in the supplementary materials.

2 Similarity Weighted Biased Coin Design

In a randomized clinical trial with m treatments, suppose that we have already assigned n patients to different arms, and a new patient arrives and is ready for treatment assignment. Let $\mathbf{X}_i$ be the p-dimensional covariate vector for the ith patient, and I_{iu} be the indicator of assigning the ith patient to treatment arm u, $u = 1, \ldots, m$.

We define a similarity measure w_i between the ith existing patient and the incoming $(n+1)$th patient, whose covariate vector is $\mathbf{X}_{n+1}$ with $X_{(n+1)k}$ denoting the kth component. For ease of exposition, we standardize all the covariate values to be within the range of $[-1, 1]$. The similarity measure between the new patient and the ith patient in the trial is defined as

$$w_i = \prod_{k=1}^{p} w_{ik}, \quad i = 1, \ldots, n. \tag{1}$$

where

$$w_{ik} = K_{h_n}(X_{ik} - X_{(n+1)k}), \tag{2}$$

$K_{h_n}(x) = K(x/h_n)/h_n$, and $K(\cdot)$ is a kernel function satisfying $K(\cdot) \geq 0$, and $K(0) = 1$, and $h_n > 0$ is a bandwidth. The classical kernel functions include the triangular kernel $K(x) = (1 - |x|)I(|x| \leq 1)$, the Epanechnikov kernel $K(x) = (1 - x^2)I(|x| \leq 1)$, the Gaussian kernel $K(x) = \exp(-x^2/2)$, and so on. Although the selection of kernels does not affect the large sample properties of our allocation procedure, in practice we recommend to use the Epanechnikov kernel for the bounded covariates, which is the most efficient one in minimizing the averaged mean squared error (Epanechnikov 1969). The similarity measure w_{ik} indicates a higher level of similarity for patients whose kth covariate values are closer to $X_{(n+1)k}$, and the similarity decreases to zero as the difference between X_{ik} and $X_{(n+1)k}$ reaches the bandwidth h_n, for $k = 1, \ldots, p$. The higher the value of h_n, the kernel opts to take into account more X_{ik}'s with larger distances from $X_{(n+1)k}$.

We propose the similarity weighted biased coin design for balancing covariates, which is described as follows.

1. Calculate the similarity measure of the new patient with each of the existing n patients in the trial to obtain $w_1, \dots, w_n$ using (1).

2. For $u = 1, \dots, m$, calculate the weighted total number of patients in treatment arm u,
$$n_u = \sum_{i=1}^{n} w_i I_{iu}, \tag{3}$$
and obtain the imbalance measure of arm u as $g_{nu} = n_u / (\sum_{u=1}^{m} n_u)$.

3. Define the allocation probability π_u to be a function of $\mathbf{g}_n = (g_{n1}, \dots, g_{n(m-1)})^{\mathrm{T}}$ that is decreasing with respect to each component g_{nu}. We assign the new patient to treatment arm u with probability $\pi_u(\mathbf{g}_n)$, $u = 1, \dots, m$.

In the construction of the similarity weighted biased coin design, a patient who is more similar to the new patient receives a larger weight, and is counted more towards the total number of patients in a specific arm. Compared with the biased coin design where $n_u = \sum_{i=1}^{n} I_{iu}$, our definition of n_u in (3) is a weighted sum of the treatment indicators I_{iu}. If the covariate vector $\mathbf{X}_i$ contains only discrete variables, our method reduces to the existing discrete covariate–adaptive randomization by choosing h_n to be smaller than the smallest difference in different categories, i.e., $h_n < \min_{X_{ik} \neq X_{i'k}} |X_{ik} - X_{i'k}|$. Such a construction leads to $w_i = 0$ whenever $X_{ik} \neq X_{(n+1)k}$ for at least one k, and w_i reaches its maximum if $\mathbf{X}_i = \mathbf{X}_{n+1}$. As a result, our method reduces to the biased coin randomization procedure within each stratum defined by the discrete covariates.

We highlight several advantages of the proposed similarity weighting scheme. First, it overcomes the difficulties caused by the high dimensionality of the covariates. To accommodate high-dimensional covariates, Yuan et al. (2011) resorted to a linear model structure, which is unfortunately subject to model misspecification. Second, even when all the covariates are discrete, we can choose the bandwidth h_n sufficiently large to avoid the situation of too many strata and too few or even zero observations within some strata. Finally, the procedure is automatic and flexible as reflected in the various ways of constructing the similarity measure.

To study the asymptotic properties of the imbalance measure $\mathbf{D}_{nu} = \sum_{i=1}^{n}(I_{iu} - \kappa_u)\mathbf{X}_i$, we explore the properties of $\mathbf{D}_{nu}^{\mathrm{T}}\mathbf{z} = \sum_{i=1}^{n}(I_{iu} - \kappa_u)\mathbf{X}_i^{\mathrm{T}}\mathbf{z}, u = 1, \ldots, m-1$, where $\mathbf{z}$ is an arbitrary p dimensional vector. We show that the allocation achieves the target ratio in the long run as follows.

Theorem 1. *Assume that Conditions (C1) – (C6) hold and let* $Q_z = E(\mathbf{z}^{\mathrm{T}}\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\mathbf{z})$. *Then in the similarity weighted biased coin design with an allocation probability* $\pi_u\{\mathbf{U}_n(\mathbf{X}_{n+1})\}$, $n^{-1/2}(\mathbf{D}_{n1}^{\mathrm{T}}\mathbf{z}, \ldots, \mathbf{D}_{n(m-1)}^{\mathrm{T}}\mathbf{z})^{\mathrm{T}}$ *converges to a zero-mean multivariate Gaussian distribution, with the variance–covariance matrix having the* u*th diagonal element* $(1+2\rho)^{-1}(1-\kappa_u)\kappa_u Q_z$ *and the* (u, v) *entry* $-(1+2\rho)^{-1}\kappa_u\kappa_v Q_z$, *for* $u, v = 1, \ldots, m-1, u \neq v$.

Obviously, due to the arbitrariness of $\mathbf{z}$, we readily obtain the asymptotic normality of $\mathbf{D}_n$. The proof relies heavily on the martingale convergence theorem (Hall & Heyde 1980), which is provided in the supplementary material.

3 Similarity Weighted Pocock–Simon Design

The proposed similarity measure can also be incorporated into the Pocock–Simon design (Pocock & Simon 1975), namely the similarity weighted Pocock–Simon design, so that continuous covariates no longer need to be discretized. Consider the hypothetical situation where we assign the new patient to treatment 1, the similarity weighted Pocock–Simon design can be implemented as follows.

1. For the kth covariate, let $n_{ku} = \sum_{i=1}^{n} w_{ik} I_{iu}$ be the weighted total number of subjects assigned to treatment arm u, where w_{ik} is defined in (2).

2. Calculate the aggregated variation in the form of

$$d_k = 1/2 \sum_{u,v \in 1,\dots,m} (n_{ku} - n_{kv})^2$$

for the kth covariate.

3. Sum over the d_k's across all the covariates, leading to the imbalance measure $g_{n1} = \sum_{k=1}^{p} d_k$.

4. Similarly, we can calculate $g_{n2}, \dots, g_{nm}$ by presuming that the new patient is assigned to treatment arms $2, \dots, m$, respectively.

5. We order the g_{nu}'s, e.g., $g_{n1} \leq \cdots \leq g_{nm}$, create the randomization probabilities satisfying $\pi_{(n+1)1} \geq \cdots \geq \pi_{(n+1)m}$, and assign the new patient to the m treatment arms with probabilities $\pi_{(n+1)1}, \dots, \pi_{(n+1)m}$.

The selection of d_k is not unique; for example, the sum of absolute differences, $d_k = 1/2 \sum_{u,v \in 1,\dots,m} |n_{ku} - n_{kv}|$, can also be used to measure the total imbalance among the treatment arms for the kth covariate. Our modified procedure can be viewed as a generalized version of the original Pocock–Simon design procedure: the former calculates the n_{ku}'s using a similarity weight w_{ik}, while the latter sets the weight $w_{ik} = 1$ if the ith patient has the same kth covariate value as the new patient, and $w_{ik} = 0$ otherwise. Our approach handles continuous covariates through a similarity-based weighting scheme and does not require discretization.

4 Numerical Studies

4.1 Simulation Study

To evaluate the finite sample properties of the proposed similarity weighted biased coin design and the similarity weighted Pocock–Simon design, we simulate 1000 two-arm clinical trials, each containing $n = 50$ subjects. We generate covariates in the form of

$$X_{ik} = \frac{2\exp(\xi_{ik})}{1 + \exp(\xi_{ik})} - 1, \quad k = 1, \ldots, p, \tag{4}$$

where the dimension p of covariates range from 1 to 8, and ξ_{ik} is a normal random variable with mean $k/2$ and standard deviation 5. To implement the biased coin design and the Pocock–Simon design, we discretize the X_{ik}'s to be 0 or 1 according to the negative or positive signs of the covariates. We use the allocation probability function $\phi_u(\mathbf{y}) = (y_u^{-1} - 1)/\sum_{u=1}^{m}(y_u^{-1} - 1)$ in Atkinson (1982), which satisfies Conditions (C1) and (C2) as shown in Smith (1984). In all the numerical studies, we take the bandwidth to be 2.1, so that the support of the kernel could completely cover all the covariates. We also experiment with other bandwidths between 2 and 2.5, and the results turned out to be similar as long as the bandwidth was chosen to be slightly larger than the covariate range.

We first make comparisons from two aspects: the imbalance of the sample sizes and the imbalance of the covariates between the two arms. To quantify the former, let n_u denote the sample size in arm $u, u = 1, 2$, and we obtain $|n_1 - n_2|$ averaged over the 1000 simulated trials for all four methods: similarity weighted biased coin design, biased coin design, similarity weighted Pocock–Simon design and Pocock–Simon design. Table 1 summarizes the imbalance measure on sample sizes, which demonstrates the similarity weighted designs tend to induce more balanced numbers of subjects between the two arms. Figure 1 shows the survival function, i.e., one minus the empirical cumulative distribution function (CDF), of $|n_1 - n_2|$ over the 1000 replicated data sets. A sharper decline of the survival function of the imbalance measure indicates a more effective procedure, and clearly the similarity weighted designs outperform their counterparts. To compare the covariate imbalance, we borrow the idea from the analysis of variance to construct an F test statistic for each covariate,

$$F_k = \frac{\mathrm{SSB}_k/(m-1)}{(\mathrm{SST}_k - \mathrm{SSB}_k)/(n-m)}, \quad k = 1, \ldots, p \tag{5}$$

where the between-arm sum of squared errors (SSB_k) is given by

$$\mathrm{SSB}_k = n_1 \left(n_1^{-1} \sum_{i=1}^{n} I_{i1} X_{ik} - n^{-1} \sum_{i=1}^{n} X_{ik} \right)^2 + n_2 \left(n_2^{-1} \sum_{i=1}^{n} I_{i2} X_{ik} - n^{-1} \sum_{i=1}^{n} X_{ik} \right)^2,$$

and the total sum of squared errors (SST_k) is given by

$$\mathrm{SST}_k \equiv \sum_{i=1}^{n} \left(X_{ik} - n^{-1} \sum_{i=1}^{n} X_{ik} \right)^2.$$

As the F statistic has the same distribution across all the covariates, we summarize the overall mean of the F statistics for all the covariates in Table 2, and plot the survival functions of the F statistics in Figure 2. Both the similarity weighted biased coin design and similarity weighted Pocock–Simon design outperform their counterparts in terms of balancing the covariates. For the biased coin designs, the improvement by using similarity weights enhances as the dimension p increases, while the opposite is true for the Pocock–Simon designs. Figure 2 further demonstrates the advantages of the proposed methods, and particularly the similarity weighted Pocock–Simon design performs the best in terms of reducing the imbalance in both the sample size and covariates.

To explore the estimation of the treatment effect under the four designs, we consider a two-arm trial, where $I_i = 1$ indicates that the ith patient is allocated to arm 1, and $I_i = 0$ otherwise. We simulate 1000 clinical trials with response Y_i generated as

$$Y_i = \mu I_i + \exp(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{X}_i/2) + \epsilon_i$$

where the true parameter values are $\mu = -5$ and $\boldsymbol{\beta} = (p/3, \ldots, p/3)^{\mathrm{T}}$, ϵ_i is a zero-mean normal random error with standard deviation 0.1, and $\mathbf{X}_i$ is generated in the same way as before. We choose $p = 1, \ldots, 8$ and sample size $n = 30$. We allocate the first patient with equal probability to each arm, and started the adaptive allocation from the second patient.

Table 3 shows the estimated treatment effect $\widehat{\mu} = \sum_{i=1}^{n} I_i Y_i / n_1 - \sum_{i=1}^{n} (1 - I_i) Y_i / n_2$ and its empirical standard deviation and mean squared error for $p = 1, \ldots, 8$. Note that $\widehat{\mu}$ is always a consistent estimator of μ regardless of the regression form (Shao et al. 2010). The biases of the estimates of μ are negligible under all four designs, while the empirical standard deviations and mean squared errors deteriorate as p grows. Again, both the similarity weighted biased coin design and the similarity weighted Pocock–Simon design outperform the unweighted counterparts, respectively, in terms of the mean squared errors.

4.2 Real Data Example

We apply all the four methods, namely the biased coin design, the Pocock–Simon design, and the corresponding similarity weighted versions, to the data from the AIDS trial A5175. To study the treatment effect, seven covariates were considered important, which should be balanced between the two arms at randomization: CD4 cell count and percentage (at screening), Karnofsky score, Hepatitis-B surface antigen reactivity, the laboratory test values including platelets, white blood cell count, absolute neutrophil count, and albumin. In the original trial, there are $n = 370$ patients with complete observations, and they are allocated to the two arms with equal probability. We take the standardized CD4 count at week 96 as the outcome, and transformed the standardized covariates via $2\exp(x)/\{1 + \exp(x)\} - 1$ to ensure that all the covariate values were within $[-1, 1]$. Let $\mathbf{X}_{iu}$ and Y_{iu} denote the covariates and response respectively for the ith patient in arm $u, u = 1, 2$. We build separate models for each arm,

$$Y_{iu} = \beta_{0u} + \boldsymbol{\beta}_u^{\mathrm{T}} \mathbf{X}_{iu} + e_{iu}, \quad u = 1, 2,$$

with $e_{iu} \sim N(0, \sigma_u^2)$. We obtain the least squared estimators $(\widehat{\beta}_{0u}, \widehat{\boldsymbol{\beta}}_u^{\mathrm{T}})$ for each arm, and use these parameter estimates as the true values to generate the outcomes in different randomization procedures.

For illustration, we select the first 50 samples to evaluate and compare the four designs. The observed difference of the mean outcomes between the two arms over these 50 samples is 0.39, which is substantially different from that using the full 370 samples, 0.22. Since the trial data are balanced in covariates for $n = 370$, we used 0.22 as a benchmark to approximate the true underlying mean difference between the two arms. Using each of the four randomization procedures, we re-randomized the 50 patients and each procedure is replicated 1000 times to obtain the average effect. The means of $|n_1 - n_2|$ under the similarity weighted biased coin design, biased coin design, similarity weighted Pocock–Simon design and Pocock–Simon design are 1.30, 1.49, 0.17 and 0.32, respectively, and the corresponding means of the F statistics in (5) summing over all the covariates are 3.24, 5.41, 2.29 and 2.56. The results show that the similarity weighted procedures outperform the original counterparts in reducing both the sample size and covariate imbalance, and overall the similarity weighted Pocock–Simon design performs the best among the four designs. In addition, the estimates of the difference of the mean responses are 0.306, 0.306, 0.306 and 0.313 using the similarity weighted biased coin design, biased coin design, similarity weighted Pocock–Simon design and Pocock–Simon design, respectively. Compared with the observed mean difference in the first 50 samples, the estimates from the four covariate–adaptive designs are closer to the benchmark value 0.22, indicating that covariates adaptation helps to improve the balance.

5 Discussion

To accommodate continuous covariates in the biased coin design and Pocock–Simon design, we develop a kernel-based similarity measure and its associated imbalance assessment criterion. We define the allocation probability function based on the new imbalance measure and show that the covariate equilibrium measure $\mathbf{D}_{nu}$ of the proposed similarity weighted biased coin design asymptotically follows a normal distribution. We choose the continuous allocation function π instead of a discrete one, because discrete allocation functions can neither discriminate between large versus small values of $|g_{n1} - g_{n2}|$ nor discriminate between large versus small numbers of subjects, hence typically yield designs with poor small sample properties (Wei 1978, Smith 1984, Hu & Zhang 2004). In terms of the bandwidth requirement, we find that as long as the bandwidth is chosen to be slightly larger than the covariate range, the results are not sensitive to the bandwidth choice. Not only does the asymptotic property of the covariate equilibrium $\mathbf{D}_{nu}$ explain the covariate discrepancy between the arms, but it is also an essential component for analyzing the hypothesis testing procedures in the linear regression problem (Shao et al. 2010, Ma et al. 2015). Our theoretical results are essential for constructing inference procedures under the similarity weighted biased coin design.

Acknowledgments

We thank the guest editor Raymond J. Carroll and Haolun Shi for many constructive comments that have led to significant improvements of the article. Ma's research was supported in part by a grant (DMS-1608540) from the National Science Foundation, and Yin's research was supported in part by a grant (17326316) from the Research Grants Council of Hong Kong.

Appendix

A.1 Allocation Probability Function

Suppose that n samples have been enrolled in the trial. The allocation probability π_u is a function of the imbalance measure vector $\mathbf{g}_n = (g_{n1}, \ldots, g_{n(m-1)})^{\mathrm{T}}$. Let $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_{m-1})^{\mathrm{T}}$. Furthermore, let $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_{m-1})^{\mathrm{T}}$. We show that $\boldsymbol{\pi}$ drives $\mathbf{g}_n$ towards $\boldsymbol{\kappa}$ under the following conditions. For notational simplicity, we surpress the subindex n in these conditions.

(C1) $\pi_u(\mathbf{g})$ is a nonnegative and monotonically decreasing function with respect to the uth element g_u. Define $|\cdot|$ to be the L_1 norm of a vector; the vector $\boldsymbol{\pi}(\mathbf{g})$ satisfies $|\boldsymbol{\pi}(\mathbf{g})| \leq 1$, for any component-wise nonnegative $m-1$ dimensional vector $\mathbf{g}$ with $|\mathbf{g}| \leq 1$. Moreover, $\pi_m(\mathbf{g}) = 1 - |\boldsymbol{\pi}(\mathbf{g})|$, $g_m = 1 - |\mathbf{g}|$, $\kappa_m = 1 - |\boldsymbol{\kappa}|$. If $g_u \geq \kappa_u$, then $\pi_u(\mathbf{g}) \leq \kappa_u$; and if $g_u < \kappa_u$, then $\pi_u(\mathbf{g}) > \kappa_u$, $u = 1, \ldots, m$.

(C2) $\pi_u(\mathbf{g})$ is a twice continuously differentiable function of $\mathbf{g}$ with a uniformly bounded Hessian matrix.

Let $\pi'_u(\mathbf{g}) = \partial \pi_u(\mathbf{g})/\partial \mathbf{g}$, $\pi'_{ur}(\mathbf{g})$ be the partial derivative of $\pi_u(\mathbf{g})$ with respect to its rth argument, and π''_u be the $(m-1) \times (m-1)$ Hessian matrix.

Remark 1. *Conditions (C1) and (C2) are also used in Smith (1984) to establish the properties of the biased coin design. Condition (C1) implies $\pi_u(\boldsymbol{\kappa}) \leq \kappa_u$. If the inequality is strict for any u, summing both sides over $u = 1, \ldots, m$, we obtain $1 < 1$, which is a contradiction. Therefore, we have $\pi_u(\boldsymbol{\kappa}) = \kappa_u$ for $u = 1, \ldots, m$.*

Remark 2. *For an arbitrary δ,*

$$\pi_m(\kappa_1 + \delta, \kappa_2 - \delta, \kappa_3, \ldots, \kappa_{m-1}) = \kappa_m + \delta\{\pi'_{m1}(\boldsymbol{\kappa}) - \pi'_{m2}(\boldsymbol{\kappa})\} + O(\delta^2).$$

Note that $\pi_m(\kappa_1 + \delta, \kappa_2 - \delta, \kappa_3, \ldots, \kappa_{m-1})$ is at most κ_m regardless of the sign of δ. As $\delta \to 0$, $O(\delta^2)$ goes to 0 faster than the leading terms. This gives $\delta\{\pi'_{m1}(\boldsymbol{\kappa}) - \pi'_{m2}(\boldsymbol{\kappa})\} \leq 0$

and $-\delta\{\pi'_{m1}(\boldsymbol{\kappa}) - \pi'_{m2}(\boldsymbol{\kappa})\} \leq 0$. *Therefore,* $\pi'_{m1}(\boldsymbol{\kappa}) = \pi'_{m2}(\boldsymbol{\kappa})$. *Similarly, for each* $u = 1, \ldots, m-1$, $\pi'_{mu}(\boldsymbol{\kappa}) = \rho$, *a constant that does not depend on* u. *Following the same argument, for any* $u < m$, $u \neq 1$,

$$\pi_u(\kappa_1 + \delta, \kappa_2, \ldots, \kappa_{m-1}) = \kappa_u + \delta\pi'_{u1}(\boldsymbol{\kappa}) + O(\delta^2) \leq \kappa_u$$

for all δ, *which implies* $\pi'_{u1}(\boldsymbol{\kappa}) = 0$ *and in turn* $\pi'_{ur}(\boldsymbol{\kappa}) = 0$ *for* $r < m, r \neq u$.

Remark 3. *Because* $\sum_{u=1}^{m} \pi_u(\mathbf{g}) = 1$ *for all* $\mathbf{g}$, $\sum_{u=1}^{m} \pi'_{ur}(\mathbf{g}) = 0$ *for any* $r = 1, \ldots, m-1$, *we have*

$$\sum_{u=1}^{m} \pi'_{ur}(\mathbf{g}) = \pi'_{rr}(\mathbf{g}) + \sum_{u=1, u\neq r}^{m} \pi'_{ur}(\mathbf{g}) = \pi'_{rr}(\mathbf{g}) + \pi'_{mr}(\mathbf{g}) = 0,$$

which implies

$$\pi'_{rr}(\boldsymbol{\kappa}) = -\pi'_{mr}(\mathbf{g}) = -\rho \leq 0,$$

for $r = 1, \ldots, m-1$. *The last inequality holds due to the fact that* $\pi_r(\mathbf{g})$ *is non-increasing at* $g_r = \kappa_r$.

Remark 4. *Combining the results in Remarks 1 to 3, we have that* $\pi_u(\boldsymbol{\kappa}) = \kappa_u$ *for all* $u = 1, \ldots, m$; $\pi'_{ur}(\boldsymbol{\kappa}) = 0$ *for* $u, r = 1, \ldots, m-1$ *and* $u \neq r$; *and* $\pi'_{mr}(\boldsymbol{\kappa}) = -\pi'_{rr}(\boldsymbol{\kappa}) = \rho \geq 0$ *for* $r = 1, \ldots, m-1$.

Recall the definition of the imbalance measure,

$$g_{nu} = U_{nu}(\mathbf{X}_{(n+1)}) = \frac{\sum_{i=1}^{n} \prod_{k=1}^{p} K_{h_n}(X_{ik} - X_{(n+1)k}) I_{iu}}{\sum_{i=1}^{n} \prod_{k=1}^{p} K_{h_n}(X_{ik} - X_{(n+1)k})}. \tag{A.1}$$

Let $\mathbf{U}_n = (U_{n1}, \ldots, U_{n(m-1)})^{\mathrm{T}}$, and then the allocation probability is $\pi_u(\mathbf{g}_n) = \pi_u\{\mathbf{U}_n(\mathbf{X}_{(n+1)})\}$, where π_u satisfies Conditions (C1) and (C2).

A.2 Asymptotic Properties

We describe additional conditions for the theoretical development as follows.

(C3) In the kernel function $K_{h_n}(t) = K(t/h_n)/h_n$, K is a second order symmetric kernel function that satisfies $\int K(t)dt = 1$, $\int K(t)^2 dt < \infty$, and $\int t^2 K(t)^2 dt < \infty$. h_n satisfies $nh_n^2 \to \infty$, and $nh_n^4 \to 0$.

(C4) The density function $f_k(X_k)$ is bounded away from zero and infinity almost surely on the support for all k.

(C5) X_k^2 is a uniformly integrable random variable.

(C6) Let $n_0 > \rho \geq 0$. Assume the first n_0 patients are randomized to arms $1, \ldots, m$ with probabilities $\kappa_1, \ldots, \kappa_m$, respectively, and the adaptive allocation process starts from the $(n_0 + 1)$th patient.

Let $I_{0u} = 0$. Under the situation where the desired allocation ratio in the long run is κ_u, $u = 1, \ldots, m$, we show that the covariate equilibrium of the similarity weighted biased coin design, defined by $\mathbf{D}_{nu} \equiv \sum_{i=1}^{n}(I_{iu} - \kappa_u)\mathbf{X}_i$, has mean zero and is asymptotically normally distributed. We first state the asymptotic property for similarity weighted biased coin design with one covariate.

Lemma 1. *Assume that Conditions (C1) – (C6) hold. Let* $Q = E(X_i^2)$, $D_{nu} = \sum_{i=1}^{n}(I_{iu} - \kappa_u)X_i$,

$$\mathbf{\Omega} = \begin{pmatrix} (1+2\rho)^{-1}(1-\kappa_1)\kappa_1 & \ldots & -(1+2\rho)^{-1}\kappa_1\kappa_{m-1} \\ \vdots & \ddots & \vdots \\ -(1+2\rho)^{-1}\kappa_1\kappa_{m-1} & \ldots & (1+2\rho)^{-1}(1-\kappa_{m-1})\kappa_{m-1} \end{pmatrix} Q$$

and $\mathbf{D}_n = (D_{n1}, \ldots, D_{n(m-1)})^{\mathrm{T}}$. *Then in the similarity weighted biased coin design with the allocation probability* $\pi_u\{\mathbf{U}_n(X_{n+1})\}$, $n^{-1/2}\mathbf{\Omega}^{-1/2}\mathbf{D}_n$ *converges to a standard multivariate normal distribution.*

The proof of Lemma 1 is given in the supplementary material.

References

Antognini, A. B. & Giovagnoli, A. (2004), 'A new biased coin design for the sequential allocation of two treatments', *Journal of the Royal Statistical Society: Series C* **53**, 651–664.

Atkinson, A. (1982), 'Optimum biased coin designs for sequential clinical trials with prognostic factors', *Biometrika* **69**, 61–67.

Campbell, T. B., Smeaton, L. M., Kumarasamy, N., Flanigan, T., Klingman, K. L., Firnhaber, C., Grinsztejn, B., Hosseinipour, M. C., Kumwenda, J., Lalloo, U., Riviere, C., Sanchez, J., Melo, M., Supparatpinyo, K., Tripathy, S., Martinez, A. I., Nair, A., Walawander, A., Moran, L., Chen, Y., Snowden, W., Rooney, J. F., Uy, J., Schooley, R. T., De Gruttola, V., Hakim, J. G. & study team of the ACTG, P. (2012), 'Efficacy and safety of three antiretroviral regimens for initial treatment of HIV-1: a randomized clinical trial in diverse multinational settings', *PLOS Medicine* **9**, e1001290.

Efron, B. (1971), 'Forcing a sequential experiment to be balanced', *Biometrika* **58**, 403–417.

Epanechnikov, V. A. (1969), 'Non-parametric estimation of a multivariate probability density', *Theory of Probability and Its Applications* **14**, 153–158.

Frane, J. W. (1998), 'A method of biased coin randomization, its implementation, and its validation', *Drug Information Journal* **32**, 423–432.

Gill, R. D. & Johansen, S. (1990), 'A survey of product–integration with a view toward application in survival analysis', *Annals of Statistics* **18**, 1501–1555.

Hall, P. (1981), ‘Laws of the iterated logarithm for nonparametric density estimators’, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **56**, 47–61.

Hall, P. (1984), ‘Asymptotic properties of integrated square error and cross-validation for kernel estimation of a regression function’, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **67**, 175–196.

Hall, P. (1992), ‘Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density’, *Annals of Statistics* **67**, 675–694.

Hall, P. (1993), ‘On edgeworth expansion and bootstrap confidence bands in nonparametric curve estimation’, *Journal of the Royal Statistical Society. Series B* **55**, 291–304.

Hall, P. & Heyde, C. (1980), *Martingale Limit Theory and Its Application*, Probability and Mathematical Statistics, Academic Press, New York.

Hall, P. & Horowitz, J. (2013), ‘A simple bootstrap method for constructing nonparametric confidence bands for functions’, *Annals of Statistics* **41**, 1892–1921.

Hall, P. & Marron, J. (1988), ‘Choice of kernel order in density estimation’, *Annals of Statistics* **16**, 161–173.

Härdle, W. (2004), *Nonparametric and Semiparametric Models*, Springer Series in Statistics, Springer, New York.

Hu, F. & Zhang, L.-X. (2004), ‘Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials’, *Annals of Statistics* **32**, 268–301.

Hu, Y. & Hu, F. (2012), ‘Asymptotic properties of covariate-adaptive randomization’, *Annals of Statistics* **40**, 1794–1815.

Ma, W., Hu, F. & Zhang, L. (2015), 'Testing hypotheses of covariate-adaptive randomized clinical trials', *Journal of the American Statistical Association* **110**, 669–680.

Ma, Z. & Hu, F. (2013), 'Balancing continuous covariates based on kernel densities', *Contemporary Clinical Trials* **34**, 262–269.

Pocock, S. J. & Simon, R. (1975), 'Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial', *Biometrics* **31**, 103–115.

Shao, J., Yu, X. & Zhong, B. (2010), 'A theory for testing hypotheses under covariate-adaptive randomization', *Biometrika* **97**, 347–360.

Silverman, B. W. (1978), 'Weak and strong uniform consistency of the kernel estimate of a density and its derivatives', *Annals of Statistics* **6**, 177–184.

Slavík, A. & Karlova, U. (2007), *Product Integration, Its History and Applications*, Matfyzpress Prague, Prague.

Smith, R. L. (1984), 'Properties of biased coin designs in sequential clinical trials', *Annals of Statistics* **12**, 1018–1034.

Stigsby, B. & Taves, D. R. (2010), 'Rank-minimization for balanced assignment of subjects in clinical trials', *Contemporary Clinical Trials* **31**, 147–150.

Su, Z. (2011), 'Balancing multiple baseline characteristics in randomized clinical trials', *Contemporary Clinical Trials* **32**, 547–550.

Taves, D. R. (1974), 'Minimization: a new method of assigning patients to treatment and control groups.', *Clinical Pharmacology and Therapeutics* **15**, 443.

Wei, L. J. (1978), 'The adaptive biased coin design for sequential experiments', *Annals of Statistics* **6**, 92–100.

Yuan, Y., Huang, X. & Liu, S. (2011), ‘A Bayesian response-adaptive covariate-balanced randomization design with application to a leukemia clinical trial’, *Statistics in Medicine* **30**, 1218–1229.

Tables and Figures

Table 1: Comparison of the sample size imbalance, $|n_1 - n_2|$, among the similarity weighted biased coin design, biased coin design, similarity weighted Pocock–Simon design and Pocock–Simon design for different dimensions (p) of covariates.

	Dimension of covariates p							
Design	1	2	3	4	5	6	7	8
Weighted biased coin	1.277	1.289	1.265	1.219	1.325	1.343	1.286	1.411
Biased coin	1.279	1.276	1.334	1.687	2.047	2.247	2.456	2.605
Weighted Pocock–Simon	0.122	0.159	0.159	0.183	0.199	0.211	0.233	0.246
Pocock–Simon	0.387	0.241	0.221	0.262	0.288	0.294	0.354	0.351

Table 2: Comparison of the covariate imbalance using the F statistics, among the similarity weighted biased coin design, biased coin design, similarity weighted Pocock–Simon design and Pocock–Simon design for different dimensions (p) of covariates.

	Dimension of covariates p							
Design	1	2	3	4	5	6	7	8
Weighted biased coin	0.278	0.297	0.299	0.311	0.326	0.345	0.373	0.389
Biased coin	0.280	0.299	0.312	0.402	0.525	0.663	0.809	0.872
Weighted Pocock–Simon	0.028	0.053	0.085	0.128	0.166	0.207	0.256	0.308
Pocock–Simon	0.149	0.145	0.162	0.193	0.225	0.271	0.322	0.358

Table 3: Comparison of the estimated treatment effect under the similarity weighted biased coin design, biased coin design, similarity weighted Pocock–Simon design and Pocock–Simon design, where $\widehat{\mu}$ is the estimate of $\mu = -5$, and SD and MSE are the corresponding empirical standard deviation and mean squared error, respectively.

p	$\widehat{\mu}$	SD	MSE	$\widehat{\mu}$	SD	MSE
	Weighted biased coin			Weighted Pocock–Simon		
1	-4.983	0.314	0.332	-4.992	0.134	0.141
2	-5.030	0.830	0.860	-4.999	0.686	0.687
3	-5.022	1.723	1.745	-4.949	1.698	1.749
4	-4.864	3.050	3.186	-5.058	3.164	3.223
5	-4.924	4.776	4.852	-4.726	5.169	5.442
6	-5.130	6.778	6.909	-4.573	7.328	7.755
7	-4.997	8.224	8.227	-4.959	9.401	9.442
8	-5.153	9.482	9.636	-5.016	10.582	10.598
	Biased coin			Pocock–Simon		
1	-4.996	0.332	0.335	-4.996	0.268	0.272
2	-5.028	0.822	0.850	-4.994	0.865	0.871
3	-4.989	1.783	1.794	-5.039	1.899	1.938
4	-5.025	3.201	3.225	-5.063	3.660	3.723
5	-5.015	5.328	5.343	-4.956	5.821	5.865
6	-5.167	7.930	8.098	-5.239	7.921	8.161
7	-5.353	9.787	10.140	-5.372	10.015	10.387
8	-5.253	11.797	12.050	-5.169	12.011	12.181

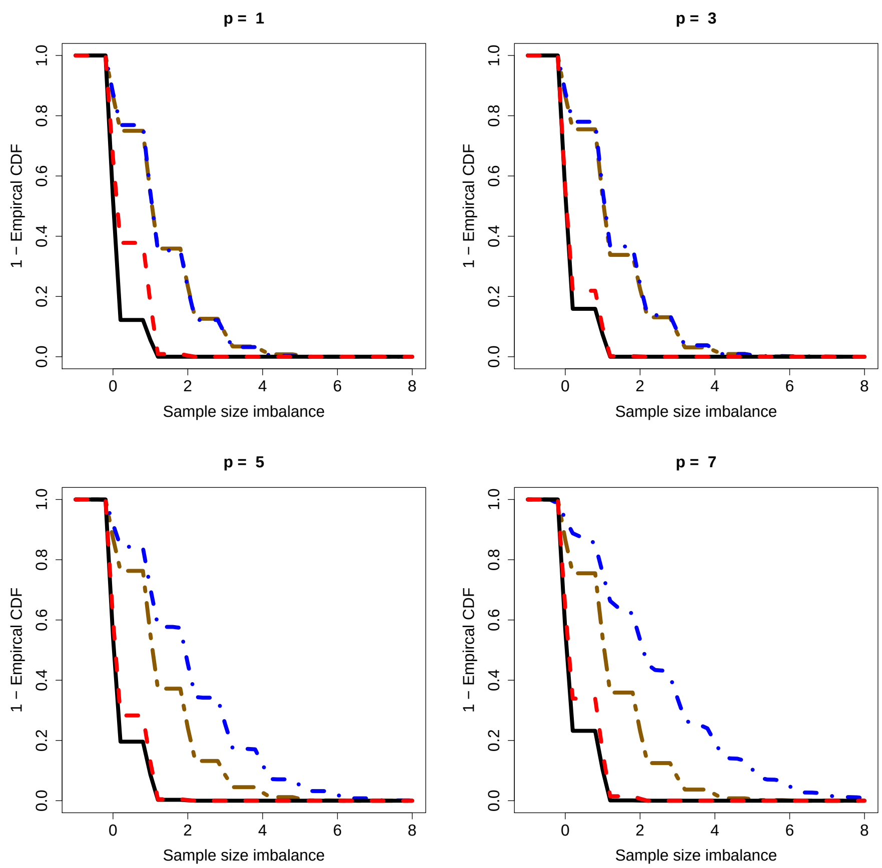

Figure 1: Survival functions (or one minus the empirical CDF) of sample size imbalance for the similarity weighted biased coin design (dotted line), biased coin design (dot-dashed line), similarity weighted Pocock–Simon design (solid line), and Pocock–Simon design (dashed line).

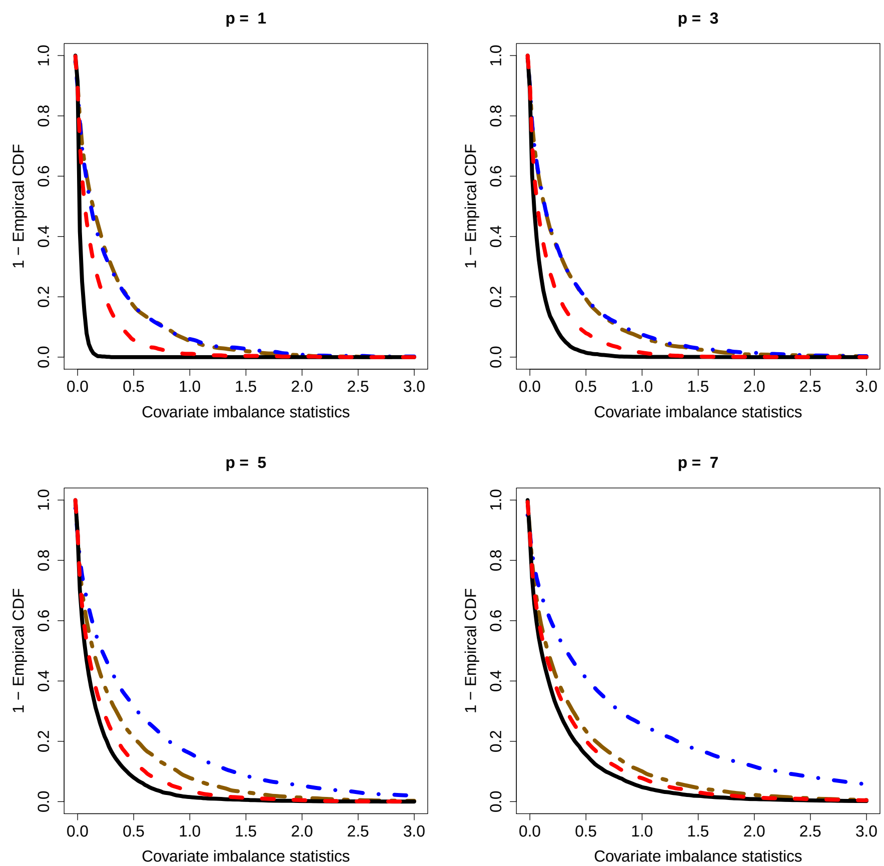

Figure 2: Survival functions (or one minus the empirical CDF) of sample size imbalance for the similarity weighted biased coin design (dotted line), biased coin design (dot-dashed line), similarity weighted Pocock–Simon design (solid line), and Pocock–Simon design (dashed line).

Supplementary Material

S.1 Necessary results for Lemma 1

For a new data point X_{n+1}, (A.1) reduces to

$$U_{nu}(X_{n+1}) = \frac{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1}) I_{iu}}{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})},$$

and $\mathbf{U}_n(X_{n+1}) = \{U_{n1}(X_{n+1}), \ldots, U_{n(m-1)}(X_{n+1})\}^{\mathrm{T}}$. We set $\pi_u\{\mathbf{U}_n(X_{n+1})\}$ to be the probability of assigning the new observation to arm u. To derive the asymptotic property of $D_{nu} = \sum_{i=1}^{n}(I_{iu} - \kappa_u)X_i$, we first state useful lemmas that support Lemma 1.

Lemma 2. *Suppose that n subjects have been enrolled in a clinical trial. For a new data point X_{n+1} and arms u and v, we have*

$$E\left\{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})(I_{iu} - \kappa_u)\right\}^2 = O(nh_n^{-1})$$

and

$$\Big|\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})(I_{iu} - \kappa_u)X_{n+1}\Big| = O_p(n^{1/2}h_n^{-1/2}).$$

Proof: Let $\mathcal{F}_j$ be a sigma field generated by all the event history up to stage j. Suppose that a new participant with covariate $X_{n+1} = x_0$ is to be allocated. We define a function $\eta(x, y) = I(x \neq y)(x - y) + I(x = y)x$, then

$$\begin{aligned}
& E\left(\left[\sum_{i=1}^{j+1} K_{h_n}\{\eta(X_i, x_0)\}(I_{iu} - \kappa_u)\right]^2 \middle| \mathcal{F}_j, X_{j+1} = x_0\right) \\
= \quad & E\left(\left[\sum_{i=1}^{j} K_{h_n}\{\eta(X_i, x_0)\}(I_{iu} - \kappa_u)\right]^2 \middle| \mathcal{F}_j, X_{j+1} = x_0\right) \\
& +2\left[\sum_{i=1}^{j} K_{h_n}\{\eta(X_i, x_0)\}(I_{iu} - \kappa_u)\right] \\
& \times E\left[K_{h_n}\{\eta(x_0, x_0)\}(I_{(j+1)u} - \kappa_u) \middle| \mathcal{F}_j, X_{j+1} = x_0\right] \\
& +E\left[\left\{K_{h_n}(x_0)(I_{(j+1)u} - \kappa_u)\right\}^2 \middle| \mathcal{F}_j, X_{j+1} = x_0\right]
\end{aligned}$$

$$
\begin{aligned}
&\le E\left(\left[\sum_{i=1}^{j} K_{h_n}\{\eta(X_i,x_0)\}(I_{iu}-\kappa_u)\right]^2 \middle| \mathcal{F}_j, X_{j+1}=x_0\right) \\
&\quad +E\left[\left\{K_{h_n}(x_0)(I_{(j+1)u}-\kappa_u)\right\}^2 \middle| \mathcal{F}_j, X_{j+1}=x_0\right] \quad a.s. \qquad \text{(S.1)}
\end{aligned}
$$

The last equality holds almost surely, because for any fixed value x_0, $X_i \neq x_0 (i = 1,\ldots,j)$ $a.s.$, which implies $\eta(X_i,x_0) = X_i - x_0$ $a.s.$ Further, with probability one,

$$
\begin{aligned}
&\left[\sum_{i=1}^{j} K_{h_n}\{\eta(X_i,x_0)\}(I_{iu}-\kappa_u)\right] E\left[K_{h_n}\{\eta(x_0,x_0)\}(I_{(j+1)u}-\kappa_u) \middle| \mathcal{F}_j, X_{j+1}=x_0\right] \\
= &\left\{\sum_{i=1}^{j} K_{h_n}(X_i-x_0)(I_{iu}-\kappa_u)\right\} E\left\{K_{h_n}(x_0)(I_{(j+1)u}-\kappa_u) \middle| \mathcal{F}_j, X_{j+1}=x_0\right\} \\
= &\left\{\sum_{i=1}^{j} K_{h_n}(X_i-x_0)\right\} \{U_j(x_0)-\kappa_u\}[\pi_u\{\mathbf{U}_j(x_0)\}-\kappa_u]K_{h_n}(x_0) \\
\le &\ 0
\end{aligned}
$$

by the fact that $U_{nu}-\kappa_u$ and $\pi_u\{\mathbf{U}_n(X_0)\}-\kappa_u$ have opposite signs according to Condition (C1). By taking the expectation on both sides of (S.1), we have

$$
\begin{aligned}
&E\left[\sum_{i=1}^{j+1} K_{h_n}\{\eta(X_i,X_{j+1})\}(I_{iu}-\kappa_u)\right]^2 \\
\le &\ E\left[\sum_{i=1}^{j} K_{h_n}\{\eta(X_i,X_{j+1})\}(I_{iu}-\kappa_u)\right]^2 + E\left\{K_{h_n}(X_{j+1})(I_{(j+1)u}-\kappa_u)\right\}^2.
\end{aligned}
$$

Summing over j from 1 to n,

$$
\begin{aligned}
&\sum_{j=1}^{n} E\left[\sum_{i=1}^{j+1} K_{h_n}\{\eta(X_i,X_{j+1})\}(I_{iu}-\kappa_u)X_{j+1}\right]^2 \\
\le &\sum_{j=1}^{n} E\left[\sum_{i=1}^{j} K_{h_n}\{\eta(X_i,X_{j+1})\}(I_{iu}-\kappa_u)X_{j+1}\right]^2 \\
&+\sum_{j=1}^{n} E\left\{K_{h_n}(X_{j+1})(I_{(j+1)u}-\kappa_u)X_{j+1}\right\}^2,
\end{aligned}
$$

and it is readily seen that the $(j-1)$th summand on the left-hand side agrees with the jth summand in the first term on the right-hand side. Further, $X_{n+1} \neq X_i$ $a.s.$, so

$K_{h_n}\{\eta(X_i, X_{n+1})\} = K_{h_n}(X_i - X_{n+1}), (i = 1, \ldots, n), a.s.$ As a result, we obtain

$$\begin{aligned} E\left[\sum_{i=1}^{n+1} K_{h_n}\{\eta(X_i, X_{n+1})\}(I_{iu} - \kappa_u)\right]^2 &\leq \sum_j^{n+1} E\left[\{K_{h_n}(X_j)(I_{ju} - \kappa_u)\}^2\right] \\ &\leq \sum_j^{n+1} E\left[\{K_{h_n}(X_j)\}^2\right] \\ &= \sum_j^{n+1} h_n^{-1} \int \{K(t)\}^2\, dt \sup|f(x)| \\ &= nO_p(h_n^{-1}). \end{aligned}$$

Also by the Minkowski triangle inequality on the L_2 space, we have

$$\begin{aligned} & E\left(\left[\sum_{i=1}^{n} K_{h_n}\{\eta(X_i, X_{n+1})\}(I_{iu} - \kappa_u)\right]^2\right)^{1/2} \\ \leq\ & E\left(\left[\sum_{i=1}^{n+1} K_{h_n}\{\eta(X_i, X_{n+1})\}(I_{iu} - \kappa_u)\right]^2\right)^{1/2} + E\left[\{K_{h_n}(X_{n+1})(I_{iu} - \kappa_u)\}^2\right]^{1/2} \\ =\ & n^{1/2} O_p(h_n^{-1/2}), \end{aligned}$$

which implies

$$E\left[\left\{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})(I_{iu} - \kappa_u)\right\}^2\right] = nO_p(h_n^{-1}),$$

and as a result, by Condition (A6), we have

$$\left\{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})(I_{iu} - \kappa_u)\right\} X_{n+1} = O_p(n^{1/2} h_n^{-1/2}).$$

This proves the result. □

Corollary 1. $\sum_{i=1}^{n}(I_{iu} - \kappa_u)X_i = O_p(nh_n^2 + n^{1/2}h_n^{-1/2})$.

Proof: First we can write

$$\left|\sum_{i=1}^{n}(I_{iu} - \kappa_u)X_i\right|$$

$$
\begin{aligned}
= & \left|\sum_{j=1}^{n}\sum_{i=1}^{n}\{nf(X_j)\}^{-1}K_{h_n}(X_i-X_j)(I_{iu}-\kappa_u)X_j\right| \\
& +\left|\sum_{i=1}^{n}(I_{iu}-\kappa_u)X_i\right|-\left|\sum_{j=1}^{n}\sum_{i=1}^{n}\{nf(X_j)\}^{-1}K_{h_n}(X_i-X_j)(I_{iu}-\kappa_u)X_j\right|.
\end{aligned}
$$

Note that

$$
\begin{aligned}
& E\left[\left|\sum_{j=1}^{n}\sum_{i=1}^{n}\{nf(X_j)\}^{-1}K_{h_n}(X_i-X_j)(I_{iu}-\kappa_u)X_j\right|\right] \\
\leq & \; n^{-1}\sum_{j=1}^{n}\left[E\left\{\left|\sum_{i=1}^{n}f(X_j)^{-1}K_{h_n}(X_i-X_j)(I_{iu}-\kappa_u)\right|\right\}^2\right]^{1/2}\{E(X_j^2)\}^{1/2} \\
= & \; O_p(n^{1/2}h_n^{-1/2})
\end{aligned}
$$

by Lemma 2.

As a result, we have

$$
\left|\sum_{j=1}^{n}\sum_{i=1}^{n}\{nf(X_j)\}^{-1}K_{h_n}(X_i-X_j)(I_{iu}-\kappa_u)X_j\right| = O_p(n^{1/2}h_n^{-1/2}). \tag{S.2}
$$

Also note that

$$
\begin{aligned}
& \left|\left|\sum_{i=1}^{n}(I_{iu}-\kappa_u)X_i\right|-\left|\sum_{j=1}^{n}\sum_{i=1}^{n}\{nf(X_j)\}^{-1}K_{h_n}(X_i-X_j)(I_{iu}-\kappa_u)X_j\right|\right| \\
\leq & \sum_{i=1}^{n}\left|(I_{iu}-\kappa_u)\left[X_i-\sum_{j=1}^{n}\{nf(X_j)\}^{-1}K_{h_n}(X_i-X_j)X_j\right]\right| \\
\leq & \sum_{i=1}^{n}\left|X_i-\sum_{j=1}^{n}\{nf(X_j)\}^{-1}K_{h_n}(X_i-X_j)X_j\right|.
\end{aligned}
$$

The last equation is of order $O_p(nh_n^2+n^{1/2}h_n^{-1/2})$, because

$$
\begin{aligned}
& \sum_{j=1}^{n}\{nf(X_j)\}^{-1}K_{h_n}(X_i-X_j)X_j \\
= & \; f(X_i)\sum_{j=1}^{n}f(X_i)^{-1}\{nf(X_j)\}^{-1}K_{h_n}(X_i-X_j)X_j \\
= & \; f(X_i)E\left[\{f(X_j)\}^{-1}X_j \mid X_j=X_i\right]+O_p\{h_n^2+(nh_n)^{-1/2}\} \\
= & \; X_i+O_p\{h_n^2+(nh_n)^{-1/2}\}.
\end{aligned}
$$

The second to the last equality holds because $\sum_{j=1}^{n} f(X_i)^{-1} \{nf(X_j)\}^{-1} K_{h_n}(X_i - X_j)X_j$ is a fixed design kernel estimator of $X_i f(X_i)^{-1} = E\left[\{f(X_j)\}^{-1} X_j \mid X_j = X_i\right]$, while its mean squared error is of order $O\{h_n^4 + (nh_n)^{-1}\}$ (Härdle 2004).

In conjunction with (S.2), we have

$$\left|\sum_{i=1}^{n}(I_{iu} - \kappa_u)X_i\right| = O_p(nh_n^2 + n^{1/2}h^{-1/2}).$$

This proves the result. □

Remark 5. *In the above derivations, X_i in $\sum_{i=1}^{n}(I_{iu}-\kappa_u)X_i$ does not affect the convergence rates. Therefore, the convergence rates are the same when considering $\sum_{i=1}^{n}(I_{iu} - \kappa_u)Z_i$ for any other integrable random variable Z_i.*

Lemma 3. *Let $f(\cdot)$ be the density of X, then we have*

$$\left| \int \frac{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})(I_{iu} - \kappa_u)}{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})} X_{n+1} f(X_{n+1}) dX_{n+1} \right.$$
$$\left. - \int \frac{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})(I_{iu} - \kappa_u)X_{n+1}}{nf(X_{n+1})} f(X_{n+1}) dX_{n+1} \right| = O_p\{(nh_n)^{-1}\}.$$

Proof: We have that

$$\begin{aligned}
&\left| \int \frac{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})(I_{iu} - \kappa_u)}{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})} f(X_{n+1}) dX_{n+1} \right. \\
&\left. - \int \frac{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})(I_{iu} - \kappa_u)}{nf(X_{n+1})} f(X_{n+1}) dX_{n+1} \right| \\
\leq\ & \left\{ \int \left| \sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})(I_{iu} - \kappa_u) \right|^2 f(X_{n+1}) dX_{n+1} \right\}^{1/2} \\
&\times \left\{ \int \left| \frac{f(X_{n+1}) - n^{-1}\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})}{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1}) f(X_{n+1})} \right|^2 f(X_{n+1}) dX_{n+1} \right\}^{1/2} \\
=\ & O_p(n^{1/2}h_n^{-1/2}) \left\{ \int \left| \frac{f(X_{n+1}) - n^{-1}\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})}{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1}) f(X_{n+1})} \right|^2 f(X_{n+1}) dX_{n+1} \right\}^{1/2} \\
=\ & O_p(n^{1/2}h_n^{-1/2}) O_p(h_n^2 + n^{-1/2}h_n^{-1/2}) O_p(n^{-1}) \\
=\ & O_p(n^{-1/2}h_n + n^{-1}h_n^{-1})
\end{aligned}$$

$$= \quad O_p(n^{-1}h_n^{-1}).$$

The first equality is a result from Lemma 2. The second equality holds because for each $X_i, i = 1, \ldots, n$,

$$\left| \sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1}) \right| = O_p(n)$$

and

$$\left| f(X_{n+1}) - n^{-1} \sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1}) \right| = O_p\{h_n^2 + (nh_n)^{-1/2}\},$$

which follows the uniform convergence of the kernel density estimator (Silverman 1978). With Condition (A6), we obtain the desired result. □

Lemma 4. *For a constant ρ_0 and $n > n_0 > \rho_0$, we define $A_n = \prod_{l=n_0}^{n}(1 - \rho_0/l)^{-1}$, then we have $\lim_{n\to\infty} n^{-\rho_0} A_n = A_0$, where $A_0 = n_0^{-\rho_0}$.*

Proof: The limiting result shown below follows the convergence of the product integral. We define $t_l = l/n, l = n_0 - 1, \ldots, n$, $n(t) = l$, for $t_l \le t < t_{l+1}$. For $t \ge t_{n_0}$, let $P(t) = \sum_{t_{n_0} \le t_l \le t} 1/n(t_l) = \int_{s\in[t_{n_0},t]} n(s)^{-1} dn(s)$. For $t < t_{n_0}$, define $P(t) = 0$. Note that $n(s)$ changes its values only when $s = t_l$, so $dn(s)$ is nonzero only at $s = t_l, l = n_0, \ldots, n$. Therefore, we can write

$$\begin{aligned}
\lim_{n\to\infty} A_n^{-1} &= \lim_{\sup_{n_0\le l\le n}|t_l - t_{l-1}|\to 0} A_{n(t_n)}^{-1} \\
&= \lim_{\sup_{n_0\le l\le n}|t_l - t_{l-1}|\to 0} \prod_{t_l=t_{n_0}}^{t_n} \{1 - \rho_0 P'(t_l) dt_l\} \\
&= \lim_{\sup_{n_0\le l\le n}|t_l - t_{l-1}|\to 0} \prod_{t_l=t_{n_0}}^{t_n} \{1 - \rho_0 \int_{t_{l-1}}^{t_l} P'(t)dt\}.
\end{aligned}$$

As $n \to \infty$, or $\sup_{n_0\le l\le n}|t_l - t_{l-1}| \to 0$, the above form is a product limit in Definition 1 in Gill & Johansen (1990). Similar to Example 2.5.6 in Slavík & Karlova (2007), this product limit can be written as

$$\exp\left(-\rho_0 \int_{t\in[t_{n_0},t_n]} dP(t)\right)$$

$$
\begin{aligned}
&= \exp\left(-\rho_0 \int_{s\in[t_{n_0},t_n]} n(t)^{-1} dn(t)\right) \\
&= \exp\left(-\rho_0[\log\{n(t_n)\} - \log\{n(t_{n_0})\}]\right) \\
&= \exp[-\rho_0\{\log(n) - \log(n_0)\}] \\
&= n_0^{\rho_0} n^{-\rho_0}.
\end{aligned}
$$

Therefore, $\lim_{n\to\infty} n^{-\rho_0} A_n = A_0$. This proves the result. □

S.2 Proof of Lemma 1

To assess the properties of D_{nu}, we first note that for $n > n_0$ and $u < m$,

$$
\begin{aligned}
& E(I_{(n+1)u}|\mathcal{F}_n, X_{n+1}) - \kappa_u \\
= \; & \pi_u\{\mathbf{U}_n(X_{n+1})\} - \kappa_u \\
= \; & \pi_u(\boldsymbol{\kappa}) + \pi_u^{'}(\boldsymbol{\kappa})\{\mathbf{U}_n(X_{n+1}) - \boldsymbol{\kappa}\} + 1/2\{\mathbf{U}_n(X_{n+1}) - \boldsymbol{\kappa}\}^{\mathrm{T}}\pi_u^{''}(\boldsymbol{\kappa})\{\mathbf{U}_n(X_{n+1}) - \boldsymbol{\kappa}\}(\{1 + o_p(1)\} - \kappa_u \\
= \; & \pi_{uu}'(\boldsymbol{\kappa})\{U_{nu}(X_{n+1}) - \kappa_u\} + \sum_{r\neq u, r=1}^{m-1} \pi_{ur}'(\boldsymbol{\kappa})\{U_{nr}(X_{n+1}) - \kappa_r\} \\
& +1/2\{\mathbf{U}_n(X_{n+1}) - \boldsymbol{\kappa}\}^{\mathrm{T}}\pi_u^{''}(\boldsymbol{\kappa})\{\mathbf{U}_n(X_{n+1}) - \boldsymbol{\kappa}\}\{1 + o_p(1)\} \\
= \; & -\rho\{U_{nu}(X_{n+1}) - \kappa_u\} + 1/2\{\mathbf{U}_n(X_{n+1}) - \boldsymbol{\kappa}\}^{\mathrm{T}}\pi_u^{''}(\boldsymbol{\kappa})\{\mathbf{U}_n(X_{n+1}) - \boldsymbol{\kappa}\}\{1 + o_p(1)\}.
\end{aligned}
$$

The third equality holds by Remark 1 that $\pi_u(\boldsymbol{\kappa}) = \kappa_u$. The last equality holds because by Remark 2, $\pi_{ur}'(\boldsymbol{\kappa}) = 0$ when $r = 1, \ldots, m-1, r \neq u$, and by Remark 3, $\pi_{uu}' = -\rho$.

Multiplying the above equation by X_{n+1} and taking expectation with respect to X_{n+1}, we have

$$
E\{(I_{(n+1)u} - \kappa_u)X_{n+1}|\mathcal{F}_n\} = -\rho E[\{U_{nu}(X_{n+1}) - \kappa_u\}X_{n+1}|\mathcal{F}_n] + \gamma_{1nu},
$$

where

$$
\gamma_{1nu} \equiv E\left[1/2\{\mathbf{U}_n(X_{n+1}) - \boldsymbol{\kappa}\}^{\mathrm{T}}\pi_u^{''}(\boldsymbol{\kappa})\{\mathbf{U}_n(X_{n+1}) - \boldsymbol{\kappa}\}X_{n+1}\middle|\mathcal{F}_n\right]\{1 + o_p(1)\}.
$$

Also, for $u = 1, \ldots, m-1$, we have

$$\begin{aligned}
& E[\{U_{nu}(X_{n+1}) - \kappa_u\}X_{n+1}|\mathcal{F}_n] \\
= & \int \{U_{nu}(X_{n+1}) - \kappa_u\}X_{n+1} f(X_{n+1}) dX_{n+1} \\
= & n^{-1}\sum_{i=1}^{n}\int K_{h_n}(X_i - X_{n+1})(I_{iu} - \kappa_u)X_{n+1}dX_{n+1} + \gamma_{2nu} \\
= & n^{-1}\sum_{i=1}^{n}(I_{iu} - \kappa_u)X_i + \gamma_{2nu},
\end{aligned}$$

where

$$\begin{aligned}
\gamma_{2nu} \equiv & \int \frac{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})(I_{iu} - \kappa_u)}{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})} X_{n+1} f(X_{n+1}) dX_{n+1} \\
& - \int \frac{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})(I_{iu} - \kappa_u)X_{n+1}}{nf(X_{n+1})} f(X_{n+1}) dX_{n+1}.
\end{aligned}$$

This gives

$$E\{(I_{(n+1)u} - \kappa_u)X_{n+1}|\mathcal{F}_n\} = -\rho n^{-1}\sum_{i=1}^{n}(I_{iu} - \kappa_u)X_i + \gamma_{1nu} - \rho\gamma_{2nu},$$

for $u = 1, \ldots, m-1$.

Define $\alpha_{nu} = 1 - \rho/n$ for $n \geq n_0$, and $\alpha_{nu} = 1$ otherwise, and let $\beta_{nu} = \gamma_{1nu} - \rho\gamma_{2nu}$ for $n \geq n_0$, and $\beta_{nu} = 0$ otherwise, for $u = 1, \ldots, m-1$. We have

$$E(D_{(n+1)u}|\mathcal{F}_n) = \alpha_{nu}D_{nu} + \beta_{nu}.$$

Combining the results of $D_{(n+1)u}, u = 1, \ldots, m-1$, we have

$$E(\mathbf{D}_{n+1}|\mathcal{F}_n) = \boldsymbol{\alpha}_n\mathbf{D}_n + \boldsymbol{\beta}_n,$$

where $\boldsymbol{\alpha}_n = \text{diag}\{\alpha_{n1}, \ldots, \alpha_{n(m-1)}\}$ and $\boldsymbol{\beta}_n = (\beta_{n1}, \ldots, \beta_{n(m-1)})^{\mathrm{T}}$. Let $A_{nu} = \prod_{l=1}^{n-1}\alpha_{lu}^{-1} = \prod_{l=n_0}^{n-1}\alpha_{lu}^{-1}, B_{nu} = \sum_{l=1}^{n-1}A_{(l+1)u}\beta_{lu} = \sum_{l=n_0}^{n-1}A_{(l+1)u}\beta_{lu}$, and define $M_{nu} = A_{nu}D_{nu} - B_{nu}$. It is easy to verify that $M_{iu} = D_{iu}$ for $i \leq n_0$. For $n > n_0$, we have

$$E(M_{(n+1)u}|\mathcal{F}_n) = A_{(n+1)u}(\alpha_{nu}D_{nu} + \beta_{nu}) - \sum_{l=n_0}^{n}A_{(l+1)u}\beta_{lu}$$

$$
\begin{aligned}
&= A_{nu} D_{nu} - \sum_{l=n_0}^{n-1} A_{(l+1)u} \beta_{lu} \\
&= M_{nu}.
\end{aligned}
$$

Further, X_i and $I_i, i = 1, \ldots, n$, and their continuous functions, D_{nu} and B_{nu}, have finite second moments by Condition (A6). Therefore, $E(|M_{nu}|) < \infty$, which implies M_{nu} is a martingale. We further define $\Delta M_{nu} = M_{nu} - M_{(n-1)u}$ to be a martingale difference. Combining the results for arm u, the vector $\mathbf{M}_n = (M_{n1}, \ldots, M_{n(m-1)})^{\mathrm{T}}$ is a martingale vector, and $\Delta \mathbf{M}_n = (\Delta M_{n1}, \ldots, \Delta M_{n(m-1)})^{\mathrm{T}}$ is a vector of martingale differences. We further define $\mathbf{A}_n = \mathrm{diag}(A_{n1}, \ldots, A_{n(m-1)})$, and $\mathbf{B}_n = (B_{n1}, \ldots, B_{n(m-1)})^{\mathrm{T}}$.

Now we assess the asymptotic properties of $\mathbf{D}_n$ through $\mathbf{M}_n$ by utilizing martingale techniques. We first derive the asymptotic properties of $\mathbf{z}^{\mathrm{T}}\mathbf{M}_n$, where $\mathbf{z}$ is an arbitrary $m-1$ dimensional vector, and then we show that the term $\mathbf{B}_n$ is ignorable because it converges faster to 0 than $\mathbf{M}_n$. Note that $\mathbf{z}^{\mathrm{T}}\mathbf{M}_n$ is a martingale while $\mathbf{z}^{\mathrm{T}}\Delta\mathbf{M}_n$ is a martingale difference, because the linear function does not alter the expectation and boundedness properties.

Let $\mathbf{s}_n = E(\mathbf{M}_n \mathbf{M}_n^{\mathrm{T}})$, according to the martingale invariance principle introduced on page 99 in Hall & Heyde (1980), if we have

$$
(\mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z})^{-1} \sum_{i=1}^{n} \mathbf{z}^{\mathrm{T}} \Delta\mathbf{M}_i (\Delta\mathbf{M}_i)^{\mathrm{T}} \mathbf{z} \xrightarrow{p} 1 \tag{S.3}
$$

$$
(\mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z})^{-1} \sum_{i=1}^{n} E[\mathbf{z}^{\mathrm{T}} \Delta\mathbf{M}_i (\Delta\mathbf{M}_i)^{\mathrm{T}} \mathbf{z} \; I\{|\mathbf{z}^{\mathrm{T}}\Delta\mathbf{M}_i| > \epsilon(\mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z})^{1/2}\}] \to 0, \forall \epsilon > 0 \tag{S.4}
$$

as $n \to \infty$, then $(\mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z})^{-1/2}\mathbf{z}^{\mathrm{T}}\mathbf{M}_n$ converges weakly to a standard normal random variable, and in turn $\mathbf{s}_n^{-1/2}\mathbf{M}_n$ converges to a multivariate standard normal vector. Thus, (S.3) readily holds by Chebyshev's inequality for the uncorrelated random variables and

$$
\begin{aligned}
& \mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z} \\
= \; & E(\mathbf{z}^{\mathrm{T}}\mathbf{M}_n\mathbf{M}_n^{\mathrm{T}}\mathbf{z})
\end{aligned}
$$

$$= E\left[\mathbf{z}^{\mathrm{T}}\left\{\begin{array}{ccc}\sum_{i=1}^{n}(\Delta M_{i1})^2 & \dots & \sum_{i=1}^{n}(\Delta M_{i1}\Delta M_{i(m-1)}) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n}(\Delta M_{i1}\Delta M_{i(m-1)}) & \dots & \sum_{i=1}^{n}(\Delta M_{i(m-1)})^2\end{array}\right\}\mathbf{z}\right]$$
$$= \sum_{i=1}^{n} E\{\mathbf{z}^{\mathrm{T}}\Delta\mathbf{M}_i(\Delta\mathbf{M}_i)^{\mathrm{T}}\mathbf{z}\}.$$

The second equality holds because for $i < j$ and arms u and v, we have $E(\Delta M_{iu}\Delta M_{jv}) = E\{\Delta M_{iu}E(\Delta M_{jv}|\mathcal{F}_{j-1})\} = 0$.

If (S.4) holds, then the martingale invariance principle allows us to show the asymptotic properties of $\mathbf{z}^{\mathrm{T}}\mathbf{M}_n$ through accessing the convergence of $\mathbf{s}_n$. Therefore, in the following, we proceed to find the exact form of $\mathbf{s}_n$ and verify (S.4).

Let $s_{nuu} = \sum_{i=1}^{n} E\{(\Delta M_{iu})^2\}$ and $s_{nuv} = \sum_{i=1}^{n} E(\Delta M_{iu}\Delta M_{ju})$, and we examine the convergence of each term s_{nuv} in the matrix $\Delta\mathbf{M}_i(\Delta\mathbf{M}_i)^{\mathrm{T}}$. Note that for $n > n_0$,

$$A_{nu}^{-1}\Delta M_{nu} = (I_{nu} - \kappa_u)X_n + \rho D_{(n-1)u}/(n-1) - \beta_{(n-1)u}. \tag{S.5}$$

By Corollary 1 and Condition (A4) that $nh_n^2 \to \infty$, we have

$$\rho D_{(n-1)u}/(n-1) = O_p(h_n^2 + n^{-1/2}h_n^{-1}) = o_p(1).$$

Next, $\gamma_{2nu} = O_p\{(nh_n)^{-1}\} = o_p(1)$ by Lemma 3. In addition, from

$$\gamma_{1nu} = E\left[1/2\{\mathbf{U}_n(X_{n+1}) - \boldsymbol{\kappa}\}^{\mathrm{T}}\pi_u^{''}(\boldsymbol{\kappa})\{\mathbf{U}_n(X_{n+1}) - \boldsymbol{\kappa}\}X_{n+1}\middle|\mathcal{F}_n\right]\{1 + o_p(1)\},$$

by the boundedness of $\pi_u^{''}$, γ_{1nu} has the same order as

$$E\left[\left\{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})\right\}^{-2}\left\{\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1})(I_{iu} - \kappa_u)\right\}^2 X_{n+1}\middle|\mathcal{F}_n\right]$$
$$= O_p(n^{-2})O_p(nh_n^{-1})$$
$$= O_p\{(nh_n)^{-1}\}$$

by Lemma 2, and the fact that $\sum_{i=1}^{n} K_{h_n}(X_i - X_{n+1}) = O_p(n)$.

These together with Lemma 3 imply $|\beta_{nu}| = o_p(1)$. Therefore,

$$A_{nu}^{-1}\Delta M_{nu} = (I_{nu} - \kappa_u)X_n + o_p(1). \tag{S.6}$$

Further note that

$$\begin{aligned}
& n^{-1-2\rho}\sum_{i=1}^{n}(\Delta M_{iu})^2 \\
= \;& n^{-1-2\rho}\sum_{i=1}^{n} A_{iu}^2(A_{iu}^{-1}\Delta M_i)^2 \\
= \;& n^{-1-2\rho}\sum_{i=1}^{n} A_{iu}^2\{(I_{iu} - \kappa_u)X_i + o_p(1)\}^2 \\
= \;& \left[n^{-1-2\rho}\sum_{i=1}^{n} A_{iu}^2\left\{(1-\kappa_u)\kappa_u X_i^2\right\}\right. \\
& \left. + n^{-1}\sum_{i=1}^{n}(A_{iu}/n^{\rho})^2\{(I_{iu} - \kappa_u)(1-2\kappa_u)X_i^2\}\right]\{1 + o_p(1)\} \\
= \;& \left\{n^{-1-2\rho}(1-\kappa_u)\kappa_u\sum_{i=1}^{n} A_{iu}^2 X_i^2 + o_p(1)\right\}\{1 + o_p(1)\}.
\end{aligned} \tag{S.7}$$

The second equality holds by directly plugging in (S.6). Strictly speaking, (S.6) can be used only when i is large. However, since the value on the first finitely many terms do not affect the final asymptotic results, we do not make distinction here. This practice also applies similarly in the remaining text. The last equality follows Remark 5 and the fact that A_{iu}/n^{ρ} is bounded due to Lemma 4. For a given $\xi > 0$

$$\lim_{n\to\infty}\Pr\left[n^{-1-2\rho}\left|\sum_{i=1}^{n} A_{iu}^2 X_i^2 - E(X_i^2)\sum_{i=1}^{n} A_{iu}^2\right| > \xi\right] \le \lim_{n\to\infty} n^{-2-4\rho}\xi^{-2}\sum_{i=1}^{n} A_{iu}^4 \mathrm{var}(X_i^2). \tag{S.8}$$

To inspect the right-hand side of (S.8), using Lemma 4 and following the same argument to that in the proof of Theorem 1 in Smith (1984), we have

$$\frac{n^{-2-4\rho}\sum_{i=1}^{n} A_{iu}^4}{n^{-2-4\rho}\sum_{i=1}^{n} A_{0u}^4 i^{4\rho}} \to 1 \tag{S.9}$$

in probability, as $n \to \infty$. Further,

$$n^{-2}\sum_{i=1}^{n}(i/n)^{4\rho} \to n^{-1}\int_0^1 x^{4\rho}dx = n^{-1}(1+4\rho)^{-1}.$$

Thus, the right-hand side of (S.8) goes to 0. This shows that

$$n^{-1-2\rho}\left|\sum_{i=1}^{n} A_{iu}^2 X_i^2 - E(X_i^2)\sum_{i=1}^{n} A_{iu}^2\right|$$

converges to 0 in probability.

Now, we assess the limit of $n^{-1-2\rho}E(X_i^2)\sum\limits_{i=1}^{n} A_{iu}^2$. Similar to the previous argument, as $n \to \infty$,

$$\frac{n^{-1-2\rho}\sum_{i=1}^{n} A_{iu}^2}{n^{-1-2\rho}A_{0u}^2\sum_{i=1}^{n} i^{2\rho}} \xrightarrow{p} 1,$$

and

$$n^{-1}\sum_{i=1}^{n}(i/n)^{2\rho} \to \int_0^1 x^{2\rho}dx = (1+2\rho)^{-1}.$$

Therefore,

$$n^{-1-2\rho}E(X_i^2)\sum_{i=1}^{n} A_{iu}^2 \to (1+2\rho)^{-1}A_{0u}^2 E(X_i^2),$$

and hence

$$n^{-1-2\rho}\sum_{i=1}^{n} A_{iu}^2 X_i^2 \xrightarrow{p} (1+2\rho)^{-1}A_{0u}^2 Q.$$

Plugging the result into (S.7), we have

$$n^{-1-2\rho}\sum_{i=1}^{n}(\Delta M_{iu})^2 \to (1+2\rho)^{-1}A_0^2(1-\kappa_u)\kappa_u E(X_i^2) = (1+2\rho)^{-1}(1-\kappa_u)\kappa_u A_{0u}^2 Q \quad \text{(S.10)}$$

in probability as $n \to \infty$.

Similarly, for $\sum_{i=1}^{n}\Delta M_{iu}\Delta M_{iv}$, $u \neq v$, we have

$$\begin{aligned}
& n^{-1-2\rho}\sum_{i=1}^{n}\Delta M_{iu}\Delta M_{iv} \\
= \;& n^{-1-2\rho}\left(\sum_{i=1}^{n} A_{iu}A_{iv}A_{iu}^{-1}\Delta M_{iu}A_{iv}^{-1}\Delta M_{iv}\right) \\
= \;& n^{-1-2\rho}\left\{\sum_{i=1}^{n} A_{iu}A_{iv}(-I_{iu}\kappa_v - I_{iv}\kappa_u + \kappa_u\kappa_v)X_i^2\right\}\{1+o_p(1)\}
\end{aligned}$$

$$
\begin{aligned}
&= \Bigg[- n^{-1-2\rho} \left(\sum_{i=1}^{n} A_{iu} A_{iv} \kappa_u \kappa_v X_i^2 \right) - n^{-1-2\rho} \left\{ \sum_{i=1}^{n} A_{iu} A_{iv} (I_{iu} - \kappa_u) \kappa_v X_i^2 \right\} \\
&\quad - n^{-1-2\rho} \left\{ \sum_{i=1}^{n} A_{iu} A_{iv} (I_{iv} - \kappa_v) \kappa_u X_i^2 \right\} \Bigg] \{1 + o_p(1)\} \\
&= \Bigg[- n^{-1} \left\{ A_{0u} A_{0v} \sum_{i=1}^{n} (i/n)^{2\rho} \kappa_u \kappa_v E(X_i^2) \right\} - n^{-1-2\rho} \left\{ \sum_{i=1}^{n} A_{iu} A_{iv} (I_{iu} - \kappa_u) \kappa_v X_i^2 \right\} \\
&\quad - n^{-1-2\rho} \left\{ \sum_{i=1}^{n} A_{iu} A_{iv} (I_{iv} - \kappa_v) \kappa_u X_i^2 \right\} \Bigg] \{1 + o_p(1)\} \\
&= -(1+2\rho)^{-1} A_{0u} A_{0v} \kappa_u \kappa_v E(X_i^2) + o_p(1).
\end{aligned}
$$

The second equality holds because $I_{iu} I_{iv} = 0$ for $u \neq v$. The third equality holds because

$$
\begin{aligned}
&\lim_{n\to\infty} \Pr \left\{ n^{-1-2\rho} \left| \sum_{i=1}^{n} A_{iu} A_{iv} X_i^2 - E(X_i^2) \sum_{i=1}^{n} A_{iu} A_{iv} \right| > \xi \right\} \\
\leq\ &\lim_{n\to\infty} n^{-2-4\rho} \sum_{i=1}^{n} A_{iu}^2 A_{iv}^2 \operatorname{var}(X_i^2) \xi^{-2} \\
\leq\ &\lim_{n\to\infty} (n^{-2-4\rho} \sum_{i=1}^{n} A_{iu}^4 + n^{-1-2\rho} \sum_{i=1}^{n} A_{iv}^4) \operatorname{var}(X_i^2) \xi^{-2}/2,
\end{aligned}
$$

which goes to 0 by (S.9). The fourth equality holds because

$$
\frac{n^{-1-2\rho} \sum_{i=1}^{n} A_{iu} A_{iv}}{n^{-1-2\rho} A_{0u} A_{0v} \sum_{i=1}^{n} i^{2\rho}} \to 1
$$

in probability, and

$$
n^{-1} \sum_{i=1}^{n} (i/n)^{2\rho} \to \int_0^1 x^{2\rho} dx = (1+2\rho)^{-1}.
$$

Finally, the last equality holds by Remark 5 and the fact that A_{iu}/n^{ρ} is bounded due to Lemma 4.

Now we proceed to show the convergence of s_{nuv}. But note that if $n^{-1-2\rho} \sum_{i=1}^{n} (\Delta M_{iu})^2$ and $n^{-1-2\rho} \sum_{i=1}^{n} \Delta M_{iu} \Delta M_{iv}$ are dominated by integrable functions, then the asymptotic properties of s_{nuu} and $s_{nuv}, u \neq v$, can be derived easily by using the dominated convergence theorem. Further, since

$$
\left| n^{-1-2\rho} \sum_{i=1}^{n} \Delta M_{iu} \Delta M_{iv} \right| \leq n^{-1-2\rho} \sum_{i=1}^{n} \left| \Delta M_{iu} \Delta M_{iv} \right|
$$

$$\leq \quad 1/2n^{-1-2\rho}\left\{\sum_{i=1}^{n}(\Delta M_{iu})^2+\sum_{i=1}^{n}(\Delta M_{iv})^2\right\},$$

we need to show the boundedness of $n^{-1-2\rho}\sum_{i=1}^{n}(\Delta M_{iu})^2$ for obtaining the convergence result. Thus, we evaluate the upper bound of $n^{-1-2\rho}\sum_{i=1}^{n}(\Delta M_{iu})^2$ as follows. Because there exists a constant $C_1<\infty$,

$$\begin{aligned}
& n^{-1-2\rho}\sum_{i=1}^{n}(\Delta M_{iu})^2 \\
= \quad & n^{-1-2\rho}\left\{\sum_{i\leq n_0}(\Delta M_{iu})^2+\sum_{i>n_0}A_{iu}^2(A_{iu}^{-1}\Delta M_{iu})^2\right\} \\
\leq \quad & \left[n_0\max_{i\leq n_0}X_i^2+C_1n^{-1}\sum_{i>n_0}^{n}(i/n)^{2\rho}\left\{(I_{iu}-\kappa_u)X_i+\rho D_{i-1u}/(i-1)-\beta_{(i-1)u}\right\}^2\right] \\
\leq \quad & \left[n_0\max_{i\leq n_0}X_i^2+C_1n^{-1}\sum_{i>n_0}^{n}\left\{(I_{iu}-\kappa_u)X_i+\rho D_{i-1u}/(i-1)-\beta_{(i-1)u}\right\}^2\right], \qquad \text{(S.11)}
\end{aligned}$$

we first show the boundedness of $\rho D_{i-1u}/(i-1)$ and $\beta_{n-1u}=\gamma_{1(n-1)u}-\rho\gamma_{2(n-1)u}$.

Clearly $|\rho D_{i-1u}/(i-1)|\leq|\rho|\max_{i<n}|X_i|$. Further, since $|U_{nu}(X_{n+1})|\leq 1$ and π_u'' is bounded by Condition (A2), there exists a constant $C_2<\infty$ so that

$$\gamma_{1(n-1)u}=E\left[1/2\{\mathbf{U}_{n-1}(X_n)-\boldsymbol{\kappa}\}^{\mathrm{T}}\pi_u''(\boldsymbol{\kappa}^*)\{\mathbf{U}_{n-1}(X_n)-\boldsymbol{\kappa}\}X_n\middle|\mathcal{F}_n\right]\leq C_2m\max_{i\leq n}|X_i|,$$

where $\boldsymbol{\kappa}^*=(\kappa_1^*,\ldots,\kappa_m^*)$ with κ_u^* defined as a point on the line connecting κ_u and $U_{nu}(X_{n+1})$. In addition,

$$\begin{aligned}
\gamma_{2(n-1)u} \quad = \quad & \int\frac{\sum_{i=1}^{n}K_{h_n}(X_i-X_n)(I_{iu}-\kappa_u)}{\sum_{i=1}^{n}K_{h_n}(X_i-X_n)}X_nf(X_n)dX_n \\
& -\int\frac{\sum_{i=1}^{n}K_{h_n}(X_i-X_n)(I_{iu}-\kappa_u)X_n}{nf(X_n)}f(X_n)dX_n \\
\leq \quad & E(|X_n|)+\max_{i\leq n}|X_i|.
\end{aligned}$$

Therefore, (S.11) implies that there exist constants $C_3,C_4<\infty$ such that

$$n^{-1-2\rho}\sum_{i=1}^{n}(\Delta M_{iu})^2 \quad\leq\quad C_3\max_{i\leq n}X_i^2+C_4,$$

almost surely. Since $C_3 \max_{i\le n} X_i^2 + C_4$ is an integrable function, by the dominated convergence theorem, we have

$$\begin{aligned} n^{-1-2\rho} s_{nuu} &= n^{-1-2\rho}\sum_{i=1}^{n} E\left\{(\Delta M_{iu})^2\right\} \to (1+2\rho)^{-1}(1-\kappa_u)\kappa_u A_{0u}^2 Q, \\ n^{-1-2\rho} s_{nuv} &= n^{-1-2\rho}\sum_{i=1}^{n} E\left(\Delta M_{iu}\Delta M_{iu}\right) \to -(1+2\rho)^{-1}\kappa_u\kappa_v A_{0u}A_{0v}Q. \end{aligned} \quad \text{(S.12)}$$

These give the limiting form of $\mathbf{s}_n$ in (S.3).

To show (S.4), we first note that (S.6), (S.12) and Lemma 4 yield

$$\begin{aligned} & |(\mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z})^{-1/2}\mathbf{z}^{\mathrm{T}}\Delta\mathbf{M}_n|^2 \\ = & |(\mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z})^{-1}||\mathbf{z}^{\mathrm{T}}\Delta\mathbf{M}_n\Delta\mathbf{M}_n^{\mathrm{T}}\mathbf{z}| \\ = & \left|\left(\sum_{u=1}^{m-1}\sum_{v=1}^{m-1} z_u z_v s_{nuv}\right)^{-1}\right| \sum_{u=1}^{m-1}\sum_{v=1}^{m-1}\left| z_u z_v A_{nu}(I_{nu}-\kappa_u)A_{nv}(I_{nv}-\kappa_v)X_n^2\{1+o_p(1)\}\right| \\ = & O_p(n^{-1})\sum_{u=1}^{m-1}\sum_{v=1}^{m-1}\left| z_u z_v A_{nu}/n^{\rho}(I_{nu}-\kappa_u)A_{nv}/n^{\rho}(I_{nv}-\kappa_v)X_n^2\{1+o_p(1)\}\right| \\ \le & O_p\left\{n^{-1}\max_{u\in 1,\ldots,m-1}(I_{nu}-\kappa_u)^2 X_n^2\right\}. \end{aligned}$$

By Condition (A6), this implies

$$(\mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z})^{-1/2}\mathbf{z}^{\mathrm{T}}\Delta\mathbf{M}_n = o_p(1).$$

Further,

$$\begin{aligned} & E[\mathbf{z}^{\mathrm{T}}\Delta\mathbf{M}_i(\Delta\mathbf{M}_i)^{\mathrm{T}}\mathbf{z}I\{|\mathbf{z}^{\mathrm{T}}\Delta\mathbf{M}_i| > \epsilon(\mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z})^{1/2}\}] \\ \le & E\{(\mathbf{z}^{\mathrm{T}}\Delta\mathbf{M}_i(\Delta\mathbf{M}_i)^{\mathrm{T}}\mathbf{z})^2\}^{1/2} E\left[I\{|\mathbf{z}^{\mathrm{T}}\Delta\mathbf{M}_i| > \epsilon(\mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z})^{1/2}\}\right]^{1/2} \\ = & E\{(\mathbf{z}^{\mathrm{T}}\Delta\mathbf{M}_i(\Delta\mathbf{M}_i)^{\mathrm{T}}\mathbf{z})^2\}^{1/2} Pr\{(\mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z})^{-1/2}|\mathbf{z}^{\mathrm{T}}\Delta\mathbf{M}_i| > \epsilon\}^{1/2} \\ \to & 0, \end{aligned}$$

by Condition (A6) and because $(\mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z})^{-1/2}\mathbf{z}^{\mathrm{T}}\Delta\mathbf{M}_n$ is $o_p(1)$. This result along with the fact that $\mathbf{s}_n = O(n^{1+2\rho})$ proves (S.4). So far, we have proven that $(\mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z})^{-1/2}\mathbf{z}^{\mathrm{T}}\mathbf{M}_n$ converges to a standard normal random variable. Since $\mathbf{z}$ is an arbitrary vector, we conclude that $\mathbf{s}_n^{-1/2}\mathbf{M}_n$ converges to a multivariate standard normal vector. Next, in order to use the martingale results to show the asymptotic property of $\mathbf{D}_n$, we first show that for each u,

$$n^{-1/2}|A_{nu}^{-1}B_{nu}| \xrightarrow{p} 0.$$

Note that there exist constants $C_5, C_6, C_7 < \infty$, such that

$$\begin{aligned}
|B_{nu}/A_{nu}| &\leq \sum_{i=1}^{n} A_{iu}A_{nu}^{-1}|\beta_{i-1}| \\
&\leq \sum_{i=1}^{n} A_{iu}A_{nu}^{-1}C_5(ih_n)^{-1} \\
&\leq C_6 n^{-1}\sum_{i=1}^{n}(i/n)^{\rho-1}h_n^{-1} \\
&\leq C_7 h_n^{-1}
\end{aligned}$$

in probability. Here, we use the definition of B_{nu} to obtain the first inequality, the definition of β_{lu} and the results on the orders of $\gamma_{1nu}, \gamma_{2nu}$ lead to the second inequality, Lemma 4 yields the third inequality, and replacing average with integration we can obtain the last inequality. Therefore, together with Condition (A4) we have

$$n^{-1/2}|A_{nu}^{-1}B_{nu}| \xrightarrow{p} 0$$

by Condition (A4), and this gives

$$n^{-1/2}|A_{nu}^{-1}M_{nu} - D_{nu}| = o_p(1).$$

This convergence in probability result for a single $u, u = 1, \ldots, m-1$ implies the joint convergence in probability of the vector constructed by these elements. So, $n^{-1/2}\mathbf{D}_n$ and $n^{-1/2}\mathbf{A}_n^{-1}\mathbf{M}_n$ converge equivalently to the same limit. Also, we have shown that $(\mathbf{z}^{\mathrm{T}}\mathbf{s}_n\mathbf{z})^{-1/2}\mathbf{z}^{\mathrm{T}}\mathbf{M}_n$ converges to a standard normal random variable for an arbitrary $\mathbf{z}$. Therefore,

$$\mathbf{s}_n^{-1/2}\mathbf{M}_n \xrightarrow{d} N(0, I).$$

Further, $\mathbf{A}_n \to \text{diag}(A_{01}n^{\rho}, \ldots, A_{0m}n^{\rho})$ implies $|n^{-1}\mathbf{A}_n^{-1}\mathbf{s}_n\mathbf{A}_n^{-1} - \Omega| = o(1)$ by (S.12), where Ω is defined in the statement of Lemma 1. Hence,

$$n^{-1/2}\mathbf{A}_n^{-1}\mathbf{s}_n^{1/2}\mathbf{s}_n^{-1/2}\mathbf{M}_n \xrightarrow{d} N(0, \Omega).$$

As a result, we have

$$n^{-1/2}\Omega^{-1/2}\mathbf{D}_n \xrightarrow{d} N(0, I).$$

□

S.3 Necessary Lemmas for Theorem 1

Lemma 5. *For a new data point $\mathbf{X}_{n+1}$, we have $E\{(\sum_{i=1}^n \prod_{k=1}^p K_{h_n}(X_{ik} - X_{(n+1)k})(I_{iu} - \kappa_u)\mathbf{X}_{n+1}^{\mathrm{T}}\mathbf{z})^2\} = O\left(nh_n^{-p}\right)$.*

Corollary 2. $|\sum_{i=1}^n (I_{iu} - \kappa_u)\mathbf{X}_i^{\mathrm{T}}\mathbf{z}| = O_p\left\{nh_n^2 + n^{1/2}h_n^{-p/2}\right\}$.

Lemma 6. *Let $f(\mathbf{X}_i)$ be the density function of $\mathbf{X}_i$. We have*

$$\begin{aligned}
& \Bigg| \int \frac{\sum_{i=1}^n \prod_{k=1}^p K_{h_n}(X_{ik} - X_{(n+1)k})(I_{iu} - \kappa_u)}{\sum_{i=1}^n \prod_{k=1}^p K_{h_n}(X_{ik} - X_{(n+1)k})} \mathbf{X}_{n+1}^{\mathrm{T}}\mathbf{z} f(\mathbf{X}_{n+1}) d\mathbf{X}_{n+1} \\
& - \int \frac{\sum_{i=1}^n \prod_{k=1}^p K_{h_n}(X_{ik} - X_{(n+1)k})(I_{iu} - \kappa_u)\mathbf{X}_{n+1}^{\mathrm{T}}\mathbf{z}}{nf(\mathbf{X}_{n+1})} f(\mathbf{X}_{n+1}) d\mathbf{X}_{n+1} \Bigg| \\
= \; & O_p\left(n^{-1}h_n^{-p}\right).
\end{aligned}$$

S.4 Proof of Theorem 1

Following Lemma 5, Corollary 2 and Lemma 6, Theorem 1 holds by using the same arguments as those leading to Lemma 1.