学位論文

Bootstrap methods for ranked set samples and homogeneity tests for functional data
(順位付集合標本に対するブートストラップ法及び関数データに対する同等性検定)

2023年3月

山口　光

Acknowledgements

First and foremost, I would like to express my sincere appreciation to my advisor, Associate Professor Hidetoshi Murakami. Although I was not an excellent student, he introduced me to the world of research and shared his immense knowledge in the field, for the enormous amount of work, time, and support. Moreover, he provided me with opportunities to present our research at both domestic and international conferences. Without his guidance and continuing help, I could not complete this paper.

In addition, I am deeply grateful to my supervisor, Professor Hiroki Hashiguchi, for his insightful comments and helpful suggestions in preparing this paper. I have learned from his insight a lot. My sincere thanks also to Assistant Professor Masato Kitani for his invaluable help with my research. His sincere approach to research inspired me a lot. Furthermore, I would like to thank everyone who has continued to support me, including all members of the Murakami Laboratory at the Tokyo University of Science.

Finally, in gratitude for their unconditional support and understanding, I would like to dedicate this paper to my family.

Hikaru Yamaguchi

Contents

Chapter 1

Introduction

In many statistical analyses, the observed data are often generated through simple random sampling (SRS). One of the advantages of the SRS method is its ease of use. Unlike more complicated sampling approaches, including stratified and clustered sampling, the SRS method does not require steps such as dividing the population into sub-populations. However, the SRS method can be time-consuming and costly compared to other methods used for sampling from large populations. Thus, cost-effective sampling methods are of significant concern in statistics, particularly when measuring the characteristic of interest is a costly and time-consuming task. McIntyre (1952) proposed the so-called ranked set sampling (RSS) to effectively estimate the pasture yield in Australia effectively. The RSS procedure is a two-stage scheme. In the first stage, simple random samples are drawn, and a certain ranking mechanism is employed to rank the units in each simple random sample. In the second stage, actual measurements of the variable of interest are conducted on the selected units based on the ranking information obtained in the first stage. Despite no theoretical rigor, McIntyre (1952) indicated that RSS outperforms SRS in terms of estimating the population mean. The notion of RSS provides an effective way to achieve an observational economy under certain particular conditions. There have been many new developments based on McIntyre's original idea, which have made the method applicable within a much more comprehensive range of fields than initially intended. A mathematical foundation for RSS was first investigated by Takahasi and Wakimoto (1968). In addition, Dell and Clutter (1972) showed that the RSS-based estimator of a population mean is at least as efficient as the SRS-based estimator with the same number of observations, even when ranking errors occur. The middle of the 1980s was a turning point in the development of the theory and methodology using RSS. Since then, various statistical procedures with RSS, both nonparametric and parametric, have been investigated; in addition, variations of the original notion of RSS have been proposed and developed, and general theoretical foundations of RSS have been laid. See Bohn (1996), Wolfe (2004) and Chen et al. (2004) for details.

Furthermore, certain quantities, such as the mean and variance of the population distribution F, can be expressed as a statistical functional of the form $\theta(F)$. A straightforward method for estimating a statistical functional is the so-called plug-in method, which estimates $\theta(F)$ using $\theta(\hat{F})$, where $\hat{F}$ is an estimator of the distribution function F. The estimation of the cumulative distribution function (CDF) with various RSS settings was considered by Stokes and Sager (1988), Kvam and Samaniego (1994), Huang (1997), and Duembgen and Zamanzade (2020). To estimate the probability density function (PDF), for example, the kernel density estimator (KDE) proposed by Chen (1999), Barabesi and Fattorini (2002), and Lim et al. (2014) can be constructed as a statistical functional of the RSS-based empirical distribution function (EDF). The RSS-based KDE has recently been used in various fields, including mode estimation (Samawi et al., 2018, 2019) and reliability estimation (Mahdizadeh and Zamanzade, 2016, 2018, 2021; Yin et al., 2016). However, because the RSS data generally have a data structure with unknown ranking errors, the asymptotic distribution of an RSS-based plug-in estimator of a statistical functional cannot be readily used for a statistical inference. We have to use other methods such as bootstrapping for this purpose. The bootstrap method for use with SRS data was proposed by Efron (1979) and has been studied by several authors. Following Silverman and Young (1987), Hall et al. (1989), and Wang (1995), the smoothed bootstrap method uses a smooth estimator of the CDF rather than the EDF. De Martini (2000) presented results on the consistency of the smoothed bootstrap method for SRS data (Alonso and Cuevas, 2003; El-Nouty and Guillou, 2000). In addition, Chen et al. (2004) and Modarres et al. (2006) have investigated several bootstrap methods for drawing a statistical inference under RSS.

Meanwhile, with advances in modern technology, it has become possible to observe various types of data. In particular, functional data analysis (FDA) deals with the analysis and theory of data that are in the form of functions, images, shapes, or more general objects. Although the term "functional data analysis" was coined by Ramsay (1982) and Ramsay and Dalzell (1991), the history of this area is much older and dates back to Grenander (1950) and Rao (1958). The objects of study in FDA are real functions that are assumed to be generated by means of a stochastic process that can be handled by viewing them as random elements from probability distributions in infinite-dimensional spaces. FDA can be used for analysis in many different fields, such as finance, genomics, medicine, and chemistry. Ramsay and Silverman (2002) presented a wide range of applications of the FDA. Several techniques of multivariate data have been adapted or generalized to the FDA context. See Ramsay and Silverman (2005), Ferraty and Vieu (2006), and Horváth and Kokoszka (2012) for details in this field.

Two-sample hypothesis testing for functional data under many different situations has been considered. Several test statistics have been proposed to detect differences in the mean functions (Chakraborty and Chaudhuri, 2009; Horváth et al., 2013; Ramsay and Silverman, 2005; Zhang et al., 2010) and covariance functions (Ferraty et al., 2007; Fremdt et al., 2013; Kraus and Panaretos, 2012). Because functional data are recorded discretely in practice, a pre-smoothing of the data is necessary in many cases. However, for all discretely observed functional data, the use of different tuning parameters for the smoothing step can mask the differences between distributions that a test attempts to locate. Hall and Van Keilegom (2007) proposed an extension of the multivariate Cramèr–von Mises test and investigated the effect of pre-smoothing on the testing procedures. Furthermore, they verified that the effect of smoothing can be reduced by using a common tuning parameter for all observed functional data. This method was developed for noisy functional data observed at dense grids of points. However, functional data observed in fields such as diffusion tensor imaging (DTI) may contain numerous missing values. As a method applicable to a more general sampling design, Pomann et al. (2016) proposed the Anderson–Darling test based on the so-called marginal functional principal analysis (Crainiceanu et al., 2009; Di et al., 2009; Goldsmith et al., 2012; Yao et al., 2005).

The remainder of this paper is organized as follows. In Chapter 2, we propose several smoothed bootstrap methods for RSS. In addition, we detail the development of a more efficient resampling method when the underlying distribution is symmetric. These results are based on Yamaguchi and Murakami (2021). In Chapter 3, we propose an interpoint distance-based test for functional data. Furthermore, we derive some asymptotic properties of the proposed test statistics. These results are based on Yamaguchi and Murakami (2022). Finally, we provide some concluding remarks in Chapter 4.

Chapter 2

Smoothed bootstrap methods for ranked set samples

In this chapter, we develop the smoothed bootstrap methods for RSS. In particular, we also propose a more efficient resampling method when the underlying distribution is symmetric. Chapter 2 is organized as follows. In Section 2.1, we introduce the imperfect ranking model for RSS. Moreover, the RSS-based kernel cumulative distribution estimator (KCDF) is defined. In Section 2.2, we describe several bootstrap methods and their asymptotic properties. In Section 2.3, we derive the asymptotic mean integrated squared error (MISE) of the RSS-based KCDE for selecting the bandwidth for the KCDE based on RSS data. Simulation studies using these bootstrap methods are discussed in Section 2.4.

2.1 Fundamentals of ranked set sampling

First, we introduce some of the notation used. The PDF and CDF of the underlying distribution are denoted by f and F, respectively. Let $F_{[i]}$ be the CDF of the i-th judgment-ranked order statistic of a random sample of size k from F. In this paper, we assume that the ranking mechanism is (so-called) consistent (Chen et al., 2004, p.12), that is, the following equality holds:

$$F = \frac{1}{k}\sum_{i=1}^{k} F_{[i]}. \tag{2.1.1}$$

Let $\mathcal{M}$ be an imperfect ranking model that satisfies the equality (2.1.1). For example, Frey (2007) introduced the imperfect ranking model

$$F_{[i]} = \sum_{j=1}^{r} p_{ij} F_{j:r},$$

where $F_{j:r}$ is the CDF of the true j-th order statistic from a set of size r ($\geq k$) and $P = [p_{ij}]_{k\times r}$ is the $k \times r$ non-negative matrix satisfying the row sums of 1 and column sums of k/r. The ranking mechanism based on Frey's imperfect ranking model is consistent since the following equation holds:

$$\frac{1}{k}\sum_{i=1}^{k} F_{[i]} = \frac{1}{k}\sum_{j=1}^{r}\sum_{i=1}^{k} p_{ij}F_{j:r} = \frac{1}{r}\sum_{j=1}^{r} F_{j:r} = F.$$

In particular, if $r = k$, Fray's imperfect ranking model corresponds to the imperfect ranking model introduced by Bohn and Wolfe (1994). Let m_i $(i = 1, \ldots, k)$ be the number of sets allocated to measure units having the i-th judgment-rank. A ranked set sample can be represented by $\{X_{[i]j} : i = 1, \ldots, k,\ j = 1, \ldots, m_i\}$, where $X_{[i]j}$ is the measurement on the j-th unit, which is judged to be the i-th ranked observation in a set of k independent samples. Throughout this paper, for each $i = 1, \ldots, k$, we assume that $\lim_{n\to\infty} m_i/n = \lambda_i \in (0, 1)$, where $n = \sum_{i=1}^{k} m_i$.

The nonparametric bootstrap method for SRS data can be interpreted as random sampling from the EDF. Some nonparametric bootstrap methods for RSS data are also based on random sampling from the RSS-based EDF defined in

$$\hat{F}_{\mathrm{RSS}}(x) = \frac{1}{k}\sum_{i=1}^{k} \hat{F}_{[i]}(x), \quad \text{where} \quad \hat{F}_{[i]}(x) = \frac{1}{m_i}\sum_{j=1}^{m_i} I(X_{[i]j} \le x)$$

and $I(\cdot)$ is the indicator function. Stokes and Sager (1988) proposed an EDF under a balanced RSS setting $(m_1 = \cdots = m_k)$, showing that the variance of $\hat{F}_{\mathrm{RSS}}(x)$ is smaller than or equal to that of the SRS-based EDF. The EDF has the advantage of always being unbiased regardless of the quality of the ranking under the imperfect ranking model $\mathcal{M}$. However, the EDF $\hat{F}_{\mathrm{RSS}}(x)$ is a distribution function for a discrete random variable. Although the discreteness of the EDF is not problematic in many applications, a continuous or smooth estimate of the variable is desirable in applications where the tails of the CDF F are of interest. For example, Polansky (1998) used the kernel estimation of the CDF to estimate the capability of a stable process using the standard process capability indices. Furthermore, a smooth estimator of the CDF can be easily applied to the smoothed bootstrap method (Hall et al., 1989; Silverman and Young, 1987; Wang, 1995). To obtain smoothed bootstrap samples for RSS data, we use the RSS-based KCDE

$$\tilde{F}_{\mathrm{RSS}}(x) = \frac{1}{k}\sum_{i=1}^{k} \tilde{F}_{[i]}(x), \quad \text{where} \quad \tilde{F}_{[i]}(x) = \frac{1}{m_i}\sum_{j=1}^{m_i} K\left(\frac{x - X_{[i]j}}{h}\right)$$

and K is a symmetric CDF with mean zero, and h is a positive real constant assumed to satisfy the conditions $h \to 0$ and $nh \to \infty$ as $n \to \infty$. The selection of an appropriate bandwidth h is described in Section 2.3.

2.2 RSS-based bootstrap methods

In this section, we consider the smoothed version of the bootstrap method for RSS data. In addition, we propose bootstrap methods for a symmetric distribution and prove the consistency of these bootstrap methods in terms of the location parameter.

2.2.1 BRSSR: bootstrap RSS by row

Bootstrap methods for RSS data have been proposed by several researchers. Chen et al. (2004) introduced the bootstrap RSS by row (BRSSR). The BRSSR method is described as follows.

Step 1. Assign to each element of the ith row a probability of m_i^{-1}, and randomly select m_i elements randomly $\{X_{[i]1}, \dots, X_{[i]m_i}\}$ with a replacement to obtain $\{X^*_{[i]1}, \dots, X^*_{[i]m_i}\}$.

Step 2. Conduct Step 1 for $i = 1, \dots, k$ to obtain a bootstrap ranked set sample $\{X^*_{[i]j} : i = 1, \dots, k,\ j = 1, \dots, m_i\}$.

For example, Hui et al. (2005) used the BRSSR method to construct a confidence interval for an estimation of the population mean through a linear regression under the RSS. Modarres et al. (2006) showed that the BRSSR sample mean is consistent in terms of the Mallows distance defined in Definition 2.2.1.

Definition 2.2.1 (Bickel and Freedman, 1981)**.** *Let $\mathcal{F}_s$ be a collection of CDFs having finite s-th moments. Let X and Y be random variables with CDFs $G, H \in \mathcal{F}_s$, respectively, and define $\rho_s(G, H) = \inf_{\tau_{X,Y}} \{\mathbb{E}[|X - Y|^s]\}^{1/s}$, where $\tau_{X,Y}$ is the collection of all possible joint distributions of the pair (X, Y), which have marginal distributions G and H, respectively.*

The following sections describe several smoothed or symmetrized versions of the BRSSR method.

2.2.2 SBRSSR: smoothed bootstrap RSS by row

With the BRSSR method, the resampling of each row ($i = 1, \dots, k$) used random sampling from the EDF $\hat{F}_{[i]}$. Herein, we consider using the KCDE $\tilde{F}_{[i]}$ instead of the EDF $\hat{F}_{[i]}$ for each $i = 1, \dots, k$. The procedure of the smoothed BRSSR method is thus as follows:

Step 1. Generate m_i elements $\{X^*_{[i]1}, \dots, X^*_{[i]m_i}\}$ randomly from $\tilde{F}_{[i]}$.

Step 2. Conduct Step 1 for $i = 1, 2, \dots, k$ to obtain a smoothed bootstrap ranked set sample $\{X^*_{[i]j} : i = 1, \dots, k,\ j = 1, \dots, m_i\}$.

Similar to the BRSSR method, we show that the SBRSSR sample mean is consistent in the sense of the Mallows distance defined through Definition 2.2.1.

Proposition 2.2.1. *If $F, K \in \mathcal{F}_s$, then $\rho_s(\tilde{F}_{\mathrm{RSS}}, F) \to 0$ along almost all sample sequences.*

Proof. By using the results of Bickel and Freedman (1981, Section 8), it suffices to show that $U_n \sim \tilde{F}_{\mathrm{RSS}}$ and $U \sim F$ imply $U_n \rightsquigarrow U$ and $\int x^s d\tilde{F}_{\mathrm{RSS}}(x) \to \int x^s dF(x)$. When $F \in \mathcal{F}_s$ is assumed, it is straightforward to show $F_{[i]} \in \mathcal{F}_s$ for each $i = 1, \dots, k$. Based on Theorem 3 of Yamato (1973), we have

$$\sup_{x \in \mathbb{R}} |\tilde{F}_{[i]}(x) - F_{[i]}(x)| \xrightarrow{a.s.} 0 \quad \text{as} \quad m_i \to \infty.$$

Because

$$\begin{aligned}\sup_{x\in\mathbb{R}}|\tilde{F}_{\mathrm{RSS}}(x)-F(x)| &= \sup_{x\in\mathbb{R}}\left|\frac{1}{k}\sum_{i=1}^{k}[\tilde{F}_{[i]}(x)-F_{[i]}(x)]\right| \\ &\leq \frac{1}{k}\sum_{i=1}^{k}\sup_{x\in\mathbb{R}}|\tilde{F}_{[i]}(x)-F_{[i]}(x)| \to 0 \quad \text{as} \quad \min_{i=1,\ldots,k} m_i \to \infty,\end{aligned}$$

$U_n \sim \tilde{F}_{\mathrm{RSS}}$ and $U \sim F$ imply that $U_n \rightsquigarrow U$ as $\min_i m_i \to \infty$. Finally, based on Khinchine's strong law of large numbers (SLLN), we also obtain

$$\begin{aligned}\int x^s d\tilde{F}_{\mathrm{RSS}}(x) &= \frac{1}{k}\sum_{i=1}^{k}\frac{1}{m_i}\sum_{j=1}^{m_i}\int_{-\infty}^{\infty}(xh+X_{[i]j})^s dK(x) \\ &= \frac{1}{k}\sum_{i=1}^{k}\sum_{t=0}^{s}\left(\frac{1}{m_i}\sum_{j=1}^{m_i}X_{[i]j}^{s-t}\right)\binom{s}{t}h^t\int_{-\infty}^{\infty}x^t dK(x) \\ &= \frac{1}{k}\sum_{i=1}^{k}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i}X_{[i]j}^{s}+\sum_{t=1}^{s}\frac{1}{m_i}\sum_{j=1}^{m_i}X_{[i]j}^{s-t}\binom{s}{t}h^t\int_{-\infty}^{\infty}x^t dK(x)\right\} \\ &\overset{a.s.}{\to} \frac{1}{k}\sum_{i=1}^{k}\int_{-\infty}^{\infty}x^s dF_{[i]}(x) = \int_{-\infty}^{\infty}x^s dF(x) \quad \text{as} \quad \min_{i=1,\ldots,k} m_i \to \infty,\end{aligned}$$

because $\int_{-\infty}^{\infty} x^s dF(x) < \infty$ and $\int_{-\infty}^{\infty} x^s dK(x) < \infty$. □

For each $i = 1, \ldots, k$, let $\mu_{[i]}$ be the mean of a random variable with the CDF $F_{[i]}$. Under the imperfect ranking model $\mathcal{M}$, the mean of a random variable with the CDF F is expressed as $\mu = \frac{1}{k}\sum_{i=1}^{k}\mu_{[i]}$. We are then interested in statistical inferences such as constructing confidence intervals for μ based on RSS data.

Proposition 2.2.2. *Let $\{X^*_{[i]j} : i = 1, \ldots, k,\ j = 1, \ldots, m_i\}$ be a smoothed bootstrap sample using the SBRSSR method. Define*

$$T_n = \sqrt{n}(\bar{X}_{\mathrm{RSS}} - \mu) \quad \textit{and} \quad T_n^* = \sqrt{n}(\bar{X}^*_{\mathrm{RSS}} - \bar{X}_{\mathrm{RSS}}),$$

where $\bar{X}_{\mathrm{RSS}} = \sum_{i=1}^{k}\sum_{j=1}^{m_i} X_{[i]j}/(km_i)$ and $\bar{X}^*_{\mathrm{RSS}} = \sum_{i=1}^{k}\sum_{j=1}^{m_i} X^*_{[i]j}/(km_i)$. If $F, K \in \mathcal{F}_2$, then,

$$\rho_2(H_{n,\tilde{F}_{\mathrm{RSS}}}, H_{n,F}) \overset{a.s.}{\to} 0 \quad \textit{as} \quad \min_{i=1,\ldots,k} m_i \to \infty,$$

where $H_{n,F}$ is the sampling distribution of T_n, and $H_{n,\tilde{F}_{\mathrm{RSS}}}$ is the sampling distribution of T_n^*.

Proof. From Proposition 2.2.1, we can show that $\rho_2(\tilde{F}_{[i]}, F_{[i]}) \overset{a.s.}{\to} 0$ for each $i = 1, \ldots, k$. Using the properties of $\rho_2(\cdot,\cdot)$ provided by Bickel and Freedman (1981, Section 8), we have

$$\rho_2\left[\sqrt{m_i}\left(\bar{X}^*_{[i]} - \bar{X}_{[i]}\right), \sqrt{m_i}\left(\bar{X}_{[i]} - \mu_{[i]}\right)\right] = \frac{1}{\sqrt{m_i}}\rho_2\left[\sum_{j=1}^{m_i}\left(X^*_{[i]j} - \bar{X}_{[i]}\right), \sum_{j=1}^{m_i}\left(X_{[i]j} - \mu_{[i]}\right)\right]$$

$$
\begin{aligned}
&\leq \sqrt{\frac{1}{m_i}\sum_{j=1}^{m_i}\rho_2\left[X^*_{[i]1}-\bar{X}_{[i]}, X_{[i]1}-\mu_{[i]}\right]^2} \\
&= \sqrt{\rho_2\left(X^*_{[i]1}, X_{[i]1}\right)^2 - \left|\mathbb{E}^*[X^*_{[i]1}] - \mathbb{E}[X_{[i]1}]\right|^2} \\
&= \sqrt{\rho_2\left(\tilde{F}_{[i]}, F_{[i]}\right)^2 - \left|\bar{X}_{[i]} - \mu_{[i]}\right|^2}
\end{aligned}
$$

based on $\mathbb{E}^*[X^*_{[i]1}] = \int_{-\infty}^{\infty} x d\tilde{F}_{[i]}(x) = \bar{X}_{[i]}$. Finally, we obtain

$$
\begin{aligned}
\rho_2\left(H_{n,\tilde{F}_{\mathrm{RSS}}}, H_{n,F}\right) &= \rho_2\left[\frac{1}{k}\sum_{i=1}^{k}\sqrt{\frac{n}{m_i}}\cdot\sqrt{m_i}(\bar{X}_{[i]}-\mu_{[i]}), \frac{1}{k}\sum_{i=1}^{k}\sqrt{\frac{n}{m_i}}\cdot\sqrt{m_i}(\bar{X}^*_{[i]}-\bar{X}_{[i]})\right] \\
&\leq \frac{1}{k}\sqrt{\sum_{i=1}^{k}\rho_2\left[\sqrt{\frac{n}{m_i}}\cdot\sqrt{m_i}(\bar{X}_{[i]}-\mu_{[i]}), \sqrt{\frac{n}{m_i}}\cdot\sqrt{m_i}(\bar{X}^*_{[i]}-\bar{X}_{[i]})\right]^2} \\
&= \frac{1}{k}\sqrt{\sum_{i=1}^{k}\frac{n}{m_i}\rho_2\left[\sqrt{m_i}(\bar{X}_{[i]}-\mu_{[i]}), \sqrt{m_i}(\bar{X}^*_{[i]}-\bar{X}_{[i]})\right]^2} \\
&\overset{a.s.}{=} o(1)
\end{aligned}
$$

because $|\bar{X}_{[i]} - \mu_{[i]}| \overset{a.s.}{=} o(1)$ $(i = 1, \ldots, k)$ from Khinchine's SLLN. □

Proposition 2.2.2 can be extended to include statistics that are regular functions of the sample mean based on RSS data. In fact, by defining g as a real differentiable function within a neighborhood of μ, such that g' is continuous in μ and $g'(\mu) \neq 0$, we obtain

$$
\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\sqrt{n}\left\{g(\bar{X}^*_{\mathrm{RSS}}) - g(\bar{X}_{\mathrm{RSS}})\right\} \leq t|\tilde{F}_{\mathrm{RSS}}\right) - \mathbb{P}\left(\sqrt{n}\left\{g(\bar{X}_{\mathrm{RSS}}) - g(\mu)\right\} \leq t\right)\right| \overset{a.s.}{=} o(1).
$$

2.2.3 SymBRSSR: symmetric bootstrap RSS by row

Let $\mathcal{F}_S^{\mu}$ be a set of all symmetric continuous distributions with a center of symmetry μ. Herein, we add the following assumption with respect to the symmetry of the underlying distribution.

(A) Let μ be a real number. Assume that $F_{[i]}(x) = 1 - F_{[k-i+1]}(2\mu - x)$ $(x \in \mathbb{R})$ for each $i = 1, \ldots, k$.

Then, let $\mathcal{M}_S$ be an imperfect ranking model that satisfies both the equality (2.1.1) and assumption (A). For example, assumption (A) is satisfied if $F \in \mathcal{F}_S^{\mu}$ and $p_{ij} = p_{k-i+1,r-j+1}$ $(i = 1, \ldots, k;\ j = 1, \ldots, r)$ under Frey's imperfect ranking model.

Even if $F \in \mathcal{F}_S^{\mu}$, the RSS-based EDF $\hat{F}_{\mathrm{RSS}}$ is generally not a symmetric distribution. Thus, it is inappropriate to obtain the bootstrap sample using the BRSSR or SBRSSR methods for applications such as testing symmetry. Here, we consider the symmetric distribution closest to the RSS-based EDF $\hat{F}_{\mathrm{RSS}}$ in the sense of the integrated squared error (L^2-distance $d_2(\cdot,\cdot)$).

From Theorem 3.1 of Drikvandi et al. (2011), the closest symmetric distribution to $\hat{F}_{\text{RSS}}(x)$ is obtained by

$$\hat{F}^{\text{Sym}_1}(x;\mu) = \frac{1}{k}\sum_{i=1}^{k} \hat{F}_{[i]}^{\text{Sym}_1}(x;\mu), \quad \text{where} \quad \hat{F}_{[i]}^{\text{Sym}_1}(x;\mu) = \frac{1}{2}\left\{\hat{F}_{[i]}(x) + 1 - \hat{F}_{[k-i+1]}(2\mu - x)\right\}$$

for each $i = 1, \ldots, k$. That is,

$$d_2(\hat{F}_{\text{RSS}}, \hat{F}^{\text{Sym}_1}(\cdot;\mu)) = \left(\int_{-\infty}^{\infty}\left\{\hat{F}_{\text{RSS}}(x) - \hat{F}^{\text{Sym}_1}(x;\mu)\right\}^2 dx\right)^{1/2} = \inf_{G\in\mathcal{F}_S^{\mu}}\left\{d_2(\hat{F}_{\text{RSS}}, G)\right\}.$$

Furthermore, if $\mu \in \mathbb{R}$ is unknown, we need to estimate μ using a suitable estimator $\hat{\mu}$, such as the sample mean $\bar{X}_{\text{RSS}}$. The following proposition shows that the Hodges–Lehmann-type estimator $\hat{\mu}_{\text{HL}}$ is optimal in terms of the L^2-distance.

Proposition 2.2.3. *Let $\{X_{[i]j} : i = 1, \ldots, k,\ j = 1, \ldots, m_i\}$ be a ranked set sample from a symmetric distribution $F \in \mathcal{F}_S^{\mu}$ with a center of symmetry $\mu \in \mathbb{R}$. Then,*

$$\hat{F}^{\text{Sym}_1}(x;\hat{\mu}_{\text{HL}}) = \frac{1}{k}\sum_{i=1}^{k}\hat{F}_{[i]}^{\text{Sym}_1}(x;\hat{\mu}_{\text{HL}})$$

is the closest symmetric distribution in $\mathcal{F}_S^{\mu}$ to $\hat{F}_{\text{RSS}}(x)$, where the closeness is measured with respect to the L^2-distance; that is,

$$d_2(\hat{F}_{\text{RSS}}, \hat{F}^{\text{Sym}_1}(\cdot;\hat{\mu}_{\text{HL}})) = \inf_{\mu\in\mathbb{R}}\inf_{G\in\mathcal{F}_S^{\mu}}\left\{d_2\left(\hat{F}_{\text{RSS}}, G(\cdot;\mu)\right)\right\},$$

and

$$\hat{\mu}_{\text{HL}} = \begin{cases} Y_{(t)} & \text{if} \quad \sum_{i=1}^{t-1} w_{(i)} < \sum_{i=t}^{n^2} w_{(i)} \text{ and } \sum_{i=1}^{t} w_{(i)} > \sum_{i=t+1}^{n^2} w_{(i)}, \\ (Y_{(t)} + Y_{(t+1)})/2 & \text{if} \quad \sum_{i=1}^{t} w_{(i)} = \sum_{i=t+1}^{n^2} w_{(i)}, \end{cases} \tag{2.2.2}$$

where $Y_{(t)}$ is the t-th order statistic of $(X_{[i_1]j_1} + X_{[i_2]j_2})/2$, $(i_1, i_2 = 1, \ldots, k;\ j_1 = 1, \ldots, m_{i_1};j_2 = 1, \ldots, m_{i_2})$, and $w_{(t)}$ is the weight corresponding to $Y_{(t)}$; that is, if $Y_{(t)} = (X_{[i_1]j_1} + X_{[i_2]j_2})/2$, then $w_{(t)} = 1/(m_{i_1}m_{i_2})$.

Proof. From Theorem 3.1 of Drikvandi et al. (2011), we obtain

$$\begin{aligned}
\inf_{G\in\mathcal{F}_S^{\mu}} d_2\left(\hat{F}_{\text{RSS}}(x), G(x;\mu)\right)^2 &= d_2\left(\hat{F}_{\text{RSS}}(x), \hat{F}^{\text{Sym}_1}(x;\mu)\right)^2 \\
&= \frac{1}{4}\int_{-\infty}^{\infty}\left\{\hat{F}_{\text{RSS}}(x) + \hat{F}_{\text{RSS}}(2\mu - x) - 1\right\}^2 dx \\
&= -\frac{1}{4k^2}\sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\frac{1}{m_{i_1}m_{i_2}}\sum_{j_1=1}^{m_{i_1}}\sum_{j_2=1}^{m_{i_2}}|X_{[i_1]j_1} - X_{[i_2]j_2}| \\
&\quad + \frac{1}{4k^2}\sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\frac{1}{m_{i_1}m_{i_2}}\sum_{j_1=1}^{m_{i_1}}\sum_{j_2=1}^{m_{i_2}}|X_{[i_1]j_1} + X_{[i_2]j_2} - 2\mu|
\end{aligned}$$

for every $\mu \in \mathbb{R}$. Define

$$\rho_{\hat{F}_{\mathrm{RSS}}}(\mu) := \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_2}} \frac{1}{m_{i_1} m_{i_2}} \cdot \left| \frac{X_{[i_1]j_1} + X_{[i_2]j_2}}{2} - \mu \right| = \sum_{i=1}^{n^2} w_{(i)} |Y_{(i)} - \mu|,$$

where $Y_{(t)}$ is the t-th order statistic of $(X_{[i_1]j_1} + X_{[i_2]j_2})/2$, $(i_1, i_2 = 1, \ldots, k;\ j_1 = 1, \ldots, m_{i_1}; j_2 = 1, \ldots, m_{i_2})$, and $w_{(t)}$ is the weight corresponding to $Y_{(t)}$. Note that the function $\rho_{\hat{F}_{\mathrm{RSS}}}(\mu)$ is convex with respect to μ. Therefore, there exists at least one μ for minimizing $\rho_{\hat{F}_{\mathrm{RSS}}}(\mu)$. From Lemma 1 of Huber (1964), the set of μ that minimizes $\rho_{\hat{F}_{\mathrm{RSS}}}(\mu)$ is convex and compact. In this paper, we use the midpoint of this set as the estimator of the center of symmetry. Except for points $Y_{(1)}, \ldots, Y_{(n^2)}$, the derivative of $\rho_{\hat{F}_{\mathrm{RSS}}}(\mu)$ is expressed as

$$\rho'_{\hat{F}_{\mathrm{RSS}}}(\mu) = \begin{cases} -\sum_{i=1}^{n^2} w_{(i)} & \text{if } \mu < Y_{(1)}, \\ \sum_{i=1}^{t-1} w_{(i)} - \sum_{i=t}^{n^2} w_{(i)} & \text{if } Y_{(t-1)} < \mu < Y_{(t)};\ t = 2, \ldots, n^2, \\ \displaystyle\sum_{i=1}^{n^2} w_{(i)} & \text{if } Y_{(n^2)} < \mu. \end{cases}$$

Therefore, we obtain

$$\hat{\mu}_{\mathrm{HL}} = \begin{cases} Y_{(t)} & \text{if } \sum_{i=1}^{t-1} w_{(i)} < \sum_{i=t}^{n^2} w_{(i)} \text{ and } \sum_{i=1}^{t} w_{(i)} > \sum_{i=t+1}^{n^2} w_{(i)}, \\ (Y_{(t)} + Y_{(t+1)})/2 & \text{if } \sum_{i=1}^{t} w_{(i)} = \sum_{i=t+1}^{n^2} w_{(i)}, \end{cases}$$

as the point minimizing $\rho_{\hat{F}_{\mathrm{RSS}}}(\mu)$. □

Under appropriate assumptions, the RSS-based Hodges–Lehmann estimator $\hat{\mu}_{\mathrm{HL}}$ is a strongly consistent estimator of μ from the asymptotic property of M-estimators (see Huber (1964) and Serfling (1980) for further details).

Equal-weighted symmetric BRSSR method

If $F \in \mathcal{F}_S^{\mu}$ with an unknown center of symmetry $\mu \in \mathbb{R}$, we generate bootstrap samples from $\hat{F}^{\mathrm{Sym}_1}(x; \hat{\mu})$, where $\hat{\mu}$ is a strongly consistent estimator for μ (e.g., $\bar{X}_{\mathrm{RSS}}$ or $\hat{\mu}_{\mathrm{HL}}$). The algorithm used for the symmetric BRSSR method is as follows:

Step 1. Generate m_i elements $\{X^*_{[i]1}, \ldots, X^*_{[i]m_i}\}$ randomly from $\hat{F}^{\mathrm{Sym}_1}_{[i]}(x; \hat{\mu})$.

Step 2. Conduct Step 1 for $r = 1, 2, \ldots, k$ to obtain a symmetric bootstrap ranked set sample $\{X^*_{[i]j} : i = 1, \ldots, k,\ j = 1, \ldots, m_i\}$.

We show that the sample mean of the symmetric BRSSR is consistent in terms of the Mallows distance.

Proposition 2.2.4. *Suppose that $F \in \mathcal{F}_s \cap \mathcal{F}_S^{\mu}$, and that $\hat{\mu}$ is a strongly consistent estimator of μ. Then, $\rho_s(\hat{F}^{\mathrm{Sym}_1}(\cdot; \hat{\mu}), F) \to 0$ along almost all sample sequences.*

Proof. Based on the Glivenko–Cantelli theorem, for each $i = 1, \dots, k$, we obtain

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}} |\hat{F}_{[i]}^{\mathrm{Sym}_1}(x;\hat{\mu}) - F_{[i]}(x)| \\
&\le \frac{1}{2}\sup_{x\in\mathbb{R}} |\hat{F}_{[i]}(x) - F_{[i]}(x)| + \frac{1}{2}\sup_{x\in\mathbb{R}} |\hat{F}_{[k-i+1]}(2\hat{\mu}-x) - F_{[k-i+1]}(2\mu - x)| \\
&= o(1) + \frac{1}{2}\sup_{t\in\mathbb{R}} |\hat{F}_{[k-i+1]}(t) - F_{[k-i+1]}(t - 2(\hat{\mu}-\mu))| \\
&\le o(1) + \frac{1}{2}\sup_{t\in\mathbb{R}} |\hat{F}_{[k-i+1]}(t) - F_{[k-i+1]}(t)| + \frac{1}{2}\sup_{t\in\mathbb{R}} |F_{[k-i+1]}(t) - F_{[k-i+1]}(t - 2(\hat{\mu}-\mu))| \\
&\overset{a.s.}{\to} 0 \quad \text{as} \quad \min_{i=1,\dots,k} m_i \to \infty
\end{aligned}
$$

because $F_{[i]}(x) = 1 - F_{[k-i+1]}(2\mu - x)$, $i = 1, 2, \dots, k$, under the imperfect ranking model $\mathcal{M}_S$. It follows that $\sup_{x\in\mathbb{R}} |\hat{F}^{\mathrm{Sym}_1}(x) - F(x)| \overset{a.s.}{\to} 0$; that is, if $U_n \sim \hat{F}^{\mathrm{Sym}_1}$ and $U \sim F$, then $U_n \rightsquigarrow U$ as $\min_i m_i \to \infty$. Finally, based on Khinchine's SLLN, we also have

$$
\begin{aligned}
&\int_{-\infty}^{\infty} x^s d\hat{F}^{\mathrm{Sym}_1}(x;\hat{\mu}) \\
&= \frac{1}{k}\sum_{i=1}^{k}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i}\frac{X_{[i]j}^s + (2\hat{\mu} - X_{[k-i+1]j})^s}{2}\right\} \\
&= \frac{1}{k}\sum_{i=1}^{k}\left\{\frac{1}{2m_i}\sum_{j=1}^{m_i} X_{[i]j}^s + \frac{1}{2m_{k-i+1}}\sum_{j=1}^{m_{k-i+1}}\sum_{t=0}^{s}\binom{s}{t}2^t(\hat{\mu}-\mu)^t(2\mu - X_{[k-i+1]j})^{s-t}\right\} \\
&\overset{a.s.}{\to} \frac{1}{k}\sum_{i=1}^{k}\left\{\frac{1}{2}\int_{-\infty}^{\infty} x^s dF_{[i]}(x) + \frac{1}{2}\int_{-\infty}^{\infty}(2\mu - x)^s dF_{[k-i+1]}(x)\right\} \\
&= \int_{-\infty}^{\infty} x^s dF(x)
\end{aligned}
$$

because $2\mu - X_{[i]1}$ and $X_{[k-i+1]1}$ follow the same distribution under the model $\mathcal{M}_S$. □

Proposition 2.2.5. *Let $\{X_{[i]j}^* : i = 1, \dots, k,\ j = 1, \dots, m_i\}$ be a bootstrap sample using the equal-weighted symmetric BRSSR method with $\hat{\mu} = \bar{X}_{\mathrm{RSS}} = \sum_{i=1}^{k}\sum_{j=1}^{m_i} X_{[i]j}/(km_i)$. Define*

$$T_n = \sqrt{n}(\bar{X}_{\mathrm{RSS}} - \mu) \quad \textit{and} \quad T_n^* = \sqrt{n}(\bar{X}_{\mathrm{RSS}}^* - \bar{X}_{\mathrm{RSS}}),$$

where $\bar{X}_{\mathrm{RSS}}^* = \sum_{i=1}^{k}\sum_{j=1}^{m_i} X_{[i]j}^*/(km_i)$. If $F \in \mathcal{F}_2 \cap \mathcal{F}_S^{\mu}$, then

$$\rho_2(H_{n,\hat{F}^{\mathrm{Sym}_1}}, H_{n,F}) \overset{a.s.}{\to} 0 \quad \textit{as} \quad \min_{i=1,\dots,k} m_i \to \infty,$$

where $H_{n,F}$ is the sampling distribution of T_n, and $H_{n,\hat{F}^{\mathrm{Sym}_1}}$ is the sampling distribution of T_n^*.

Proof. From Proposition 2.2.4, we can show that $\rho_2(\hat{F}_{[i]}^{\mathrm{Sym}_1}(\cdot;\bar{X}_{\mathrm{RSS}}), F_{[i]}) \overset{a.s.}{\to} 0$ for each $i = 1, \dots, k$. Note that

$$\bar{X}_{\mathrm{RSS}} = \frac{1}{k}\sum_{i=1}^{k}\bar{X}_{[i]}^{\mathrm{Sym}_1}, \quad \text{where} \quad \bar{X}_{[i]}^{\mathrm{Sym}_1} = \frac{\bar{X}_{[i]} + 2\bar{X}_{\mathrm{RSS}} - \bar{X}_{[k-i+1]}}{2}.$$

By using the properties of $\rho_2(\cdot,\cdot)$ provided in Section 8 of Bickel and Freedman (1981), we obtain

$$\begin{aligned}
\rho_2\left[\sqrt{m_i}\left(X^*_{[i]}-\bar{X}^{\mathrm{Sym}_1}_{[i]}\right),\sqrt{m_i}\left(\bar{X}_{[i]}-\mu_{[i]}\right)\right] &= \frac{1}{\sqrt{m_i}}\rho_2\left[\sum_{j=1}^{m_i}\left(X^*_{[i]j}-\bar{X}^{\mathrm{Sym}_1}_{[i]}\right),\sum_{j=1}^{m_i}\left(X_{[i]j}-\mu_{[i]}\right)\right] \\
&\le \sqrt{\frac{1}{m_i}\sum_{j=1}^{m_i}\rho_2\left[X^*_{[i]1}-\bar{X}^{\mathrm{Sym}_1}_{[i]},X_{[i]1}-\mu_{[i]}\right]^2} \\
&= \sqrt{\rho_2\left(X^*_{[i]1},X_{[i]1}\right)^2-\left|\mathbb{E}^*[X^*_{[i]1}]-\mathbb{E}[X_{[i]1}]\right|^2} \\
&= \sqrt{\rho_2\left(\hat{F}^{\mathrm{Sym}_1}_{[i]},F_{[i]}\right)^2-\left|\bar{X}^{\mathrm{Sym}_1}_{[i]}-\mu_{[i]}\right|^2}
\end{aligned}$$

because $\mathbb{E}^*[X^*_{[i]1}]=\int_{-\infty}^{\infty}xd\hat{F}^{\mathrm{Sym}_1}_{[i]}(x)=\bar{X}^{\mathrm{Sym}_1}_{[i]}$. Finally, we obtain

$$\begin{aligned}
\rho_2\left(H_{n,\tilde{F}_{\mathrm{RSS}}},H_{n,F}\right) &= \rho_2\left[\frac{1}{k}\sum_{i=1}^{k}\sqrt{\frac{n}{m_i}}\cdot\sqrt{m_i}(\bar{X}^*_{[i]}-\bar{X}^{\mathrm{Sym}_1}_{[i]}),\frac{1}{k}\sum_{i=1}^{k}\sqrt{\frac{n}{m_i}}\cdot\sqrt{m_i}(\bar{X}_{[i]}-\mu_{[i]})\right] \\
&\le \frac{1}{k}\sqrt{\sum_{i=1}^{k}\rho_2\left[\sqrt{\frac{n}{m_i}}\cdot\sqrt{m_i}(\bar{X}^*_{[i]}-\bar{X}^{\mathrm{Sym}_1}_{[i]}),\sqrt{\frac{n}{m_i}}\cdot\sqrt{m_i}(\bar{X}_{[i]}-\mu_{[i]})\right]^2} \\
&= \frac{1}{k}\sqrt{\sum_{i=1}^{k}\frac{n}{m_i}\rho_2\left[\sqrt{m_i}(\bar{X}^*_{[i]}-\bar{X}^{\mathrm{Sym}_1}_{[i]}),\sqrt{m_i}(\bar{X}_{[i]}-\mu_{[i]})\right]^2} \\
&\overset{a.s.}{=} o(1)
\end{aligned}$$

from Khinchine's SLLN. □

Unequal-weighted symmetric BRSSR method

Here, we construct the new symmetric bootstrap method using the unequal-weighted symmetric kernel estimator introduced by Lim et al. (2014). We define the unequal-weighted symmetric EDF as follows:

$$\hat{F}^{\mathrm{Sym}_2}(x;\mu)=\frac{1}{k}\sum_{i=1}^{k}\hat{F}^{\mathrm{Sym}_2}_{[i]}(x;\mu),$$

where

$$\hat{F}^{\mathrm{Sym}_2}_{[i]}(x;\mu)=\frac{m_i}{m_i+m_{k-i+1}}\hat{F}_{[i]}(x)+\frac{m_{k-i+1}}{m_i+m_{k-i+1}}\left(1-\hat{F}_{[k-i+1]}(2\mu-x)\right).$$

This estimator can be computed by estimating $F_{[i]}(x)$ using both $\{X_{[i]1},\ldots,X_{[i]m_i}\}$ and $\{2\mu-X_{[k-i+1]1},\ldots,2\mu-X_{[k-i+1]m_{k-i+1}}\}$. Under the model $\mathcal{M}_S$, we have

$$\mathrm{Var}(\hat{F}^{\mathrm{Sym}_2}(x;\mu))=\frac{1}{k^2}\sum_{i=1}^{k}\frac{m_i}{(m_i+m_{k-i+1})^2}\,\mathrm{Var}(I(2\mu-x\le X_{[i]1}\le x))$$

$$
\begin{aligned}
&= \frac{1}{k^2}\sum_{i=1}^{k}\frac{1}{2(m_i+m_{k-i+1})}\operatorname{Var}(I(2\mu - x \le X_{[i]1} \le x))\\
&\le \frac{1}{k^2}\sum_{i=1}^{k}\frac{1}{4m_i}\operatorname{Var}(I(2\mu - x \le X_{[i]1} \le x))\\
&= \operatorname{Var}(\hat{F}^{\mathrm{Sym}_1}(x;\mu)),
\end{aligned}
$$

because $\operatorname{Var}(I(2\mu - x \le X_{[i]1} \le x)) = \operatorname{Var}(I(2\mu - x \le X_{[k-i+1]1} \le x))$. This means that $\hat{F}^{\mathrm{Sym}_2}(x;\mu)$ may be superior in the sense that the variance of $\hat{F}^{\mathrm{Sym}_2}(x;\mu)$ is smaller than that of $\hat{F}^{\mathrm{Sym}_1}(x;\mu)$. Thus, it is suggested that the bootstrap method based on $\hat{F}^{\mathrm{Sym}_2}(x;\mu)$ achieves a better performance than the equal-weighted symmetric BRSSR method. The algorithm for the unequal-weighted symmetric BRSSR method is as follows:

Step 1. Generate m_i elements $\{X^*_{[i]1},\ldots,X^*_{[i]m_i}\}$ randomly from $\hat{F}^{\mathrm{Sym}_2}_{[i]}(x;\hat{\mu})$.

Step 2. Conduct Step 1 for $i = 1,2,\ldots,k$ to obtain a symmetric bootstrap ranked set sample

$$\{X^*_{[1]1}, X^*_{[1]2},\ldots,X^*_{[1]m_1},\ldots,X^*_{[k]1},X^*_{[k]2},\ldots,X^*_{[k]m_k}\}.$$

Because $X_{[i]1}$ and $2\mu - X_{[k-i+1]1}$ follow the same distribution under the imperfect ranking model $\mathcal{M}_S$ and $F \in \mathcal{F}^{\mu}_S$, the following asymptotic properties of the unequal-weighted symmetric BRSSR sample mean are immediately proved from Propositions 2.2.4 and 2.2.5.

Proposition 2.2.6. *Suppose that $F \in \mathcal{F}_s \cap \mathcal{F}^{\mu}_S$, and that $\hat{\mu}$ is a strongly consistent estimator of μ. Then, $\rho_s(\hat{F}^{\mathrm{Sym}_2}(\cdot;\hat{\mu}), F) \to 0$ along almost all sample sequences.*

Proof. Note that $F_{[i]}(x) = 1 - F_{[k-i+1]}$ $(i = 1,2,\ldots,k)$ under the imperfect ranking model $\mathcal{M}_S$. Based on the Glivenko–Cantelli theorem, we obtain

$$
\begin{aligned}
\sup_{x\in\mathbb{R}} |\hat{F}^{\mathrm{Sym}_2}_{[i]}(x;\hat{\mu}) - F_{[i]}(x)| &\le \frac{m_i}{m_i+m_{k-i+1}}\sup_{x\in\mathbb{R}}|\hat{F}_{[i]}(x) - F_{[i]}(x)|\\
&\quad + \frac{m_{k-i+1}}{m_i+m_{k-i+1}}\sup_{x\in\mathbb{R}}|\hat{F}_{[k-i+1]}(2\hat{\mu}-x) - F_{[k-i+1]}(2\mu - x)|\\
&= o(1) + \frac{m_{k-i+1}}{m_i+m_{k-i+1}}\sup_{t\in\mathbb{R}}|\hat{F}_{[k-i+1]}(t) - F_{[k-i+1]}(t-2(\hat{\mu}-\mu))|\\
&\le o(1) + \frac{m_{k-i+1}}{m_i+m_{k-i+1}}\sup_{t\in\mathbb{R}}|\hat{F}_{[k-i+1]}(t) - F_{[k-i+1]}(t)|\\
&\quad + \frac{m_{k-i+1}}{m_i+m_{k-i+1}}\sup_{t\in\mathbb{R}}|F_{[k-i+1]}(t) - F_{[k-i+1]}(t-2(\hat{\mu}-\mu))|\\
&\overset{a.s.}{\to} 0 \quad \text{as} \quad \min_{i=1,\ldots,k} m_i \to \infty
\end{aligned}
$$

for each $i = 1,\ldots,k$. It follows that $\sup_{x\in\mathbb{R}}|\hat{F}^{\mathrm{Sym}_2}(x) - F(x)| \overset{a.s.}{\to} 0$; that is, if $U_n \sim \hat{F}^{\mathrm{Sym}_2}$ and $U \sim F$, then $U_n \rightsquigarrow U$ as $\min_{i=1,\ldots,k} m_i \to \infty$. Finally, based on Khinchine's SLLN, we also have

$$\int_{-\infty}^{\infty} x^s d\hat{F}^{\mathrm{Sym}_2}(x;\hat{\mu})$$

$$= \frac{1}{k}\sum_{i=1}^{k}\left\{\frac{1}{m_i+m_{k-i+1}}\sum_{j=1}^{m_i}X_{[i]j}^s + \frac{1}{m_i+m_{k-i+1}}\sum_{j=1}^{m_{k-i+1}}(2\hat{\mu}-X_{[k-i+1]j})^s\right\}$$

$$= \frac{1}{k}\sum_{i=1}^{k}\left\{\frac{1}{m_i+m_{k-i+1}}\sum_{j=1}^{m_i}X_{[i]j}^s + \frac{1}{m_i+m_{k-i+1}}\sum_{j=1}^{m_i}\sum_{t=0}^{s}\binom{s}{t}2^t(\hat{\mu}-\mu)^t(2\mu-X_{[k-i+1]j})^{s-t}\right\}$$

$$\xrightarrow{a.s.} \frac{1}{k}\sum_{i=1}^{k}\int_{-\infty}^{\infty}x^s dF_{[i]}(x) = \int_{-\infty}^{\infty}x^s dF(x)$$

because $2\mu - X_{[i]1}$ and $X_{[k-i+1]1}$ follow the same distribution under the model $\mathcal{M}_S$. □

Proposition 2.2.7. *Let $\{X_{[i]j}^* : i = 1,\dots,k,\ j = 1,\dots,m_i\}$ be a bootstrap sample using the unequal-weighted symmetric BRSSR method with $\hat{\mu} = \bar{X}_{\text{RSS}} = \sum_{i=1}^{k}\sum_{j=1}^{m_i}X_{[i]j}/(km_i)$. Define*

$$T_n = \sqrt{n}(\bar{X}_{\text{RSS}} - \mu) \quad \textit{and} \quad T_n^* = \sqrt{n}(\bar{X}_{\text{RSS}}^* - \bar{X}_{\text{RSS}}),$$

where $\bar{X}_{\text{RSS}}^* = \sum_{i=1}^{k}\sum_{j=1}^{m_i}X_{[i]j}^*/(km_i)$. If $F \in \mathcal{F}_2 \cap \mathcal{F}_S^{\mu}$, then

$$\rho_2(H_{n,\hat{F}^{\text{Sym}_2}}, H_{n,F}) \xrightarrow{a.s.} 0 \quad \textit{as} \quad \min_{i=1,\dots,k} m_i \to \infty,$$

where $H_{n,F}$ is the sampling distribution of T_n, and $H_{n,\hat{F}^{\text{Sym}_2}}$ is the sampling distribution of T_n^*.

Proof. From Proposition 2.2.4, we can show $\rho_2(\hat{F}_{[i]}^{\text{Sym}_2}(\cdot;\bar{X}_{\text{RSS}}), F_{[i]}) \xrightarrow{a.s.} 0$ for each $i = 1,\dots,k$. Note that

$$\bar{X}_{\text{RSS}} = \frac{1}{k}\sum_{i=1}^{k}\bar{X}_{[i]}^{\text{Sym}_2}, \quad \text{where} \quad \bar{X}_{[i]}^{\text{Sym}_2} = \frac{m_i\bar{X}_{[i]} + m_{k-i+1}(2\bar{X}_{\text{RSS}} - \bar{X}_{[k-i+1]})}{m_i + m_{k-i+1}}.$$

The rest of the proof is the same as that in Proposition 2.2.5. □

If the RSS design with $m_i = m_{k-i+1}$ $(i = 1,\dots,k)$, the unequal-weighted symmetric BRSSR method is the same as the equal-weighted version.

2.2.4 SymSBRSSR: symmetric smoothed bootstrap RSS by row

In this subsection, we consider the smoothed version of the SymBRSSR methods described in Section 2.2.3. The asymptotic properties of these methods are omitted because they are proved in much the same way as in Propositions 2.2.4–2.2.7.

Equal-weighted symmetric SBRSSR method

Even if $F \in \mathcal{F}_S^{\mu}$, the RSS-based KCDE $\tilde{F}_{\text{RSS}}$ is generally not symmetric. Similar to the description in Section 2.2.3, we consider a symmetric distribution that is close to the KCDE $\tilde{F}_{\text{RSS}}$ in terms of the L^2-distance to obtain smoothed bootstrap samples from a symmetric distribution. From Theorem 3.1 of Drikvandi et al. (2011), the symmetric CDF closest to $\tilde{F}_{\text{RSS}}$ is denoted by

$$\tilde{F}^{\text{Sym}_1}(x;\mu) = \frac{1}{k}\sum_{i=1}^{k}\tilde{F}_{[i]}^{\text{Sym}_1}(x;\mu), \quad \text{where} \quad \tilde{F}_{[i]}^{\text{Sym}_1}(x;\mu) = \frac{1}{2}\left\{\tilde{F}_{[i]}(x) + 1 - \tilde{F}_{[k-i+1]}(2\mu - x)\right\}$$

for each $i = 1, \ldots, k$. Therefore,

$$d_2\left(\tilde{F}_{\text{RSS}}, \tilde{F}^{\text{Sym}_1}(\cdot;\mu)\right) = \inf_{G\in\mathcal{F}_S^{\mu}} \left\{d_2\left(\tilde{F}_{\text{RSS}}, G\right)\right\}.$$

The algorithm for the symmetric SBRSSR method is as follows:

Step 1. Generate m_i elements $\{X^*_{[i]1}, \ldots, X^*_{[i]m_i}\}$ randomly from $\tilde{F}^{\text{Sym}_1}_{[i]}(x;\hat{\mu})$.

Step 2. Conduct Step 1 for $i = 1, 2, \ldots, k$ to obtain a symmetric smoothed bootstrap ranked set sample $\{X^*_{[i]j} : i = 1, \ldots, k,\ j = 1, \ldots, m_i\}$.

Unequal-weighted symmetric SBRSSR method

Here, we describe the construction of the symmetric smoothed bootstrap method using the unequal-weighted symmetric kernel estimator introduced by Lim et al. (2014). We define the unequal-weighted symmetric KCDE as

$$\tilde{F}^{\text{Sym}_2}(x;\mu) = \frac{1}{k}\sum_{i=1}^{k} \tilde{F}^{\text{Sym}_2}_{[i]}(x;\mu),$$

where

$$\tilde{F}^{\text{Sym}_2}_{[i]}(x;\mu) = \frac{m_i}{m_i + m_{k-i+1}}\tilde{F}_{[i]}(x) + \frac{m_{k-i+1}}{m_i + m_{k-i+1}}\left(1 - \tilde{F}_{[k-i+1]}(2\mu - x)\right).$$

This estimator $\tilde{F}^{\text{Sym}*}$ is constructed by estimating $F_{[i]}(x)$ using both $\{X_{[i]1}, \ldots, X_{[i]m_i}\}$ and $\{2\mu - X_{[k-i+1]1}, \ldots, 2\mu - X_{[k-i+1]m_{k-i+1}}\}$.

The algorithm for the unequal-weighted symmetric SBRSSR method is as follows:

Step 1. Generate m_i elements $\{X^*_{[i]1}, \ldots, X^*_{[i]m_i}\}$ randomly from $\tilde{F}^{\text{Sym}_2}_{[i]}(x;\hat{\mu})$.

Step 2. Conduct Step 1 for $i = 1, 2, \ldots, k$ to obtain a symmetric smoothed bootstrap ranked set sample $\{X^*_{[i]j} : i = 1, \ldots, k,\ j = 1, \ldots, m_i\}$.

2.3 Asymptotic MISE and bandwidth selection

Bandwidth selection is an important step in estimating the CDF using the kernel method. Lim et al. (2014) proposed a plug-in estimator of the bandwidth to minimize the asymptotic MISE for the RSS-based kernel density estimator. In this section, we derive the optimal bandwidth that minimizes the asymptotic MISE of the RSS-based KCDE $\tilde{F}_{\text{RSS}}$. Our approach to the bandwidth selection rule is based on the idea by Polansky and Baker (2000). Furthermore, we derive the optimal bandwidth by asymptotically minimizing the MISE of the symmetrized kernel estimators.

2.3.1 Non-symmetrized case

The typical measure of accuracy used for estimates of F is the weighted MISE, which is defined in

$$\mathrm{MISE}[\tilde{F}] = \mathbb{E}\left[\int_{-\infty}^{\infty}\left\{\tilde{F}(x) - F(x)\right\}^2 w(x)dx\right], \tag{2.3.3}$$

where $w(\cdot)$ is a weight function, and $\tilde{F}$ is an estimator of F. The measure $\mathrm{MISE}(\tilde{F})$ gives a global assessment of the closeness of $\tilde{F}$ to F for a given random sample. The MISE is often used as an average measure of the performance of $\tilde{F}$. Several authors have derived asymptotic expressions for the MISE (2.3.3). In particular, Swanepoel (1988) derived an expression for $w(x) = f(x)$, and Jones (1990) derived an expression for $w(x) = 1$. Here, we specifically examine the case in which $w(x) = 1$. To evaluate the MISE of the RSS-based KCDE, we assume that the density f is continuous and differentiable with a finite mean and has a square-integrable derivative.

For each $i = 1, \ldots, k$, the variance and bias of $\tilde{F}_{[i]}(x)$ are

$$\mathrm{Var}[\tilde{F}_{[i]}(x)] = \frac{1}{m_i}F_{[i]}(x)\{1 - F_{[i]}(x)\} - \frac{2h}{m_i}A_1(K)F'_{[i]}(x) + o(hm_i^{-1})$$

and

$$\mathrm{bias}[\tilde{F}_{[i]}(x)] = \frac{1}{2}h^2 F''_{[i]}(x)A_2(K) + o(h^2),$$

where

$$A_1(K) = \int xK(x)dK(x) \quad \text{and} \quad A_2(K) = \int x^2 dK(x).$$

Therefore, the variance and bias of $\tilde{F}_{\mathrm{RSS}}(x)$ is calculated by

$$\begin{aligned}\mathrm{Var}[\tilde{F}_{\mathrm{RSS}}(x)] &= \frac{1}{k^2}\sum_{i=1}^{k}\mathrm{Var}[\tilde{F}_{[i]}(x)]\\ &= \frac{1}{k^2}\sum_{i=1}^{k}\left\{\frac{1}{m_i}F_{[i]}(x)\{1 - F_{[i]}(x)\} - \frac{2h}{m_i}A_1(K)F'_{[i]}(x)\right\} + o\left(\frac{h}{\min_i(m_i)}\right)\end{aligned}$$

and

$$\mathrm{bias}[\tilde{F}_{\mathrm{RSS}}(x)]^2 = \left\{\frac{1}{k}\sum_{i=1}^{k}\mathrm{bias}[\tilde{F}_{[i]}(x)]\right\}^2 = \frac{1}{4}h^4 F''(x)^2 A_2(K)^2 + o(h^4).$$

Thus, the MISE of $\tilde{F}_{\mathrm{RSS}}$ is expressed as

$$\begin{aligned}\mathrm{MISE}(\tilde{F}_{\mathrm{RSS}}) = \frac{1}{4}h^4 A_2(K)^2 D_1(F) + \frac{1}{k^2}\sum_{i=1}^{k}\frac{1}{m_i}D_2(F_{[i]}) - \frac{2h}{k^2}A_1(K)\sum_{i=1}^{k}\frac{1}{m_i}\\ + o\left(\max\left\{h^4, \frac{h^2}{\min_i(m_i)}\right\}\right),\end{aligned}$$

where

$$D_1(F) = \int F''(x)^2 dx \quad \text{and} \quad D_2(F_{[i]}) = \int F_{[i]}(x)\{1 - F_{[i]}(x)\}dx.$$

Thus, the asymptotically optimal bandwidth h that minimizes the asymptotic MISE is

$$h_{\text{RSS}} = \left(\frac{n}{k^2} \sum_{i=1}^{k} \frac{1}{m_i} \right)^{1/3} \left(\frac{2A_1(K)}{A_2(K)^2 D_1(F)} \right)^{1/3} n^{-1/3}.$$

The optimal bandwidth h_{RSS} depends on the unknown parameter

$$D_1(F) = \int F''(x)^2 dx = \int f^{(2)}(x) f(x) dx.$$

We then unbiasedly estimate $D_1(F)$ as

$$\hat{D}_1(F) = \frac{1}{k^2} \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \frac{1}{m_{i_1} m_{i_2}} \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_2}} \frac{1}{g^3} L^{(2)} \left(\frac{X_{[i_1]j_1} - X_{[i_2]j_2}}{g} \right),$$

where L is the kernel density function satisfying $\int L(x)dx = 1$, and g is the bandwidth estimated using the plug-in method developed by Sheather and Jones (1991). Similarly to Lim et al. (2014), the bandwidth h_{RSS} can be evaluated by treating the RSS data as SRS data.

2.3.2 Symmetrized case

In this subsection, let μ be the center of symmetry, and assume that the imperfect ranking model is $\mathcal{M}_S$ defined in Section 2.2.3. First, we derive the optimal bandwidth for the asymptotically minimizing MISE of the symmetrized kernel density estimators $\tilde{f}_{\text{RSS}}^{\text{Sym}_1}(x;\mu) = d\tilde{F}_{\text{RSS}}^{\text{Sym}_1}(x;\mu)/dx$ and $\tilde{f}_{\text{RSS}}^{\text{Sym}_2}(x;\mu) = d\tilde{F}_{\text{RSS}}^{\text{Sym}_2}(x;\mu)/dx$. Using a calculation similar to that presented in Kraft et al. (1985), we obtain

$$\begin{aligned}
\text{MISE}[\tilde{f}_{\text{RSS}}^{\text{Sym}_1}(\cdot;\mu)] &= \int \text{bias}[\tilde{f}_{\text{RSS}}^{\text{Sym}_1}(x;\mu)]^2 dx + \int \text{Var}[\tilde{f}_{\text{RSS}}^{\text{Sym}_1}(x;\mu)] dx \\
&= \frac{h^4}{4} A_2(K)^2 \int \{f^{(2)}(t)\}^2 dt + \frac{1}{nh} \left(\frac{n}{2k^2} \sum_{i=1}^{k} \frac{1}{m_i} \right) \int K'(t)^2 dt \\
&\quad + o\left(\max\left\{ h^4, \frac{h}{\min_i(m_i)} \right\} \right)
\end{aligned}$$

and

$$\begin{aligned}
\text{MISE}[\tilde{f}_{\text{RSS}}^{\text{Sym}_2}(\cdot;\mu)] &= \int \text{bias}[\tilde{f}_{\text{RSS}}^{\text{Sym}_2}(x;\mu)]^2 dx + \int \text{Var}[\tilde{f}_{\text{RSS}}^{\text{Sym}_2}(x;\mu)] dx \\
&= \frac{h^4}{4} A_2(K)^2 \int \{f^{(2)}(t)\}^2 dt + \frac{1}{nh} \left\{ \frac{n}{k^2} \sum_{i=1}^{k} \frac{1}{m_i + m_{k-i+1}} \right\} \int K'(t)^2 dt \\
&\quad + o\left(\max\left\{ h^4, \frac{h}{\min_i(m_i)} \right\} \right).
\end{aligned}$$

Therefore, the optimal bandwidth h_{opt_1} asymptotically minimizing $\text{MISE}(\tilde{f}_{\text{RSS}}^{\text{Sym}_1}(\cdot;\mu))$ is

$$h_{\text{opt}_1} = \left(\frac{n}{2k^2}\sum_{i=1}^{k}\frac{1}{m_i}\right)^{1/5}\left\{\frac{\int K'(t)^2 dt}{A_2(K)^2\int\{f^{(2)}(t)\}^2 dt}\right\}^{1/5} n^{-1/5},$$

and the optimal bandwidth h_{opt_2} asymptotically minimizing $\text{MISE}(\tilde{f}_{\text{RSS}}^{\text{Sym}_2}(\cdot;\mu))$ is

$$h_{\text{opt}_2} = \left\{\frac{n}{k^2}\sum_{i=1}^{k}\frac{1}{m_i+m_{k-i+1}}\right\}^{1/5}\left\{\frac{\int K'(t)^2 dt}{A_2(K)^2\int\{f^{(2)}(t)\}^2 dt}\right\}^{1/5} n^{-1/5}.$$

Herein, let the asymptotic MISE (AMISE) of the kernel density estimators $\tilde{f}_{\text{RSS}}^{\text{Sym}_1}$ and $\tilde{f}_{\text{RSS}}^{\text{Sym}_2}$ be

$$\text{AMISE}_h(\tilde{f}_{\text{RSS}}^{\text{Sym}_1}(\cdot;\mu)) = \frac{h^4}{4}A_2(K)^2\int\{f^{(2)}(t)\}^2 dt + \frac{1}{nh}\left(\frac{n}{2k^2}\sum_{i=1}^{k}\frac{1}{m_i}\right)\int K'(t)^2 dt$$

and

$$\text{AMISE}_h(\tilde{f}_{\text{RSS}}^{\text{Sym}_2}(\cdot;\mu)) = \frac{h^4}{4}A_2(K)^2\int\{f^{(2)}(t)\}^2 dt + \frac{1}{nh}\left\{\frac{n}{k^2}\sum_{i=1}^{k}\frac{1}{m_i+m_{k-i+1}}\right\}\int K'(t)^2 dt,$$

respectively. By substituting the optimum bandwidths h_{opt_1} and h_{opt_2} into the corresponding AMISE, we have

$$\text{AMISE}_{h_{\text{opt}_1}}(\tilde{f}_{\text{RSS}}^{\text{Sym}_1}(\cdot;\mu)) = h_{\text{opt}_1}^5\left\{\frac{5}{4}A_2(K)^2\int\{f^{(2)}(t)\}^2 dt\right\}$$

and

$$\text{AMISE}_{h_{\text{opt}_2}}(\tilde{f}_{\text{RSS}}^{\text{Sym}_2}(\cdot;\mu)) = h_{\text{opt}_2}^5\left\{\frac{5}{4}A_2(K)^2\int\{f^{(2)}(t)\}^2 dt\right\}.$$

Then, the ratio of these AMISEs is

$$\frac{\text{AMISE}_{h_{\text{opt}_2}}(\tilde{f}_{\text{RSS}}^{\text{Sym}_2}(\cdot;\mu))}{\text{AMISE}_{h_{\text{opt}_1}}(\tilde{f}_{\text{RSS}}^{\text{Sym}_1}(\cdot;\mu))} = \left(\frac{h_{\text{opt}_2}}{h_{\text{opt}_1}}\right)^5 = \frac{\sum_{i=1}^{k}1/(m_i+m_{k-i+1})}{\sum_{i=1}^{k}1/(2m_i)} \leq 1,$$

where the equality holds if and only if $m_i = m_{k-i+1}$ for each $i = 1,\dots,k$.

Next, we derive the optimal bandwidth that asymptotically minimizes the MISE of the symmetrized kernel distribution estimators $\tilde{F}_{\text{RSS}}^{\text{Sym}_1}(x;\mu)$ and $\tilde{F}_{\text{RSS}}^{\text{Sym}_2}(x;\mu)$. Through a simple calculation, we have

$$\begin{aligned}
&\text{MISE}[\tilde{F}_{\text{RSS}}^{\text{Sym}_1}(\cdot;\mu)] \\
&= \int \text{bias}[\tilde{F}_{\text{RSS}}^{\text{Sym}_1}(x;\mu)]^2 dx + \int \text{Var}[\tilde{F}_{\text{RSS}}^{\text{Sym}_1}(x;\mu)] dx \\
&= \frac{h^4}{4}A_2(K)^2 D_1(F) + \frac{1}{2k^2}\sum_{i=1}^{k}\int \text{Var}[\tilde{F}_{[i]}(x)]dx - \frac{1}{2k^2}\sum_{i=1}^{k}\int \text{Cov}[\tilde{F}_{[i]}(x), \tilde{F}_{[i]}(2\mu - x)]dx
\end{aligned}$$

$$
\begin{aligned}
&= \frac{h^4}{4} A_2(K)^2 D_1(F) + \frac{1}{2k^2} \sum_{i=1}^{k} \frac{1}{m_i} D_2(F_{[i]}) - A_1(K) \frac{h}{k^2} \sum_{i=1}^{k} \frac{1}{m_i} \\
&\quad - \frac{1}{2k^2} \sum_{i=1}^{k} \frac{1}{m_i} \left\{ \int_0^\infty F_{[i]} \left(\mu - \frac{x}{2} \right) dx - \int_{-\infty}^{\infty} F_{[i]}(x) F_{[i]}(2\mu - x) dx + o(1) \right\} \\
&\quad + o\left(\max\left\{ h^4, \frac{h}{\min_i(m_i)} \right\} \right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{MISE}[\tilde{F}_{\mathrm{RSS}}^{\mathrm{Sym}_2}(\cdot;\mu)] \\
&= \int \mathrm{bias}[\tilde{F}_{\mathrm{RSS}}^{\mathrm{Sym}_2}(x;\mu)]^2 dx + \int \mathrm{Var}[\tilde{F}_{\mathrm{RSS}}^{\mathrm{Sym}_2}(x;\mu)] dx \\
&= \frac{h^4}{4} A_2(K)^2 D_1(F) + \frac{2}{k^2} \sum_{i=1}^{k} \left(\frac{m_i}{m_i + m_{k-i+1}} \right)^2 \int \mathrm{Var}[\tilde{F}_{[i]}(x)] dx \\
&\quad - \frac{2}{k^2} \sum_{i=1}^{k} \left(\frac{m_i}{m_i + m_{k-i+1}} \right)^2 \int \mathrm{Cov}[\tilde{F}_{[i]}(x), \tilde{F}_{[i]}(2\mu - x)] dx \\
&= \frac{h^4}{4} A_2(K)^2 D_1(F) + \frac{2}{k^2} \sum_{i=1}^{k} \frac{m_i D_2(F_{[i]})}{(m_i + m_{k-i+1})^2} - A_1(K) \frac{h}{k^2} \sum_{i=1}^{k} \frac{4 m_i}{(m_i + m_{k-i+1})^2} \\
&\quad - \frac{2}{k^2} \sum_{i=1}^{k} \frac{m_i}{(m_i + m_{k-i+1})^2} \left\{ \int_0^\infty F_{[i]} \left(\mu - \frac{x}{2} \right) dx - \int_{-\infty}^{\infty} F_{[i]}(x) F_{[i]}(2\mu - x) dx + o(1) \right\} \\
&\quad + o\left(\max\left\{ h^4, \frac{h}{\min_i(m_i)} \right\} \right).
\end{aligned}
$$

Therefore, the optimal bandwidth h_{opt_1} that asymptotically minimizes $\mathrm{MISE}[\tilde{F}_{\mathrm{RSS}}^{\mathrm{Sym}_1}(\cdot;\mu)]$ is

$$
h_{\mathrm{opt}_1} = \left(\frac{n}{2k^2} \sum_{i=1}^{k} \frac{1}{m_i} \right)^{1/3} \left(\frac{2 A_1(K)}{A_2(K)^2 D_1(F)} \right)^{1/3} n^{-1/3},
$$

and the optimal bandwidth h_{opt_2} that asymptotically minimizes $\mathrm{MISE}[\tilde{F}_{\mathrm{RSS}}^{\mathrm{Sym}_2}(\cdot;\mu)]$ is

$$
h_{\mathrm{opt}_2} = \left(\frac{n}{k^2} \sum_{i=1}^{k} \frac{1}{m_i + m_{k-i+1}} \right)^{1/3} \left(\frac{2 A_1(K)}{A_2(K)^2 D_1(F)} \right)^{1/3} n^{-1/3}.
$$

Similarly to Lim et al. (2014), by regarding RSS data as SRS data, the parameters that depend on the true distribution F, including the asymptotically optimal bandwidth $D_1(F)$, are estimated using a plug-in method based on Sheather and Jones (1991).

2.4 Simulation studies

Let F_n be an estimator of the CDF F, and construct a bootstrap empirical distribution F_n^* using any one of the several available bootstrap methods. Given a statistic of interest $\hat{\theta}_n = \theta(F_n)$, the corresponding bootstrap replicate is defined by $\hat{\theta}_n^* = \theta(F_n^*)$. To approximate the sampling distribution $H_{n,F}(t) = \mathbb{P}_F(\hat{\theta}_n \leq t)$, in principle, we use the bootstrap estimate $H_{n,F_n}(t) = \mathbb{P}_{F_n}(\hat{\theta}_n^* \leq t)$. However, for practicality, we employ a Monte Carlo simulation to estimate $H_{n,F_n}(t)$ by taking B independent bootstrap samples from F_n. Thus, we have $F_{n,1}^*, \ldots, F_{n,B}^*$ and the corresponding bootstrap replicates $\hat{\theta}_{n,1}^*, \ldots, \hat{\theta}_{n,B}^*$. The Monte Carlo approximation of $H_{n,F_n}(t)$ is defined by $\hat{H}_{n,F_n}(t) = (1/B)\sum_{b=1}^{B} I(\hat{\theta}_{n,b}^* \leq t)$.

In this section, we assume the imperfect ranking model introduced by Bohn and Wolfe (1994). For the set sizes $k = 2, 3, 4$, we consider the following misplacement probability matrix $P_k = [p_{ij}]_{k\times k}$:

$$P_2 = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}, \quad P_3 = \begin{pmatrix} 0.9 & 0.1 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0 & 0.1 & 0.9 \end{pmatrix}, \quad P_4 = \begin{pmatrix} 0.9 & 0.1 & 0 & 0 \\ 0.1 & 0.8 & 0.1 & 0 \\ 0 & 0.1 & 0.8 & 0.1 \\ 0 & 0 & 0.1 & 0.9 \end{pmatrix}.$$

2.4.1 Confidence intervals for a population mean

In this subsection, we use a Monte Carlo simulation based on $10,000$ repetitions to study an RSS-based sample mean. We compare the performance of the BRSSR, SBRSSR, SymBRSSR, and SymSBRSSR methods with the estimated coverage frequency of a 95% confidence interval for the population mean.

In the estimation, we use the Gaussian kernel distribution $K = \Phi$, where Φ is the standard normal distribution function. The bandwidths for the SRS data using the procedure developed by Sheather and Jones (1991) and Polansky and Baker (2000) are represented as $h_{\mathrm{SRS,SJ}}$ and $h_{\mathrm{SRS,PB}}$, respectively. For the non-symmetrized RSS-based kernel estimators, we use the bandwidth

$$h_{\mathrm{SJ}} = \left(\frac{n}{k^2}\sum_{i=1}^{k}\frac{1}{m_i}\right)^{1/5} h_{\mathrm{SRS,SJ}} \quad \text{and} \quad h_{\mathrm{PB}} = \left(\frac{n}{k^2}\sum_{i=1}^{k}\frac{1}{m_i}\right)^{1/3} h_{\mathrm{SRS,PB}}.$$

For the symmetrized RSS-based kernel estimators, we use the bandwidths

$$h_{\mathrm{SJ}}^{\mathrm{Sym}} = \left(\frac{n}{k^2}\sum_{i=1}^{k}\frac{1}{m_i + m_{k-i+1}}\right)^{1/5} h_{\mathrm{SRS,SJ}}$$

and

$$h_{\mathrm{PB}}^{\mathrm{Sym}} = \left(\frac{n}{k^2}\sum_{i=1}^{k}\frac{1}{m_i + m_{k-i+1}}\right)^{1/3} h_{\mathrm{SRS,PB}}.$$

We consider the following different bootstrap methods in this simulation study:

- BRSSR: bootstrap RSS by row (Chen et al., 2004),
- SymBRSSR: unequal-weighted symmetric BRSSR,

- SB(h): SBRSSR with bandwidth h,
- SymSB(h): unequal-weighted symmetric SBRSSR with bandwidth h.

We compute the coverage frequency of the 95% percentile intervals of the sample mean using the above bootstrap methods based on $B = 1,000$ bootstrap samples. We use three distributions as the true underlying distributions: (i) a standard normal distribution; (ii) a t-distribution with three degrees of freedom; and (iii) a uniform distribution on $[0, 1]$. We consider $k = 2, 3, 4$, and the misplacement probability matrix uses P_2, P_3, and P_4, respectively.

Table 2.1 shows that the smoothed bootstrap methods are generally conservative. Because the bandwidths h_{SJ} and $h_{\mathrm{SJ}}^{\mathrm{Sym}}$ tend to be larger than h_{PB} and $h_{\mathrm{PB}}^{\mathrm{Sym}}$, the confidence interval becomes too conservative. In addition, it can be seen that the symmetric bootstrap methods improve the accuracy of the coverage probability.

2.4.2 Application for testing symmetry

In this subsection, we introduce a bootstrap-based test statistic for symmetry under the balanced and unbalanced RSS settings and investigate the influence of the smoothing for the bootstrap test.

We devise a nonparametric test for

$$H_0 : F \in \mathcal{F}_S^{\mu} \quad \text{against} \quad H_1 : F \notin \mathcal{F}_S^{\mu},$$

where μ is an unknown center of symmetry. From Proposition 2.2.3, we propose an Öztürk (2001) test statistic for symmetry

$$T(\hat{\mu}_{\mathrm{HL}}) = n \left\{ -1 + \frac{\sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \frac{1}{m_{i_1} m_{i_2}} \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_2}} |X_{[i_1]j_1} + X_{[i_2]j_2} - 2\hat{\mu}_{\mathrm{HL}}|}{\sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \frac{1}{m_{i_1} m_{i_2}} \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_2}} |X_{[i_1]j_1} - X_{[i_2]j_2}|} \right\}$$

under the unbalanced RSS settings. To calculate the p-value of the test, we need to estimate the distribution of T under H_0. Using the same method as Öztürk (2001), we compare the Type I error rates of the tests for symmetry based on the SymBRSSR method and the unequal-weighted SymSBRSSR method. To estimate the Type I error rates, we generate 100 samples for each sample size, unbalanced RSS design, and distribution combination. For each of the 100 samples, we generate $B = 5,000$ symmetric bootstrap samples from the symmetrized CDF estimator and calculate the test statistic T. Based on T^b ($b = 1, 2, \ldots, B$), we estimate the critical values t_i^*, $i = 1, \ldots, 100$, for an α-percent test; that is,

$$\frac{1}{B} \sum_{b=1}^{B} I(T^b \geq t_i^*) = \alpha.$$

The bootstrap estimate of the critical value of an α-percent test is taken as the average of these t_i^*, $i = 1, \ldots, 100$, $\bar{t}^* = \sum_{i=1}^{100} t_i^*/100$. To investigate the accuracy of the estimated critical value

Table 2.1. Coverage probability of the 95% percentile intervals of the sample means for symmetric distributions

Dist.	$\{m_1,\dots,m_k\}$	BRSSR	SymBRSSR	SB(h_{SJ})	SymSB($h_{\mathrm{SJ}}^{\mathrm{Sym}}$)	SB(h_{PB})	SymSB($h_{\mathrm{PB}}^{\mathrm{Sym}}$)
$N(0,1)$	$\{10,10\}$	0.920	0.920	0.957	0.951	0.964	0.952
	$\{20,20\}$	0.933	0.937	0.962	0.960	0.961	0.956
	$\{40,40\}$	0.945	0.944	0.964	0.962	0.961	0.957
	$\{10,10,10\}$	0.928	0.932	0.967	0.965	0.969	0.962
	$\{10,20,10\}$	0.926	0.931	0.962	0.961	0.963	0.957
	$\{20,10,20\}$	0.927	0.934	0.966	0.968	0.964	0.962
	$\{20,20,20\}$	0.936	0.939	0.968	0.963	0.965	0.958
	$\{20,40,20\}$	0.938	0.938	0.966	0.959	0.961	0.956
	$\{40,20,40\}$	0.941	0.941	0.972	0.967	0.967	0.958
	$\{40,40,40\}$	0.941	0.947	0.965	0.964	0.958	0.958
	$\{10,10,10,10\}$	0.925	0.929	0.973	0.966	0.970	0.961
	$\{10,20,20,10\}$	0.925	0.932	0.961	0.966	0.958	0.964
	$\{20,10,10,20\}$	0.932	0.935	0.976	0.962	0.973	0.956
	$\{20,20,20,20\}$	0.939	0.940	0.974	0.968	0.968	0.960
	$\{20,40,40,20\}$	0.941	0.940	0.965	0.959	0.960	0.952
	$\{40,20,20,40\}$	0.947	0.939	0.978	0.967	0.971	0.959
	$\{40,40,40,40\}$	0.935	0.945	0.964	0.966	0.956	0.956
$t(3)$	$\{10,10\}$	0.910	0.927	0.945	0.950	0.951	0.951
	$\{20,20\}$	0.932	0.942	0.951	0.958	0.951	0.955
	$\{40,40\}$	0.933	0.946	0.948	0.955	0.946	0.952
	$\{10,10,10\}$	0.914	0.937	0.949	0.958	0.951	0.957
	$\{10,20,10\}$	0.922	0.935	0.948	0.953	0.948	0.950
	$\{20,10,20\}$	0.930	0.945	0.961	0.964	0.959	0.961
	$\{20,20,20\}$	0.930	0.943	0.952	0.958	0.950	0.955
	$\{20,40,20\}$	0.928	0.948	0.946	0.958	0.942	0.955
	$\{40,20,40\}$	0.935	0.950	0.954	0.963	0.949	0.960
	$\{40,40,40\}$	0.935	0.951	0.948	0.959	0.945	0.955
	$\{10,10,10,10\}$	0.918	0.942	0.952	0.961	0.951	0.958
	$\{10,20,20,10\}$	0.924	0.937	0.948	0.954	0.945	0.950
	$\{20,10,10,20\}$	0.928	0.942	0.963	0.966	0.960	0.960
	$\{20,20,20,20\}$	0.930	0.943	0.951	0.959	0.948	0.955
	$\{20,40,40,20\}$	0.929	0.947	0.944	0.956	0.939	0.952
	$\{40,20,20,40\}$	0.941	0.947	0.964	0.963	0.958	0.957
	$\{40,40,40,40\}$	0.938	0.945	0.951	0.957	0.947	0.952
$U(0,1)$	$\{10,10\}$	0.921	0.919	0.955	0.949	0.961	0.951
	$\{20,20\}$	0.935	0.934	0.961	0.954	0.961	0.952
	$\{40,40\}$	0.949	0.942	0.963	0.955	0.962	0.953
	$\{10,10,10\}$	0.928	0.933	0.966	0.962	0.968	0.960
	$\{10,20,10\}$	0.926	0.931	0.964	0.960	0.965	0.957
	$\{20,10,20\}$	0.930	0.936	0.961	0.959	0.962	0.955
	$\{20,20,20\}$	0.936	0.942	0.964	0.963	0.963	0.959
	$\{20,40,20\}$	0.941	0.940	0.965	0.961	0.963	0.956
	$\{40,20,40\}$	0.939	0.942	0.959	0.957	0.958	0.954
	$\{40,40,40\}$	0.941	0.947	0.960	0.961	0.958	0.956
	$\{10,10,10,10\}$	0.925	0.931	0.970	0.964	0.971	0.961
	$\{10,20,20,10\}$	0.934	0.937	0.971	0.968	0.970	0.962
	$\{20,10,10,20\}$	0.927	0.931	0.964	0.962	0.964	0.957
	$\{20,20,20,20\}$	0.936	0.941	0.967	0.965	0.964	0.959
	$\{20,40,40,20\}$	0.940	0.941	0.969	0.965	0.963	0.957
	$\{40,20,20,40\}$	0.940	0.941	0.961	0.958	0.959	0.954
	$\{40,40,40,40\}$	0.944	0.949	0.966	0.961	0.961	0.958

$\bar{t}^*$, we generate $10,000$ independent samples for each combination of sample size and distribution and calculate

$$\alpha(\bar{t}^*) = \frac{\sum_{i=1}^{10000} I(T^i(\hat{\mu}_{\mathrm{HL}}) > \bar{t}^*)}{10000},$$

where $T^i(\hat{\mu}_{\mathrm{HL}})$ is the test statistic evaluated in an independent sample. Let T_B be the bootstrap test based on the unequal-weighted SymBRSSR method, and let T_{SB} be the bootstrap test based on the unequal-weighted SymSBRSSR method.

Table 2.2. Type I error rate for T_B and T_{SB} at the nominal significance level 0.05

		$N(0,1)$		$t(3)$		$U(0,1)$	
k	$\{m_1,\ldots,m_k\}$	T_B	T_{SB}	T_B	T_{SB}	T_B	T_{SB}
2	$\{10,10\}$	0.0235	0.0432	0.0263	0.0458	0.0320	0.0710
2	$\{10,20\}$	0.0257	0.0454	0.0306	0.0480	0.0371	0.0699
2	$\{20,20\}$	0.0299	0.0478	0.0332	0.0504	0.0366	0.0676
2	$\{20,40\}$	0.0342	0.0468	0.0392	0.0515	0.0374	0.0612
2	$\{40,40\}$	0.0357	0.0467	0.0369	0.0487	0.0414	0.0649
3	$\{10,10,10\}$	0.0299	0.0466	0.0312	0.0478	0.0345	0.0618
3	$\{10,20,10\}$	0.0328	0.0464	0.0298	0.0421	0.0380	0.0613
3	$\{20,10,20\}$	0.0309	0.0463	0.0347	0.0499	0.0358	0.0622
3	$\{20,20,20\}$	0.0349	0.0472	0.0379	0.0501	0.0394	0.0608
3	$\{20,40,20\}$	0.0369	0.0474	0.0375	0.0489	0.0402	0.0593
3	$\{40,20,40\}$	0.0383	0.0487	0.0405	0.0511	0.0360	0.0565
3	$\{40,40,40\}$	0.0395	0.0482	0.0391	0.0477	0.0419	0.0605
4	$\{10,10,10,10\}$	0.0363	0.0468	0.0358	0.0490	0.0395	0.0593
4	$\{10,20,20,10\}$	0.0426	0.0545	0.0524	0.0685	0.0397	0.0556
4	$\{20,10,10,20\}$	0.0399	0.0501	0.0343	0.0434	0.0576	0.0776
4	$\{20,20,20,20\}$	0.0389	0.0474	0.0416	0.0522	0.0419	0.0575
4	$\{20,40,40,20\}$	0.0461	0.0549	0.0512	0.0602	0.0420	0.0545
4	$\{40,20,20,40\}$	0.0439	0.0505	0.0349	0.0420	0.0592	0.0760
4	$\{40,40,40,40\}$	0.0426	0.0481	0.0435	0.0507	0.0465	0.0592

Table 2.2 shows that the test T_{SB} has the Type I error rate close to the nominal $\alpha = 0.05$ level for $N(0,1)$ and $t(3)$. However, T_{SB} is anti-conservative for $U(0,1)$. By contrast, T_B is too conservative overall compared with T_{SB}. In addition, we verify that the smoothed bootstrap method works effectively for unbalanced RSS designs. For asymmetric distributions, the critical value $\bar{t}^*$ still estimates the critical value of a test because it is calculated based on symmetric bootstrap samples through the use of the unequal-weighted SymBRSSR method and the unequal-weighted SymSBREER method. Therefore, $\alpha(\bar{t}^*)$ yields the empirical power of the test if the independent RSS data are generated from an asymmetric distribution. The values of empirical power $\alpha(\bar{t}^*)$ are presented in Table 2.3 for several asymmetric distributions. The distributions include the skew-normal distribution with shape parameter 5, SN(5); the chi-square distribution with 3 degrees of freedom, χ^2_3; and the beta distribution with shape parameters 4 and 2, Beta$(4,2)$.

Table 2.3. Percentage of rejection for T_B and T_{SB} at the nominal significance level 0.05

		SN(5)		χ^2_5		Beta(4, 2)	
k	$\{m_1, \ldots, m_k\}$	T_B	T_{SB}	T_B	T_{SB}	T_B	T_{SB}
2	$\{10, 10\}$	0.1187	0.1958	0.1886	0.2807	0.0645	0.1077
2	$\{10, 20\}$	0.2110	0.2809	0.2790	0.3694	0.0815	0.1259
2	$\{20, 20\}$	0.3301	0.4100	0.4756	0.5646	0.1469	0.1995
2	$\{20, 40\}$	0.4836	0.5543	0.6736	0.7321	0.2066	0.2528
2	$\{40, 40\}$	0.7048	0.7508	0.8706	0.8977	0.3544	0.4053
3	$\{10, 10, 10\}$	0.2460	0.3172	0.3719	0.4489	0.1133	0.1539
3	$\{10, 20, 10\}$	0.3174	0.3827	0.4462	0.5132	0.1505	0.1896
3	$\{20, 10, 20\}$	0.3955	0.4775	0.5777	0.6563	0.1804	0.2302
3	$\{20, 20, 20\}$	0.5503	0.6066	0.7433	0.7867	0.2598	0.3057
3	$\{20, 40, 20\}$	0.6574	0.7000	0.8351	0.8613	0.3253	0.3686
3	$\{40, 20, 40\}$	0.7749	0.8167	0.9297	0.9475	0.3852	0.4410
3	$\{40, 40, 40\}$	0.9017	0.9172	0.9805	0.9857	0.5558	0.5913
4	$\{10, 10, 10, 10\}$	0.3853	0.4375	0.5329	0.5909	0.1744	0.2107
4	$\{10, 20, 20, 10\}$	0.5402	0.5854	0.6996	0.7404	0.2623	0.3016
4	$\{20, 10, 10, 20\}$	0.5461	0.5937	0.7281	0.7716	0.2597	0.2954
4	$\{20, 20, 20, 20\}$	0.7400	0.7724	0.8956	0.9145	0.3858	0.4202
4	$\{20, 40, 40, 20\}$	0.8756	0.8923	0.9677	0.9733	0.5430	0.5747
4	$\{40, 20, 20, 40\}$	0.8915	0.9055	0.9790	0.9838	0.5367	0.5691
4	$\{40, 40, 40, 40\}$	0.9752	0.9790	0.9982	0.9986	0.7294	0.7548

From the results of Table 2.3, it can be seen that the unequal-weighted SymSBRSSR method improves the power compared to the unequal-weighted SymBRSSR method. In the case of the balanced RSS settings, it is shown that the power tends to increase as the set size k increases. Under the skew-normal and chi-square distributions, the central RSS designs such as $(10, 20, 20, 10)$ and $(20, 40, 40, 20)$ tend to have less power than the extreme RSS designs such as $(20, 10, 10, 20)$ and $(40, 20, 20, 40)$. From the results for the uniform distribution shown in Table 2.2, it should be noted that the Type I error rate of T_{SB} may be anti-conservative for distribution with a compact support such as the beta distribution.

2.4.3 Data example

In this section, we compare the performance of the proposed bootstrap methods using the real dataset given by Hollander et al. (2014, pp. 709–713). The dataset includes the weight of 224 sheep on a research farm at Ataturk University in Erzurum, Turkey, which includes two variables: birth weight and weight at 7 months after birth. The frequency distributions of the birth weight and the 7th-month weight of the sheep population are all approximately symmetric. The variable of interest is the weight (Y) at 7 months after birth. Birth weight (X) is available in the data frame with no additional cost. The correlation coefficient between X and Y is $\text{Corr}(X, Y) = 0.843$. Thus, this auxiliary variable X, which is positively correlated with the variable of interest Y, can be used to complete the ranking process in the RSS samples.

From the weight measurements of 224 sheep, the mean of variable Y is computed as $\mu_Y = 28.111$ kg. We regard the bivariate empirical distribution from this dataset as the population distribution and treat the mean μ_Y as the true parameter of the population model. An RSS sample is taken from this empirical distribution, a confidence interval for the population mean μ_Y is constructed using the same procedure as described in Section 2.4.1, and the performances of various bootstrap methods are compared. Because this model is a finite population, the bootstrap methods should be modified. However, for simplicity, we consider the case in which the total sample size is $\sum_{r=1}^{k} m_r \leq 25$, and analyze it in the same way as the infinite population. These results are shown in Table 2.4.

Table 2.4. Coverage probability of the 95% percentile intervals of the sample means for the weight of 224 sheep data

$\{m_1, \ldots, m_k\}$	BRSSR	SymBRSSR	SB(h_{SJ})	SymSB($h_{\mathrm{SJ}}^{\mathrm{Sym}}$)	SB(h_{PB})	SymSB($h_{\mathrm{PB}}^{\mathrm{Sym}}$)
$\{5, 5\}$	0.891	0.892	0.947	0.938	0.963	0.949
$\{10, 10\}$	0.934	0.934	0.968	0.961	0.973	0.962
$\{5, 5, 5\}$	0.911	0.923	0.965	0.963	0.973	0.966
$\{5, 10, 5\}$	0.911	0.920	0.961	0.958	0.967	0.960
$\{10, 5, 10\}$	0.929	0.941	0.973	0.972	0.976	0.972
$\{5, 5, 5, 5\}$	0.927	0.936	0.973	0.972	0.980	0.974

These results indicate that the symmetric bootstrap methods improve the coverage probability of the confidence intervals. The BRSSR and SymBRSSR methods tend to be anti-conservative, and the smoothed bootstrap methods were verified to be more conservative than the EDF-based approaches.

Chapter 3

Interpoint distance-based two-sample tests for functional data

In this chapter, we propose two-sample tests based on the interpoint distance for functional data. Chapter 3 is organized as follows. For infinite-dimensional data in a Banach space, the test statistics are proposed in Section 3.1. In addition, we derive the asymptotic properties of the test statistic and proposed the p-value approximation based on the jackknife variance estimators and the Welch–Satterthwaite equation. In Section 3.3, we verify the accuracy of the proposed p-value approximation of the test and compare the power of the proposed tests and the functional Anderson–Darling test based on Pomann et al. (2016).

3.1 Test statistics for homogeneity of infinite-dimensional data on a Banach space

In this section, we first construct the test statistics for ideal curves and discuss the properties of the test, such as the limiting distribution and asymptotic power. However, there are few cases in which true smooth curves are available in practice. Therefore, pre-smoothing is necessary in many cases. Hall and Van Keilegom (2007) and Pomann et al. (2016) suggested that the individual smoothing of each curve significantly affects the performance of the test. The methods of pre-smoothing and the implementation of the test statistic calculation used to reduce the influence are described in Section 3.3.

Let $\mathcal{X}$ be a Banach space. If $\mathcal{X} = L_p[a,b]$ for some $p > 1$, which is the Banach space of all functions $x : [a,b] \to \mathbb{R}$ satisfying $\int_a^b |x(s)|^p ds < \infty$, then

$$\lim_{t \to 0} t^{-1}(\|x + th\|_p - \|x\|_p) = \int_a^b |x(s)|^{p-2} x(s) h(s) ds / \|x\|_p^{p-1} \tag{3.1.1}$$

for any $x \neq 0$ and $h \in L_p[a,b]$. If $\mathcal{X} = L_1[a,b]$, then

$$\lim_{t \to 0} t^{-1}(\|x + th\|_1 - \|x\|_1) = \int_a^b \mathrm{sign}(x(s)) h(s) ds$$

for any $x \overset{a.e.}{\neq} 0$ and $h \in L_1[a,b]$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space: Ω is the set where the random experiment takes place, $\mathcal{F}$ is a σ-algebra of subsets of Ω, and $\mathbb{P}$ is a probability measure over $\mathcal{F}$. For simplicity, we assume that this space is complete (i.e., $\mathcal{F}$ contains the $\mathbb{P}$-negligible sets). Let X_1 and X_2 be random elements in a Banach space $\mathcal{X}$ with the probability measures P_1 and P_2 on the measurable space $(\mathcal{X}, \mathcal{A})$, where $\mathcal{A}$ is the σ-field generated by the open sets induced by the norm $\|\cdot\|$. Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ be independent copies of X_1 and X_2, respectively. Suppose that $\mathcal{L} : \mathcal{X}^2 \to \mathbb{R}$ is a continuous function. Let $\mathcal{B}_{\mathcal{L}}$ be the set of all probability measures P on $(\mathcal{X}, \mathcal{A})$ under condition $\int_{\mathcal{X}} \int_{\mathcal{X}} \mathcal{L}(x,y) dP(x) dP(y) < \infty$.

We state that function $\mathcal{L}$ is a negative definite kernel if for any $n \in \mathbb{N}$, arbitrary points $x_1, \ldots, x_n \in \mathcal{X}$, and any complex numbers $c_1, \ldots, c_n$ under condition $\sum_{j=1}^n c_j = 0$, the following inequality holds:

$$\sum_{i=1}^n \sum_{j=1}^n \mathcal{L}(x_i, x_j) c_i \bar{c}_j \leq 0,$$

where $\bar{c}$ denotes the complex conjugate of a complex number c. We state that a negative definite kernel $\mathcal{L}$ is strictly negative definite if the above equality is true for any $x_1, \ldots, x_n$ only if $c_1 = \cdots = c_n = 0$. The important property of the negative definite kernel $\mathcal{L}$ is shown by the following lemma.

Lemma 3.1.1 (Klebanov 2005, Theorem 1.8)**.** *Let $\mathcal{L}$ be a real continuous function on $\mathcal{X}^2$ under conditions $\mathcal{L}(x,y) = \mathcal{L}(y,x)$ and $\mathcal{L}(x,x) = 0$ for any $x, y \in \mathcal{X}$. The inequality*

$$\begin{aligned} \mathfrak{N}(P_1, P_2) := 2 &\int_{\mathcal{X}} \int_{\mathcal{X}} \mathcal{L}(x,y) dP_1(x) dP_2(y) \\ &- \int_{\mathcal{X}} \int_{\mathcal{X}} \mathcal{L}(x,y) dP_1(x) dP_1(y) - \int_{\mathcal{X}} \int_{\mathcal{X}} \mathcal{L}(x,y) dP_2(x) dP_2(y) \geq 0 \end{aligned} \tag{3.1.2}$$

holds $\forall\ P_1, P_2 \in \mathcal{B}_{\mathcal{L}}$ if and only if $\mathcal{L}$ is a negative definite kernel.

One of the main notions in the theory of the $\mathfrak{N}$-distance $\mathfrak{N}(\cdot,\cdot)$ is a strong negative definiteness. Let Q be a measure on $(\mathcal{X}, \mathcal{A})$, and h be a function integrable with respect to Q, such that $\int_{\mathcal{X}} h(x) dQ(x) = 0$. We state that $\mathcal{L}$ is a strongly negative definite kernel if $\mathcal{L}$ is a negative definite, and equality $\int_{\mathcal{X}} \int_{\mathcal{X}} \mathcal{L}(x,y) h(x) h(y) dQ(x) dQ(y) = 0$ implies that $h(x) = 0$ Q-almost everywhere for any measure Q. If $\mathcal{X} = L_p[a,b]$, for example, the kernel $\mathcal{L}_p^\alpha(x,y) = \|x-y\|_p^\alpha$ $(0 < \alpha \leq p < 2)$ is a strongly negative definite kernel, but not a strongly negative definite kernel for $p = 2$. The following lemma is necessary to show the consistency of the interpoint distance-based tests.

Lemma 3.1.2 (Klebanov 2005, Theorem 1.9)**.** *Let $\mathcal{L}$ be a real continuous function on $\mathcal{X}^2$ under conditions $\mathcal{L}(x,y) = \mathcal{L}(y,x)$ and $\mathcal{L}(x,x) = 0$ for any $x, y \in \mathcal{X}$. The inequality* (3.1.2)

$$\mathfrak{N}(P_1, P_2) \geq 0$$

holds for all measures $P_1, P_2 \in \mathcal{B}_{\mathcal{L}}$ with equality in the case of $P_1 = P_2$ only, if and only if $\mathcal{L}$ is a strongly negative definite kernel.

Another possible expression of $\mathfrak{N}(P_1, P_2)$ can be given in terms of random elements. Let $X_{11}, X_{12} \overset{\text{i.i.d.}}{\sim} P_1$ and $X_{21}, X_{22} \overset{\text{i.i.d.}}{\sim} P_2$ be independent random elements. We can now write $\mathfrak{N}(P_1, P_2)$ in the following form:

$$\mathfrak{N}(P_1, P_2) = 2\mathbb{E}\mathcal{L}(X_{11}, X_{21}) - \mathbb{E}\mathcal{L}(X_{11}, X_{12}) - \mathbb{E}\mathcal{L}(X_{21}, X_{22}),$$

where $\mathbb{E}(\cdot)$ is the expectation in the Bochner sense (see Section 2, Chapter 3 in Araujo and Giné (1980)). We now assume that $\mathbb{E}\mathcal{L}(X_{i1}, X_{i2})^2 < \infty$ $(i = 1, 2)$. By applying the ideas of Baringhaus and Franz (2004) and Biswas and Ghosh (2014) to functional or infinite-dimensional data, we consider testing the null hypothesis $H_0 : P_1 = P_2$ versus the alternative hypothesis $H_1 : P_1 \neq P_2$. Using a strongly negative definite kernel $\mathcal{L}$, we suggest the Baringhaus–Franz type test T^{BF} and the Biswas–Ghosh type test T^{BG} as follows:

$$T^{BF} = 2\hat{\mu}_{12} - \hat{\mu}_{11} - \hat{\mu}_{22} \quad \text{and} \quad T^{BG} = (\hat{\mu}_{12} - \hat{\mu}_{11})^2 + (\hat{\mu}_{12} - \hat{\mu}_{22})^2,$$

where

$$\hat{\mu}_{11} = \frac{1}{\binom{n_1}{2}} \sum\sum_{1 \le j < k \le n_1} \mathcal{L}(X_{1j}, X_{1k}), \quad \hat{\mu}_{22} = \frac{1}{\binom{n_2}{2}} \sum\sum_{1 \le j < k \le n_2} \mathcal{L}(X_{2j}, X_{2k})$$

and

$$\hat{\mu}_{12} = \frac{1}{n_1 n_2} \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} \mathcal{L}(X_{1j}, X_{2k}).$$

Here, the quantities $\hat{\mu}_{12}$, $\hat{\mu}_{11}$, and $\hat{\mu}_{22}$ are U-statistics for the expectations $\mu_{11} := \mathbb{E}\mathcal{L}(X_{11}, X_{12})$, $\mu_{22} := \mathbb{E}\mathcal{L}(X_{21}, X_{22})$ and $\mu_{12} := \mathbb{E}\mathcal{L}(X_{11}, X_{21})$, respectively. In particular, the Baringhaus–Franz type statistic T^{BF} is called the energy statistic (see Klebanov (2005), Bakshaev (2009), and Székely and Rizzo (2013) for details). Clearly, the test statistic T^{BF} is not a distribution-free test. To obtain the critical values, Baringhaus and Franz (2004) proposed using the bootstrap approach and showed that the bootstrap procedure is consistent against any fixed alternative $P_1 \neq P_2$ with finite expectations $\int \|x_1\| dP_1(x_1)$ and $\int \|x_2\| dP_2(x_2)$ for finite-dimensional data. If $\mathcal{X} = L_p[a, b]$ $(0 < p < 2)$, the test statistics T_p^{BF} and T_p^{BG} based on the L_p-norm $\|x\|_p = \left(\int_a^b |x(t)|^p dt\right)^{1/p}$ hold the consistency of the test, although the consistency of the L_2-norm-based statistics T_2^{BF} and T_2^{BG} are not guaranteed by Lemma 3.1.2. However, because the actually obtained functional data are not ideal curves but discretely observed curves, the test for infinite-dimensional data is reduced to a test for finite-dimensional data using pre-smoothing methods described by Pomann et al. (2016). Therefore, the test statistics T_2^{BF} and T_2^{BG} are justified in the sense that the hypothesis approximated in a finite dimension is tested. In addition, since $P_1 = P_2$ if and only if $\mu_{11} = \mu_{22} = \mu_{12}$ from Lemma 3.1.2, the hypothesis $H_0 : P_1 = P_2$ v.s. $H_1 : P_1 \neq P_2$ can be rewritten as follows:

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \quad \text{v.s.} \quad H_1 : \boldsymbol{\mu} \neq \mathbf{0},$$

where $\boldsymbol{\mu} = (\mu_{12} - \mu_{11}, \mu_{12} - \mu_{22})^\top$. In addition, the test statistic T^{BG} can be expressed as $T^{BG} = \hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\mu}}$, where $\hat{\boldsymbol{\mu}} = (\hat{\mu}_{12} - \hat{\mu}_{11}, \hat{\mu}_{12} - \hat{\mu}_{22})^\top$.

3.2 Asymptotic properties of the Biswas–Ghosh type test

Note that the Biswas–Ghosh type statistic T^{BG} can be expressed as

$$NT^{BG} = \frac{1}{2}\left\{\sqrt{N}(\hat{\mu}_{11} - \hat{\mu}_{22})\right\}^2 + \frac{1}{2}\left\{\sqrt{N}T^{BF}\right\}^2. \tag{3.2.3}$$

In the case of $\mathcal{X} = \mathbb{R}^d$ $(d \geq 1)$, Biswas and Ghosh (2014) proved the limiting null distribution of T^{BG} by using $\sqrt{N}T^{BF} = o_p(1)$ under H_0. For the functional data, the limiting null distribution of T^{BG} is obtained by asymptotically ignoring the second term in (3.2.3). We give the proof including the results in contiguous alternatives.

Theorem 3.2.1. *Let $X_{11}, \ldots, X_{1n_1} \overset{\text{i.i.d.}}{\sim} P_1$ and $X_{21}, \ldots, X_{2n_2} \overset{\text{i.i.d.}}{\sim} P_2$ be two independent random functions with $\mathbb{E}\mathcal{L}(X_{11}, X_{12})^2, \mathbb{E}\mathcal{L}(X_{21}, X_{22})^2 < \infty$. In addition, assume that $n_1/N \to \gamma \in (0,1)$ as $N \to \infty$.*

(i) Under $H_0 : \boldsymbol{\mu} = \mathbf{0}$, then

$$NT^{BG} \rightsquigarrow \frac{2\sigma_0^2}{\gamma(1-\gamma)}\chi_1^2 \quad \text{as} \quad N \to \infty,$$

where $\sigma_0^2 = \mathrm{Cov}(\mathcal{L}(X_{11}, X_{12}), \mathcal{L}(X_{11}, X_{13})|H_0)$, and χ_1^2 denotes the chi-square random variable with one degree of freedom.

(ii) Suppose that $\sup_{N\in\mathbb{N}} \mathbb{E}[\mathcal{L}(X_{z1}, X_{z'2})^{2+\varepsilon}|H_{1N}] < \infty$ $(z, z' = 1, 2)$ for some $\varepsilon > 0$ and the sequence of alternatives $H_{1N} : \boldsymbol{\mu} = N^{-1/2}\boldsymbol{\delta}$ $(\boldsymbol{\delta} = (\delta_1, \delta_2)^\top \neq \mathbf{0})$. Then

$$NT^{BG} \rightsquigarrow \frac{2\sigma_0^2}{\gamma(1-\gamma)}\chi_1^2\left(\frac{\gamma(1-\gamma)(\delta_1-\delta_2)^2}{4\sigma_0^2}\right) + \frac{(\delta_1+\delta_2)^2}{2} \quad \text{as} \quad N \to \infty,$$

where $\chi_1^2(\delta)$ is the non-central chi-square random variable with one degree of freedom and the noncentrality parameter δ.

Proof. (i) This proof is the same as the proof of Theorem 4.1 by Biswas and Ghosh (2014).

(ii) For every $x \in \mathbb{R}$,

$$\lim_{N\to\infty} \mathbb{P}\left(\mathcal{L}(X_{11}, X_{21}) \leq x|H_{1N}\right) = \mathbb{P}(\mathcal{L}(X_{11}, X_{21}) \leq x|H_0)$$

holds from the continuity of $\mathcal{L}(\cdot,\cdot)$ and the continuous mapping theorem. Furthermore, note that

$$\begin{aligned}
\mathrm{Cov}(\hat{\mu}_{12}, \hat{\mu}_{11}) &= \frac{2}{n_1}\mathrm{Cov}(\mathcal{L}(X_{11}, X_{21}), \mathcal{L}(X_{11}, X_{12})), \\
\mathrm{Cov}(\hat{\mu}_{12}, \hat{\mu}_{22}) &= \frac{2}{n_2}\mathrm{Cov}(\mathcal{L}(X_{11}, X_{21}), \mathcal{L}(X_{21}, X_{22})), \\
\mathrm{Var}(\hat{\mu}_{11}) &= \frac{4(n_1-2)}{n_1(n_1-1)}\mathrm{Cov}(\mathcal{L}(X_{11}, X_{12}), \mathcal{L}(X_{11}, X_{13})) + \frac{2\mathrm{Var}(\mathcal{L}(X_{11}, X_{12}))}{n_1(n_1-1)},
\end{aligned}$$

$$\mathrm{Var}(\hat{\mu}_{22}) = \frac{4(n_2-2)}{n_2(n_2-1)}\mathrm{Cov}(\mathcal{L}(X_{21}, X_{22}), \mathcal{L}(X_{21}, X_{23})) + \frac{2\mathrm{Var}(\mathcal{L}(X_{21}, X_{22}))}{n_2(n_2-1)}$$

and

$$\begin{aligned}\mathrm{Var}(\hat{\mu}_{12}) = &\frac{n_2-1}{n_1 n_2}\mathrm{Cov}(\mathcal{L}(X_{11}, X_{21}), \mathcal{L}(X_{11}, X_{22})) \\ &+ \frac{n_1-1}{n_1 n_2}\mathrm{Cov}(\mathcal{L}(X_{11}, X_{21}), \mathcal{L}(X_{12}, X_{21})) + \frac{1}{n_1 n_2}\mathrm{Var}(\mathcal{L}(X_{11}, X_{21})).\end{aligned}$$

From the uniform integrability of $\mathcal{L}(X_{z1}, X_{z'2})^2$ $(z, z' = 1, 2)$, we have

$$\begin{aligned}\mathrm{Var}[\sqrt{N}(\hat{\mu}_{11} - \hat{\mu}_{12})|H_{1N}] &= N\{\mathrm{Var}(\hat{\mu}_{11}|H_{1N}) + \mathrm{Var}(\hat{\mu}_{12}|H_{1N}) - 2\mathrm{Cov}(\hat{\mu}_{11}, \hat{\mu}_{12}|H_{1N})\} \\ &\to \frac{4\sigma_0^2}{\gamma} + \frac{\sigma_0^2}{\gamma} + \frac{\sigma_0^2}{1-\gamma} - \frac{4\sigma_0^2}{\gamma} = \frac{\sigma_0^2}{\gamma(1-\gamma)}, \\ \mathrm{Var}[\sqrt{N}(\hat{\mu}_{22} - \hat{\mu}_{12})|H_{1N}] &= N\{\mathrm{Var}(\hat{\mu}_{22}|H_{1N}) + \mathrm{Var}(\hat{\mu}_{12}|H_{1N}) - 2\mathrm{Cov}(\hat{\mu}_{22}, \hat{\mu}_{12}|H_{1N})\} \\ &\to \frac{4\sigma_0^2}{1-\gamma} + \frac{\sigma_0^2}{\gamma} + \frac{\sigma_0^2}{1-\gamma} - \frac{4\sigma_0^2}{1-\gamma} = \frac{\sigma_0^2}{\gamma(1-\gamma)}\end{aligned}$$

as $N \to \infty$. Thus, we obtain

$$\begin{aligned}\sqrt{N}\,\{(\hat{\mu}_{11} - \hat{\mu}_{12}) - (\mu_{11} - \mu_{12})\} &\rightsquigarrow Y_1 \sim N(0, \sigma_0^2/(\gamma(1-\gamma))), \\ \sqrt{N}\,\{(\hat{\mu}_{22} - \hat{\mu}_{12}) - (\mu_{22} - \mu_{12})\} &\rightsquigarrow Y_2 \sim N(0, \sigma_0^2/(\gamma(1-\gamma)))\end{aligned}$$

from the asymptotic normality of the U-statistics (Lee, 1990). Because the covariance between Y_1 and Y_2 is

$$\begin{aligned}\mathrm{Cov}(Y_1, Y_2|H_0) &= \lim_{N\to\infty} \mathrm{Cov}(\sqrt{N}(\hat{\mu}_{11} - \hat{\mu}_{12}), \sqrt{N}(\hat{\mu}_{22} - \hat{\mu}_{12})|H_{1N}) \\ &= \lim_{N\to\infty} N\{\mathrm{Var}(\hat{\mu}_{12}) - \mathrm{Cov}(\hat{\mu}_{12}, \hat{\mu}_{11}) - \mathrm{Cov}(\hat{\mu}_{12}, \hat{\mu}_{22})\} \\ &= \frac{\sigma_0^2}{\gamma} + \frac{\sigma_0^2}{1-\gamma} - \frac{2\sigma_0^2}{\gamma} - \frac{2\sigma_0^2}{1-\gamma} = -\frac{\sigma_0^2}{\gamma(1-\gamma)},\end{aligned}$$

the correlation coefficient between Y_1 and Y_2 is $\mathrm{Corr}(Y_1, Y_2) = -1$. Thus, the relation $Y_1 + Y_2 = 0$ is given. We then obtain

$$\begin{aligned}\lim_{N\to\infty} NT^{BG} &= \lim_{N\to\infty} N(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} + \boldsymbol{\mu})^\top(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} + \boldsymbol{\mu}) \\ &\overset{d}{=} (Y_1 + \delta_1)^2 + (Y_2 + \delta_2)^2 \\ &\overset{d}{=} 2\left(Y_1 + \frac{\delta_1 - \delta_2}{2}\right)^2 + \frac{(\delta_1 + \delta_2)^2}{2} \\ &\overset{d}{=} \frac{2\sigma_0^2}{\gamma(1-\gamma)}\chi_1^2\left(\frac{\gamma(1-\gamma)(\delta_1 - \delta_2)^2}{4\sigma_0^2}\right) + \frac{(\delta_1 + \delta_2)^2}{2},\end{aligned}$$

where "$\overset{d}{=}$" denotes that the random elements on either side have the same distribution.

□

Using certain consistent estimators $\hat{\gamma}$ and $\hat{\sigma}_0^2$ for γ and σ_0^2, the test statistic $N\hat{\gamma}(1-\hat{\gamma})T^{BG}/(2\hat{\sigma}_0^2)$ turns out to be an asymptotically distribution-free test, that is,

$$\mathbb{P}\left(\frac{\hat{\gamma}(1-\hat{\gamma})}{2\hat{\sigma}_0^2}NT^{BG} > \chi^2_{1,\alpha} \;\middle|\; H_0\right) \to \alpha \quad \text{as} \quad N \to \infty,$$

where $\chi^2_{1,\alpha}$ denotes the upper α-quantile of the chi-square distribution with one degree of freedom. As simple consistent estimators of γ and σ_0^2, Biswas and Ghosh (2014) gave $\hat{\gamma} = n_1/N$ and $S_0^2 = (n_1 S_1^2 + n_2 S_2^2)/N$, where

$$S_z^2 = \frac{1}{3\binom{n_z}{3}} \sum\sum_{1\le j<k\le n_z} \sum_{i\ne j,k} \mathcal{L}(X_{zi}, X_{zj})\mathcal{L}(X_{zi}, X_{zk}) - \hat{\mu}_{zz}^2 \quad (z = 1, 2).$$

By using these estimators, the approximated p-value of the asymptotic test T^{BG} is given by

$$\hat{p}(S_0^2) = 1 - F_{\chi_1^2}\left(\frac{\hat{\gamma}(1-\hat{\gamma})}{2S_0^2}NT^{BG}\right),$$

where $F_{\chi_1^2}$ denotes the distribution function of the chi-square distribution with one degree of freedom. Although the second term in (3.2.3) is asymptotically ignored, because $(T^{BF})^2 \ge 0$ for a finite sample size, the approximated p-value $\hat{p}(S_0^2)$ tends to be anti-conservative. To improve this anti-conservative tendency, we propose a p-value approximation using the jackknife variance estimator with a positive bias instead of S_0^2 (Efron and Stein, 1981). Here, the jackknife variance estimators of $\hat{\mu}_{11}$ and $\hat{\mu}_{22}$ are given by

$$\widehat{\mathrm{Var}}_J(\hat{\mu}_{11}) = \frac{n_1 - 1}{n_1}\sum_{i=1}^{n_1}(\hat{\mu}_{11}^{(i)} - \hat{\mu}_{11})^2 \quad \text{and} \quad \widehat{\mathrm{Var}}_J(\hat{\mu}_{22}) = \frac{n_2 - 1}{n_2}\sum_{j=1}^{n_2}(\hat{\mu}_{22}^{(j)} - \hat{\mu}_{22})^2,$$

where $\hat{\mu}_{11}^{(i)}$ is the estimator $\hat{\mu}_{11}$ when the i-th sample X_{1i} is deleted, and $\hat{\mu}_{22}^{(j)}$ is the estimator $\hat{\mu}_{22}$ when the j-th sample X_{2j} is removed. In addition, $\sqrt{n_1}\{\widehat{\mathrm{Var}}_J(\hat{\mu}_{11}) - \mathrm{Var}(\hat{\mu}_{11})\}$ and $\sqrt{n_2}\{\widehat{\mathrm{Var}}_J(\hat{\mu}_{22}) - \mathrm{Var}(\hat{\mu}_{22})\}$ converge to zero with probability one (Lee, 1990, Section 5.1.1). We use $N(\mathrm{Var}_J(\hat{\mu}_{11}) + \mathrm{Var}_J(\hat{\mu}_{22}))$ as an estimator of $\mathrm{Var}(\sqrt{N}\{\hat{\mu}_{11} - \hat{\mu}_{22}\})$. Because

$$\mathrm{Var}(\sqrt{N}(\hat{\mu}_{11} - \hat{\mu}_{22})|H_0) = N\mathrm{Var}(\hat{\mu}_{11}|H_0) + N\mathrm{Var}(\hat{\mu}_{22}|H_0) \to \frac{4\sigma_0^2}{\gamma} + \frac{4\sigma_0^2}{1-\gamma} = \frac{4\sigma_0^2}{\gamma(1-\gamma)},$$

we consider the jackknife-based consistent estimator of σ_0^2 by

$$\begin{aligned}\hat{\sigma}_J^2 &= \frac{\hat{\gamma}(1-\hat{\gamma})}{4} \cdot N(\mathrm{Var}_J(\hat{\mu}_{11}) + \mathrm{Var}_J(\hat{\mu}_{22})) \\ &= \frac{N\hat{\gamma}(1-\hat{\gamma})}{4}\left\{\frac{n_1 - 1}{n_1}\sum_{i=1}^{n_1}(\hat{\mu}_{11}^{(i)} - \hat{\mu}_{11})^2 + \frac{n_2 - 1}{n_2}\sum_{j=1}^{n_2}(\hat{\mu}_{22}^{(j)} - \hat{\mu}_{22})^2\right\}.\end{aligned}$$

Thus, the p-value approximation based on the jackknife variance estimator is given by

$$\hat{p}(\hat{\sigma}_J^2) = 1 - F_{\chi_1^2}\left(\frac{\hat{\gamma}(1-\hat{\gamma})}{2\hat{\sigma}_J^2}NT^{BG}\right).$$

Furthermore, similar to the nonparametric Behrens–Fisher problem described in Brunner et al. (2018), we consider approximating the null distribution of $(\hat{\mu}_{11} - \hat{\mu}_{22})/\sqrt{\mathrm{Var}_J(\hat{\mu}_{11}) + \mathrm{Var}_J(\hat{\mu}_{22})}$ using a t-distribution based on the Welch–Satterthwaite approximation instead of the standard normal distribution for small sample sizes. Let $\tilde{\mu}_{11}$ and $\tilde{\mu}_{22}$ be the Hájek projections of $\hat{\mu}_{11} - \mu_{11}$ and $\hat{\mu}_{22} - \mu_{22}$, respectively. That is, note that

$$\tilde{\mu}_{11} = \frac{2}{n_1}\sum_{i=1}^{n_1} h_1(X_{1i}) \quad \text{and} \quad \tilde{\mu}_{22} = \frac{2}{n_2}\sum_{i=1}^{n_2} h_2(X_{2i}),$$

where $h_1(X_{1i}) = \mathbb{E}[\mathcal{L}(X_{1i}, X_{1j}) - \mu_{11}|X_{1i}]$ and $h_2(X_{2i}) = \mathbb{E}[\mathcal{L}(X_{2i}, X_{2j}) - \mu_{22}|X_{2i}]$. In particular, the variances of $\tilde{\mu}_{11}$ and $\tilde{\mu}_{22}$ are expressed as

$$\mathrm{Var}(\tilde{\mu}_{11}) = \frac{4}{n_1}\mathrm{Var}(h_1(X_{11})) \quad \text{and} \quad \mathrm{Var}(\tilde{\mu}_{22}) = \frac{4}{n_2}\mathrm{Var}(h_2(X_{21})).$$

Herein, we consider the quantities

$$\tilde{\sigma}_1^2 = \frac{4}{n_1 - 1}\sum_{i=1}^{n_1}(h_1(X_{1i}) - \bar{h}_1)^2 \quad \text{and} \quad \tilde{\sigma}_2^2 = \frac{4}{n_2 - 1}\sum_{i=1}^{n_2}(h_2(X_{2i}) - \bar{h}_2)^2,$$

where $\bar{h}_1 = \sum_{i=1}^{n_1} h_1(X_{1i})/n_1$ and $\bar{h}_2 = \sum_{i=1}^{n_2} h_2(X_{2i})/n_2$. Analogous to the derivation of the approximate t-test for unequal variances, the distribution of $\tilde{\sigma}_1^2/n_1 + \tilde{\sigma}_2^2/n_2$ is approximated by that of a random variable $g \cdot Z_f$, where $g > 0$ and $Z_f \sim \chi_f^2$. These constants f and g are selected such that the first two moments match in the following manner:

$$\mathbb{E}[\tilde{\sigma}_1^2/n_1 + \tilde{\sigma}_2^2/n_2] = \mathbb{E}[gZ_f] = gf \quad \text{and} \quad \mathrm{Var}[\tilde{\sigma}_1^2/n_1 + \tilde{\sigma}_2^2/n_2] = \mathrm{Var}[gZ_f] = 2g^2 f.$$

Here, we obtain

$$\mathbb{E}[\tilde{\sigma}_1^2/n_1 + \tilde{\sigma}_2^2/n_2] = \frac{4\mathrm{Var}(h_1(X_{11}))}{n_1} + \frac{4\mathrm{Var}(h_2(X_{21}))}{n_2} = gf. \tag{3.2.4}$$

Regarding the variance of $\tilde{\sigma}_1^2/n_1 + \tilde{\sigma}_2^2/n_2$, we assume that the variables $h_1(X_{1i})$ $(i = 1, \ldots, n_1)$ and $h_2(X_{2i})$ $(i = 1, \ldots, n_2)$ follow approximately normal distributions in the same way as in Brunner et al. (2018, Section 3.5.2), and we approximate the variance as

$$\begin{aligned}\mathrm{Var}[\tilde{\sigma}_1^2/n_1 + \tilde{\sigma}_2^2/n_2] &\approx \frac{2}{n_1^2(n_1-1)}\mathrm{Var}(2h_1(X_{11}))^2 + \frac{2}{n_2^2(n_2-1)}\mathrm{Var}(2h_2(X_{21}))^2 \\ &= \frac{2}{n_1 - 1}\left(\frac{4\mathrm{Var}(h_1(X_{11}))}{n_1}\right)^2 + \frac{2}{n_2 - 1}\left(\frac{4\mathrm{Var}(h_2(X_{21}))}{n_2}\right)^2 = 2g^2 f. \qquad (3.2.5)\end{aligned}$$

By solving the above two equations (3.2.4) and (3.2.5), we obtain

$$f = \frac{\left\{\frac{4\mathrm{Var}(h_1(X_{11}))}{n_1} + \frac{4\mathrm{Var}(h_2(X_{21}))}{n_2}\right\}^2}{\frac{1}{n_1-1}\left(\frac{4\mathrm{Var}(h_1(X_{11}))}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{4\mathrm{Var}(h_2(X_{21}))}{n_2}\right)^2} = \frac{\{\mathrm{Var}(\tilde{\mu}_{11}) + \mathrm{Var}(\tilde{\mu}_{22})\}^2}{\frac{1}{n_1-1}\mathrm{Var}(\tilde{\mu}_{11})^2 + \frac{1}{n_2-1}\mathrm{Var}(\tilde{\mu}_{22})^2}.$$

Then, the distribution of $(\tilde{\sigma}_1^2/n_1 + \tilde{\sigma}_2^2/n_2)/(gf)$ is approximated by that of the random variable Z_f/f. Because $\sqrt{N}\{\tilde{\mu}_{11} - (\hat{\mu}_{11} - \mu_{11})\} \xrightarrow{p} 0$ and $\sqrt{N}\{\tilde{\mu}_{22} - (\hat{\mu}_{22} - \mu_{22})\} \xrightarrow{p} 0$ as $N \to \infty$, we estimate f with

$$\hat{f} = \frac{\left\{\frac{n_1-1}{n_1}\sum_{i=1}^{n_1}(\hat{\mu}_{11}^{(i)} - \hat{\mu}_{11})^2 + \frac{n_2-1}{n_2}\sum_{j=1}^{n_2}(\hat{\mu}_{22}^{(j)} - \hat{\mu}_{22})^2\right\}^2}{\frac{\{\frac{n_1-1}{n_1}\sum_{i=1}^{n_1}(\hat{\mu}_{11}^{(i)} - \hat{\mu}_{11})^2\}^2}{n_1-1} + \frac{\{\frac{n_2-1}{n_2}\sum_{j=1}^{n_2}(\hat{\mu}_{22}^{(j)} - \hat{\mu}_{22})^2\}^2}{n_2-1}}$$

by replacing $\mathrm{Var}(\tilde{\mu}_{11})$ and $\mathrm{Var}(\tilde{\mu}_{22})$ with $\mathrm{Var}_J(\hat{\mu}_{11})$ and $\mathrm{Var}_J(\hat{\mu}_{22})$, respectively. Therefore, the null distribution of $(\hat{\mu}_{11} - \hat{\mu}_{22})^2/\{\mathrm{Var}_J(\hat{\mu}_{11}) + \mathrm{Var}_J(\hat{\mu}_{22})\}$ is approximated by the F-distribution with 1 and $\hat{f}$ degrees of freedom. The approximated p-value is then defined by

$$\hat{p}(\hat{\sigma}_J^2, \hat{f}) = 1 - F_{F(1,\hat{f})}\left(\frac{\hat{\gamma}(1-\hat{\gamma})}{2\hat{\sigma}_J^2} N T^{BG}\right),$$

where $F_{F(1,\hat{f})}$ denotes the distribution of the F-distribution with 1 and $\hat{f}$ degrees of freedom. It can easily be seen that $\hat{f} \to \infty$ $(N \to \infty)$, and that the $F(1,\hat{f})$-distribution converges to the chi-square distribution with one degree of freedom. The performance of these p-value approximations is numerically verified in Section 3.3.1.

Furthermore, under the sequence of alternatives $H_{1N} : \boldsymbol{\mu} = N^{-1/2}\boldsymbol{\delta}$, the asymptotic power of the test statistic T^{BG} is given by

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{P}\left(\frac{\hat{\gamma}(1-\hat{\gamma})}{2\hat{\sigma}_0^2} N T^{BG} > \chi^2_{1,\alpha} \;\middle|\; H_{1N}\right) \\
&= \mathbb{P}\left(\chi_1^2\left(\frac{\gamma(1-\gamma)(\delta_1 - \delta_2)^2}{4\sigma_0^2}\right) > \chi^2_{1,\alpha} - \frac{\gamma(1-\gamma)(\delta_1 + \delta_2)^2}{4\sigma_0^2}\right)
\end{aligned}$$

from Theorem 3.2.1 (ii). Here, δ_1 and δ_2 are expressed as

$$\begin{aligned}
\delta_1 &= \lim_{N\to\infty} N^{1/2}\{\mathbb{E}\mathcal{L}(X_{11}, X_{21}) - \mathbb{E}\mathcal{L}(X_{11}, X_{12})\}, \\
\delta_2 &= \lim_{N\to\infty} N^{1/2}\{\mathbb{E}\mathcal{L}(X_{11}, X_{21}) - \mathbb{E}\mathcal{L}(X_{21}, X_{22})\}.
\end{aligned}$$

Here, we assume that $\mathcal{X} = L_p[a,b]$ $(p \geq 1)$ and $\mathcal{L}(x,y) = \|x - y\|_p$. In addition, we assume the exchangeability of the limit and expectation. For a simple location-shift model, we consider the sequence of alternatives

$$H_{1N}^L : X_2 \stackrel{d}{=} X_1 + \frac{\Delta}{\sqrt{N}},$$

where Δ is a non-zero constant function. Denote $X_{11}, X_{12} \stackrel{\text{i.i.d.}}{\sim} P_1$ and $X_{21}, X_{22} \stackrel{\text{i.i.d.}}{\sim} P_2$. We then obtain

$$\begin{aligned}
\delta_1 &= \lim_{N\to\infty} N^{1/2}\{\mathbb{E}\|X_{11} - X_{21}\|_p - \mathbb{E}\|X_{11} - X_{12}\|_p\} \\
&= \lim_{N\to\infty} \int_{L_p[a,b]}\int_{L_p[a,b]} N^{1/2}\left\{\left\|x_1 - x_2 - \frac{\Delta}{\sqrt{N}}\right\|_p - \|x_1 - x_2\|_p\right\} dP_1(x_1)dP_1(x_2)
\end{aligned}$$

$$
\begin{aligned}
&= \int_{L_p[a,b]} \int_{L_p[a,b]} \frac{\int_a^b |x_1(s) - x_2(s)|^{p-2}\{x_1(s) - x_2(s)\}(-\Delta) ds}{\|x_1 - x_2\|_p^{p-1}} dP_1(x_1) dP_1(x_2) \\
&= -\Delta \cdot \mathbb{E}\left[\frac{\int_a^b |X_{11}(s) - X_{12}(s)|^{p-2}\{X_{11}(s) - X_{12}(s)\} ds}{\|X_{11} - X_{12}\|_p^{p-1}} \right] = 0
\end{aligned}
$$

from the Gâteaux derivative (3.1.1). Similarly, the equation $\delta_2 = 0$ also holds because $X_{11} - X_{12} \overset{d}{=} X_{21} - X_{22}$ under H_{1N}^L. This means that the asymptotic power of the test statistic T_p^{BG} for H_{1N}^L coincides with the size of the test. This result suggests that T_p^{BG} has lower power than other distribution homogeneity tests such as T_p^{BF} for the location-shift model. Meanwhile, for a simple scale-shift model, we consider the sequence of alternatives

$$
H_{1N}^S : X_2 \overset{d}{=} \left(1 + \frac{\Delta}{\sqrt{N}}\right) X_1,
$$

where Δ is non-zero constant function. We then have

$$
\begin{aligned}
\delta_1 &= \lim_{N\to\infty} N^{1/2}\{\mathbb{E}\|X_{11} - X_{21}\|_p - \mathbb{E}\|X_{11} - X_{12}\|_p\} \\
&= \lim_{N\to\infty} \int_{L_p[a,b]} \int_{L_p[a,b]} N^{1/2} \left\{ \left\| x_1 - \left(1 + \frac{\Delta}{\sqrt{N}}\right) x_2 \right\|_p - \|x_1 - x_2\|_p \right\} dP_1(x_1) dP_1(x_2) \\
&= \int_{L_p[a,b]} \int_{L_p[a,b]} \frac{\int_a^b |x_1(s) - x_2(s)|^{p-2}\{x_1(s) - x_2(s)\} \cdot \{-\Delta \cdot x_2(s)\} ds}{\|x_1 - x_2\|_p^{p-1}} dP_1(x_1) dP_1(x_2) \\
&= -\Delta \cdot \mathbb{E}\left[\frac{\int_a^b |X_{11}(s) - X_{12}(s)|^{p-2}\{X_{11}(s) X_{12}(s) - X_{12}^2(s)\} ds}{\|X_{11} - X_{12}\|_p^{p-1}} \right]
\end{aligned}
$$

and

$$
\begin{aligned}
\delta_2 &= \lim_{N\to\infty} N^{1/2}\{\mathbb{E}\|X_{11} - X_{21}\|_p - \mathbb{E}\|X_{21} - X_{22}\|_p\} \\
&= \lim_{N\to\infty} N^{1/2}\{\mathbb{E}\|X_{11} - X_{21}\|_p - \mathbb{E}\|X_{11} - X_{12}\|_p + \mathbb{E}\|X_{11} - X_{12}\|_p - \mathbb{E}\|X_{21} - X_{22}\|_p\} \\
&= \delta_1 + \lim_{N\to\infty} N^{1/2}\{\mathbb{E}\|X_{11} - X_{12}\|_p - \mathbb{E}\|X_{21} - X_{22}\|_p\} \\
&= \delta_1 - \lim_{N\to\infty} \iint_{L_p[a,b]^2} N^{1/2} \left\{ \left\| x_1 - x_2 + \frac{\Delta}{\sqrt{N}}(x_1 - x_2) \right\|_p - \|x_1 - x_2\|_p \right\} dP_1(x_1) dP_1(x_2) \\
&= \delta_1 - \Delta \cdot \iint_{L_p[a,b]^2} \frac{\int_a^b |x_1(s) - x_2(s)|^p ds}{\|x_1 - x_2\|_p^{p-1}} dP_1(x_1) dP_1(x_2) \\
&= \delta_1 - \Delta \cdot \iint_{L_p[a,b]^2} \|x_1 - x_2\|_p dP_1(x_1) dP_1(x_2) \\
&= \delta_1 - \Delta \cdot \mathbb{E}\|X_{11} - X_{12}\|.
\end{aligned}
$$

Therefore, we obtain the equations $\delta_1 - \delta_2 = \Delta \cdot \mathbb{E}\|X_{11} - X_{12}\|$ and

$$
\delta_1 + \delta_2 = 2\delta_1 - \Delta \cdot \mathbb{E}\|X_{11} - X_{12}\|
$$

$$
\begin{aligned}
&= -\Delta \cdot \mathbb{E}\left[\frac{\int_a^b |X_{11}(s) - X_{12}(s)|^{p-2}\{2X_{11}(s)X_{12}(s) - 2X_{12}^2(s)\}ds}{\|X_{11} - X_{12}\|_p^{p-1}}\right] \\
&\quad - \Delta \cdot \mathbb{E}\left[\frac{\int_a^b |X_{11}(s) - X_{12}(s)|^{p-2}\{X_{11}(s) - X_{12}(s)\}^2 ds}{\|X_{11} - X_{12}\|_p^{p-1}}\right] \\
&= -\Delta \cdot \mathbb{E}\left[\frac{\int_a^b |X_{11}(s) - X_{12}(s)|^{p-2}\{X_{11}^2(s) - X_{12}^2(s)\}ds}{\|X_{11} - X_{12}\|_p^{p-1}}\right] = 0.
\end{aligned}
$$

The asymptotic power is then given by

$$
\mathbb{P}\left(\chi_1^2\left(\frac{\Delta^2\gamma(1-\gamma)\mu_{11}^2}{4\sigma_0^2}\right) > \chi_{1,\alpha}^2\right),
$$

and this value increases as γ reaches close to $1/2$, and decreases as the coefficient of variation of $\mathbb{E}[\|X_{11} - X_{12}\|_p | X_{11}]$ (i.e., σ_0/μ_{11}) increases.

3.3 Simulation studies

Extending the test procedure described in Section 3.1 to practical applications is not straightforward because the true smooth trajectories cannot be directly observed. For example, in the DTI (diffusion tensor imaging) study as discussed in Pomann et al. (2016), the data on the subject are noisy and discretely observations. Hence, when using certain methods, it is necessary to carry out a pre-smoothing of discretely observed functional data. Suppose that we observe data arising from two groups, $\{(t_{1ij}, Y_{1ij}) : j = 1, \ldots, m_{1i}\}_{i=1}^{n_1}$ and $\{(t_{2ij}, Y_{2ij}) : j = 1, \ldots, m_{2i}\}_{i=1}^{n_2}$, where $t_{1ij}, t_{2ij} \in [0, 1]$. The notation of the time points, t_{1ij} and t_{2ij}, allows for different observation points within the two groups. It is assumed that Y_{1ij}'s and Y_{2ij}'s are independent realizations of two underlying (stochastic) processes observed with noise on a finite grid of points. Specifically, assume that

$$
Y_{1ij} = X_{1i}(t_{1ij}) + \varepsilon_{1ij} \quad \text{and} \quad Y_{2ij} = X_{2i}(t_{2ij}) + \varepsilon_{2ij}, \tag{3.3.6}
$$

where $X_{11}, \ldots, X_{1n_1} \overset{\text{i.i.d.}}{\sim} P_1$ and $X_{21}, \ldots, X_{2n_2} \overset{\text{i.i.d.}}{\sim} P_2$ are independent and square-integrable random functions over $[0, 1]$. The measurement errors $\{\varepsilon_{1ij}\}$ and $\{\varepsilon_{2ij}\}$ are independent and identically distributed with mean 0, and variance σ_1^2 and σ_2^2, respectively. We assume that $\sigma_1^2 = \sigma_2^2 =: \sigma^2$. By applying the common functional principal component analysis (FPCA) techniques (Benko et al., 2009) and the R package "refund", the functional data are reconstructed as

$$
\hat{X}_{zi}(t) = \hat{\mu}(t) + \sum_{k=1}^{K} \hat{\xi}_{zik}\hat{\phi}_k(t) \quad (z = 1, 2;\ i = 1, \ldots, m_z),
$$

where $\hat{\xi}_{zik}$ $(z = 1, 2; i = 1, \ldots, m_z; k = 1, \ldots, K)$ are the estimated FPC scores, $\hat{\mu}(\cdot)$ is the estimated mean function and $\hat{\phi}_k(\cdot)$ $(k = 1, \ldots, K)$ are the eigenfunctions subject to $\int_0^1 \hat{\phi}_k(t)^2 dt = 1$ and $\int_0^1 \hat{\phi}_k(t)\hat{\phi}_\ell(t)dt = 0$ for $k \neq \ell$. Here, the truncation K is a suitably large integer; for example, it is selected by the percentage of the explained variance (such as $\tau = 95\%$). See Pomann et al. (2016) for further details. If the overall sets of pooled observed points are $\mathbb{T} = \{(2i-1)/(2M) : i = 1, 2, \ldots, M\}$, the L_p-distance can be approximated as

$$\|\hat{X}_{zi} - \hat{X}_{z'j}\|_p \approx \left(\frac{1}{2M} \sum_{t=1}^{M} \left| \sum_{k=1}^{K} (\hat{\xi}_{zik} - \hat{\xi}_{z'jk}) \hat{\phi}_k \left(\frac{2t-1}{2M} \right) \right|^p \right)^{1/p}.$$

In the case of $p = 2$, the L_2-distance is approximated by $\|\hat{X}_{zi} - \hat{X}_{z'j}\|_2 \approx \sqrt{\sum_{k=1}^{K} (\hat{\xi}_{zik} - \hat{\xi}_{z'jk})^2}$, and which is the K-dimensional Euclidean distance for FPC scores. Then, the test statistics T_2^{BF} and T_2^{BG} correspond to those of Baringhaus and Franz (2004) and Biswas and Ghosh (2014) for finite-dimensional data, respectively. Furthermore, the null hypothesis $H_0 : P_1 = P_2$ reduces to $H_0^K : \{\xi_{1k}\}_{k=1}^K \overset{d}{=} \{\xi_{2k}\}_{k=1}^K$, where the superscript K in H_0^K emphasizes the dependence of the reduced null hypothesis on the finite truncation K. The L_2-norm-based kernel $\mathcal{L}_2(x, y) = \|x - y\|_2$ is not a strongly negative definite kernel, and Theorem 3.1.2 does not hold. However, because the Euclidean norm-based kernel is a strongly negative definite kernel, the test statistics T_2^{BF} and T_2^{BG} based on the L_2-norm are justified in the sense of testing the approximated null hypothesis H_0^K.

3.3.1 Type I error rate for the asymptotic Biswas–Ghosh type tests

First, we set the threshold parameter value to $\tau = 0.95$ and compare the accuracy of the Type I error rates for various p-value approximations of the asymptotic Biswas–Ghosh type test T_q^{BG} $(0 < q \leq 2)$, which is the test based on the kernel $\mathcal{L}(x, y) = \|x - y\|_q$. We construct data sets $\{(t_{1ij}, Y_{1ij}) : j = 1, \ldots, m_{1i}\}_{i=1}^{n_1}$ and $\{(t_{2ij}, Y_{2ij}) : j = 1, \ldots, m_{2i}\}_{i=1}^{n_2}$ using model (3.3.6) for $t_{1ij} = t_{2ij} = t_j$ observed points within $[0, 1]$. In this section, we set $\mathbb{T} = \{(2i-1)/(2M) : i = 1, \ldots, M\}$ $(M = 100)$, and consider the dense sampling design $m_{1i} = m_{2i} = 100$ and the sparse sampling design $m_{1i} = m_{2i} = 20$, respectively. Throughout this study, it is assumed that $\varepsilon_{1ij}, \varepsilon_{2ij}$ are independent Gaussian variables with mean zero and variance 0.25. Here,

$$X_{1i}(t) = \mu_1(t) + \sum_{k=1}^{\infty} \xi_{1ik} \phi_{1k}(t) \quad \text{and} \quad X_{2i}(t) = \mu_2(t) + \sum_{k=1}^{\infty} \xi_{2ik} \phi_{2k}(t),$$

where $\phi_{1k}(t)$ and $\phi_{2k}(t)$ $(k \geq 1)$ are Fourier basis functions, for example, $\phi_{11}(t) = \phi_{21}(t) = \sqrt{2}\sin(2\pi t)$, $\phi_{12}(t) = \phi_{22}(t) = \sqrt{2}\cos(2\pi t)$; and ξ_{1ik} and ξ_{2ik} are uncorrelated random variables, respectively, with $\mathrm{Var}(\xi_{1ik}) = \lambda_{1k}$ and $\mathrm{Var}(\xi_{2ik}) = \lambda_{2k}$. Setting $\phi_{zk}(t) = \phi_k(t)$ allows us to study different types of departure from the null hypothesis in which the underlying processes of the two data sets are the same.

For simplicity, we set $\mu_1(t) = \mu_2(t) = 0$ for all values of t, and $\lambda_1 = 10$, $\lambda_2 = 5$, $\lambda_3 = 2$, and $\lambda_k = 0$ for $k \geq 4$. As some configurations of ξ_{1ik} and ξ_{2ik}, we apply the following distributions:

(A) set $\xi_{1ik}, \xi_{2ik} \sim N(0, \lambda_k)$, where $N(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2

(B) set $\xi_{1ik}, \xi_{2ik} \sim T_4(0, \lambda_k)$, where $T_\nu(\mu, \sigma^2)$ denotes the Student T-distribution with ν degrees of freedom with mean μ and the variance σ^2.

In addition, we add the following simulation setting (C):

(C) generates the data from the stochastic process on $[0, 1]$, introduced through

$$X_{zi}(t) = \sum_{n=1}^{K} \frac{\sqrt{2}\sin\left(\left(n - \frac{1}{2}\right)\pi t\right)}{\left(n - \frac{1}{2}\right)\pi} Z_{zn}, \quad (z = 1, 2),$$

where $\{Z_{1n}\}_{n=1}^{K}$ and $\{Z_{2n}\}_{n=1}^{K}$ are independent Gaussian variables with mean zero and variance one, and $K = 100$.

We provide the empirical size of the L_q-norm-based Biswas–Ghosh type test T_q^{BG} and the accuracy of several p-value approximations. As the p-value approximations, we consider the simple approximated p-value of T_q^{BG}, i.e.,

$$\hat{p}_q(S_0^2) = 1 - F_{\chi_1^2}\left(\frac{\hat{\gamma}(1-\hat{\gamma})}{2S_0^2} N T_q^{BG}\right),$$

and the proposed p-value approximations

$$\hat{p}_q(\hat{\sigma}_J^2) = 1 - F_{\chi_1^2}\left(\frac{\hat{\gamma}(1-\hat{\gamma})}{2\hat{\sigma}_J^2} N T_q^{BG}\right),$$
$$\hat{p}_q(\hat{\sigma}_J^2, \hat{f}) = 1 - F_{F(1,\hat{f})}\left(\frac{\hat{\gamma}(1-\hat{\gamma})}{2\hat{\sigma}_J^2} N T_q^{BG}\right),$$

where S_0^2, $\hat{\sigma}_J^2$, and $\hat{f}$ are defined in Section 3.1. All results described in this section are based on the nominal level $\alpha = 0.05$ and $10,000$ Monte Carlo replications.

Tables 3.1 and 3.2 show that the empirical size of the asymptotic tests when the total sample size ranges from 20 to 400. The simulation results for the dense sampling design (Table 3.1) showed a similar pattern as that of the sparse sampling design (Table 3.2). Under all settings, the accuracy of the proposed p-value approximations based on the jackknife variance estimators is verified to be close to the nominal level $\alpha = 0.05$. Because the jackknife variance estimator has a positive bias, the asymptotic tests incorporating it tend to obtain conservative results. This property is confirmed by the fact that $\hat{p}_q(\hat{\sigma}_J^2)$ and $\hat{p}_q(\hat{\sigma}_J^2, \hat{f})$ are more conservative than $\hat{p}_q(S_0^2)$ in all cases listed in Tables 3.1 and 3.2. In particular, the Welch–Satterthwaite approximation $\hat{p}(\hat{\sigma}_J^2, \hat{f})$ works effectively when the sample size is moderate and large. However, the Type I error rates of the asymptotic tests may not be controlled at the nominal level when the sample sizes n_1 and n_2 are small. In such cases, the approximate permutation test is recommended instead of the asymptotic test, although it is more computationally expensive.

3.3.2 Power comparison

To describe the characteristics of the interpoint distance-based tests, we conducted a simulation study with respect to the empirical power of the test statistics. We consider some functional samples defined in $[0, 1]$, which are considered the realizations of a stochastic process that has continuous trajectories within the interval $[0, 1]$. Furthermore, we assume the sparse sampling

Table 3.1. Type I error rate of the asymptotic test T_q^{BG} in the dense sampling design

Setting	(n_1, n_2)	$\hat{p}_1(S_0^2)$	$\hat{p}_2(S_0^2)$	$\hat{p}_1(\hat{\sigma}_J^2)$	$\hat{p}_2(\hat{\sigma}_J^2)$	$\hat{p}_1(\hat{\sigma}_J^2, \hat{f})$	$\hat{p}_2(\hat{\sigma}_J^2, \hat{f})$
A	$(10, 10)$	0.262	0.261	0.070	0.071	0.051	0.050
	$(20, 20)$	0.120	0.119	0.059	0.059	0.050	0.050
	$(40, 40)$	0.076	0.075	0.054	0.053	0.050	0.049
	$(60, 60)$	0.065	0.064	0.052	0.051	0.050	0.048
	$(100, 100)$	0.063	0.064	0.055	0.055	0.054	0.054
	$(200, 200)$	0.056	0.057	0.053	0.053	0.052	0.052
	$(5, 15)$	0.291	0.291	0.096	0.095	0.064	0.063
	$(10, 30)$	0.135	0.133	0.077	0.077	0.060	0.060
	$(20, 60)$	0.081	0.080	0.058	0.058	0.051	0.050
	$(30, 90)$	0.065	0.065	0.056	0.056	0.050	0.050
	$(50, 150)$	0.062	0.063	0.055	0.055	0.051	0.051
	$(100, 300)$	0.054	0.053	0.051	0.050	0.049	0.048
B	$(10, 10)$	0.207	0.206	0.063	0.063	0.041	0.040
	$(20, 20)$	0.098	0.098	0.055	0.054	0.046	0.046
	$(40, 40)$	0.068	0.068	0.050	0.050	0.046	0.045
	$(60, 60)$	0.063	0.063	0.052	0.053	0.050	0.050
	$(100, 100)$	0.056	0.057	0.050	0.051	0.048	0.049
	$(200, 200)$	0.052	0.053	0.050	0.050	0.049	0.049
	$(5, 15)$	0.237	0.237	0.095	0.094	0.064	0.064
	$(10, 30)$	0.114	0.112	0.078	0.077	0.062	0.062
	$(20, 60)$	0.072	0.072	0.067	0.067	0.060	0.059
	$(30, 90)$	0.067	0.068	0.066	0.065	0.061	0.060
	$(50, 150)$	0.057	0.057	0.059	0.060	0.056	0.056
	$(100, 300)$	0.054	0.055	0.055	0.055	0.053	0.053
C	$(10, 10)$	0.277	0.275	0.078	0.079	0.057	0.057
	$(20, 20)$	0.121	0.119	0.059	0.059	0.049	0.049
	$(40, 40)$	0.080	0.078	0.056	0.054	0.053	0.051
	$(60, 60)$	0.069	0.068	0.054	0.054	0.051	0.053
	$(100, 100)$	0.061	0.060	0.054	0.051	0.053	0.050
	$(200, 200)$	0.055	0.054	0.051	0.051	0.053	0.050
	$(5, 15)$	0.292	0.291	0.112	0.110	0.083	0.080
	$(10, 30)$	0.138	0.137	0.086	0.086	0.071	0.069
	$(20, 60)$	0.086	0.087	0.069	0.069	0.061	0.060
	$(30, 90)$	0.073	0.073	0.064	0.062	0.057	0.056
	$(50, 150)$	0.062	0.061	0.055	0.055	0.053	0.052
	$(100, 300)$	0.053	0.054	0.052	0.051	0.051	0.050

Table 3.2. Type I error rate of the asymptotic test T_q^{BG} in the sparse sampling design

Setting	(n_1, n_2)	$\hat{p}_1(S_0^2)$	$\hat{p}_2(S_0^2)$	$\hat{p}_1(\hat{\sigma}_J^2)$	$\hat{p}_2(\hat{\sigma}_J^2)$	$\hat{p}_1(\hat{\sigma}_J^2, \hat{f})$	$\hat{p}_2(\hat{\sigma}_J^2, \hat{f})$
A	$(10, 10)$	0.273	0.272	0.066	0.066	0.047	0.047
	$(20, 20)$	0.121	0.120	0.060	0.060	0.052	0.051
	$(40, 40)$	0.076	0.075	0.052	0.052	0.048	0.048
	$(60, 60)$	0.067	0.068	0.053	0.052	0.051	0.050
	$(100, 100)$	0.056	0.056	0.049	0.050	0.048	0.048
	$(200, 200)$	0.056	0.056	0.052	0.052	0.052	0.052
	$(5, 15)$	0.306	0.307	0.095	0.094	0.062	0.061
	$(10, 30)$	0.134	0.133	0.073	0.073	0.058	0.057
	$(20, 60)$	0.082	0.081	0.060	0.059	0.052	0.051
	$(30, 90)$	0.071	0.071	0.059	0.059	0.054	0.053
	$(50, 150)$	0.062	0.062	0.054	0.054	0.051	0.052
	$(100, 300)$	0.054	0.053	0.052	0.052	0.051	0.051
B	$(10, 10)$	0.217	0.215	0.062	0.061	0.041	0.041
	$(20, 20)$	0.104	0.103	0.056	0.057	0.047	0.047
	$(40, 40)$	0.068	0.068	0.050	0.051	0.046	0.046
	$(60, 60)$	0.061	0.062	0.051	0.051	0.048	0.048
	$(100, 100)$	0.057	0.057	0.051	0.051	0.048	0.049
	$(200, 200)$	0.052	0.053	0.049	0.050	0.049	0.049
	$(5, 15)$	0.250	0.248	0.096	0.095	0.066	0.066
	$(10, 30)$	0.117	0.117	0.080	0.079	0.066	0.066
	$(20, 60)$	0.075	0.075	0.065	0.065	0.058	0.058
	$(30, 90)$	0.063	0.064	0.062	0.062	0.057	0.057
	$(50, 150)$	0.055	0.056	0.055	0.055	0.053	0.053
	$(100, 300)$	0.055	0.055	0.055	0.055	0.054	0.053
C	$(10, 10)$	0.274	0.272	0.072	0.073	0.053	0.053
	$(20, 20)$	0.127	0.127	0.064	0.064	0.053	0.054
	$(40, 40)$	0.077	0.078	0.054	0.054	0.050	0.050
	$(60, 60)$	0.069	0.070	0.054	0.054	0.052	0.051
	$(100, 100)$	0.061	0.060	0.053	0.053	0.052	0.051
	$(200, 200)$	0.055	0.054	0.051	0.050	0.050	0.050
	$(5, 15)$	0.300	0.297	0.114	0.112	0.082	0.082
	$(10, 30)$	0.140	0.139	0.081	0.081	0.067	0.066
	$(20, 60)$	0.088	0.087	0.069	0.067	0.060	0.058
	$(30, 90)$	0.071	0.071	0.060	0.061	0.057	0.055
	$(50, 150)$	0.063	0.064	0.057	0.058	0.054	0.054
	$(100, 300)$	0.059	0.059	0.055	0.055	0.054	0.054

design used in Section 3.3.1. However, we can obtain similar results even for the dense sampling design. In this section, we provide a numerical comparison of the proposed tests and the functional Anderson–Darling (FAD) test introduced by Pomann et al. (2016). The FAD test is based on the Bonferroni correction of the p-value p_k $(k = 1, \ldots, K)$ of the univariate two-sample Anderson–Darling test (Pettitt, 1976) for each common FPC scores $\{\hat{\xi}_{1ik}\}_{i=1}^{n_1}$ and $\{\hat{\xi}_{2ik}\}_{i=1}^{n_2}$.

First, we generate data $\{(t_{1ij}, Y_{1ij}) : j = 1, \ldots, m_{1i}\}_{i=1}^{n_1}$ and $\{(t_{2ij}, Y_{2ij}) : j = 1, \ldots, m_{2i}\}_{i=1}^{n_2}$ using model (3.3.6). We consider the underlying stochastic model

$$X_{zi}(t) = \mu_z(t) + \xi_{zi1}\sqrt{2}\sin(2\pi t) + \xi_{zi2}\sqrt{2}\cos(2\pi t), \quad (z = 1, 2).$$

Here, we consider the following settings:

(A-1) *location shift*: Set the mean functions as $\mu_1(t) = t$ and $\mu_2(t) = t + 2\delta t^3$. Generate the coefficients as $\xi_{1i1}, \xi_{2i1} \sim N(0, 10)$ and $\xi_{1i2}, \xi_{2i2} \sim N(0, 5)$. The index δ controls the departure in the mean behavior of the two distributions.

(A-2) *scale shift*: Set $\mu_1(t) = \mu_2(t) = 0$. Generate the coefficients $\xi_{1i1} \sim N(0, 10)$, $\xi_{2i1} \sim N(0, 10(1+2\delta))$ and $\xi_{1i2}, \xi_{2i2} \sim N(0, 5)$. The index δ controls the difference in the variance of the first basis coefficient between the two data sets.

(B-1) *location shift*: Set the mean functions as $\mu_1(t) = t$ and $\mu_2(t) = t + 2\delta t^3$. Generate the coefficients as $\xi_{1i1}, \xi_{2i1} \sim T_4(0, 10)$ and $\xi_{1i2}, \xi_{2i2} \sim T_4(0, 5)$.

(B-2) *scale shift*: Set $\mu_1(t) = \mu_2(t) = 0$. Generate the coefficients $\xi_{1i1} \sim T_4(0, 10)$, $\xi_{2i1} \sim T_4(0, 10(1 + 2\delta))$ and $\xi_{1i2}, \xi_{2i2} \sim T_4(0, 5)$.

Next, we consider the stochastic processes, for each $z = 1, 2$ and $i = 1, \ldots, n_z$,

$$X_{zi}(t) = \mu_z t + \sigma_z \sum_{n=1}^{K} \frac{\sqrt{2}\sin\left(\left(n - \frac{1}{2}\right)\pi t\right)}{\left(n - \frac{1}{2}\right)\pi} Z_{zin},$$

where $\{Z_{zin}\}_{n=1}^{K}$ are independent Gaussian variables with mean zero and variance one. For each $z = 1, 2$, the stochastic processes $X_{zi}(t)$'s converge to the Wiener processes with drift μ_z and infinitesimal variance σ_z^2 as $K \to \infty$. Here, we set $K = 100$ and consider the following settings:

(C-1) *location shift*: Set $\mu_1 = 0,\ \mu_2 = \delta$ and $\sigma_1 = \sigma_2 = 1$.

(C-2) *scale shift*: Set $\mu_1 = \mu_2 = 0$ and $\sigma_1 = 1,\ \sigma_2 = 1 + \delta$.

All the results in this section are based on $\alpha = 0.05$ level of significance and $10,000$ Monte Carlo replications. We conducted the asymptotic tests based on $\hat{p}_1(\hat{\sigma}_J^2, \hat{f})$ and $\hat{p}_2(\hat{\sigma}_J^2, \hat{f})$ under these settings. The simulation results of $\hat{p}_q(S_0^2)$ and $\hat{p}_q(\hat{\sigma}_J^2)$ are omitted from the viewpoint of controlling the Type I error rate. For comparison, we conducted the approximate permutation tests of T_q^{BF} and T_q^{BG} $(q = 1, 2)$. The distributions of T_q^{BF} and T_q^{BG} are approximated through empirical distributions based on $1,000$ permutations.

The results of the numerical comparison are listed in Tables 3.3 and 3.4. In the location-shift models (Table 3.3), because the estimator $\hat{\sigma}_J^2$ is unaffected by shifts, the results of the asymptotic

Table 3.3. Empirical powers for location-shift models

Setting	(n_1, n_2)	δ	FAD	Permutation T_1^{BF}	T_2^{BF}	T_1^{BG}	T_2^{BG}	Asymptotic $\hat{p}_1(\hat{\sigma}_J^2, \hat{f})$	$\hat{p}_2(\hat{\sigma}_J^2, \hat{f})$
A-1	$(10, 10)$	2.0	0.419	0.097	0.102	0.065	0.068	0.057	0.058
		3.0	0.803	0.245	0.319	0.120	0.154	0.116	0.138
		4.0	0.962	0.558	0.746	0.261	0.384	0.223	0.311
	(5, 15)	2.0	0.353	0.120	0.126	0.069	0.069	0.079	0.081
		3.0	0.628	0.178	0.205	0.102	0.109	0.125	0.136
		4.0	0.879	0.334	0.484	0.168	0.224	0.185	0.223
	$(20, 20)$	2.0	0.464	0.157	0.185	0.064	0.068	0.062	0.065
		3.0	0.904	0.596	0.794	0.176	0.256	0.154	0.220
		4.0	0.997	0.976	0.996	0.446	0.716	0.421	0.677
	$(10, 30)$	2.0	0.395	0.116	0.130	0.066	0.068	0.074	0.076
		3.0	0.817	0372	0.515	0.121	0.162	0.145	0.174
		4.0	0.978	0.832	0.960	0.314	0.518	0.283	0.415
B-1	$(10, 10)$	2.0	0.448	0.141	0.155	0.066	0.070	0.058	0.059
		3.0	0.800	0.407	0.539	0.139	0.181	0.129	0.156
		4.0	0.943	0.738	0.869	0.325	0.485	0.222	0.325
	$(5, 15)$	2.0	0.324	0.111	0.122	0.055	0.063	0.082	0.083
		3.0	0.661	0.226	0.295	0.111	0.138	0.130	0.142
		4.0	0.864	0.511	0.650	0.194	0.283	0.193	0.245
	$(20, 20)$	2.0	0.498	0.228	0.278	0.066	0.075	0.062	0.066
		3.0	0.903	0.772	0.864	0.176	0.273	0.143	0.204
		4.0	0.986	0.981	0.988	0.457	0.712	0.416	0.607
	$(10, 30)$	2.0	0.416	0.192	0.213	0.074	0.079	0.085	0.088
		3.0	0.771	0.514	0.655	0.114	0.147	0.120	0.165
		4.0	0.953	0.886	0.934	0.313	0.490	0.269	0.398
C-1	$(10, 10)$	1.0	0.326	0.386	0.357	0.222	0.209	0.234	0.206
		1.5	0.641	0.746	0.720	0.482	0.450	0.516	0.482
		2.0	0.884	0.941	0.933	0.785	0.768	0.790	0.744
	$(5, 15)$	1.0	0.279	0.283	0.267	0.167	0.154	0.205	0.195
		1.5	0.488	0.597	0.557	0.354	0.325	0.378	0.360
		2.0	0.761	0.815	0.792	0.639	0.605	0.603	0.571
	$(20, 20)$	1.0	0.681	0.731	0.686	0.338	0.292	0.366	0.329
		1.5	0.959	0.974	0.969	0.779	0.742	0.784	0.723
		2.0	0.998	0.998	0.998	0.971	0.959	0.981	0.971
	$(10, 30)$	1.0	0.505	0.555	0.518	0.252	0.233	0.265	0.235
		1.5	0.885	0.908	0.886	0.649	0.599	0.608	0.574
		2.0	0.992	0.992	0.986	0.921	0.897	0.900	0.855

Table 3.4. Empirical powers for scale-shift models

Setting	(n_1, n_2)	δ	FAD	Permutation				Asymptotic	
				T_1^{BF}	T_2^{BF}	T_1^{BG}	T_2^{BG}	$\hat{p}_1(\hat{\sigma}_J^2, \hat{f})$	$\hat{p}_2(\hat{\sigma}_J^2, \hat{f})$
A-2	$(10, 10)$	2.0	0.065	0.114	0.121	0.440	0.445	0.428	0.428
		3.0	0.097	0.178	0.178	0.633	0.636	0.588	0.586
		4.0	0.110	0.231	0.232	0.733	0.731	0.678	0.677
	(5, 15)	2.0	0.042	0.038	0.038	0.140	0.143	0.461	0.460
		3.0	0.038	0.036	0.036	0.197	0.194	0.630	0.629
		4.0	0.027	0.055	0.055	0.274	0.273	0.750	0.749
	(20, 20)	2.0	0.144	0.305	0.306	0.838	0.842	0.823	0.824
		3.0	0.262	0.530	0.523	0.961	0.960	0.943	0.944
		4.0	0.340	0.633	0.637	0.975	0.974	0.974	0.974
	(10, 30)	2.0	0.043	0.123	0.126	0.578	0.575	0.825	0.826
		3.0	0.059	0.195	0.195	0.772	0.772	0.944	0.943
		4.0	0.052	0.284	0.284	0.859	0.859	0.982	0.992
B-2	(10, 10)	2.0	0.075	0.125	0.125	0.315	0.311	0.274	0.274
		3.0	0.082	0.153	0.157	0.461	0.459	0.388	0.386
		4.0	0.086	0.195	0.199	0.561	0.553	0.457	0.455
	(5, 15)	2.0	0.036	0.033	0.033	0.072	0.072	0.377	0.376
		3.0	0.018	0.032	0.030	0.101	0.099	0.502	0.501
		4.0	0.047	0.044	0.049	0.142	0.143	0.583	0.581
	(20, 20)	2.0	0.136	0.279	0.280	0.619	0.621	0.578	0.578
		3.0	0.194	0.399	0.404	0.781	0.784	0.736	0.733
		4.0	0.241	0.541	0.542	0.868	0.869	0.826	0.825
	(10, 30)	2.0	0.028	0.088	0.085	0.304	0.303	0.624	0.623
		3.0	0.053	0.169	0.171	0.458	0.460	0.780	0.781
		4.0	0.053	0.257	0.247	0.563	0.562	0.847	0.845
C-2	(10, 10)	1.0	0.093	0.130	0.132	0.512	0.539	0.515	0.545
		1.5	0.105	0.224	0.234	0.766	0.796	0.708	0.731
		2.0	0.121	0.336	0.357	0.900	0.920	0.845	0.867
	(5, 15)	1.0	0.012	0.048	0.050	0.170	0.192	0.536	0.549
		1.5	0.012	0.069	0.076	0.362	0.395	0.689	0.712
		2.0	0.006	0.064	0.078	0.509	0.565	0.832	0.846
	(20, 20)	1.0	0.152	0.310	0.332	0.872	0.891	0.880	0.893
		1.5	0.265	0.603	0.625	0.986	0.991	0.987	0.990
		2.0	0.443	0.866	0.897	0.999	1.000	0.999	1.000
	(10, 30)	1.0	0.028	0.117	0.126	0.634	0.666	0.816	0.839
		1.5	0.029	0.267	0.291	0.886	0.913	0.966	0.971
		2.0	0.042	0.500	0.536	0.982	0.990	0.998	1.000

tests based on $\hat{p}_1(\hat{\sigma}_J^2, \hat{f})$ and $\hat{p}_2(\hat{\sigma}_J^2, \hat{f})$ are close to the approximate permutation tests of T_1^{BG} and T_2^{BG}, respectively. For the location-shift models (A-1) and (B-1), the FAD test shows the best performance, and the Baringhaus–Franz type test T_q^{BF} was better than the Biswas–Ghosh type test T_q^{BG}, whereas the L_1-norm-based tests are inferior to the L_2-norm-based versions. This result reflects the property of asymptotic power presented in Section 3.1. In the setting (C-1) with a large number of eigenfunctions, the test T_q^{BF} is higher than the power of the FAD test, and the L_1-norm-based tests are superior to the L_2-norm-based tests.

For the scale-shift models (Table 3.4) and the equal sample sizes $n_1 = n_2$, the asymptotic tests based on $\hat{p}_1(\hat{\sigma}_J^2, \hat{f})$ and $\hat{p}_2(\hat{\sigma}_J^2, \hat{f})$ are close to the approximate permutation tests of T_1^{BG} and T_2^{BG}, respectively. For the unequal sample sizes $n_1 \neq n_2$, note that the asymptotic and permutation tests of T_1^{BG} and T_2^{BG} have different powers because the estimator $\hat{\sigma}_J^2$ is affected by the shifts in scale. For the scale-shift models (A-2), (B-2), and (C-2), the power of T_q^{BG} is superior to that of the other tests. In the setting (C-2) with a large number of eigenfunctions, the power of the L_2-norm-based tests is higher than that in the L_1-norm-based tests.

3.3.3 Application to diffusion tensor image data analysis

One of the purposes of the diffusion tensor image (DTI) study is to formally assess whether several imaging modalities vary differently between healthy controls and patients with multiple sclerosis (MS). MS is a disease that affects the central nervous system and, in particular, damages white matter tracts in the brain through lesions, a loss of myelin, and axonal damage. As one of the approaches used to visualize white matter tracts, DTI is well known as the magnetic resonance imaging method that measures water diffusivity in the brain. To characterize the microstructure of tissue, one of the measures (referred to as modalities) provided by DTI is fractional anisotropy (FA), which describes the degree of anisotropy of the water diffusion process.

In this section, we investigate the FA profiles along the corpus callosum (CCA-FA profiles) and test the null hypothesis that these profiles have the same distribution for both the MS patients and the control subjects. The data used are 100 subjects with MS and 42 healthy controls available in the R package "refund." For the MS patients, the number of visits per subject ranged from 2 to 8, and a total of 340 visits were recorded. These MRI/DTI data were collected at Johns Hopkins University and the Kennedy Krieger Institute. Each curve is observed on 93 grids along the corpus callosum. We used the CCA-FA profile data at the time of the first visit and conducted a two-sample homogeneity test with the following settings:

Case 1: CCA-FA profiles for the 100 MS patients and 42 controls in all data.

Case 2: CCA-FA profiles for the 66 MS patients and 30 controls in males.

Case 3: CCA-FA profiles for the 34 MS patients and 12 controls in females.

Case 4: CCA-FA profiles for 66 men and 34 women among the MS patients.

Case 5: CCA-FA profiles for 30 men and 12 women among the healthy controls.

For these cases, we conducted some of the two-sample tests described in Section 3.3.2. The p-values for the Baringhaus–Franz and Biswas–Ghosh type tests are calculated using $100,000$ permutations. In Cases 1–3, the p-values of all tests for both MS patients and the control subjects were close to zero, indicating a difference between them. The p-values of the asymptotic Biswas–Ghosh type tests defined in Section 3.1 are also close to zero. For Case 4, the p-values of the test statistics FAD, T_1^{BF}, T_2^{BF}, T_1^{BG} and T_2^{BG} are 1.00, 0.77, 0.73, 0.91, and 0.83, respectively. Furthermore, the p-values of these test statistics in Case 5 are 0.49, 0.50, 0.70, 0.90, and 0.91, respectively. Therefore, the results in Cases 4 and 5 show that the CCA-FA profiles are not statistically different between men and women. In addition, the p-values of the asymptotic Biswas–Ghosh type tests in Cases 4 and 5 are as listed in Table 3.5.

Table 3.5. The p-values of the Biswas–Ghosh type tests in Cases 4 and 5

	Permutation		Asymptotic					
	T_1^{BG}	T_2^{BG}	$\hat{p}_1(S_0^2)$	$\hat{p}_2(S_0^2)$	$\hat{p}_1(\hat{\sigma}_J^2)$	$\hat{p}_2(\hat{\sigma}_J^2)$	$\hat{p}_1(\hat{\sigma}_J^2, \hat{f})$	$\hat{p}_2(\hat{\sigma}_J^2, \hat{f})$
Case 4	0.913	0.831	0.841	0.771	0.847	0.780	0.847	0.781
Case 5	0.898	0.907	0.734	0.758	0.746	0.753	0.749	0.755

From Table 3.5, the p-values of the asymptotic Biswas–Ghosh type tests also showed large values similar to those of the approximate permutation tests. Furthermore, the p-values of the asymptotic Biswas–Ghosh type tests are slightly smaller than those of the corresponding approximate permutation tests. These results may be affected by the non-conservative tendency of the Type I error rate investigated in Section 3.3.1 when the sample sizes n_1 and n_2 are unequal.

Chapter 4

Conclusion

In this paper, we proposed several smoothed bootstrap methods for RSS and investigated their asymptotic properties. Furthermore, we proposed a two-sample test for functional data and investigated its asymptotic properties.

In Chapter 2, we proposed several smoothed bootstrap methods based on RSS. We also examined several bootstrap methods based on RSS data and derived asymptotic properties of these resampling methods under an imperfect ranking model. First, we proposed smoothed and symmetrized bootstrap methods, and showed that each bootstrap method has consistency for the sample mean. Second, we computed the asymptotic MISE for the RSS-based KCDE, as well as an asymptotic optimal bandwidth that minimizes the MISE. Finally, we investigated the influence of smoothing and symmetrizing of the bootstrap methods using simulation studies. Moreover, we also believe that it is necessary to conduct a theoretical study of various bootstrap methods in the context of the stratified ranked set sampling discussed in Samawi et al. (2017, 2019). In this study, although we considered a bandwidth that minimizes MISE, it is also important to select the optimum bandwidth according to each purpose.

In Chapter 3, we constructed interpoint distance-based two-sample tests for random elements on a Banach space. In particular, we considered the homogeneity tests for functional data. We derived the limiting distribution of the Biswas–Ghosh type test (Biswas and Ghosh, 2014) under contiguous alternatives and proposed a p-value approximation based on a jackknife variance estimator and the Welch–Satterthwaite equation. In addition, we compared the powers of the proposed tests and the functional Anderson–Darling test (Pomann et al., 2016). Through a simulation study, we verified that the proposed p-value approximation achieves a better performance than the p-value approximation based on the naive estimator (Biswas and Ghosh, 2014), and demonstrated that the Biswas–Ghosh type tests are better than the others for the scale-shift model. In particular, the proposed tests were shown to have better power than the FAD test for functional data with numerous basis functions. The testing approach based on the interpoint distance can be easily extended to testing the null hypothesis in which multiple (more than two) groups of curves have identical distributions. In addition, we can construct a consistency test using a strongly negative definite kernel for multivariate functional data. Such detailed studies will be a topic of future research.

Bibliography

A. M. Alonso and A. Cuevas. On smoothed bootstrap for density functionals. *Journal of Nonparametric Statistics*, 15(4–5):467–477, 2003.

A. Araujo and E. Giné. *The Central Limit Theorem for Real and Banach Valued Random Variables.* John Wiley & Sons, New York, 1980.

A. Bakshaev. Goodness of fit and homogeneity tests on the basis of N-distances. *Journal of Statistical Planning and Inference*, 139(11):3750–3758, 2009.

L. Barabesi and L. Fattorini. Kernel estimators for probability density functions by ranked-set sampling. *Communications in Statistics - Theory and Methods*, 31(11):597–610, 2002.

L. Baringhaus and C. Franz. On a new multivariate two-sample test. *Journal of Multivariate Analysis*, 88(1):190–206, 2004.

M. Benko, W. Härdle, and A. Kneip. Common functional principal components. *The Annals of Statistics*, 37(1):1–34, 2009.

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

M. Biswas and A. K. Ghosh. A nonparametric two-sample test applicable to high dimensional data. *Journal of Multivariate Analysis*, 123:160–171, 2014.

L. L. Bohn. A review of nonparametric ranked-set sampling methodology. *Communications in Statistics - Theory and Methods*, 25(11):2675–2685, 1996.

L. L. Bohn and D. A. Wolfe. The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the Mann–Whitney–Wilcoxon statistic. *Journal of the American Statistical Association*, 89(425):168–175, 1994.

E. Brunner, A. C. Bathke, and F. Konietschke. *Rank and Pseudo-Rank Procedures for Independent Observations in Factorial Designs: using R and SAS.* Springer International Publishing, 2018.

A. Chakraborty and P. Chaudhuri. A Wilcoxon–Mann–Whitney-type test for infinite-dimensional data. *Biometrika*, 37(1):1–34, 2009.

Z. Chen. Density estimation using ranked set sampling data. *Environmental and Ecological Statistics*, 6(2):135–146, 1999.

Z. Chen, Z. Bai, and B. K. Sinha. *Ranked Set Sampling: Theory and Applications*. Springer, New York, 2004.

C. M. Crainiceanu, A. M. Staicu, and C.-Z. Di. Generalized multilevel functional regression. *Journal of the American Statistical Association*, 104(488):1550–1561, 2009.

D. De Martini. Smoothed bootstrap consistency through the convergence in mallows metric of smooth estimates. *Journal of Nonparametric Statistics*, 12(6):819–835, 2000.

T. R. Dell and J. L. Clutter. Ranked set sampling theory with order statistics background. *Biometrics*, 28(2):545–555, 1972.

C.-Z. Di, C. M. Crainiceanu, B. S. Caffo, and N. M. Punjabi. Multilevel functional principal component analysis. *The Annals of Applied Statistics*, 3(1):458–488, 2009.

R. Drikvandi, R. Modarres, and A. H. Jalilian. A bootstrap test for symmetry based on ranked set samples. *Computational Statistics & Data Analysis*, 55(4):1807–1814, 2011.

L. Duembgen and E. Zamanzade. Inference on a distribution function from ranked set samples. *Annals of the Institute of Statistical Mathematics*, 72(1):157–185, 2020.

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

B. Efron and C. Stein. The jackknife estimate of variance. *The Annals of Statistics*, 9(3): 586–596, 1981.

C. El-Nouty and A. Guillou. On the smoothed bootstrap. *Journal of Statistical Planning and Inference*, 83(1):203–220, 2000.

F. Ferraty and P. Vieu. *Nonparametric functional data analysis: theory and practice*. Springer-Verlag, New York, 2006.

F. Ferraty, P. Vieu, and S. Viguier-Pla. Factor-based comparison of groups of curves. *Computational Statistics & Data Analysis*, 51(10):4903–4910, 2007.

S. Fremdt, L. Horvath, P. Kokoszka, and J. Steinebach. Testing the equality of covariance operators in functional samples. *Scandinavian Journal of Statistics*, 40(1):138–152, 2013.

J. Frey. New imperfect rankings models for ranked set sampling. *Journal of Statistical Planning and Inference*, 137(4):1433–1445, 2007.

J. Goldsmith, C. M. Crainiceanu, B. Caffo, and D. Reich. Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(3):453–469, 2012.

U. Grenander. Stochastic processes and statistical inference. *Arkiv f ör Matematik*, 1(3):195–277, 1950.

P. Hall and I. Van Keilegom. Two-sample tests in functional data analysis starting from discrete data. *Statistica Sinica*, 17(4):1511–1531, 2007.

P. Hall, T. J. DiCiccio, and J. P. Romano. On smoothing and the bootstrap. *The Annals of Statistics*, 17(2):692–704, 1989.

M. Hollander, D. Wolfe, and E. Chicken. *Nonparametric Statistical Methods.* John Wiley and Sons, Inc., New York, 3rd edition, 2014.

L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications.* Springer, New York, 2012.

L. Horváth, P. Kokoszka, and R. Reeder. Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):103–122, 2013.

J.. Huang. Asymptotic properties of the NPMLE of a distribution function based on ranked set sampling. *The Annals of Statistics*, 25(3):1036–1049, 1997.

P. J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

T. P. Hui, R. Modarres, and G. Zheng. Bootstrap confidence interval estimation of mean via ranked set sampling linear regression. *Journal of Statistical Computation and Simulation*, 75 (7):543–553, 2005.

M. C. Jones. The performance of kernel density functions in kernel distribution function estimation. *Statistics and Probability Letters*, 9(2):129–132, 1990.

L. B. Klebanov. $\mathfrak{N}$-*Distances and Their Applications.* The Karolinum Press, Charles University in Prague, 2005.

C. H. Kraft, Y. Lepage, and C. Van Eeden. Estimation of a symmetric density function. *Communications in Statistics - Theory and Methods*, 14(2):273–288, 1985.

D. Kraus and V. Panaretos. Dispersion operators and resistant second-order analysis of functional data. *Biometrika*, 99(4):813–832, 2012.

P. H. Kvam and F. J. Samaniego. Nonparametric maximum likelihood estimation based on ranked set samples. *Journal of the American Statistical Association*, 89(426):526–537, 1994.

A. J. Lee. *U-Statistics: Theory and Practice.* Marcel Dekker, New York, 1990.

J. Lim, M. Chen, S. Park, X. Wang, and L. Stokes. Kernel density estimator from ranked set samples. *Communications in Statistics - Theory and Methods*, 43(10–12):2156–2168, 2014.

M. Mahdizadeh and E. Zamanzade. Kernel-based estimation of $P(X > Y)$ in ranked set sampling. *SORT*, 1(2):243–266, 2016.

M. Mahdizadeh and E. Zamanzade. Smooth estimation of a reliability function in ranked set sampling. *Statistics*, 52(4):750–768, 2018.

M. Mahdizadeh and E. Zamanzade. Smooth estimation of the area under the ROC curve in multistage ranked set sampling. *Statistical Papers*, 62(4):1753–1776, 2021.

G. A. McIntyre. A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agriculture Research*, 3(4):385–390, 1952.

R. Modarres, T. P. Hui, and G. Zheng. Resampling methods for ranked set samples. *Computational Statistics & Data Analysis*, 51(2):1039–1050, 2006.

Ö. Öztürk. A nonparametric test of symmetry versus asymmetry for ranked-set samples. *Communications in Statistics - Theory and Methods*, 30(10):2117–2133, 2001.

A. N. Pettitt. A two-sample Anderson–Darling rank statistic. *Biometrika*, 63(1):161–168, 1976.

A. M. Polansky. A smooth nonparametric approach to process capability. *Quality and Reliability Engineering International*, 14(1):43–48, 1998.

A. M. Polansky and E. R. Baker. Multistage plug-in bandwidth selection for kernel distribution function estimates. *Journal of Statistical Computation and Simulation*, 65(1–4):63–80, 2000.

G.-M. Pomann, A.-M. Staicu, and S. Ghosh. A two-sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 65(3):395–414, 2016.

J. O. Ramsay. When the data are functions. *Psychometrika*, 47(4):379–396, 1982.

J. O. Ramsay and C. Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):539–572, 1991.

J. O. Ramsay and B. W. Silverman. *Applied Functional Data Analysis*. Springer, New York, 2002.

J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, New York, 2nd edition, 2005.

C. R. Rao. Some statistical methods for comparison of growth curves. *Biometrics*, 14(1):1–17, 1958.

H. Samawi, A. Chatterjee, J. J. Yin, and H. Rochani. On kernel density estimation based on different stratified sampling with optimal allocation. *Communications in Statistics - Theory and Method*, 46(22):10973–10990, 2017.

H. Samawi, H. Rochani, J. Yin, D. Linder, and R. Vogel. Notes on kernel density based mode estimation using more efficient sampling designs. *Computational Statistics*, 33:1071–1090, 2018.

H. Samawi, H. Rochani, J. Yin, and R. Vogel. On kernel-based mode estimation using different stratified sampling design. *Journal of Statistical Theory and Practice*, 13:1–23, 2019.

R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley and Sons, Inc., New York, 1980.

S. J. Sheather and M. C. Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3): 683–690, 1991.

B. W. Silverman and G. A. Young. The bootstrap: To smooth or not to smooth? *Biometrika*, 74(3):469–479, 1987.

S. L. Stokes and T. W. Sager. Characterization of a ranked-set sample with application to estimating distribution functions. *Journal of the American Statistical Association*, 83(402): 374–381, 1988.

J. W. H. Swanepoel. Mean integrated squared error properties and optimal kernels when estimating a distribution function. *Communications in Statistics - Theory and Methods*, 17: 3785–3799, 1988.

G. J. Székely and M. L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, 2013.

K. Takahasi and K. Wakimoto. On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, 20 (1):1–31, 1968.

S. Wang. Optimizing the smoothed bootstrap. *Annals of the Institute of Statistical Mathematics*, 47(1):65–80, 1995.

D. A. Wolfe. Ranked set sampling: An approach to more efficient data collection. *Statistical Science*, 19(4):636–643, 2004.

H. Yamaguchi and H. Murakami. Symmetric smoothed bootstrap methods for ranked set samples. *Journal of Nonparametric Statistics*, 33(3–4):435–463, 2021.

H. Yamaguchi and H. Murakami. Interpoint distance-based two-sample tests for functional data. *Communications in Statistics - Theory and Methods*, 2022. doi: 10.1080/03610926. 2022.2148475.

H. Yamato. Uniform convergence of an estimator of a distribution function. *Bulletin of Mathematical Statistics*, 15(3/4):69–78, 1973.

F. Yao, H.-G. Müller, and J.-L. Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.

J. H. Yin, Y. Hao, H. Samawi, and H. Rochani. Rank-based kernel estimation of the area under the ROC curve. *Statistical Methodology*, 32:91–106, 2016.

J.-T. Zhang, X. Liang, and S. Xiao. On the two-sample Behrens–Fisher problem for functional data. *Journal of Statistical Theory and Practice*, 4(4):571–587, 2010.