ON EXISTENCE OF A CHANGE IN MEAN OF FUNCTIONAL DATA

Buddhananda Banerjee\textsuperscript{0a} and Satyaki Mazumder \textsuperscript{1b}

Department of Mathematics & Statistics
Indian Institute of Science Education and Research, Kolkata\textsuperscript{0,1}
buddha.banerjee@iiserkol.ac.in\textsuperscript{a}, satyaki@iiserkol.ac.in\textsuperscript{b}

Abstract. Functional data often arise as sequential temporal observations over a continuous state-space. A set of functional data with a possible change in its structure may lead to a wrong conclusion if it is not taken into account. So, sometimes, it is crucial to know about the existence of change point in a given sequence of functional data before doing any further statistical inference. We develop a new methodology to provide a test for detecting a change in the mean function of the corresponding data. To obtain the test statistic we provide an alternative estimator of the covariance kernel. The proposed estimator is asymptotically unbiased under the null hypothesis and, at the same time, has smaller amount of bias than that of the existing estimator. We show here that under the null hypothesis the proposed test statistic is pivotal asymptotically. Moreover, it is shown that under alternative hypothesis the test is consistent for large enough sample size. It is also found that the proposed test is more powerful than the available test procedure in the literature. From the extensive simulation studies we observe that the proposed test outperforms the existing one with a wide margin in power for moderate sample size. The developed methodology performs satisfactorily for the average daily temperature of central England and monthly global average anomaly of temperatures.

Keywords: Change point detection, functional data analysis, covariance kernel.

1. Introduction

Functional data analysis (FDA) is becoming increasingly popular because of its wide applicability in various fields of statistics. The natural proximity of functional data to feature some real life observations is more appealing over its finite dimensional representation and at the same time it is often noticed that FDA leads to more accurate inference in this regard. \cite{RamseySilverman:05} has enriched the literature with a detailed discussions on several techniques and usefulness of FDA. Some recent developments in many more aspects

\footnotesize
\textsuperscript{0}corresponding author: buddha.banerjee@iiserkol.ac.in.
of FDA can be found in Ferraty (2011). However, the inference and especially the prediction may alter if there exists an inherent change in the stochastic structure of the functional data observed temporally. The change may occur at an unknown point of time within the chronological sequence of data but it is always challenging to test, statistically, whether the change has occurred or not. For the cases of scalar and vector data a considerable amount of contributions can be found from the works by Cobb (1978), Inclán and Tiao (1994), Davis et al. (1995), Antoch et al. (1997), Horváth et al. (1999), Kokoszka and Leipus (2000), Kirch et al. (2014) and references therein, among many others. In the context of functional data a change may occur in the mean function or in the covariance kernel of the data or both. This paper shades light on the discussion about the change in the mean function in particular. Recently, Berkes et al. (2009) and Aue et al. (2009) have proposed a method for detecting changes in the mean functions of an observed set of functional data. Berkes et al. (2009), in their pioneering work in this context, have provided an elegant test procedure to decide the existence of a significant amount of change in the mean function, whereas Aue et al. (2009) following the method of Berkes et al. (2009) have dealt with the detection of the position of the change in the mean function. In practice, both are equally important to judge whether there is a change in the mean function of the data and if there is a significant change at all then detecting the location of it. For example, while analyzing the temperature of a certain region over a long period of time, it is very important to environmentalist to identify the time point after which a significant change in the mean temperature is observed as a possible effect of global warming. In this paper we come up with a different methodology to analyze the functional data subject to a possible change point and propose a new statistical test, which is more powerful than the existing one(s), for detecting the presence of a change in the mean function of the data. Here we show that under the null hypothesis, i.e. with no change in the data, the proposed test statistic converges in distribution to a functional of the Brownian bridges, as shown in Berkes et al. (2009). Moreover, we prove here that the test is consistent under alternative hypothesis when the number of the observations becomes large enough. We provide an estimator of the covariance kernel which not only enjoys its property of consistency under the null hypothesis but also has less asymptotic bias compared to that of the estimator provided by Berkes et al. (2009) or Aue et al. (2009) under the alternative hypothesis. Because of the reduction in the asymptotic bias while estimating the covariance kernel, we successfully obtain that the proposed test has better power than the existing
method by Berkes et al. (2009). The outcomes of an extensive simulation study reflects the same. It is also noted that our method outperforms the existing method in a wide margin for small samples. Therefore, it is more advantageous to use the proposed method in practice for deciding with the presence of significant change in the mean of the functional data, specially when the data size is not big enough.

The organization of the paper is as follows. In Section 2, we introduce the required notation and definitions for introducing the subject. The details of the model, discussed in the paper, are described in this section. Section 3 deals with the testing methodology and main results of the paper. In this section we provide the theorems about the consistency of the proposed estimator of covariance kernel, asymptotic null distribution and asymptotic consistency of the test procedure. In Section 4 simulations results are provided in great detail where we show that our method substantially improves over the existing method in terms of power of the test. In Section 5 we show the performance of our test in real data. Remarks and conclusion of the work are given in the Section 6. Finally we provide the required proofs of the results of section 3 in the Appendix (Section 7).

2. Preliminaries and assumptions

Let $X_i(t)$ for $i = 1, \ldots, N$, be Hilbert-valued random functions defined over a compact set $\tau = [0, 1]$. We assume that $X_i$s are independent. We are interested to check the equality of the mean functions of $X_i$ for all $i = 1, 2, \cdots, N$. More precisely, the null hypothesis to test will be

$$H_0: E(X_1(t)) = E(X_2(t)) = \cdots = E(X_N(t)).$$

It is important to note that nothing is presumed about any property of the common mean under the null hypothesis.

Under alternative hypothesis we assume that the null hypothesis $H_0$ does not hold. We deal with the situation when the data contains at most one change point, however, in case of applications we elaborate how to implement this method with multiple change points case. In particular, in Section 5, we specifically deal with the situation with more than one change points. There the data can be subdivided into several consecutive parts and within each part the mean function remains constant but it deviates between different contiguous parts. The details of the model with single change point is discussed in the sub-Section 2.1.
Under the null hypothesis we express $X_i, i = 1, \ldots, N,$ in the following manner.

$$X_i(t) = \mu(t) + Y_i(t)$$

$$E(Y_i(t)) = 0. \tag{2.1}$$

Now we specify the assumptions about mean function $\mu$ and random element $Y_i,$ based on which the asymptotic behaviour of the test statistic can be determined. From here on words all integrations are computed over the compact set $\tau,$ unless otherwise mentioned.

2.1. **Assumptions.**

A1. The mean function is square integrable that is, $\mu \in L^2(\tau),$ and the unobservable random component $Y_i$s, are independent and identically distributed random elements in $L^2(\tau)$ with

$$E(Y_i(t)) = 0 \forall t \in \tau,$$

for $i = 1, \ldots, N$ and

$$E||Y_i||^2 = \int E(Y_i^2(t)) \, dt < \infty. \tag{2.2}$$

The covariance kernel is defined as

$$c(t, s) = E(Y_i(t)Y_i(s)) \quad t, s \in \tau \tag{2.3}$$

with the assumption that $c(t, s) \in L^2(\tau \times \tau).$ Assumption 1 implies that the covariance operator of $Y,$ which is a positive definite symmetric Hilbert-Schmidt (H-S) operator mapping from $L^2(\tau)$ to itself, will be of the form

$$C(x) = E[(Y, x)Y]. \tag{2.4}$$

The evaluation of $C(x)$ at $t,$ i.e., $C(x)(t),$ is given by

$$C(x)(t) = \int c(t, s)x(s) \, ds \quad \forall t \in \tau.$$

Moreover, Mercer’s theorem in (Indritz, 1963, Chapter 4) implies that $c(t, s)$ has the following spectral decomposition:

$$c(t, s) = \sum_{l=1}^{\infty} \lambda^l v^l(t)v^l(s) \quad t, s \in \tau, \tag{2.5}$$
where each real scalar $\lambda^l$ and function $\nu^l$ (in $L^2(\tau)$) are defined, for $t \in \tau$, as

$$C(\nu^l)(t) = \lambda^l \nu^l(t), \quad l = 1, 2, \ldots,$$

i.e.,

$$\int c(t, s) \nu^l(s) = \lambda^l \nu^l(t), \quad l = 1, 2, \ldots.$$  \hspace{1cm} (2.6)

In other words, $\lambda^l$s and $\nu^l$s are the eigenvalues and the corresponding eigenfunctions respectively, of the operator $C(.)$. Since the eigenfunctions of the positive definite symmetric operator, $C(.)$, form a complete orthonormal basis of $L^2(\tau)$ and eigenvalues are positive, Karhunen-Loève representation of $Y_i$ holds good in $L^2(\tau)$ and is given by

$$Y_i(t) = \sum_{l=1}^{\infty} \sqrt{\lambda^l} \delta^l_i \nu^l(t),$$  \hspace{1cm} (2.7)

where $\sqrt{\lambda^l} \delta^l_i = (Y_i, \nu^l) = \int Y_i(s) \nu^l(s)$ is known as $l$th functional principal component score. By construction, the elements of the sequence $\{\delta^l_i\}^l$ are uncorrelated random variables with zero mean and unit variance and $\{\delta^l_i\}^l$ and $\{\delta^l_j\}^l$ are independent for $i \neq j$.

A2. There exists some positive integer $d$, such that the eigenvalues $\lambda^l$ satisfy

$$\lambda^1 > \lambda^2 > \ldots > \lambda^d > \lambda^{d+1}.$$  

A3. $Y_i, i = 1, \ldots, N$, satisfy

$$E(||Y_i||^4) = \int E(Y_i(t))^4 dt < \infty.$$  

A4. Under the alternative, with an existence of single change point the observations, $X_i, i = 1, \ldots, N$ can be represented as follows

$$X_i(t) = \begin{cases} 
\mu_1(t) + Y_i(t), & 1 \leq i \leq k^* \\
\mu_2(t) + Y_i(t), & k^* < i \leq N
\end{cases}$$  \hspace{1cm} (2.8)

where $Y_i, i = 1, \ldots, N$ satisfy the assumption A1, $\mu_j(t), j = 1, 2$ are in $L^2(\tau)$ and $k^* = \lfloor N\theta \rfloor$, with $\theta \in (0, 1)$. Therefore, we assume that under the alternative hypothesis of single change point a change may occur in the mean function but the covariance kernel remains the same before and after the change in the data. Keeping this in consideration we estimate the covariance kernel in the following section and develop a new methodology to test $H_0$. 

3. Methodology and Main results

To estimate the covariance kernel let us define the piecewise sample means for two segments

$$\hat{\mu}_k(t) = \frac{1}{k} \sum_{i=1}^{k} X_i(t),$$  \hspace{1cm} (3.1)

$$\tilde{\mu}_k(t) = \frac{1}{N-k} \sum_{i=k+1}^{N} X_i(t),$$  \hspace{1cm} (3.2)

where $k = \lfloor Nu \rfloor$ with $u \in (0, 1)$, implying $1 \leq k < N$. For $u = 1$ we define $\hat{\mu}_N(t) = \frac{1}{N} \sum_{i=1}^{N} X_i(t)$. With the help of equations (3.1) and (3.2), the newly proposed estimator of covariance kernel is

$$\hat{c}_u(t, s) = \frac{1}{N} \left[ \sum_{i=1}^{k} (X_i(t) - \hat{\mu}_k(t)) (X_i(s) - \hat{\mu}_k(s)) + \sum_{i=k+1}^{N} (X_i(t) - \tilde{\mu}_k(t)) (X_i(s) - \tilde{\mu}_k(s)) \right].$$  \hspace{1cm} (3.3)

For $u = 1$, we define $\hat{c}_1(t, s) = \frac{1}{N} \left[ \sum_{i=1}^{N} (X_i(t) - \hat{\mu}_N(t)) (X_i(s) - \hat{\mu}_N(s)) \right]$, which is commonly used as estimator of covariance kernel, see for example, Berkes et al. (2009) and Aue et al. (2009). With the newly proposed estimator of the covariance kernel we obtain the most important finding of this paper which is narrated in the following theorem.

**Theorem 3.1.** Defining $c_u(t, s) := c(t, s) + \theta (1 - \theta) \Delta(t) \Delta(s) f_\theta(u)$, under assumption A4,

$$\int \int [\hat{c}_u(t, s) - c_u(t, s)]^2 dt ds \overset{P}{\to} 0, \text{ as } N \uparrow \infty,$$

where,

$$f_\theta(u) = \frac{\max \{u, \theta\} - \min \{u, \theta\}}{\max \{u, \theta\} (1 - \min \{u, \theta\})} \in [0, 1]$$

with $\theta \in (0, 1), u \in (0, 1]$ and $\Delta(t) = \mu_1(t) - \mu_2(t)$.

Proof: The proof of the theorem is provided in the Appendix.  \hfill \square

**Corollary 3.2.** If null hypothesis is true then $\hat{c}_u(t, s) \overset{P}{\to} c(t, s)$ for all $u \in (0, 1]$.

Some more interesting observations, which show the greater applicability of Theorem 3.1, are immediate from it.
Remark 3.3. It can be easily checked that $c_u(t, s)$ is a positive definite, symmetric satisfying
\[ \int \int c_u^2(t, s) \, dt \, ds < \infty, \]
and hence is a covariance kernel.

Remark 3.4. If $u = 1$, that is, if commonly used estimator of $c(t, s)$ is used then it is readily observable that, under alternative, $\hat{c}_1(t, s) \overset{P}{\rightarrow} c(t, s) + \theta(1 - \theta)\Delta(t)\Delta(s) = \tilde{c}(t, s)$, say, which is also proved by Berkes et al. (2009). We note here that whenever $H_0$ is false, $\hat{c}_1(t, s)$ has a constant bias $\theta(1 - \theta)\Delta(t)\Delta(s)$. Therefore, for any $u \in (0, 1)$, the asymptotic bias of the estimator $\hat{c}_u(t, s)$ is less than that of $\hat{c}_1(t, s)$ under alternative hypothesis.

Remark 3.5. If $u = \theta$, that is, when the data is partitioned in true position, then $\hat{c}_\theta(t, s) \overset{P}{\rightarrow} c(t, s)$ and in that case asymptotic bias of $\hat{c}_\theta(t, s)$ is zero whereas asymptotic bias of $\hat{c}_1(t, s)$ remains $\theta(1 - \theta)\Delta(t)\Delta(s)$.

A few more notations and definitions are needed to be introduced here to state the further results

Definition 3.6. The orthonormal functions $\omega^l_u(t)$ in $L^2(\tau)$ corresponding to real scalars $\gamma^l_u$ are defined as orthonormal eigenfunctions corresponding to eigenvalues $\gamma^l_u$ of the covariance operator $C_u(.)$ from $L^2(\tau)$ to $L^2(\tau)$, defined as $C_u(x)(t) = \int c_u(t, s)x(s) \, ds$, satisfying the relation
\[ \int c_u(t, s)\omega^l_u(s) \, ds = \gamma^l_u \omega^l_u(t) \]  

(3.4)

Definition 3.7. The estimates of the eigenvalues $\gamma^l_u$ and $\omega^l_u$ are denoted as $\hat{\gamma}^l_u$ and $\hat{\omega}^l_u$, satisfying the relation
\[ \int \hat{c}_u(t, s)\hat{\omega}^l_u(s) \, ds = \hat{\gamma}^l_u \hat{\omega}^l_u(t). \]  

(3.5)

With the above two definitions we have the following important observations can be noted

Corollary 3.8. Under the assumption A4, for every $1 \leq l \leq d$ and $u \in (0, 1)$, we have
\[ \hat{\gamma}^l_u \overset{P}{\rightarrow} \gamma^l_u \]  

(3.6)

\[ \int [\hat{\gamma}^l_u(t) - \hat{c}_u(t)\omega^l_u(t)]^2 \, dt \overset{P}{\rightarrow} 0, \]  

(3.7)

where $\hat{c}_u = \text{sgn}(\omega^l_u, \hat{\omega}^l_u)$.
Proof: The proof follows from the Theorem \ref{thm1} and lemmas 4.2 and 4.3 of \cite{?}.

\textbf{Remark 3.9.} Under \(H_0\), for all \(1 \leq l \leq d\) and \(u \in (0, 1]\), \(\hat{\lambda}_u \xrightarrow{P} \lambda^l\) and \(\hat{v}_u \rightarrow v^l\) in probability, in \(L^2(\tau)\). Moreover, under alternative hypothesis, if \(u = \theta\) then for all \(1 \leq l \leq d\), \(\hat{\lambda}_\theta \xrightarrow{P} \lambda^l\) and \(\hat{v}_\theta \rightarrow v^l\) in probability, in \(L^2(\tau)\). It, in fact, can be easily seen that

\[
\sup_{0 < u \leq 1} \int \left[ \hat{v}_u(t) - v^l_u(t) \right]^2 dt \xrightarrow{P} 0,
\]

In the direction of the eigenfunctions \(\hat{v}_u\) corresponding to the largest \(d\) eigenvalues \(\hat{\lambda}_u\) the noncentral scores can be obtained as

\[
\hat{\eta}_i,l(u) = \int X_i(t) \hat{v}_u(t) dt, \quad i = 1, \ldots, N, \quad l = 1, \ldots d.
\]  

(3.8)

Utilizing the score functions, as defined above, we provide a statistic and its distributional convergence in the following theorem which will important to know to construct the test statistic and perform the asymptotic test. First we define the statistic based on the self normalized partial sums in \(d\) dimensions

\[
R_N(u) = \frac{1}{N} \sum_{l=1}^{d} \frac{1}{\hat{\lambda}_u} \left( \sum_{i=1}^{[Nu]} \hat{\eta}_i,l(u) - u \sum_{i=1}^{N} \hat{\eta}_i,l(u) \right)^2.
\]

(3.9)

Further denoting \(B_1(\cdot), \ldots, B_d(\cdot)\) be the standard independent Brownian bridges, the theorem is provided

\textbf{Theorem 3.10.} Let the assumptions \(A1\) to \(A3\) hold. Then with the proper embedding of Skorohod topology in \(D[0, 1]\), under \(H_0\)

\[
R_N(u) \xrightarrow{d} \sum_{l=1}^{d} B^2_l(u), \quad 0 \leq u \leq 1.
\]

(3.10)

Proof: Proof of the theorem is given in Appendix, \cite{?}

Finally we define the test statistic as follows:

\[
H_{N,d} := \frac{1}{N^2} \sum_{l=1}^{d} \sum_{[Nu]=1}^{N} \frac{1}{\hat{\lambda}_u} \left( \sum_{i=1}^{[Nu]} \hat{\eta}_i,l(u) - u \sum_{i=1}^{N} \hat{\eta}_i,l(u) \right)^2.
\]

(3.11)

Using the Theorem \ref{thm10} it is immediate to see that \(H_{N,d} \xrightarrow{d} \int_{\tau} \sum_{l=1}^{d} B^2_l(u) du\) under \(H_0\), because integral is a continuous functional and \(U(R_N(\cdot)) \xrightarrow{d} U\left( \sum_{l=1}^{d} B^2_l(\cdot) \right)\) for any continuous
functional $U : D[0, 1] \rightarrow \mathbb{R}$ (see [Berkes et al. (2009)] for further details). The distribution of the limiting random variable can be found in [Kiefer (1959)] and its $(1 - \alpha)$th quantile are given in Table 1 of [Berkes et al. (2009)]. We use this asymptotic critical values for performing the tests and $H_0$ is rejected at $100(1 - \alpha)$% confidence level if the observed value of $H_{N,d}$ is bigger than the tabulated $(1 - \alpha)$th quantile $K_d(\alpha)$ in [Berkes et al. (2009)].

Now we show that the proposed test is consistent under the alternative hypothesis. Basically we show here that $H_{N,d} \overset{P}{\to} \infty$ under the hypothesis of single change point. The following theorem assures the claim.

**Theorem 3.11.** Under the assumption A4,

$$
\frac{1}{N}H_{N,d} \overset{P}{\to} \sum_{l=1}^{d} \int_{0}^{1} \frac{g_l^2(u)}{\gamma_l u} du,
$$

where $g_l(u) = \min\{\theta, u\} (1 - \max\{\theta, u\}) \int_{t}^{\tau} \Delta(t)\omega'_l(t)dt$.

Proof: The proof follows from Theorem 3.1 and the following lemma. □

**Lemma 3.12.** Under the assumption A4, $\sup_{0 \leq u \leq 1} |N^{-1}R_N(u) - \sum_{l=1}^{d} \frac{g_l^2(u)}{\gamma_l u}| = o_P(1)$.

Proof: Proof of the lemma follows from the proof of the Theorem 2 of [Berkes et al. (2009)]. □

Clearly from Theorem 3.11, if $\int_{0}^{1} \frac{g_l^2(u)}{\gamma_l u} du > 0$ for some $1 \leq l \leq d$, then $H_{N,d} \overset{P}{\to} \infty$.

Similar to [Berkes et al. (2009)], the change point $\theta$ is estimated by finding the value of $u$ which maximizes the function $R_N(u)$. For uniqueness we define the estimator formally as

$$
\hat{\theta}_N = \inf\{u' : R_N(u') = \sup_{0 \leq u \leq 1} R_N(u)\}. 
$$

(3.12)

It can be easily shown that (using lemma 3.12), under the assumption A4, $\hat{\theta}_N \overset{P}{\to} \theta$ provided $< \Delta, \omega'_u > \neq 0$ for all $u \in (0, 1]$ (see for example the proposition 1 and its proof of [Berkes et al. (2009)]).

4. Simulation studies

In this section we report a summary of the extensive simulation studies that we have conducted for moderate and large sample sizes. As proposed in Section 3, we reject the null hypothesis when the observed value of $H_{N,d}$ exceeds the corresponding critical value.
The critical values that are available in \cite{Berkes2009} (Table 1). Without loss of generality initial mean function is considered to be zero. For the first set of simulation studies the samples are generated from the standard Brownian motion (BM) over the interval $[0, 1]$ and a drift of amount $t$ and $\sin(t)$ are considered after the presumed locations of change point. The same is done for the standard Brownian bridge over $[0, 1]$ and the mean shift after the change point is considered to be a quadratic function $0.8t(1 - t)$. To generate a sample from each of such Gaussian processes 1000 equidistant grid points are used. 750 B-spline basis functions are used to convert the grid data to functional data and first $3(= d)$ eigenfunctions are used to execute the testing procedures. For a pre-decided sample size and a specific change point the entire process is replicated 10000 times to assess the power of the test. The considered sample sizes ($N$) are 50, 100, 150, 200, 300, 500. For any particular sample size different possible locations of change points ($k^*$) are chosen, to cover a wide range, which are summarized in Table 1 and Figure 1, Figure 2. For all practical purposes, we use the complete data together for computing the estimated covariance kernel when $[Nu] = 1$ or $[Nu] = N - 1$, otherwise as proposed in equation (3.3).

4.1. Small sample bias correction: For small sample size (less than or equals to 100, say) we observe some fluctuations in the empirical size of the proposed test based on $H_{N,d}$. To overcome this instability we propose a bias correction which helps us to get empirical size reasonably close to 0.05. Under the null, it is easy to observe that

$$E[\hat{c}_u(t, s)] = \left(1 - \frac{2}{N}\right) c(t, s).$$

So we suggest to multiply the correction factor with $(1 - 2/N)^{-1}$ with $\hat{c}_u(t, s)$ to obtain the satisfactory results. Indeed for the large sample the effect of the correction factor vanishes automatically and it hardly matters whether we use it or not.

4.2. Simulation findings: In all of the cases we find that the power curves for the proposed test based on $H_{N,d}$ strictly dominates that of the $S_{N,d}$ proposed by \cite{Berkes2009}. For large sample (200 and above, say) the two power curves get very close to each other. But for small sample we observe a remarkable gap between these two. In particular, we provide the details of power for $N = 100$ and $d = 3$ at different point of change points starting from 15 to 85 for Brownian motion and Brownian bridge in Table 1. We add two different functions, namely $t$, $\sin t$ with the mean of Brownian motion and add $0.8t(1 - t)$ with the mean of
standard Brownian bridge. In all of the above cases it is found that the proposed method has more power than that of the method by Berkes et al. (2009) for all different locations of change points. The Figure 1 and the Figure 2 show the powers of two methods for sample size 50 (= N) at different point of changes, where the data have been simulated from standard Brownian motion and two different functions, t and sin t are added separately with its mean at different locations of change for illustration purposes. It can be clearly observed that if sample size is small then our method is outperforming the method of Berkes et al. (2009) with much larger difference. We also have done simulations with different sample sizes and varieties of functions, e. g. $t^2$, $\sqrt{t}$, $\exp(t)$, $\cos(t)$ etc, being added to the mean of Brownian motion and Brownian bridge, and in all cases we have found that our method has a better power than that of existing method. This finding is quite intuitive because both test are asymptotic tests (both converging to the same asymptotic distribution) and the proposed one always has higher power than that of Berkes et al. (2009), mainly because the bias in the newly proposed estimate of covariance kernel under alternative is smaller than that in the usual estimate of covariance kernel used elsewhere. This satisfies the desirable quality of a better asymptotic test. We also observe quite good performance of the test statistic when the location of change point is $\leq N/4$ and $\geq 3N/4$.

5. Real Data analysis

The findings of real data analysis to show the performance of proposed test is demonstrated in this section. Two temperature data have been analyzed. One data consists of average daily temperatures of central England for 228 years, from 1780 to 2007. The data has been taken from the website of British Atmospheric Data Centre. The second data, taken from Carbon Dioxide Information Analysis Center, consists of monthly global average anomaly of the temperatures from 1850 to 2012. Thus, these two data sets can be viewed as 228 curves with 365 measurements on each curve and 163 curves with 12 measurements on each curve, respectively. These two data sets are converted to functional data using 12 B-spline basis functions and 8 B-spline basis functions, respectively. Now we discuss the performance of the test statistics on these two temperature data sets individually.

To use the proposed test statistic for temperature data of the central England we use first 8 ($= d$) eigenfunctions explaining about 85% of the total variability. Given the test indicates a change, the change point is estimated by calculating $\hat{\theta}_N$ as described in the Lemma (3.12).
Thereafter dividing the data set into two parts the procedure is repeated for each part until the test fails to reject the null hypothesis. The outcome of our method on this data has been provided in Table 2. It can be seen that the change points detected by our method and by the method of Berkes et al. (2009) are very much adjacent. Both of the methods have detected 1850 and 1926 as possible change points. In case of other years of change point it is observed that the timings are very close, for example our method has detected a change in 1810 whereas Berkes et al. (2009) has detected a change in 1808 and in the recent years our method has detected a change in 1989 and Berkes et al. (2009) has detected 1993 as possible change point. Overall, it is important to note that both these methods have detected four change points in the given data. Table 2 also shows the p-values corresponding to the observed value of the statistic for both of the methods. From the p-values it is noted that the p-values of proposed test are much more smaller than the p-values of existing method showing the greater power of our test. The mean functions for each partitioned data sets are provided in the Figure 3. The picture clearly shows that there is an upward trend in the structure of the mean function from one period to another.

For the monthly average anomaly of the global temperature data of 163 years, first 3 ($d$) eigenfunctions are used which explains about 96% variability of the total variation. We apply the same procedure as as done in the case of the previous data set to detect the changes. Table 3 shows the outcomes of the test. The functional data representation of the complete data and segment wise mean functions are shown in Figure 4 which reflects the prominent changes around the mentioned period of year. From the analysis of the second the data set we clearly observe that the global temperature is changing (more specifically increasing) significantly over the period of time.

6. DISCUSSIONS AND CONCLUSIONS

In this paper we have proposed a new test for testing the existence of a change point in a given sequence of independent functional data. It is shown that the null distribution of proposed test is asymptotically pivotal. We have proven that under the null hypothesis the distribution of the test statistics is a functional of the sum of squares of Brownian bridges. Moreover, it has been established that under alternative hypothesis of single change point the power of the proposed test goes to unity when sample sizes increases to infinity. While developing the test statistic we have proposed an alternative estimator of the covariance
kernel, which is not only a consistent estimator of the true covariance kernel under the null hypothesis but also it has lesser bias than the existing usual estimate of covariance kernel under the alternative hypothesis. In fact it is successfully shown that even under the alternative hypothesis, if the data is divided at the true point of change then our estimate has zero asymptotic bias whereas the existing estimate of covariance kernel mostly used in change point literature in functional data has a constant asymptotic bias. Because of the fact that our used estimate of covariance kernel has a smaller bias than the existing one under any circumstances, we are able to show that our test has greater power than the existing one for testing the presence of change point in a given sequence of functional data. The extensive simulation studies support such a claim also. Specially when the data size is not very big then our method outperforms the existing one with a great margin.

We have used our method in two real data to see the performance of our test in practice. One of these data is central England temperature which is also used in Berkes et al. (2009), and the other one is the global temperature data. In case of first data, it is seen that our method and the method of Berkes et al. (2009) both, have pointed four changes in the data sequence. Two time points have exactly matched for two methods, namely 1850 and 1926. For two other change points two methods differ marginally. Berkes et al. (2009) has detected 1808 as possible change point whereas our method detected 1810 as possible change point. For the other one Berkes et al. (2009) detected 1993 as a possible change point and our method indicated 1989 as a possible change point. We have plotted the mean function for each of the different segments which clearly shows an upward trend in the mean temperature over the said periods. The mean curves of different time segments are very similar to that of Berkes et al. (2009) which make sure the little observed difference in change points among two methods in this particular real data are not major. For the second data, which is global monthly temperature data from 1850 to 2012, is analyzed based on our method. It is found that there exists three change points around 1933, 1986 and 1996. The analysis of global temperature in terms of finding change points will help the scientists working on the global temperature. It clearly shows that in last three decades the temperature has increased significantly over the past.

To conclude we evince that the proposed method has asymptotic null pivotal distribution with greater power than the existing method for testing the presence of change in a sequence of functional data and hence can be used in practice with more confidence.
CHANGE IN MEAN OF FUNCTIONAL DATA

ACKNOWLEDGMENT

The authors are thankful to British Atmospheric Data Centre and carbon dioxide information analysis center for real data. The Daily Central England Temperature data has been taken from [NCAS British Atmospheric Data Centre (2007)] and monthly global average anomaly of temperatures is taken from [Jones et al. (2013)].

7. Appendix

Proof of the Theorem 3.1 Define \( \hat{\mu}_k(t) = \frac{1}{k} \sum_{i=1}^{k} X_i(t) \) and \( \bar{\mu}_k(t) = \frac{1}{N - k} \sum_{i=k+1}^{N} X_i(t) \) for some \( k = \lfloor Nu \rfloor \) and \( k^* = \lfloor N\theta \rfloor \) to express the estimated covariance kernel as

\[
\hat{c}_u(t, s) = \frac{1}{N} \left[ \sum_{i=1}^{k} \{X_i(t) - \hat{\mu}_k(t)\}\{X_i(s) - \hat{\mu}_k(s)\} + \sum_{i=k+1}^{N} \{X_i(t) - \bar{\mu}_k(t)\}\{X_i(s) - \bar{\mu}_k(s)\} \right]
\]

It immediately gives

\[
\hat{c}_u(t, s) = \frac{1}{N} \sum_{i=1}^{N} \{X_i(t) - \hat{\mu}_N(t)\}\{X_i(s) - \hat{\mu}_N(s)\} - \frac{k}{N}\{\hat{\mu}_k(t) - \hat{\mu}_N(t)\}\{\hat{\mu}_k(s) - \hat{\mu}_N(s)\}
\]

\[
- \frac{k}{N}\{\bar{\mu}_k(t) - \hat{\mu}_N(t)\}\{\bar{\mu}_k(s) - \hat{\mu}_N(s)\}
\]

For \( k \leq k^* \), note that

\[
\hat{\mu}_k(t) = \widehat{Y}_k(t) + \mu_1(t) \text{ where, } \widehat{Y}_k(t) = \frac{1}{k} \sum_{i=1}^{k} Y_i(t),
\]

\[
\bar{\mu}_k(t) = \tilde{Y}_k(t) + \mu_2(t) + \left( \frac{k^* - k}{N - k} \right) \Delta(t) \text{ where, } \tilde{Y}_k(t) = \frac{1}{k} \sum_{i=k+1}^{N} Y_i(t),
\]

and

\[
\hat{\mu}_N(t) = \widehat{Y}_N(t) + \left( \frac{k^*}{N} \right) \mu_1(t) + \left( \frac{N - k^*}{N} \right) \mu_2(t)
\]

Now observe that

\[
\hat{\mu}_k(t) - \hat{\mu}_N(t) = \widehat{Y}_k(t) - \widehat{Y}_N(t) + \left( 1 - \frac{k^*}{N} \right) \Delta(t)
\]

and

\[
\bar{\mu}_k(t) - \bar{\mu}_N(t) = \tilde{Y}_k(t) - \tilde{Y}_N(t) - \frac{k(N - k^*)}{(N - k)N} \Delta(t)
\]
to get the following deductions,

\[ \hat{c}_u(t, s) = \frac{1}{N} \sum_{i=1}^{N} \{ X_i(t) - \hat{\mu}_N(t) \} \{ X_i(s) - \hat{\mu}_N(s) \} - \Delta(t) \Delta(s) \left( \frac{N - k^*}{N} \right)^2 \left( \frac{k}{N - k} \right) \]

\[ - \frac{k}{N} \{ \hat{Y}_k(t) - \hat{Y}_N(t) \} \{ \hat{Y}_k(s) - \hat{Y}_N(s) \} - \left( 1 - \frac{k}{N} \right) \{ \hat{Y}_k(t) - \hat{Y}_N(t) \} \{ \hat{Y}_k(s) - \hat{Y}_N(s) \} \]

\[ - \frac{k}{N} \left( 1 - \frac{k^*}{N} \right) \left[ \{ \hat{Y}_k(t) - \hat{Y}_k(t) \} \Delta(s) + \{ \hat{Y}_k(s) - \hat{Y}_k(s) \} \Delta(t) \right] \]

Again,

\[ \hat{c}_1(t, s) = \frac{1}{N} \sum_{i=1}^{N} \{ Y_i(t) - \hat{\mu}_N(t) \} \{ Y_i(s) - \hat{\mu}_N(s) \} \]

\[ = \frac{1}{N} \sum_{i=1}^{N} \{ Y_i(t) - \hat{Y}_N(t) \} \{ Y_i(s) - \hat{Y}_N(s) \} + \frac{k^*}{N} \left( 1 - \frac{k^*}{N} \right) \Delta(t) \Delta(s) \]

\[ + \frac{k^*}{N} \left( 1 - \frac{k^*}{N} \right) \left[ \{ \hat{Y}_k(t) - \hat{Y}_k(t) \} \Delta(s) + \{ \hat{Y}_k(s) - \hat{Y}_k(s) \} \Delta(t) \right] \]

gives,

\[ \hat{c}_u(t, s) = \frac{1}{N} \sum_{i=1}^{N} \{ Y_i(t) - \hat{Y}_N(t) \} \{ Y_i(s) - \hat{Y}_N(s) \} + \frac{k^*}{N} \left( 1 - \frac{k^*}{N} \right) \left[ 1 - \frac{(N - k^*)k^*}{(N - k)k^*} \right] \Delta(t) \Delta(s) \]

\[ + \left( 1 - \frac{k^*}{N} \right) \Delta(s) \left[ \frac{k^*}{N} \{ \hat{Y}_k^*(t) - \hat{Y}_k^*(t) \} - \frac{k}{N} \{ \hat{Y}_k(t) - \hat{Y}_k(t) \} \right] \]

\[ + \left( 1 - \frac{k^*}{N} \right) \Delta(t) \left[ \frac{k^*}{N} \{ \hat{Y}_k^*(s) - \hat{Y}_k^*(s) \} - \frac{k}{N} \{ \hat{Y}_k(s) - \hat{Y}_k(s) \} \right] \]

\[ - \frac{k}{N} \left( 1 - \frac{k}{N} \right) \{ \hat{Y}_k(t) - \hat{Y}_k(t) \} \{ \hat{Y}_k(s) - \hat{Y}_k(s) \} \]

\[ = \frac{1}{N} \sum_{i=1}^{N} \{ Y_i(t) - \hat{Y}_N(t) \} \{ Y_i(s) - \hat{Y}_N(s) \} + \theta(1 - \theta) \Delta(t) \Delta(s) f_\theta(u) \]

\[ + r_1(t, s) + r_2(t, s) + r_3(t, s), \text{ say} \]

(7.1)

Using the law of large numbers for independent, identically distributed Hilbert-space-valued random variables (see for example theorem 2.4 of ?), we obtain

\[ \int_\tau \int_\tau r_1^2(t, s) dt ds \xrightarrow{P} 0 \text{ and } \int_\tau \int_\tau r_2^2(t, s) dt ds \xrightarrow{P} 0 \text{ as } N \to \infty. \]
At the same time using theorem 5.1 of Horváth and Kokoszka (2012) we get

\[ N^2 \int_\tau \int_\tau r_3^2(t,s)dt\,ds \overset{d}{\to} \left( \int_\tau \Gamma^2(t)dt \right)^2, \]

where \( \{\Gamma(t) : t \in \tau\} \) is a Gaussian process with \( E(\Gamma(t)) = 0 \) and \( E(\Gamma(t)\Gamma(s)) = c(t,s) \), which in turn implies that

\[ \int_\tau \int_\tau r_3^2(t,s)dt\,ds \overset{P}{\to} 0 \quad \text{as} \quad N \to \infty. \]

These help to conclude that

\[ \int_\tau \int_\tau [\hat{c}_u(t,s) - c_u(t,s)]^2dt\,ds \overset{P}{\to} 0 \quad \text{as} \quad N \to \infty. \]

The similar proof holds when \( k > k^* \). It is easy to see that under the null hypothesis

\[ \hat{c}_u(t,s) \overset{P}{\to} c(t,s) \quad \forall u \in (0,1] \quad \text{as} \quad N \to \infty \]

\[ \Box \]

Proof of Theorem 3.10:

The proof follows from the Theorem 3.1, Corollary 3.8 and the proof of Theorem 6.1 of Horváth and Kokoszka (2012).

References

J. Antoch, M. Husková, and Z. Prášková. Effect of dependence on statistics for determination of change. *Journal of Statistical Planning and Inference*, 60:291–310, 1997.

A. Aue, R. Gabrys, L. Horváth, and P. Kokoszka. Estimation of a change-point in the mean function of functional data. *J. Multivariate Anal.*, 100(10):2254–2269, 2009. ISSN 0047-259X. doi: 10.1016/j.jmva.2009.04.001. URL http://dx.doi.org/10.1016/j.jmva.2009.04.001

I. Berkes, R. Gabrys, L. Horváth, and P. Kokoszka. Detecting changes in the mean of functional observations. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 71(5):927–946, 2009. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2009.00713.x. URL http://dx.doi.org/10.1111/j.1467-9868.2009.00713.x

NCAS British Atmospheric Data Centre. Hadley centre for climate prediction and research. *Daily Central England Temperature series*, 2007. URL http://catalogue.ceda.ac.uk/uuid/b621ef77b07d3c8e116b5b31fd5eb92b

G. Cobb. The problem of the nile: Conditional solution to a change-point problem. *Biometrika*, 65:243–251, 1978.

R. Davis, D. Huang, and Y.-C. Yao. Testing for a change in the parameter values and order of an autoregressive model. *The Annals of Statistics*, 23:282–304, 1995.

F. Ferraty. *Recent Advances in Functional Data Analysis and Related Topics*. Springer-Verlag Berlin Heidelberg, 2011.

L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer Science & Business Media, 2012.

L. Horváth, P. Kokoszka, and J. Steinebach. Testing for changes in multivariate dependent observations with applications to temperature changes. *Journal of Multivariate Analysis*, 68:96–119, 1999.
C. Inclán and G. C. Tiao. Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Amer. Statist. Assoc.*, 89(427):913–923, 1994. ISSN 0162-1459. URL http://links.jstor.org/sici?sid=sici=0162-1459(199409)89:427<913:UOCSOS>2.0.CO;2-H&origin=MSN

J. Indritz. *Methods in analysis*. The Macmillan Co., New York; Collier-Macmillan Ltd., London, 1963.

P. Jones, D. Parker, T. Osborn, and K. Briffa. Global and hemispheric temperature anomalies land and marine instrumental records. in *trends: A compendium of data on global change*. *Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.*, 2013. doi: 10.3334/CDIAC/cli.002. URL http://cdiac.ornl.gov/trends/temp/jonescru/jones.html

J. Kiefer. K-sample analogues of the kolmogorov-smirnov and cramér-v. mises tests. *The Annals of Mathematical Statistics*, pages 420–447, 1959.

C. Kirch, B. Mushal, and H. Ombao. Detection of changes in multivariate time series with applications to eeg data. *Journal of the American Statistical Association*, 2014.

P. Kokoszka and R. Leipus. Change-point estimation in arch models. *Bernoulli*, 6:513–539, 2000.

J. O. Ramsey and B. Silverman. *Functional Data Analysis*. New York: Springer, 2005.
### Table 1. Power comparison of two tests with test statistics $S_{N,d}$ and $H_{N,d}$ for different $k^*$

| $k^*$ | $S_{N,d}$ | $H_{N,d}$ | $S_{N,d}$ | $H_{N,d}$ | $S_{N,d}$ | $H_{N,d}$ |
|---|---|---|---|---|---|---|
| 0  | 4.6  | 5.5  | 4.6  | 5.5  | 4.6  | 5.0  |
| 15 | 36.3 | 39.9 | 30.9 | 35.4 | 11.2 | 12.9 |
| 20 | 57.3 | 62.0 | 44.6 | 49.5 | 15.3 | 16.5 |
| 25 | 72.0 | 75.6 | 61.2 | 64.7 | 19.5 | 21.3 |
| 35 | 92.9 | 94.2 | 80.1 | 83.4 | 28.0 | 31.8 |
| 50 | 94.9 | 95.8 | 88.0 | 90.1 | 34.7 | 37.4 |
| 65 | 91.0 | 92.9 | 81.5 | 83.7 | 31.1 | 33.9 |
| 75 | 74.3 | 78.1 | 59.0 | 64.4 | 21.9 | 23.8 |
| 80 | 58.8 | 64.3 | 46.1 | 50.2 | 13.7 | 16.1 |
| 85 | 36.4 | 40.1 | 27.8 | 33.0 | 12.9 | 14.1 |

* The values are reported from the tables provided by Berkes et. al (2009, Table 3).

### Table 2. Comparisons of the performance of $S_{N,d}$ and $H_{N,d}$ for UK temperature data

| Year Segment | Observed $S_{N,d}$ | Obtained P-value | Estimated Change point | Year Segment | Observed $H_{N,d}$ | Obtained P-value | Estimated Change point |
|---|---|---|---|---|---|---|---|
| 1780-2007 | 8.020593 | 0.00000 | 1926 | 1780-2007 | 9.820036 | 0.00000 | 1926 |
| 1780-1925 | 3.252796 | 0.00088 | 1808 | 1780-1926 | 3.764348 | 0.00011 | 1850 |
| 1808-1925 | 2.351132 | 0.02322 | 1850 | 1780-1850 | 2.403308 | 0.01900 | 1810 |
| 1926-2007 | 2.311151 | 0.02643 | 1993 | 1927-2007 | 2.649414 | 0.00797 | 1989 |

* The values are reported from the tables provided by Berkes et. al (2009, Table 4).

### Table 3. Change points for average anomaly global temperature data

| Year Segment | Observed $H_{N,d}$ | Obtained P-value | Estimated Change point |
|---|---|---|---|
| 1850-2012 | 23.63304 | 0.00000 | 1933 |
| 1934-2012 | 13.46585 | 0.00000 | 1986 |
| 1987-2012 | 4.34103 | 0.00000 | 1996 |
Figure 1. Power comparison of $H_{n,d}$ and $S_{n,d}$ for $N = 50$ and $d = 3$ with $\Delta(t) = t$.

Figure 2. Power comparison of $H_{n,d}$ and $S_{n,d}$ for $N = 50$ and $d = 3$ with $\Delta(t) = \sin(t)$. 
Figure 3. Segment wise mean functions of central England temperature data

Figure 4. Segment wise mean functions of average anomaly of global temperature data