CHANGE-POINT ESTIMATION WITH SAMPLE FOURIER COEFFICIENTS

JAEHEE KIM [1]

ABSTRACT. In this paper we propose a change-point estimator with left and right regressions using the sample Fourier coefficients on the orthonormal bases. The asymptotic properties of the proposed change-point estimator are established. The limiting distribution and the consistency of the estimator are derived.

KEYWORDS: Change-point model, Sample Fourier coefficients, stochastic process,

1. Introduction

Consider a sequence $\{y_i\}$ of observations with finite fourth moment. The model considered is

$$Y_i = f(x_i) + \epsilon_i \tag{1}$$

where the errors are independent $N(0,1)$ and $x_i = i/n$, $i = 1, 2, \ldots, n$. Also f is right continuous and left continuous except at an unknown change-point or discontinuity point $\tau \in (0,1)$. For definiteness, we suppose that τ is an event time i.e. $\tau = x_i$ for some i.

Left and right local regressions with the sample Fourier coefficients are used to estimate the left and right limit of event times. Hinkley(1970) proposed the maximum likelihood estimator of the change-point estimator with the normal errors in the mean shift model. McDonald and Owen(1986) and Hall and Titterington(1992) investigated change-point estimation based on three smoothed estimates of the function. Loader(1996) proposed an estimate of the location of the discontinuity based on one-sided nonparametric

1 Department of Statistics, Duksung Women's University, Seoul 132-714, S.Korea
This work was supported by grant No. R04-2001-000-00135 from the Korea Science & Engineering Foundation .

regression estimates of the mean function using the kernel function. Kim and Hart(1998) proposed a test for change using the sample Fourier coefficients with the dependent data.

2. Left and Right Function Estimation using the Sample Fourier Coefficients

We assume that the underlying function f has the following Fourier series representation with the cosine system $\{\cos \pi jx\}$ and $x \in (0,1)$:

$$f(x) = \phi_0 + 2\sum_{j=1}^{\infty} \phi_j \cos \pi jx \tag{2}$$

where $\phi_j = \int_0^1 f(x) \cos \pi jx dx$, $j = 0, 1, \dots$, which are the Fourier coefficients of f. The sample Fourier coefficients using a cosine system are defined by

$$\hat{\phi}_j = \frac{1}{n} \sum_{i=1}^{n} y_i \cos \pi j x_i, \qquad j = 0, 1, \cdots, n-1 \tag{3}$$

where $x_i = i/n$. Consider the following Fourier series estimator of the left and the right function estimation with the $(K+1)$ sample Fourier coefficients and the bandwidth h. With the data in bandwidth at each point, left and right estimators can be obtained as follows:

$$\begin{aligned} \hat{f}_{K+}(x) &= \hat{\phi}_0 + 2\sum_{j=1}^{K} \hat{\phi}_j \cos \pi j t_1 & (4) \\ \hat{f}_{K-}(x) &= \hat{\psi}_0 + 2\sum_{j=1}^{K} \hat{\psi}_j \cos \pi j t_{nh} & (5) \end{aligned}$$

with $t_i = (i-0.5)/nh$, and $0 \le t_i \le 1$ in $x \in (h, 1-h)$ that is, $nh \le m = nx \le n - nh$, where

$$\hat{\phi}_0 = \frac{1}{nh}\sum_{i=1}^{nh} y_{m+i}$$
$$\hat{\phi}_j = \frac{1}{nh}\sum_{i=1}^{nh} y_{m+i}\cos \pi j t_i$$
$$\hat{\psi}_0 = \frac{1}{nh}\sum_{i=1}^{nh} y_{m-nh+i}$$
$$\hat{\psi}_j = \frac{1}{nh}\sum_{i=1}^{nh} y_{m-nh+i}\cos \pi j t_i.$$

Define the size of change is measured by

$$\Delta = \Delta_\tau = f(\tau+) - f(\tau-).$$

The estimator of Δ_x at the point x can be defined as

$$\hat{\Delta}_x = \hat{f}_{K+}(x) - \hat{f}_{K-}(x), \qquad h \leq x \leq 1-h. \tag{6}$$

The estimate $\hat{\tau}$ of τ is the value of x which maximizes $\hat{\Delta}_x^2$ over $h \leq x \leq 1-h$. That is,

$$\hat{\tau} = \arg\max_{h \leq x \leq 1-h} \hat{\Delta}_x^2. \tag{7}$$

One could also consider the maximizer of $|\hat{\Delta}_x|$ or the maximizer of $\hat{\Delta}_x$ if $\Delta > 0$. The choice of the number of the sample Fourier coefficients K and the bandwidth h depend on the selection criteria. We do not discuss the selection criteria in this paper.

Theorem 2.1. Let $\ldots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots$ be independent $N(0,1)$ random variables. Then we have the following limiting process as

$$\lim_{n\to\infty} P\left(n(\hat{\tau} - \tau) = s\right) = P\left(S_\Delta = s\right), \tag{8}$$

where S_Δ is the location of the maximum of the process

$$Z_i = \begin{cases} -i\Delta^2 + 2\Delta\sum_{u=1}^{i}\epsilon_u + \Delta\sum_{u=1}^{i}(\epsilon_{i+u} + \epsilon_{-u}), & i > 0 \\ 0, & i = 0 \\ -|i|\Delta^2 + 2\Delta\sum_{u=i}^{-1}\epsilon_u + \Delta\sum_{u=1}^{i}(\epsilon_u + \epsilon_{i-u}), & i < 0. \end{cases} \tag{9}$$

The limit distribution of the proposed change-point estimator has the same form of the asymptotic distribution of the change-point estimator with nonparametric regression considered in Loader(1996). The local regression estimates will require a larger n and nh for the estimation and for the asymptotics to be applicable.

Theorem 2.2. $\hat{\tau}$ is a consistent estimator of τ, i.e. for $i_0 = o(nh)$ and $i_0 > 0$

$$P\left(|\hat{\tau} - \tau| \geq \frac{i_0}{n}\right) \xrightarrow{\text{P}} 0. \tag{10}$$

3. Proofs

proof of Theorem 2.1. We consider $nh(\hat{\Delta}_{\tau+i/n} - \hat{\Delta}_\tau)$ in one level change model in which the amount of change is Δ.

Let $m = n\tau$, consider the mean level change model

$$y_i = \begin{cases} \mu + \epsilon_i, & i = 1, \ldots, \tau \\ \mu + \Delta + \epsilon_i, & i = \tau + 1, \ldots, n \end{cases}$$

in which the difference about the change-point is Δ.

The difference of the right estimators is

$$\hat{f}_{K+}(\tau + i/n) - \hat{f}_{K+}(\tau) = \hat{\phi}_0 - \hat{\phi}_{0\tau} + 2\sum_{j=1}^{K}(\hat{\phi}_j - \hat{\phi}_{j\tau}) \cos \pi j t_1.$$

$$\begin{aligned} nh(\hat{f}_{K+}(\tau + i/n) \quad &- \quad \hat{f}_{K+}(\tau)) = \sum_{u=nh-i+1}^{nh} \epsilon_{m+i+u}(1 + 2\sum_{j=1}^{K} \cos \pi j t_u \cos \pi j t_1) \\ &- \quad \sum_{u=1}^{i} \epsilon_{m+u}(1 + 2\sum_{j=1}^{K} \cos \pi j t_u \cos \pi j t_1) \end{aligned}$$

And we get the similar result for the difference of the left estimators. Therefore

$$\begin{aligned} nh(\hat{\Delta}_{\tau+i/n} \quad &- \quad \hat{\Delta}_\tau) = -i\Delta(1 + 2K^*) \\ &+ \quad (1 + 2K^*)\left[\sum_{u=nh-i+1}^{nh} \epsilon_{m+i+u} - 2\sum_{u=1}^{i} \epsilon_{m+u} + \sum_{u=1}^{i} \epsilon_{m-nh+u}\right] + o(1) \end{aligned}$$

where $K^* = O(K)$ since for $u \leq i$, $t^* \in (t_1, t_u)$,

$$\cos \pi j t_u - \cos \pi j t_1 = (\pi j i/nh) \sin \pi j t^* = o(1)$$

and for $u \geq nh - i + 1$, $t^* \in (t_u, t_{nh})$,

$$\cos \pi j t_u - \cos \pi j t_{nh} = (\pi j i/nh) \sin \pi j t^*(i/nh) = o(1).$$

Since $\hat{\Delta}_{\tau+i/n} + \hat{\Delta}_\tau \to 2\Delta$ for $|i - m| \leq i_0$

We achieve

$$\frac{nh(\hat{\Delta}^2_{\tau-i/n} - \hat{\Delta}^2_\tau)}{2(1+2K^*)} = -i\Delta^2 - \Delta \sum_{u=nh-i+1}^{nh} \epsilon_{m+i+u} + \Delta \sum_{u=1}^{i} \epsilon_{m+u} - \Delta \sum_{u=1}^{i} \epsilon_{m-nh+u} + o(1).$$

proof of Theorem 2.2. We assume $\Delta > 0$; the case $\Delta < 0$ is similar.

$$\begin{aligned} P\left(|\hat{\tau} - \tau| \geq \frac{i_0}{n}\right) &= 1 - P\left(\tau - \frac{i_o}{n} \leq \hat{\tau} \leq \tau + \frac{i_0}{n}\right) \\ &= P\left(\tau + \frac{i_0}{n} \leq \hat{\tau} \leq \tau + h\right) + P\left(\tau - h \leq \hat{\tau} \leq \tau - \frac{i_0}{n}\right). \end{aligned}$$

We only consider that

$$P\left(\tau + \frac{i_0}{n} \leq \hat{\tau} \leq \tau + h\right) \leq \sum_{i=i_0}^{nh} P\left(\hat{\Delta}^2_{\tau+\frac{i}{n}} \geq \hat{\Delta}^2_\tau\right)$$

$$\leq \sum_{i=i_0}^{nh} P\left(\hat{\Delta}_{\tau+\frac{i}{n}} - \hat{\Delta}_\tau > 0\right) + \sum_{i=i_0}^{nh} P\left(\hat{\Delta}_{\tau+\frac{i}{n}} + \hat{\Delta}_\tau < 0\right).$$

Then we have $P\left(\hat{\Delta}_{\tau+\frac{i}{n}} + \hat{\Delta}_\tau < 0\right) \to 0$ by the assumption and

$$\begin{aligned} \sum_{i=i_0}^{nh} P\left(\hat{\Delta}_{\tau+i/n} - \hat{\Delta}_\tau > 0\right) &= \sum_{i=i_0}^{nh} P\left(nh(\hat{\Delta}_{\tau+i/n} - \hat{\Delta}_\tau) > 0\right) \\ &\leq \sum_{i=i_0}^{nh} \frac{1}{C_2\sqrt{i}} \exp\left(-\frac{C_2^2 i}{2}\right) \\ &\to 0 \end{aligned}$$

using $\phi(a)/a \geq 1 - \Phi(a)$ for $a > 0$, as $nh \to \infty$ and $i_0/K^2 \to \infty$, where

$$\Delta^* = E\left[nh(\hat{\Delta}_{\tau+i/n} - \hat{\Delta}_\tau)\right] = O(Ki)$$

and

$$M = Var\left[nh(\hat{\Delta}_{\tau+i/n} - \hat{\Delta}_\tau)\right] = O(K^2 i).$$

Note that

$$\begin{aligned} \Delta^* &= E\left[\hat{\Delta}_{\tau+\frac{i}{n}} - \hat{\Delta}_\tau\right] \\ &= E\left[(\hat{\Delta}_{\tau+\frac{i}{n}} - \Delta_{\tau+\frac{i}{n}}) - (\hat{\Delta}_\tau - \Delta_\tau)\right] + \Delta_{\tau+\frac{i}{n}} - \Delta_\tau \xrightarrow{\text{p}} -\Delta_\tau \end{aligned}$$

since $\hat{\Delta}_{\tau+\frac{i}{n}} \xrightarrow{\text{p}} \Delta_{\tau+\frac{i}{n}} = 0$ at the continuity point.

4. Concluding remark

The objective of this research is to find a consistent change-point or discontinuity estimator. The left and right estimators with the sample Fourier coefficients were used for the proposed change-point estimator. The selection of the bandwidth and the truncated number of the Fourier series affects the estimation. The appropriate selection will increase the power of detection. Also we expect further development with wavelet function estimation in change-point estimation.

References

[1] R. Hall and D.M. Titterington, Edge-preserving and peak-preserving smoothing, *Technometrics* 34, (1992), 429-440.

[2] D. V. Hinkley, Inference about the change-point in a sequence of random variables, *Biometrika* 57, (1970), 1-17.

[3] J. H. Kim and J. D. Hart, Tests for change in a mean function when the data are dependent, *J. Time Series Anal.* 19, (1998), 399-424.

[4] C. R. Loader, Change point estimation using nonparametric regression, *Ann. Statist.* 24, (1996), 1667-1678.

[5] F. Lombard, Detecting change points by Fourier analysis, *Technometrics* 30 (1988), 305-310.

[6] J. A. McDonald and A. B. Owen, Smoothing with split linear fits, *Technometrics* 28, (1986), 195-208.

[7] G. P. Tolstov, *Fourier Series*, Dover, New York, 1962.