A Use of Kernel Density Estimates for Change Point Problems

Daehak Kim
Department of Statistical Information, Catholic University of Daegu
330, Kumrak 1, Hayang-up, Kyungsan, Kyungbuk 712-702, Korea
dhkim@cuth.cataegu.ac.kr

Kwangsik Oh
Department of Statistical Information, Catholic University of Daegu
330, Kumrak 1, Hayang-up, Kyungsan, Kyungbuk 712-702, Korea
ohkwang@cuth.cataegu.ac.kr

Changwan Kang
Dongeui University, Department of Computer Science and Statistics
24, Kaya-Dong, PusanJin-Gu, Pusan 614-714, Korea
cwkang@hyomin.dongeui.ac.kr

1. Introduction

Let $X_{1:j}, X_{2:j}, \cdots, X_{\tau:j}$ where $j = 1, \cdots, n_i, i = 1, \cdots, \tau$ be independent, continuous random samples from some distribution F and $X_{\tau+1:j}, \cdots, X_{n:j}$ where $j = 1, \cdots, n_i, i = \tau+1, \cdots, n$ be independent, continuous random samples from some distribution G. That means we have multiple observations for each point $i, i = 1, 2, \cdots, n$. Assume F and G are continuous and unknown. When the integer τ is known and $n_i = 1$ for all $i = 1, 2, \cdots, n$, this is nothing more than the standard two sample problem of testing $H_0 : F = G$ against $H_1 : F \neq G$. However the model where the integer τ is unknown, is referred to as a well known change point problem with at most one change. Sen and Srivastava(1975) studied detecting change points with a mean shift model. Hsu(1977) proposed a locally most powerful test for variance shift with a fixed mean.

Nonparametric kernel density estimation has many applications and is widely used in many areas of statictics. For details, see Silverman(1996). In particular, we consider the use of kernel density estimates in hypothesis testing $H_0 : F_1 = F_2 = \cdots = F_n$ against $H_1 : F_1 = F_2 = \cdots = F_{\tau-1} \neq F_\tau = \cdots = F_n$. David(1997) considered a two sample test based on the L_1 distance between kernel density estimates for F and G. Ferger(1994) studied nonparametric change-point tests of the Kolmogorov-Smirnov type. We want to propose a testing procedure for change point problems including all possible distribution changes based on nonparametric kernel density estimates. We not only consider a change in mean or variance shift, but also a change of distribution by expanding David's idea.

2. Testing procedure for change point

We use a kernel density estimate defined by

$$\hat{f}(x) = \sum K((x - Y_i)/h)/nh$$

for given random samples $Y_1, Y_2, \cdots, Y_n$ where $K(\cdot)$ is kernel function and h is called bandwdith or smoothing parameter which goes to 0 but nh goes to ∞ as n increases. We will obtain the suspected change point τ by maximizing the integral of absolute difference of two estimated kernel densities

$$\tau = \max_{1 \le k \le n} \int \mid \hat{f}_k - \hat{g}_{n-k} \mid$$

where $\hat{f}_k$ are density estimates based on the first k samples, and $\hat{g}_{n-k}$ are density estimates based on the other $n-k$ samples. We then apply a testing procedure of David for that suspected point τ based on the bootstrap distribution of

$$\int \mid \hat{f}^*_{\tau-1} - \hat{g}^*_{n-\tau+1} \mid$$

where $\hat{f}^*_{\tau-1}$ and $\hat{g}^*_{n-\tau+1}$ are density estimates based on bootstrap samples.

3. Simulations

To show the validity of proposed change point tests, a small sample size monte carlo simulation is stuided for various distributions.

REFERENCES

David, L.A.(1997) Hypothesis testing using an L_1 distance bootstrap. *American Statitsical Association.*,**51**, No. 2, 145-150

Efron, B. and Tibshirani, R.(1993). *An Introduction to the Bootstrap.* Newyork : Chapman and Hall

Ferger, D.(1994). *Nonparametric change-point tests of the Kolmogorov-Smirnov type.* IMS Lecture notes -Monograph Series, **23**

Hsu, D.A.(1977). Tests for variance shift at an unknown time point. *Appl. Statist.* **26.**, No. 3, 279-284

Sen, A. and Srivastava, M.S.(1975). On tests for detecting changes in mean. *Ann. of Satist.* **3.**,98-108

Silverman, B.W.(1986). *Density Estimation for statistics and data analysis.* London : Chapman and Hall.