The Korean Communications
in Statistics Vol. 12 No. 2, 2005
pp. 349-357

Change-point Estimation with Loess of Means[1]

Jaehee Kim[2]

Abstract

We suggest a functional technique with loess smoothing for estimating the change-point when there is one change-point in the mean model. The proposed change-point estimator is consistent. Simulation study shows a good performance of the proposed change-point estimator in comparison with other parametric or nonparametric change-point estimators.

Keywords : change-point model, loess, consistent

1. Introduction

Almost all classic statistical inference is based upon the assumption that there exists a fixed probabilistic mechanism of data generation. Unlike classic statistical inference, we consider about the statistical analysis of data about complex objects with more than one probabilistic mechanism of data generation. Also more than one data generation process is the most important characteristic of complex systems. Given a sequence of random variables, suppose that at some unknown point in the sequence the process governing their distribution changes abruptly, and consider the problem of the unknown change-point estimation.

The problem of concern is to detect changes of probabilistic characteristics of data from the whole observed sample. We are dealing with change-point problem in the mean change with only one change-point. Any problem of detection of changes in probabilistic characteristics can be reduced to the problem of the detection the change of the mathematical expectation for some diagnostic sequence formed from the initial sequence. Therefore it is possible to formulate the problem with comparing each two parts.

Let $X_1, X_2, \cdots, X_n$ be independent variables with

$$\begin{aligned} &X_1, X_2, \cdots, X_\tau \text{ identically distributed with cdf } F, \qquad (1.1)\\ &X_{\tau+1}, \cdots, X_n \text{ identically distributed with cdf } G. \end{aligned}$$

The unknown parameter τ is the change-point to be estimated.

[1] This research was supported by 2005 Duksung Women's University Research Fund.

[2] Associate Professor, Dept. of Statistics, Duksung Women's University, Ssangmun-Dong Tobong-Gu, Seoul, Korea
E-mail : jaehee@duksung.ac.kr

In section 2, some previous change-point estimation methods are explained to understand the change-point problem approach. Section 3 contains the proposed change-point estimation method with loess. Section 4 gives some results of comparison simulation studies. And section 4 presents concluding remarks.

2. Parametric Change-point Estimation

Hinkley(1970) used the maximum likelihood to estimate the change-point where F and G are from the same parametric family. The model is of the form

$$X_t = \begin{cases} \theta_0(t) + \epsilon_t & t = 1,\cdots,\tau \\ \theta_1(t) + \epsilon_t & t = \tau+1,\cdots,n \end{cases} \tag{2.1}$$

where $\{\epsilon_t\}$ is a sequence of uncorrelated error terms with zero mean, $\theta(t)$ is a continuous mean function and the change-point τ is unknown. A generalization (2.1) to $(p+1)$ submodels with p unknown change-points can be considered as an extensive model. Let $(X_1,\cdots,X_n)$ be a sequence of independent continuous random variables such that $\mathrm{X_i}$ has probability density function $f(x,\theta_0)$ $(i = 1,\cdots,\tau)$ and X_i has the probability density function $f(x,\theta_1)$ $(i = \tau+1,\cdots,n)$, where θ_0 *and* θ_1 are known $(\theta_0 \neq \theta_1)$ but τ is unknown. To obtain the maximum likelihood estimate $\hat{\tau}$, the log likelihood function $L(t)$ is considered where

$$L(t) = \sum_{i=1}^{t} \log f(x_i,\theta_0) + \sum_{i=t+1}^{n} \log f(x_i,\theta_1).$$

A more convenient form for $L(t)$ is obtained by defining the log likelihood increments

$$U_j = \log\ f(X_j,\theta_0) - \log\ f(X_j,\theta_1). \tag{2.2}$$

There the maximum likelihood estimate $\hat{\tau}$ is the value which maximizes the sequence of partial sums $\sum_{j=1}^{t} U_j$.

Hinkley(1972) generalized this method to the case where F and G may be arbitrary known distributions, or alternatively where a sensible discriminant function is known. The model may be written as

$$P(X_i \le x) = \begin{cases} F(x,\theta) & i = 1,\cdots,\lambda \\ G(x,\psi) & i = \lambda+1,\cdots,n \end{cases} \tag{2.3}$$

Each random variable X_i may be multidimensional, as may be θ *and* ψ. For fixed values of θ *and* ψ, the log likelihood function can be written as

$$L(\tau) = \sum_{j=1}^{\tau} l(X_j) + \sum_{j=1}^{n} c(X_j) \tag{2.4}$$

where

$$l(X_j) = log\{dF(X_j,\theta)/dG(X_j,\psi)\}, \qquad c(X_j) = log\ dG(X_j,\psi). \tag{2.5}$$

When θ and ψ are unknown, the log likelihood can be very inconvenient for actual data analysis so that a general class of discriminate functions $d(X)$ are considered to discriminate between $F(x,\theta)$ and $G(x,\psi)$, so that $D(\tau) = \sum_{j=1}^{\tau} d(X_j)$ corresponds to $L(\tau)$ in (2.4).

In the normal case with both θ_0 and θ_1 are unknown, the log-likelihood becomes

$$L(t) = -\frac{1}{2}\left\{\sum_{i=1}^{n}(X_i - \overline{X_n})^2 - t(n-t)(\overline{X_t} - \overline{X_t^*})^2/n\right\} \tag{2.6}$$

so that the change-point estimator is

$$T_{Hink} = argmax_{1 \leqq t < n} Z_t^2 \tag{2.7}$$

where

$$Z_t^2 = t(n-t)(\overline{X_t} - \overline{X_t^*})^2/n, \qquad t = 1,2,\cdots,n-1 \tag{2.8}$$

and

$$\overline{X_t} = \frac{S(t)}{t}, \qquad \overline{X_t^*} = \frac{S(n) - S(t)}{n-t}, \qquad t^* = n - t$$

and $S(t) = \sum_{i=1}^{t} X_i$, $t = 1,2,\cdots,n$.

Gombay and Horvath(1990) considered the test statistic for a change in the mean of independent random variables with $\overline{X_t}$, $\overline{X_t^*}$ and $\overline{X_n}$. Their test statistics are based on

$$Z_{g,t} = 2\left\{tg(\overline{X_t}) + t^* g(\overline{X_t^*}) - ng(\overline{X_n})\right\} \tag{2.9}$$

where g is a given function with the second derivative $g^{(2)} \neq 0$. The choice $g(t) = \frac{1}{2}t^2$ in (2.9) gives

$$Z_t^* = \frac{\{nS(t) - tS(n)\}^2}{nt(n-t)}, \qquad 1 < m_1 \leq t \leq m_2 < n. \tag{2.10}$$

Their change-point estimator is

$$T_{GH} = argmax_{1 \leqq t < n} Z_{g,t}^2. \tag{2.11}$$

Gombay and Horvath(1996) considered the maximum likelihood change-point estimator when the observations are from the exponential family and obtained the asymptotic distribution if there is a change in the parameters at an unknown time. They considered the parameter vectors and observation vectors.

After Hinkley's research nonparametric change-point estimators are developed. Here we briefly review the following nonparametric estimators for simulation later.

Schechtman(1982) considered two samples $(X_1, X_2, \cdots, X_j), (X_{j+1}, \cdots, X_n)$ and

$$U_{(t,n-t)} = \frac{1}{2}\left\{\sum_{i=1}^{t}\ \sum_{k=t+1}^{n} sgn(X_i - X_k) + t(n-t)\right\} \tag{2.12}$$

and standardized version of U

$$V_t = \frac{\dfrac{U_{(t,n-t)}}{t(n-t)} - 0.5}{\left[\dfrac{(n+1)}{12t(n-t)}\right]^{0.5}}, \quad t = 1,2,\cdots,n-1. \tag{2.13}$$

Schechtman(1982) suggested the change-point estimator as

$$T_{Sche} = argmax_{1 \leqq t < n} V_t \ . \tag{2.14}$$

Carlstein(1988) considered the change-point estimator which maximizes the distance between the two distributions. The pre-t empirical cdf ${}_t h(x)$ and the post-t empirical cdf $h_t(x)$ are defined respectively as follows:

$${}_t h(x) = \sum_{i=1}^{nt} I\{X_i \le x\}/nt, \text{ and } h_t(x) = \sum_{i=nt+1}^{n} I\{X_i \le x\}/n(1-t) \tag{2.15}$$

where $I(\ \cdot\ )$is the indicator function as

$$I(X \le a) = \begin{cases} 1, & X \le a \\ 0, & X > a. \end{cases}$$

Using ${}_t h(x)$ and $h_t(x)$, Carlstein(1988) considered three criterion function and suggested the following change-point estimators:

$$T_{C1} = argmax_{1 \leqq t < n} D_1(t) \text{ where } D_1(t) = t^{0.5}(1-t)^{0.5} n^{-1} \sum_{i=1}^{n} |\, h(x_i) - h_t(x_i)\,|.$$

$$T_{C2} = argmax_{1 \leqq t < n} D_2(t) \text{ where } D_2(t) = t^{0.5}(1-t)^{0.5}\left[\frac{1}{n}\sum_{i=1}^{n}(h(x_i) - h_t(x_i))^2\right]^{0.5}$$

$$T_{C3} = argmax_{1 \leqq t < n} D_3(t) \text{ where } D_3(t) = t^{0.5}(1-t)^{0.5} sup_{1 \le i \le n} |{}_t h(x_i) - h_t(x_i)|. \tag{2.16}$$

Carlstein(1988) estimators is known to perform well when F and G share the same mean, variance and skewness.

3. Change-point Estimation with the Loess of Means

As was seen in Section 2, the change-point estimator can be a function of $\overline{X_t}$, $\overline{X_t^*}$ and $\overline{X_n}$. The function should measure the divergence of the difference at each point t. We consider a loess smoothing function of $\overline{X_t}$ and $\overline{X_t^*}$ and compare their difference.

Local polynomial regression has been systematically studied by Stone(1977) and Cleveland(1979). Loess(local regression smoothing) is a modern popular local regression technique globally modelled by a polynomial and locally fitted with the kernel function.

A locally weighted polynomial regression at x is modelled as

$$\sum_{i=1}^{n}\{Y_i-\beta_0-\beta_1(t-x)\}^2K_h(t-x) \tag{3.1}$$

where $K(\cdot)$ denotes a kernel function and h is a bandwidth. $\widehat{\beta}_j$, $j=0,1$ is the minimizer of (3.1). The whole curve $s(\cdot)$ is obtained by running the above local polynomial regression with t varying in an appropriate estimation domain. With the degree of the local polynomial $p=1$, the estimator $s(x)$ is termed a local linear regression or a local linear fit. The estimator can be explicitly expressed as

$$s(t)=\frac{\sum_{i=1}^{n}w_iX_i}{\sum_{i=1}^{n}w_i},\quad w_i=K_h(t_i-t)\{S_{n,2}-(t_i-t)S_{n,1}\} \tag{3.2}$$

where $S_{n,j}=\sum_{i=1}^{n}K_h(t_i-x)(t_i-x)^j$. With this estimator,

$$Z_L^2(t)=t(n-t)[s(\overline{X_t})-s(\overline{X_t^*})]^2/n,\qquad t=1,2,\cdots,n-1. \tag{3.3}$$

can be considered. At the point t, Z_L^2 would reflect the existence of the change-point with the bigger values. The change-point estimator is proposed as

$$T_L=argmax_{1\leq t<n}Z_L^2(t). \tag{3.4}$$

Proposition 3.1 T_L is a consistent change-point estimator.

proof. At the point $t\neq\tau$, $P(|\overline{X_t}-\overline{X_t^*}|>\delta)<\epsilon$ and by the continuity of the smoothing function, $P(|s(\overline{X_t})-s(\overline{X_t^*})|>\delta)<\epsilon$. At $t=\tau$, let $|\overline{X_t}-\overline{X_t^*}|=\Delta$.

If $s(\cdot)$ is sufficiently smooth, then Taylor's expansion implies, $k<\tau$

$$s(\overline{X_k})=s(\mu)+s'(\mu)(\overline{X_k}-\mu)+op(1),$$

$$s(\overline{X_k^*})=s(\mu)+s'(\mu)(\overline{X_k^*}-\mu)+op(1)$$

and therefore $(s(\overline{X_k})-s(\overline{X_k^*}))^2=op(1)$.

At the true change-point at τ , with $\mu^*=\mu+\Delta$

$$s(\overline{X_\tau})=s(\mu)+s'(\mu)(\overline{X_\tau}-\mu)+op(1),$$

$$s(\overline{X_\tau^*})=s(\mu^*)+s'(\mu^*)(\overline{X_\tau^*}-\mu^*)+op(1)$$

and therefore $(s(\overline{X_\tau})-s(\overline{X_\tau^*}))^2=\Delta^2+op(1)$.

To prove $P(|\hat{\tau}-\tau|>\delta)\to0$ as $n\to\infty$, we consider the probability

$$
\begin{aligned}
&P\left[k(n-k)(s(\overline{X_k})-s(\overline{X_k^*}))^2>\tau(n-\tau)(s(\overline{X_\tau})-s(\overline{X_\tau^*}))^2\right]\\
&=P\left[\frac{k(n-k)(s(\overline{X_k})-s(\overline{X_k^*}))^2}{\tau(n-\tau)(s(\overline{X_\tau})-s(\overline{X_\tau^*}))^2}>1\right]\\
&=P\left[(s(\overline{X_k})-s(\overline{X_k^*}))^2>\frac{\tau}{k}\Delta^2\right]\to 0.
\end{aligned}
\tag{3.5}
$$

Since $(s(\overline{X_k})-s(\overline{X_k^*}))^2=op(1)$. For $k>\tau$, the same procedure is done and the result follows.

4. Simulation

A simulation study was conducted to investigate the behavior of the proposed change-point estimator and to compare the previously suggested estimators. The data are generated from the one change-point model with iid errors with mean 0 and variance 1. The mean level change model with one change-point is as follows:

$$
X_i=\begin{cases}\mu_0+\epsilon_i, & i=1,\cdots,\tau\\ \mu_1+\epsilon_i, & i=\tau+1,\cdots,n\end{cases}
\tag{4.1}
$$

where $\mu_0=0$, $\Delta=\mu_1-\mu_0=1$. And the errors ϵ_i's are from the normal, double exponential, uniform distributions.

The observations are randomly generated from normal, double exponential and uniform distributions. The mean, MSE(mean squared error) of the change-point estimator and the proportion as the estimated probability of $P(|\hat{\tau}-\tau|\leq 2)$, 95% confidence interval for τ are calculated.

<Table 1> shows that the proposed estimator has smaller MSE than Hinkley(1972) estimator T_{Hink} and other nonparametric estimators. Gombay and Horvath(1990) procedure is applied to the estimators as T_{GH1} with $g(t)=\frac{1}{2}t^2$ and T_{GH2} with $g(t)=e^t$. The bandwidth as a smoothing parameter affects the result. The bandwidth 0.2 and 0.3 are considered and slightly different MSE's are obtained accordingly. <Table 1> shows that the proposed change-point estimator gives smaller MSE than others especially in the normal and uniform distributions. It tells that smoothing of means works better for mean estimation and mean change-point estimation. When there are outliers, for example, one outlier outside of 2 standard deviation, the proposed estimation gives no better result than the nonparametric estimators which are less sensitive to outliers, but better result than Hinkley estimator in the sense of MSE as shown in <Table 2>. Because the effect of outliers are smoothed via loess smoothing, the proposed method gives smaller MSE than the parametric estimation method.

5. Conclusion

The parametric change-point estimation problem was considered and the new method with the loess smoothing was proposed. Loess smoothing of means gave the improvement of the estimation less sensitive with the outliers. As suggested, the change-point estimation method combined with parametric and nonparametric smoothing made another technique for the change analysis. Also other smoothing techniques can be applied in the change analysis.

References

[1] Carlstein, E. (1988) Nonparametric Change-point Estimation, *Annals of Statistics*, 16,188-197.

[2] Cleveland, W. S. (1979) Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of American Statistical Association*, 74, 829-836.

[3] Efromovich, S. (1999) *Nonparametric Curve Estimation*, Springer, New York.

[4] Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*, Chapman & Hall, New York.

[5] Gombay, E. and Horvath, L. (1990) Asymptotic Distributions of Maximum Likelihood Test for Change in the Mean, *Biometrika*, 77, 2, 411-414.

[6] Gombay, E. and Horvath, L. (1994) An Application of the Maximum Likelihood Test to the Change-point Problem, *Stochastic Processes and their Applications,* 50,161-171.

[7] Hinkley, D. V. (1970) Inference about the Change-point in a Sequence of Random Variables, *Biometrika*, 57, 1-17.

[8] Hinkley, D. V. (1972) Time-ordered Classification, *Biometrika,* 59, 509-522.

[9] Schechtman, E. (1982) Nonparametric Test for Detecting Change in Location, *Communication in Statistics-Theory and Method*, A 11(13), 1475-1482.

[10] Stone, C. J. (1977) Consistent Nonparametric Regression, *Annals of Statistics,* 14, 590-606.

[Received March 2005, Accepted April 2005]

<Table 1> Comparison of change-point estimators with n=100, the change-point τ=50, τ=30 in 1,000 repetitions

change-point		τ=50				τ=30			
		mean	MSE	propor-tion	95% CI	mean	MSE	propor-tion	95% CI
Normal $\mu_0 = 0,$ $\mu_1 = 1$	T_{Hink}	49.690	36.478	0.630	(37, 61)	31.184	73.574	0.602	(18, 57)
	T_{Sche}	49.639	37.385	0.633	(36, 61)	31.428	72.590	0.593	(19, 54)
	T_{C1}	48.323	56.893	0.609	(23, 59)	29.537	55.487	0.576	(13, 43)
	T_{C2}	48.531	56.351	0.602	(23, 60)	29.648	61.434	0.570	(13, 44)
	T_{C3}	46.992	285.048	0.367	(10, 83)	32.079	248.541	0.377	(10, 83)
	T_{GH1}	49.690	36.478	0.630	(37, 61)	31.184	73.574	0.5602	(18, 57)
	T_{GH2}	51.970	55.930	0.613	(41, 74)	33.346	117.164	0.581	(21, 68)
	$T_{L,h=0.2}$	49.881	28.825	0.591	(38, 60)	31.398	60.280	0.581	(21, 54)
	$T_{L,h=0.3}$	50.005	27.335	0.500	(40, 59)	31.794	72.730	0.467	(20, 56)
double exp $\mu_0 = 0,$ $\mu_1 = 1$	T_{Hink}	49.655	34.393	0.640	(34, 62)	30.737	52.741	0.607	(17, 50)
	T_{Sche}	49.773	20.623	0.689	(40, 60)	30.825	27.723	0.681	(22, 42)
	T_{C1}	49.201	27.835	0.680	(37, 59)	29.633	27.769	0.658	(15, 40)
	T_{C2}	49.311	22.991	0.700	(38, 58)	29.756	21.726	0.680	(17, 40)
	T_{C3}	48.278	253.672	0.443	(10, 87)	32.346	281.580	0.395	(10, 88)
	T_{GH1}	49.655	34.393	0.640	(34, 62)	30.737	52.741	0.607	(17, 50)
	T_{GH2}	51.742	58.666	0.616	(41, 74)	33.185	118.391	0.588	(21, 69)
	$T_{L,h=0.2}$	49.883	33.093	0.573	(37, 63)	31.225	45.059	0.576	(21, 49)
	$T_{L,h=0.3}$	49.888	35.414	0.505	(38, 61)	31.567	61.497	0.476	(18, 51)
Uniform $\mu_0 = 0,$ $\mu_1 = 1$	T_{Hink}	49.538	46.100	0.598	(32, 62)	30.768	45.726	0.620	(19, 48)
	T_{Sche}	49.455	61.505	0.568	(29, 66)	31.294	57.112	0.593	(19, 51)
	T_{C1}	47.957	74.561	0.555	(24, 61)	29.372	52.010	0.572	(13, 46)
	T_{C2}	48.158	84.290	0.534	(22, 66)	29.697	72.475	0.552	(13, 50)
	T_{C3}	47.101	273.117	0.370	(10, 81)	31.399	231.093	0.354	(10, 80)
	T_{GH1}	49.538	46.100	0.598	(32, 62)	30.768	45.726	0.620	(19, 48)
	T_{GH2}	51.946	62.052	0.581	(39, 72)	32.840	87.696	0.600	(22, 59)
	$T_{L,h=0.2}$	49.636	34.064	0.606	(35, 61)	31.088	39.628	0.595	(21, 45)
	$T_{L,h=0.3}$	49.726	36.244	0.503	(37, 60)	31.277	41.973	0.483	(21, 45)

<Table 2> Comparison of change-point estimators with n=100, the change-point τ=50, τ=30 in 1,000 repetitions with one outlier outside 2 standard deviation

change-point		τ=50				τ=30			
		Mean	MSE	propor-tion	95% CI	Mean	MSE	propor-tion	95% CI
Normal $\mu_0 = 0$, $\mu_1 = 1$	T_{Hink}	49.918	55.418	0.626	(29, 66)	30.627	77.495	0.567	(16, 55)
	T_{Sche}	49.997	54.955	0.628	(29, 68)	31.253	89.683	0.559	(16, 58)
	T_{C1}	48.118	90.338	0.596	(18, 64)	28.457	78.477	0.521	(10, 47)
	T_{C2}	48.193	91.137	0.595	(18, 66)	28.814	80.004	0.527	(11, 50)
	T_{C3}	45.400	345.452	0.330	(10, 84)	31.169	316.845	0.310	(10, 88)
	T_{GH1}	49.918	55.418	0.626	(29, 66)	30.627	77.495	0.567	(16, 55)
	T_{GH2}	52.478	76.578	0.602	(40, 79)	33.615	129.657	0.553	(18, 68)
	$T_{L,h=0.2}$	50.115	46.301	0.591	(31, 66)	31.259	73.549	0.556	(16, 56)
	$T_{L,h=0.3}$	50.257	49.701	0.511	(32, 64)	31.319	89.165	0.371	(14, 55)
double exp $\mu_0 = 0$, $\mu_1 = 1$	T_{Hink}	49.659	54.265	0.654	(27, 64)	30.356	62.854	0.601	(15, 50)
	T_{Sche}	49.869	28.059	0.694	(38, 60)	30.382	41.562	0.664	(17, 46)
	T_{C1}	48.880	43.870	0.682	(27, 59)	28.410	44.098	0.625	(12, 39)
	T_{C2}	49.370	31.344	0.706	(32, 58)	28.967	36.517	0.661	(14, 39)
	T_{C3}	46.781	342.731	0.390	(10, 90)	31.795	325.403	0.351	(10; 88)
	T_{GH1}	49.659	54.265	0.654	(27, 64)	30.356	62.854	0.601	(15, 50)
	T_{GH2}	51.976	74.048	0.608	(38, 78)	33.388	133.004	0.586	(18, 72)
	$T_{L,h=0.2}$	49.640	47.462	0.587	(29, 62)	30.924	53.988	0.577	(16, 51)
	$T_{L,h=0.3}$	49.877	48.863	0.506	(31, 61)	30.669	60.257	0.355	(13, 49)
Uniform $\mu_0 = 0$, $\mu_1 = 1$	T_{Hink}	49.918	46.264	0.600	(32, 64)	30.255	65.103	0.569	(13, 48)
	T_{Sche}	49.730	57.824	0.566	(27, 65)	31.259	94.745	0.545	(13, 57)
	T_{C1}	47.779	98.609	0.550	(18, 63)	28.164	80.772	0.498	(10, 45)
	T_{C2}	47.633	109.599	0.520	(15, 65)	28.577	96.319	0.484	(11, 47)
	T_{C3}	46.948	302.332	0.352	(10, 82)	31.844	273.860	0.311	(10, 79)
	T_{GH1}	49.918	46.264	0.600	(32, 64)	30.255	65.103	0.569	(13, 48)
	T_{GH2}	52.219	59.451	0.580	(41, 74)	33.051	113.965	0.550	(18, 66)
	$T_{L,h=0.2}$	49.822	42.500	0.612	(31, 61)	31.022	54.916	0.549	(15, 49)
	$T_{L,h=0.3}$	49.991	42.645	0.533	(33, 61)	30.871	57.595	0.341	(15, 48)