Applied Mathematical Sciences, Vol. 5, 2011, no. 43, 2177 - 2185

Cusum Procedure Using Transformed Observations

Reza Habibi

Department of Statistics, Central Bank of Iran
habibi1356@gmail.com

Abstract

Cusum statistic is very important statistic in change point analysis. In this paper, we apply the cusum method to transformed observations. We show this method works better than the usual cusum procedure which uses the original data. Two main transformations are the null distribution function and score function. Test statistics are given and their limiting null distributions are studied.

Keywords: Brownian bridge; Change point; Cusum; Rank; Score function; Transformation

1 Introduction. Suppose that $X_1, \cdots, X_n$ is a sequence of independent random variables such that the first k_0 observations $X_1, \cdots, X_{k_0}$ are distributed as F_0 and the remaining observations come from another unknown distribution F_1 where $F_0 \neq F_1$. Integer k_0 is called change point. It is interested to test if there is any change point in distribution of observations, i.e., $H_0 : k_0 = n$ against $H_1 : k_0 = 1, ..., n-1$.

During the last decade, change point analysis has been received considerable attentions in statistical literatures. Basseville and Nikiforov (1993) detected change points using quality control techniques. Brodsky and Darkhovsky (1993) studied nonparametric change point analysis. The limiting distributions of test statistics are given by Csorgo and Horvath (1997). Procedures based on ranks of observations proposed by Huskova and Gombay (1998). Chen and Gupta (2000) considered change point problems in parametric families of distributions. Einmahl and McKeague (2003) used empirical likelihood ration to detect change points. For more references, see Shaban (1980). For change point detection in time series models see Bai (1994) and references therein among the others. Khodadadi and Asgharian (2006) prepared an annotated bibliography about change point analysis in regression and related models.

The cusum statistic is a powerful tool to detect change point. In this paper, we suggest to replace the original data with transformed data in cusum statistic. Three reasons motivated us. They are 1) Sometimes, transformed observations have smaller variance than the original data and therefore change point is detected better. 2) Sometimes, the original data have infinite variance whilst the transformed data are finite variance. 3) Some of transformations contain all available information in the problem e.g. score functions (see subsection 3.2) and using them are much better than the original data. This paper is organized as follows. In section 2, we review the cusum procedure and describe our idea by considering three examples. In section 3, two main transformations are given. They are the null distribution function and the null score function of observations.

2 Cusum statistic. The cusum statistic plays important role in change point analysis. It is used to test if there is a shift in the means of observations (see e.g. Hinkley, 1971). To describe more, suppose that expectation of X_i is μ_0 before k_0 and μ_1 after it where $\mu_0 \neq \mu_1$. Let $e_i = X_i - \overline{X}$ and $E_k = \sum_{i=1}^{k} e_i$. The cusum test statistic is

$$\max_{1 \leq k \leq n-1} |E_k|.$$

Under H_1, the change point estimator is the maximizer of $|E_k|$, i.e.,

$$\widehat{k} = \arg\max |E_k|.$$

The cusum test statistic works well because the means of residuals e_i changes (in sign and absolute value) after k_0. It works much better if we can modify e_i's such that their variance are relatively small. The idea is to replace X_i with some transformed data like $Y_i = g(X_i)$ (for some measurable functions g) such that means of Y_i change after k_0 and $var(Y_i - \overline{Y}) < var(X_i - \overline{X})$. The last condition holds if and only if (iff)

$$var(Y_i) < var(X_i).$$

In this way, test procedure works better under H_0 and H_1. For example, suppose that (X_i, Y_i), $i = 1, ..., n$ are independent random vectors such that a change exits in marginal means of X_i. For $i = 1, ..., n$, let

$$Z_i = E(X_i | Y_i).$$

It is seen $E(Z_i) = E(X_i)$, that is, the means of Z_i changes after change point as well as $var(Z_i) \leq var(X_i)$. Therefore, it is reasonable to construct cusum statistic by Z_i. The next example illustrates this idea more.

Example 1. Suppose that X_i are independent exponential random variables with parameters $\beta_i = 1$ for $i = 1, ..., k_0$ and $\beta_i = 10$ for $i = k_0 + 1, ..., 100$.

The true value of k_0 is 40. For $i = 1, ..., 100$, let $Y_i = \sqrt{X_i}$. Applying the Monte Carlo method, the mean and standard deviation (sd) of change point estimators $\widehat{k}_x$ and $\widehat{k}_y$ are given as follows. It is seen, $\widehat{k}_y$ works better than $\widehat{k}_x$.

$$\begin{aligned} E(\widehat{k}_x) &= 42.74, \; sd(\widehat{k}_x) = 4.01 \\ E(\widehat{k}_y) &= 40.11, \; sd(\widehat{k}_x) = 1.92. \end{aligned}$$

In cases where both mean and variance do not exist the cusum procedure is not applicable. By a transformation, we can change the infinite variance case to finite variance problem. The following example describes this idea more.

Example 2. Suppose that, under H_0, the density of X_i's are

$$f_\eta(x) = \eta/x^2,$$

for $x > \eta > 0$ and zero otherwise. The mean and variance of X_i do not exist and the cusum procedure can not be applied. Let $Y_i = \log(X_i)$. The Y_i's have exponential distribution with two parameters $1, \log(\eta)$. Under H_1, the expectation of transformed data are changed and we can apply the cusum procedure.

3 Two transformations. Here, we introduce two well-known transformations. They are the null distribution function and null score function of observations. Test statistics are given and their asymptotic behavior are studied.

3.1 Distribution function. First, suppose that F_0 is continuous and completely known. Let

$$U_i = F_0(X_i),$$

$i = 1, \cdots, n$ (that is, $g = F_0$). This transformation works in infinite variance cases. Under the null hypothesis U_i's are independent and identically distributed (iid) random variables distributed uniformly on $(0, 1)$ and then $E(U_i) = 0.5$. Under the alternative hypothesis, we have

$$E(U_i) = \begin{cases} 0.5 & i = 1, \cdots, k_0, \\ \theta & i = k_0 + 1, \cdots, n, \end{cases}$$

where $\theta = E_{F_1}(F_0(X_{k_0+1}))$. Therefore, under H_1, the mean of U_i's has changed and the cumulative sum (cusum) test procedure may be applied to detect the change. It is easy to see that

$$\theta = P(X_{k_0} \leq X_{k_0+1}).$$

In the following example, we present some formulas for θ for some special parametric distributions before and after k_0. We also survey when θ is (is not) 0.5.

Example 3. For the following distributions before change point (dist. bcp) and after (dist. acp) θ's are given in Table 1. Notations exp, Γ, N, U stands for exponential, gamma, normal and uniform distributions.

Table 1: Some formulas for θ

No.	dist. bcp	dist. acp	θ
1	$\exp(\beta_0)$	$\exp(\beta_1)$	$\beta_1/(\beta_0+\beta_1)$
2	$\Gamma(\alpha_0,\beta)$	$\Gamma(\alpha_1,\beta)$	$P(Z<0.5)$
3	$N(\mu_0,\sigma_0^2)$	$N(\mu_1,\sigma_1^2)$	$\Phi(\frac{\mu_1-\mu_0}{\sqrt{\sigma_0^2+\sigma_1^2}})$
4	$\exp(\beta)$	$N(\mu,\sigma^2)$	$\exp\{(b/2)-a\}$
5	$U(a,b)$	$N(\mu,\sigma^2)$	$\frac{a^*\Phi(a^*)-b^*\Phi(b^*)}{a^*-b^*}$

Here, for each number, we briefly describe how θ is derived. First, consider no 1. Then

$$\begin{aligned}\theta &= 1-E(e^{-\beta_0^{-1}X_n})\\ &= 1-M_{X_n}(-\beta_0^{-1})=\beta_1/(\beta_0+\beta_1),\end{aligned}$$

where $M_{X_n}(\cdot)$ is the moment generating function of X_n. Then $\theta=0.5$ iff $\beta_0=\beta_1$ i.e., H_0 is satisfied. In no. 2, $\theta=P(Z<0.5)$, where

$$Z=X_1/(X_1+X_n),$$

has beta distribution with parameters α_0,α_1. Then $\theta=0.5$ iff $\alpha_0=\alpha_1$. In no. 3 deriving θ is too simple. Note that $\Phi(\cdot)$ is the standard normal distribution function. One can see that $\theta=0.5$ iff $\mu_0=\mu_1$. This shows that our test statistic can not detect change in variance if $\mu_0=\mu_1$. In no. 4 to obtain θ, we use the moment generating function of exponential distribution. Here,

$$a=\mu/\beta \text{ and } b=\sigma^2/\beta.$$

It is seen $\theta=0.5$ iff $2a-b=\log(2)$. Finally, in no. 5, $a^*=(\mu-a)/\sigma$ and $b^*=(\mu-b)/\sigma$.

Let $S_k^*=\sum_{i=1}^{k}(U_i-\overline{U})$, where $\overline{U}=\sum_{i=1}^{n}U_i/n$. The test statistic is given by

$$T_n=\max_{1\le k\le n-1}|S_k^*|.$$

The null distribution of T_n does not depend on F_0. If $\theta \geq 0.5$ (or $\theta \leq 0.5$), we use $T_n^+ = \max S_k^*$(or $T_n^- = \min S_k^*$). Three quantiles of $n^{-0.5}T_n$ ($n = 100$ and $F_0 = \exp(1)$) are $q_{0.9} = 0.3350$, $q_{0.95} = 0.3733$ and $q_{0.99} = 0.4578$.

We may consider another cusum test statistic. Let $\Phi^{-1}(\cdot)$ be the inverse of distribution function of standard normal distribution and consider transformed data

$$Y_i = \Phi^{-1}(U_i),$$

$i = 1, \cdots, n$. We apply cusum to transformed data Y_i and then test statistic is

$$V_n = \max_{1 \leq k \leq n-1} |S_k|,$$

where $S_k = \sum_{i=1}^{k}(Y_i - \overline{Y})$. Variables V_n^+ and V_n^- are defined like T_n^+ and T_n^-. The null distributions of V_n^+ and V_n^- are the same. The limiting null distribution of test statistics are given in Table 2. The suprimum is taken over $t \in (0, 1)$. Conniffe and Spencer (1999) presented a chi-square approximation to null distribution of $4V_n^{+2}/n$. Habibi (2008) derived the empirical quantiles of $n^{-0.5}V_n$ for various sample size n.

Table 2: Limiting null distributions

Statistic	Limiting dist.
$n^{-0.5}T_n$	$\sqrt{12}\sup\|B(t)\|$
$n^{-0.5}T_n^+$	$\sqrt{12}\sup B(t)$
$n^{-0.5}V_n$	$\sup\|B(t)\|$
$n^{-0.5}V_n^+$	$\sup B(t)$

Example 2 (*cont.*). Consider again infinite variance case in example 2. The distribution function is

$$F_\eta(x) = 1 - \eta/x,$$

for $x > \eta$ and zero otherwise. Suppose that η_0 and η_1 are the parameters of before and after change point. It is easy to see that

$$\theta = 1 - (1/2)\frac{\min(\eta_0, \eta_1)}{\max(\eta_0, \eta_1)}.$$

Remark 1. Let $X_1, ..., X_{100}$ be a sequence of independent exponential distribution. Suppose that $X_1, ..., X_{40}$ come form exponential exp(1) and $X_{41}, ..., X_{100}$ are distributed as exp(0.5). In the level of 0.05, the powers of $n^{-0.5}T_n$ ($n = 100$) and KS tests are 0.7176 and 0.265, respectively. This is seen our test procedure works much better than KS test in term of power.

Remark 2. Suppose that F_0 is unknown. Then, it is replaced by its empirical estimate F_n. One should note that $F_n(X_i) = R_i/n$,where $(R_1, ..., R_n)$ is the vector of ranks of observations $X_1, \cdots, X_n$. Then the test statistic T_n is changed to the following rank based statistic

$$\widehat{T}_n = \max_{1 \le k \le n-1} |\sum_{i=1}^{k} (R_i - \overline{R})|,$$

where $\overline{R} = (n+1)/2$. Reader should be able to define $\widehat{T}_n^+$ and $\widehat{T}_n^-$. The above test statistic is distribution free under the H_0. Empirical quantiles of $\widehat{T}_n$ are given in Habibi (2008). The extended version of $\widehat{T}_n$ based on linear functions of R_i is given by Huskova and Gombay (1998). Here, this fact that T_n reduces to $\widehat{T}_n$ is interesting.

3.2 Score function. In this part, interested transformation is the null score function. Score function contains all available information in inferential problem. Under the parametric setting, suppose that $f_\xi(x)$ is the density of X_i. Let ξ_0 and ξ_1 be the parameters before and after change point and suppose that ξ_0 is known. The score function for $i-th$ observation is

$$s_i(\xi_0) = \frac{\partial}{\partial \xi_0} \log(f_{\xi_0}(X_i)).$$

It is seen $E(s_i(\xi_0)) = 0$ under H_0. Therefore, $W = \max |\sum_{i=1}^{k} s_i(\xi_0)|$ is the test statistic. Habibi *et al.* (2005) studied the properties of the weighted sum of $s_i(\widehat{\eta})$ in the general class of distributions. Let $\gamma^2 = I(\xi_0)$ the Fisher information function. It is seen, as $n \to \infty$,

$$\frac{1}{\sqrt{n\gamma^2}} W \xrightarrow{d} \sup B(t).$$

Remark 3. If ξ_0 is unknown, then it is replaced by $\widehat{\xi}_0$ (null maximum likelihood estimate of ξ_0). Then the test statistic is given by

$$\widehat{W} = \max_{1 \le k \le n-1} |\sum_{i=1}^{k} s_i(\widehat{\xi}_0)|.$$

Since $\widehat{\xi}_0 \xrightarrow{p} \xi_0$, the null limiting distribution of $\frac{1}{\sqrt{n\widehat{\gamma}^2}}\widehat{W}$ and $\frac{1}{\sqrt{n\gamma^2}}W$ are the same where $\widehat{\gamma}^2 = I(\widehat{\xi}_0)$. In the following examples, we compute this statistic for some rich families of distributions.

Example 4. Let $f_\xi(x)$ belongs to exponential family of distributions, i.e.,

$$f_\xi(x) = c(x)\exp\{\eta(\xi)d(x) - \eta^*(\xi)\}.$$

Let $d_i = d(x_i)$, $i = 1, ..., n$. It is straightforward to see that

$$\eta(\widehat{\xi}_0) \max_{1\leq k\leq n-1} |\sum_{i=1}^{k}(d_i - \overline{d})|.$$

When the distribution before change point is $N(\xi_0, 1)$, this statistic reduces to V_n, the usual cusum test statistic. If $f_\xi(x) = f(x - \xi)$ the location family. One can see that the cusum is

$$\max |\sum_{i=1}^{k} l'_f(x_i - \widehat{\xi}_0)|,$$

where $l'_f = \frac{d}{dx}\log(f(x))$. Habibi (2009) extended the work of Inclan and Tiao (1994) to detecting change point in variance to scale family distributions.

Example 5. Hansen (1992) considered testing parameter instability in linear models. Let

$$y_t = \beta_1 x_{1t} + ... + \beta_m x_{mt} + \epsilon_t,$$

$t = 1, ..., n$ denote the linear model. He presented $f_{it} = x_{it}\widehat{\epsilon}_t$ as scores in a maximum likelihood context where $\widehat{\epsilon}_t$ are estimated errors. Let $s_{it} = \sum_{j=1}^{t} f_{ij}$ and $L_t = (1/n)\sum_{i=1}^{n} s_{it}^2/\sum_{i=1}^{n} f_{it}^2$. He suggested to use L_t to test instability of β_t and a similar test statistic to test instability of parameters $(\beta_1, ..., \beta_m)$ jointly. In our work, his test statistic is

$$h_n = (1/n)\frac{\sum_{i=1}^{n} E_i^2}{\sum_{i=1}^{n} e_i^2}.$$

One can see that for large n,

$$h_n \xrightarrow{d} \int_0^1 B^2(t)dt.$$

If X_i's are iid come form $N(0, 1)$, then the exact distribution of h_n is F-distribution. Unfortunately, using this version of Hansen cusum statistic, we can not estimate the location of change point. Instead, we suggest to use

$$H_n = \max_{1\leq k\leq n-1} \frac{|E_k|}{\sqrt{\sum_{i=1}^{n} e_i^2}}.$$

The limiting null distribution of $\sqrt{n}H_n$ is $\sup|B(t)|$ where supremum is taken over $t \in (0,1)$. The change point estimator is the maximizer of $|E_k|/\sqrt{\sum_{i=1}^n e_i^2}$. Then, we can apply this idea to AR(1) models. Suppose that under the null hypothesis of no change point in parameter α, the AR(1) is given by

$$x_t = \alpha x_{t-1} + \varepsilon_t,$$

where $\varepsilon_t, t = 1, ..., n$ are zero mean white noises with common variance σ^2. Therefore, the cusum test statistic is given by

$$\max_{1\leq k\leq n-1} \frac{\sum_{i=1}^k x_{i-1}\widehat{\varepsilon}_i}{\sqrt{\sum_{i=1}^n (x_{i-1}\widehat{\varepsilon}_i)^2}}.$$

Remark 4. The previous transformation does not depend on x (the support of X). Let g also depend on x, i.e., $g = g(x, X)$. The first condition is satisfies iff

$$F_0(\cdot) \neq F_1(\cdot) \Longrightarrow E_{F_0}(g(\cdot, X)) \neq E_{F_1}(g(\cdot, X)),$$

or equivalently, $E_{F_0}(g(\cdot, X)) = E_{F_1}(g(\cdot, X))$ implies that $F_0(\cdot) = F_1(\cdot)$. One of the best choice for g is $I_{(-\infty,x]}(X)$ where $I_{(-\infty,x]}(X)$ is one if $X \leq x$ and zero otherwise. Using this g, our test statistic reduces to KS test statistic.

References

[1] J. Bai, Weak convergence of the sequential empirical processes of residuals in ARMA models, Annals of Statistics, 22 (1994), 2051-2061.

[2] M. Basseville, N. Nikiforov, The detection of abrupt change- theory and applications, Prentice Hall, 1993.

[3] B. E. Brodsky, B. S. Darkhovsky, Nonparametric methods in change-point problems, Kluwer Academic Publishers, 1993.

[4] J. Chen, A. K. Gupta, Parametric statistical change point analysis, Birkhäser, 2000.

[5] D. Conniffe, J. E. Spencer, Approximating the distribution of the maximum partial sum of normal deviates. The Economic and Social Research Institute, Dublin, Working Paper No. 102 (1999).

[6] M. Csorgo, L. Horvath, Limit theorems in change point analysis. Wiley, 1997.

[7] J. H. J. Einmahl, I. W. McKeague, Empirical likelihood based hypothesis testing, Bernoulli 9 (2003), 267-290.

[8] R. Habibi, S.M. Sadooghi-Alvandi, and A. R. Nematollahi, Change point detection in a general class of distributions, Communications in Statistics, Theory and Methods 34 (2005), 1935-1938.

[9] R. Habibi, A simple method to detect a change point, Tech. Report 33 (2008), Department of Statistics, Central Bank of Iran.

[10] R. Habibi, Change point detection in scale family distributions, Statistica Neerlandica 63 (2009), 347-352.

[11] D. V. Hinkley, Inference about the change-point from cumulative sum tests, Biometrika 58 (1971), 509-523.

[12] M. Huskova, E. Gombay, Rank based estimators of the change point, Journal of Statistical Planning and Inference 67 (1998), 137-154.

[13] C. Inclan, G. C. Tiao, Use of cumulative sums of squares for retrospective detection of changes of variances, J. Amer. Statist. Assoc. 89 (1994), 913-923.

[14] A. Khodadadi, M. Asgharian, Change-point problems and regression: an annotated bibliography. Tech. Report (2006). Web address:
http:// biostats.bepress.com/cobra/ps/art44/

[15] S. A. Shaban, Change point problem and two-phase regression: An annotated bibliography. International Statistical Review 48 (1980), 83-93.