Sequential Detection and Estimation of Change-Points

Boris Brodsky
(CEMI RAS, Moscow, Russia)

In this report the problem of sequential detection and estimation of change-points is considered. The sense of this problem is as follows: in practice after raising an alarm signal about a change it is often required to divide the whole obtained sequential sample into subsamples of observations before and after an unknown change-point. In this report asymptotically optimal methods are proposed for sequential detection and estimation of a change-point in this problem.

On the probability space $(\Omega, \mathcal{F}, P)$ let us consider the following model of observations:

$$x(n) = a + hI(n > m) + \xi_n, \tag{1}$$

where m is an unknown change-point; $h \geq \delta > |a|$ is an unknown change in the mathematical expectation of a centered random sequence ξ_n ($E\xi_n = 0$, $E\xi_n^2 = \sigma^2$) which satisfies the uniform Cramer and ψ-mixing conditions:

- *the uniform Cramer's condition*: for every $i = 1, 2, \dots$

$$\exists H > 0 : \sup_i E \exp(t\xi_i) < \infty, \ |t| < H$$

- *ψ-mixing condition*:

$$\psi(n) = \sup_{t \geq 1} \sup_{A \in \mathcal{F}_1^t, B \in \mathcal{F}_{t+n}^{\infty}, P(A)P(B) \neq 0} \left| \frac{P(AB)}{P(A)P(B)} - 1 \right| \to 0 \quad \text{as } n \to \infty,$$

where $\mathcal{F}_1^t = \sigma\{\xi_1, \dots, \xi_t\}$, $\mathcal{F}_{t+n}^{\infty} = \sigma\{\xi_{t+n}, \xi_{t+n+1}, \dots\}$.

For **sequential detection of the change-point** m, let us consider the following *nonparametric cumulative sums (CUSUM) method*:

$$y_n = (y_{n-1} + x(n))^+, \quad y_0 \equiv 0, \tag{2}$$

where $b^+ = \max(b, 0)$.

The decision function of CUSUM method is $d_N(n) = I(y_n > N)$, where N is the threshold of detection which coincides with a "large parameter" N for CUSUM method.

Denote by P_0, P_m measures corresponding to a sequence of observations without a change-point and with the change-point m, respectively. We introduce the following performance measures of sequential change-point detection:

- the supremal probability of a "false decision":

$$\alpha_N = \sup_n P_0\{d_N(n) = 1\} \tag{3}$$

- the stopping time

$$\tau_N = \inf\{n : \; d_N(n) = 1\}. \tag{4}$$

For the nonparametric CUSUM method, the following theorem holds

Theorem 1.
Suppose the sequence of observations satisfies the uniform Cramer and ψ-mixing conditions. Then for any $a < 0$ and any $m > 0$:

$$\alpha_N \leq L_1 \exp(-L_2 \frac{N|a|}{\sigma^2}), \tag{5}$$

where $L_1, L_2 > 0$ depend on parameters of Cramer's and ψ-mixing conditions only and for any $h > |a|$, $\epsilon > 0$ and $\gamma_N = (\tau_N - m)^+/N$:

$$\lim_{N\to\infty} N^{-1} \ln P_m\{|\gamma_N - \frac{1}{h-|a|}| > \epsilon \;|(\tau_N > m)\} \leq -\frac{\epsilon^2}{2\sigma^2}(h-|a|)^3. \tag{6}$$

From this theorem it follows that the normalized delay time γ_N tends a.s. to $(h-|a|)^{-1}$ as $N \to \infty$ with the rate of convergence estimated in (6).

For **estimation of a detected change-point**, we use the retrospective sample of the last M observations: $\{x(\tau_N - M + 1), \ldots, x(\tau_N)\}$. The volume M of this sample can be obtained from (6):

$$M = \frac{N}{\delta - |a|} + \sigma\sqrt{2N}\frac{|\ln\alpha|^{1/2}}{(\delta-|a|)^{3/2}}. \tag{7}$$

This choice of M guarantees that the retrospective sample "covers" the true change-point m with the confidence probability $1-\alpha$ for any $0 < \alpha < 1$. We can choose $\alpha = \exp(-\beta N)$ for $\beta > 0$ and then $M = N(\frac{1}{\delta-|a|} + \frac{\sigma(2\beta)^{1/2}}{(\delta-|a|)^{3/2}})$.

So $P_m(H_\beta) \geq 1 - \exp(-\beta N)$, where H_β is the hypothesis that the retrospective sample contains the true change-point m.

For estimation of the change-point m, the following statistic is used:

$$T_M(n) = \sqrt{\frac{n(M-n)}{M}}(\frac{1}{n}\sum_{i=1}^{n} x(i+\tau_N - M) - \frac{1}{M-n}\sum_{i=n+1}^{M} x(i+\tau_N-M)) \tag{8}$$

for $n = 1, \ldots, M$.

Then the estimate of the change-point m can be constructed as follows: $\hat{m} = \hat{n} + \tau_N - M$, where $\hat{n}$ is the minimal point of the set $arg\max_n |T_N(n)|$.

Define the following values: $\rho = 1 - \frac{(\tau_N - m)^+}{M}$ and $\hat{\rho} = \frac{\hat{n}}{M}$. Then the following theorem holds:

Theorem 2.

Suppose the sequence of observations satisfies the uniform Cramer and ψ-mixing conditions. Then for any $0 < \epsilon < 1$:

$$P_m\{|\hat{\rho} - \rho| > \epsilon | H_\beta \cap \{\tau_N > m\}\} \le L_1 \exp(-L_2 \epsilon^2 M), \qquad (9)$$

where the constants $L_1, L_2 > 0$ do not depend on M.

From this theorem it follows that the proposed estimate $\hat{m}$ will be in the $[\epsilon M]$ neighborhood of the true change-point m with the probability increasing to 1 as $M \to \infty$ for any $0 < \epsilon < 1$.

Asymptotic optimality

The proposed methods are asymptotically optimal for sequential detection and estimation of a change-point. We prove this fact for a sequence of independent r.v.'s $X = \{x_1, x_2, \dots\}$ with the d.f.

$$f(x_n) = \begin{cases} f_0(x_n), & n \le m \\ f_1(x_n), & n > m \end{cases}$$

At any step of decision making, we test the null hypothesis H_0 of no change in the d.f. of observations against the alternative H_1 of a change occurred in this d.f. Suppose a certain method with the decision function $d_C(\cdot)$ (depending on some "large parameter" C) is used:

$$d_C(\cdot) = \begin{cases} 0, & \text{assume } H_0 \text{ and continue} \\ 1, & \text{assume } H_1 \text{ and stop} \end{cases}$$

Consider the following characteristics:

1) Supremal probability of a false decision:

$$\alpha_C = \sup_n P_0\{d_C(n) = 1\}, \qquad (10)$$

2) Stopping time τ_C and the normed delay time γ_C:

$$\tau_C = \inf\{n : \ d_C(n) = 1\}, \quad \gamma_C = (\tau_C - m)^+ / C. \qquad (11)$$

Then for any $m \ge 1$ and $C \to \infty$, the following asymptotical inequality holds true:

$$J \cdot E_m(\gamma_C | \tau_C > m) \ge (1 + o(1)) \frac{|\ln(\alpha_C(T(C) + m))|}{C}, \qquad (12)$$

where $J = \int f_1(x) \ln \dfrac{f_1(x)}{f_0(x)} dx$ and

$$T(C) = \min(n : \sum_{i=n}^{\infty} P_0(\tau_C = i) \le \alpha_C).$$

A method of sequential change-point detection is called the *1st order asymptotically optimal* if the equality sign in (12) is attained for this method as $C \to \infty$. For the proposed nonparametric CUSUM test, $C \equiv N$. This test is the 1st order asymptotically optimal.

Besides the 1st order asymptotical optimality, we consider the *2nd order asymptotical optimality* of sequential tests in the class of methods for which $\gamma_C \to \gamma(\theta)$ in probability as $C \to \infty$, where θ is the parameter of the d.f. of observations after the change-point m, i.e.:

$$f(x_n) = \begin{cases} f_0(x_n), & n \le m \\ f_\theta(x_n), & n > m \end{cases}$$

Suppose $I(\theta) = \int \frac{(f'_\theta(z))^2}{f_\theta(z)} dz$. Then for $\epsilon \to 0$ the following asymptotical inequality holds true:

$$\lim_{C\to\infty} C^{-1} \ln P_m\{|\gamma_C - \gamma(\theta)| > \epsilon | \tau_C > m\} \ge -\frac{\epsilon^2(1+o(1))}{2[\gamma'(\theta)]^2}\gamma(\theta)I(\theta). \quad (13)$$

A method of sequential change-point detection is called the *2nd order asymptotically optimal* if the equality sign in (13) is attained for this method as $C \to \infty$. The nonparametric CUSUM test is the 2nd order asymptotically optimal.

The 2nd order asymptotical optimality in sequential detection is closely connected with the asymptotical optimality of change-point estimation at the 2nd stage of the proposed two-stage procedure.

Monte-Carlo tests

In this section results of a simulation study of the proposed methods are presented. The following examples were studied:

1) Change in mean: independent Gaussian r.v.'s with an unknown change in the mathematical expectation

2) Change in dispersion: independent Gaussian r.v.'s with an unknown change in dispersion

1) Change in mean

The sequence of Gaussian observations $X = (x(1), x(2), ...)$ with the change-point $m = 1000$ was modeled. Both m and a change in mean h ($N(0,1) \to N(h,1)$) were unknown to the proposed algorithm.

For detection of a change-point, the nonparametric CUSUM test was used:

$$y_n = (y_{n-1} + x(n) + a)^+ > N, \quad y_0 = 0,$$

The parameters of this test were chosen as follows: $N = 12$, $a = -0.5$. The supremal probability of a false decision for these parameters is $\exp(-2|a|N/\sigma^2) = \exp(-12) = 6.14 * 10^{-6}$ and corresponds to the average time before a "false alarm" $E_0\,\tau = 162750$.

Let $\tau = \inf\{n \geq 1 : d_N = 1\}$ be the instant when the change was detected. Then $E_m\tau - m$ is the average delay time in change-point detection. Besides the average delay time, we compute the value $\sigma(\tau - m)$ - the square root of the dispersion of the delay time.

At the 2nd stage we estimate the change-point using the retrospective sample of the last M observations. The choice of M is made by formula (7) for $\alpha = 0.05$. Here $\beta = 0.25$ and for $N = 12, a = -0.5$ and $\delta = 0.55$, we obtain $M = 1000$.

In experiments the following measure of the estimation error was computed:

$$\Delta = (\frac{1}{k-1}\sum_{i=1}^{k}(\hat{m}_k - m)^2)^{1/2},$$

where $k = 5000$ is the number of independent trials.

The results obtained are reported in Table 1.

Table 1.

h		0.55	0.6	0.8	1.0	1.5	2.0	2.5
CUSUM	$E_m\tau - m$	113.1	82.1	36.4	23.3	12.2	8.2	5.9
	$\sigma(\tau - m)$	88.5	61.5	18.4	12.5	3.8	2.9	1.3
K-S	Δ	79.3	45.5	13.9	13.5	4.6	3.9	2.1

2) Change in dispersion

In this group of tests the following model of observations was considered: the sequence of i.r.v.'s $X = (x(1), x(2), ...)$ with the change-point $m = 1000$. The d.f. of observations changes at the instant $m = 1000$: $N(0,1) \to N(0,(1+h)^2)$. Both m and h were unknown to the algorithm.

For detection of the change-point m, the nonparametric CUSUM test was used:

$$y_n = (y_{n-1} + x^2(n) + a)^+ > N, \quad y_0 = 0.$$

The parameters of this test were chosen as follows: $N = 20$, $a = -1.25$. The supremal probability of a false decision for these parameters is $\exp(-10) = 4.54 * 10^{-5}$ and corresponds to the average time before a "false alarm" $E_0\,\tau = 22026$.

The choice of M is made by formula (7) for $\alpha = 0.05$. Here $\beta = 0.25$ and for $N = 20, a = -1.25$ and $\delta = 1.69$, we obtain $M = 150$. Results are presented in Table 2.

Table 2.

h		0.3	0.4	0.5	0.7	0.9	1.0
CUSUM	$E_m\tau - m$	65.7	37.8	26.7	15.9	10.8	9.4
	$\sigma(\tau - m)$	54.3	28.3	18.9	10.4	7.1	6.1
K-S	Δ	70.9	37.1	24.5	19.3	12.5	11.6

References

[1] Srivastava M.S. and Wu Ya. (1999) Quasi-stationary biases of change-point and change magnitude estimation after sequential CUSUM test. *Sequential Analysis.* 18, p. 203-216.

[2] Ding K. (2003) A lower confidence bound for the change-point after a sequential CUSUM test. *Journal of Statistical Planning and Inference.* 115, p.311-326.

[3] Gombay E. (2003) Sequential Change-Point Detection and Estimation. *Sequential Analysis.* 22, is.3, pp.203-222.