Change-point problem for high-order Markov chain under composite hypotheses

Boris Darkhovsky

Insitute for Systems Analysis RAS,
9, pr. 60-letya Oktyabria
117312 Moscow, Russia
darbor@isa.ru

Abstract. The change-point problem for high-order Markov chain under composite hypotheses before and after the change is considered. A non-asymptotic inequality for the new optimality criterion is proposed.

Keywords. Change-point problem, high-order Markov chain, prior inequality.

1 Problem description

Let $X^0 = \{x_n^0\}_{n=1}^{\infty}$, $X^1 = \{x_n^1\}_{n=1}^{\infty}$ are independent strictly stationary ergodic k-order Markov chains with real values ($k = 0$ corresponds to independent variables). The observations have the form $Z = \{z_n\}_{n=1}^{\infty}$, where

$$z_n = \begin{cases} x_n^0, \text{ if } 1 \leq n < m \\ x_n^1, \text{ if } n \geq m \end{cases}$$

and m is the change-point.

Let Θ_0, Θ_1 are certain sets in finite dimensional parametric spaces. Suppose that for $X^i, i = 0, 1$ there exist families (over θ_i) of k-dimensional initial (stationary) distribution density functions (d.f.) $f_i(\theta_i, x), x \in \mathbb{R}^k, \theta_i \in \Theta_i, i = 0, 1$ with respect to some σ-finite measure μ on $\mathbb{R}^k$. Suppose that there exist conditional (transition) density functions (c.d.f.) $\varphi_i(\theta_i, x_n^i = u | x_{n-1}^i, \dots, x_{n-k}^i) = \varphi_i(\theta_i, u | x_{n-k}^{n-1}) = \varphi_i(\theta_i, u|x), x \in \mathbb{R}^k$, $n \geq k+1$, $i = 0, 1$ with respect to one-dimensional marginal μ_1 of μ. D.f. and c.d.f. are known and defined for all parameter values from some open neighbourhoods of parametric sets. It is assumed that for any $\theta_i \in \Theta_i$, $i = 0, 1$ the measure $f_i(\theta_i, x)d\mu$ is the unique invariant measure and that $\mathcal{P}_0 \bigcap \mathcal{P}_1 = \emptyset$, where $\mathcal{P}_i$, $i = 0, 1$ is the collection (over the parameter $\theta_i \in \Theta_i$) of measures defined by functions $\{\varphi_i(\theta_i, u|x),\ f_i(\theta_i, x)\}$.

In what follows we denote by $\mathbf{P}_{m,\vartheta}(\mathbf{E}_{m,\vartheta})$ the measure (mathematical expectation) corresponding to the sequence $\{z_n\}_{n=1}^{\infty}$ with the change-point at the instant m and the fixed value of the parameter $\vartheta = (\theta_0, \theta_1) \in \Theta \stackrel{\text{def}}{=} \Theta_0 \times \Theta_1$. In case $m = 1$ we use the symbols $\mathbf{P}_{1,\theta_1}(\mathbf{E}_{1,\theta_1})$. Symbols $\mathbf{P}_{\infty,\vartheta}(\mathbf{E}_{\infty,\vartheta})$ correspond to the observed sequence without the change-point. In this case we use the symbols $\mathbf{P}_{\infty,\theta_0}(\mathbf{E}_{\infty,\theta_0})$.

The problem consists in sequential detection of the (a priori unknown) change-point m based on the observations $\{z_n\}$.

Denote by $d(n)$ the decision function of a change-point procedure (i.e., a function measurable w.r.t. the natural flow of σ-algebras generated by observations and such that $d(n) = 1$ corresponds to the decision about the presence of the change at the instant n and $d(n) = 0$ corresponds to the decision about the absence of the change). For example, the decision function $d^{CUS}(n)$ in classical CUSUM procedure for the case of independent observations $\{\eta_n\}$ with d.f. $f_0(\cdot)$ before and $f_1(\cdot)$ after the change is equal to

$$d^{CUS}(n) = \mathbb{I}\left(\max_{1 \leq k \leq n} \sum_{i=k}^{n} \ln \frac{f_1(\eta_i)}{f_0(\eta_i)} > C\right)$$

where $\mathbb{I}(A)$ denotes the indicator function of the set A.

The stopping time τ corresponding to the decision function $d(n)$ is equal to $\tau = \inf\{n : d(n) = 1\}$.

As a rule, any change-point detection procedure depends on a certain "large parameter" C such that the probability of false decision tends to zero as the "large parameter" goes to infinity. Very often the “large parameter” is simply a decision threshold (as in CUSUM procedure), but it is not necessary. We suppose that change-point detection procedures contain such “large parameter” C and will equip corresponding decision functions and stopping times with index C.

For an arbitrary point $\theta_1 \in \Theta_1$ put

$$\Theta_0^*(\theta_1) = \arg\max_{z\in\Theta_0} \left(\int \ln \frac{\varphi_1(\theta_1, u|x)}{\varphi_0(z, u|x)} \varphi_1(\theta_1, u|x) f_1(\theta_1, x) d\mu(x) d\mu_1(u) \right)^{-1}$$

Obviously, $\Theta_0^*(\theta_1)$ is the set of the *least favorable alternative to θ_1 before the change-point*. For an arbitrary point $\theta_1 \in \Theta_1$ put

$$\mathbf{J}(\theta_1) = \int \ln \frac{\varphi_1(\theta_1, u|x)}{\varphi_0(\theta_0^*, u|x)} \varphi_1(\theta_1, u|x) f_1(\theta_1, x) d\mu(x) d\mu_1(u),$$

where $\theta_0^* \in \Theta_0^*(\theta_1)$.

Define the *supremal probability of false decision* (SPFD) $\gamma(\theta_1, \tau_C)$ for the stopping time τ_C generated by the decision function $d_C(n)$, and an arbitrary point $\theta_1 \in \Theta_1$

$$\gamma(\theta_1, \tau_C) = \sup_{\theta_0^* \in \Theta_0^*(\theta_1)} \sup_n \mathbf{P}_{\infty, \theta_0^*} \{ d_C(n) = 1 \}$$

and put

$$\alpha(\theta_1, \tau_C) = \sup_{\theta_0^* \in \Theta_0^*(\theta_1)} \sup_n \mathbf{P}_{\infty, \theta_0^*} \{ \tau_C = n \}$$

Obviously, $\alpha(\theta_1, \tau_C)$ is the *lower bound* for SPFD and therefore, usually $\alpha(\theta_1, \tau_C) \to 0$ as $C \to \infty$.

Stopping time τ_C is called *nondegenerate* if $0 < \alpha(\theta_1, \tau_C) < 1$ for any θ_1.

For an arbitrary pair $\vartheta = (\theta_0^*, \theta_1),\ \theta_1 \in \Theta_1,\ \theta_0^* \in \Theta_0^*(\theta_1)$ and any fixed m consider the following *criterion of optimality*

$$\mathcal{K}(\tau_C, \theta_1, m) = \frac{\sup\limits_{\theta_0^* \in \Theta_0^*(\theta_1)} \mathbf{E}_{m,\vartheta} \Big(\tau_C - (m+k) \Big| \tau_C \ge m+k \Big)}{|\ln \alpha(\theta_1, \tau_C)|}$$

It is known that if a random variable ξ takes positive integer values and satisfies the following condition

$$\sup_n \mathbf{P}\{\xi = n\} \le \alpha$$

then

$$\mathbf{E}\xi \ge (2\alpha)^{-1}(1 + o(1)) \quad \text{as } \alpha \to 0$$

and this lower bound is sharp.

From this it follows that for any $\theta_1 \in \Theta_1,\ \theta_0 \in \Theta_0^*(\theta_1)$

$$\mathbf{E}_{\infty,\theta_0} \tau_C \ge (\alpha(\theta_1, \tau_C))^{-1}.$$

Therefore, $\mathcal{K}(\tau_C, \theta_1, m)$ is the *ratio (in an appropriate scale) of the maximal conditional average delay time for the parameter θ_1 after the change to the worst (i.e. the minimal) average time before a false decision for the least favorable alternative to θ_1*.

In our opinion, the proposed performance index characterizes the quality of a change-point detection method no worse than the conventional criteria and corresponds well to the pragmatic sense of the change-point detection problem.

2 Assumptions

For any $\vartheta = (\theta_0, \theta_1) \in \Theta$ put

$$h(n,\vartheta) \stackrel{\text{def}}{=} \ln \frac{\varphi_1(\theta_1, z_n|z_{n-k}^{n-1})}{\varphi_0(\theta_0, z_n|z_{n-k}^{n-1})},\ n \geq k+1$$
$$g(m,\vartheta) \stackrel{\text{def}}{=} \ln \frac{f_1(\theta_1, z_m^{m+k-1})}{\prod_{j=m}^{m+k-1} \varphi_0(\theta_0, z_j|z_{j-k}^{j-1})},\ k \geq 1,\ m \geq k+1$$

To formulate the results we will use assumptions from the following list.

A1. Any two measures $\nu_0 \in \mathcal{P}_0,\ \nu_1 \in \mathcal{P}_1$ are equivalent.

A2. The functions $h(n,\vartheta),\ g(m,\vartheta)$ are well defined random variables w.r.t. any measure $\mathbf{P}_{m,\vartheta},\ 1 \leq m \leq \infty$.

A3.

$$\infty > \sup_{\tilde{\theta}_1 \in \Theta_1} \sup_{\vartheta \in \Theta} \mathbf{E}_{1,\tilde{\theta}_1} h(n,\vartheta) \geq \inf_{\tilde{\theta}_1 \in \Theta_1} \inf_{\vartheta \in \Theta} \mathbf{E}_{1,\tilde{\theta}_1} h(n,\vartheta) > 0.$$

A4. Θ is a compact set.

A5. The function

$$\left(\int \ln \frac{\varphi_1(\theta_1, u|x)}{\varphi_0(\theta_0, u|x)} \varphi_1(\theta_1, u|x) f_1(\theta_1, x) d\mu(x) d\mu_1(u) \right)^{-1}$$

is continuous w.r.t. $\theta_0 \in \Theta_0$ for any $\theta_1 \in \Theta_1$ (and therefore, the set $\Theta_0^*(\theta_1)$ is well defined for any $\theta_1 \in \Theta_1$).

A6. $\mathbf{E}_{m,\vartheta} g(m,\vartheta) < \infty$ for any m, ϑ.

A7. The function $\Phi(\vartheta, u, z) \stackrel{\text{def}}{=} \dfrac{\varphi_1(\theta_1, u|z)}{\varphi_0(\theta_0, u|z)}$ is continuous w.r.t. $\vartheta = (\theta_0, \theta_1)$ for μ-a.e. (u, z).

A8.

$$0 > \sup_{\tilde{\theta}_0 \in \Theta_0} \sup_{\vartheta \in \Theta} \mathbf{E}_{\infty,\tilde{\theta}_0} h(n,\vartheta) \geq \inf_{\tilde{\theta}_0 \in \Theta_0} \inf_{\vartheta \in \Theta} \mathbf{E}_{\infty,\tilde{\theta}_0} h(n,\vartheta) > -\infty.$$

A9. For the random sequence $\{h(n,\vartheta)\}$ denote by $\psi^{(m)}(\tilde{\vartheta}, \vartheta, s)$ the coefficient of ψ-mixing (between σ-algebras $\sigma\{z_1^n\}$ and $\sigma\{z_{n+s}^{\infty}\}$ with supremum over n) under measure $\mathbf{P}_{m,\tilde{\vartheta}}$. Then (uniform ψ-mixing):

$$\sup_{\tilde{\vartheta} \in \Theta} \sup_{\vartheta \in \Theta} \sum_s \sqrt{\psi^{(m)}(\tilde{\vartheta}, \vartheta, s)} < \infty,\ m = 1,\ \infty.$$

A10. (Uniform Cramer condition):

$$\sup_n \sup_{\tilde{\vartheta} \in \Theta} \sup_{\vartheta \in \Theta} \mathbf{E}_{m,\tilde{\vartheta}} \exp\{t h(n,\vartheta)\} < \infty \quad \text{for} \quad |t| < H, m = 1,\ \infty;$$

A11. There exists the uniform w.r.t. $\vartheta = (\theta_0, \theta_1),\ \tilde{\theta}_1$ the limit

$$\sigma^2(\vartheta, \tilde{\vartheta}) = \lim_{T \to \infty} \mathbf{E}_{1,\tilde{\theta}_1} (T-k)^{-1} \left(\sum_{n=k+1}^{T} (h(n,\vartheta) - \mathbf{E}_{1,\tilde{\theta}_1} h(n,\vartheta)) \right)^2.$$

A12. For any $\vartheta, \tilde{\theta}_0$ the function

$$\varkappa(t, \vartheta, \tilde{\theta}_0) \stackrel{\text{def}}{=} \ln \mathbf{E}_{\infty,\tilde{\theta}_0} \exp\{t h(n,\vartheta)\}$$

has only two zeros: 0 and $t^*(\vartheta, \tilde{\theta}_0) > 0$, the function $t^*(\cdot,\cdot)$ is continuous and $\min\limits_{\vartheta,\tilde{\theta}_0} t^*(\cdot,\cdot) > 0$.

Let us make some comments on the assumptions. Consider the equation

$$y_{n+1} = F(y_n, y_{n-1}, \dots, y_{n-k+1}, \xi_{n+1}),\ n \geq 0, \tag{1}$$

where $\mathbf{y}_0 = (y_0, y_{-1}, \dots, y_{-k+1})$ is the initial random vector, $\{\xi\}_{n=1}^{\infty}$ is the sequence of i.i.d. random variables independent of $\mathbf{y}$.

It is well known that $\{y_n\}$ is a k-order Markov chain. In the literature one can easily find conditions on the function $F(\cdot)$ and the random sequence $\{\xi_n\}$ such that the sequence $\{y_n\}$ will be a strictly stationary ergodic k-order Markov chain with integrable correlation function (the latter corresponds to assumption A11). For example, let $F(\cdot) = \sum_{s=0}^{k-1} a_s y_{n-s} + \xi_{n+1}$, the corresponding polynomial is stable one, ξ_n is square integrable and the distribution of $\mathbf{y}_0$ coincides with the stationary distribution of $(y_n, y_{n-1}, \dots, y_{n-k+1})$. Then all such assumptions hold.

In (Blum et al., 1963) one can find checkable necessary and sufficient conditions for ψ-mixing of strictly stationary ergodic Markov chain and these conditions guarantee that the ψ-mixing coefficient tends to zero exponentially (see assumption A9).

Suppose that X^i, $i = 0, 1$ is strictly stationary ergodic exponentially ψ-mixing high-order Markov chain with integrable correlation function for any parameter ϑ. Then the same is true for $\{h(n, \vartheta)\}$ because $h(n, \vartheta)$ depends only of finite number of previous observations.

3 Main results

For an arbitrary $\theta_1 \in \Theta_1$ and $0 < r < 1$ define positive integers $L(\tau_C, \theta_1, r)$ and $\mathbb{L}(\tau_C, r)$ as follows

$$L(\tau_C, \theta_1, r) = \min\{n : \sup_{\theta_0^* \in \Theta_0^*(\theta_1)} \sum_{k=n}^{\infty} \mathbf{P}_{\infty,\theta_0^*}(\tau_C = k) \le r\}$$
$$\mathbb{L}(\tau_C, r) = \sup_{\theta_1 \in \Theta_1} L(\tau_C, \theta_1, r)$$

Theorem 1. *Let $\vartheta = (\theta_0^*, \theta_1)$, $\theta_1 \in \Theta_1$, $\theta_0^* \in \Theta_0^*(\theta_1)$ be an arbitrary point. Let τ_C be a nondegenerate stopping time such that $\sup_{\theta_0^* \in \Theta_0^*(\theta_1)} \mathbf{E}_{m,\vartheta}(\tau_C - (m+k))^+ < \infty$ for any fixed C, m and any fixed point $\theta_1 \in \Theta_1$. Let assumptions A1 - A6 hold. Then for any fixed m and any point $\theta_1 \in \Theta_1$ the following inequality holds*

$$\mathcal{K}(\tau_C, \theta_1, m) \ge \mathbf{J}^{-1}(\theta_1) \left[1 - \frac{\ln\left(\dfrac{m + k + \mathbb{L}(\tau_C, r^*)}{\sup_{\theta_0^* \in \Theta_0^*(\cdot)} \mathbf{P}_{\infty,\theta_0^*}(\tau_C \ge m+k)} + 1 \right)}{|\ln \alpha(\theta_1, \tau_C)|} - \right.$$
$$\left. - \frac{\mathbf{J}(\theta_1) + B}{\sup_{\theta_0^* \in \Theta_0^*(\cdot)} \mathbf{P}_{\infty,\theta_0^*}(\tau_C \ge m+k) \, |\ln \alpha(\theta_1, \tau_C)|} \right] \tag{2}$$

where

$$r^* = \alpha(\theta_1, \tau_C) \sup_{\theta_0 \in \Theta_0^*(\theta_1)} \mathbf{P}_{\infty,\theta_0}(\tau_C \ge m+k),$$
$$B = \begin{cases} \sup_{\theta_0 \in \Theta_0^*(\theta_1)} \mathbf{E}_{m,\vartheta} g(m, \vartheta), & k \ge 1 \\ 0, & k = 0 \end{cases}$$

Remark 1. We use the independence assumption of the sequences X^0 and X^1 only for a simplicity. In general, inequality (2) must be slightly changed, but the main term $\mathbf{J}^{-1}(\theta_1)$ will be the same.

Consider the following class of stopping times

$$\mathfrak{M} = \{\tau_C : \sup_{\theta_0 \in \Theta_0^*(\theta_1)} \mathbf{E}_{m,\vartheta}(\tau_C - (m+k))^+ < \infty,$$
$$\limsup_{r \to 0} \frac{\ln \mathbb{L}(\tau_C, r)}{|\ln r|} = 0 \text{ for any fixed } m, \theta_1, C\}$$

In particular, stopping times such that their mathematical expectations under the change-point are finite (for all values of parameters θ_0, θ_1 and any finite value of the "large parameter" C) and their distributions without the change-point have exponential tails belong to $\mathfrak{M}$.

From Theorem 1 we have the following

Corollary 1. *Let for $\tau_C \in \mathfrak{M}$ the following condition holds*

$$\limsup_{C\to\infty} \alpha(\theta_1, \tau_C) = \limsup_{C\to\infty} \sup_{\theta_0 \in \Theta_0^*(\theta_1)} \sup_n \mathbf{P}_{\infty,\theta_0}\big\{\tau_C = n\big\} = 0$$

Then for any fixed m and $\theta_1 \in \Theta_1$

$$\lim_{C\to\infty} \mathcal{K}(\tau_C, \theta_1, m) = \lim_{C\to\infty} \frac{\sup\limits_{\theta_0\in\Theta_0^*(\theta_1)} \mathbf{E}_{m,\vartheta}\Big(\tau_C - (m+k)\big|\tau_C \geq m+k\Big)}{|\ln \alpha(\theta_1, \tau_C)|} \geq \mathbf{J}^{-1}(\theta_1) \qquad (3)$$

A method of sequential change-point detection is called *adaptive asymptotically optimal* if the lower bound in inequality (3) is attained for it as $C \to \infty$. The term "adaptive" in this context means that the asymptotic optimality of a test does not depend on a true (and unknown) distribution function of observations from the family $\mathcal{P}_1$.

Define the *minimax CUSUM* stopping time as follows:

$$T(C) = \inf\{n \geq k+1 : \min_{\theta_0\in\Theta_0} \max_{\theta_1\in\Theta_1} \max_{k+1\leq j\leq n} \sum_{i=j}^{n} \ln \frac{\varphi_1(\theta_1, z_i|z_{i-k}^{i-1})}{\varphi_0(\theta_0, z_i|z_{i-k}^{i-1})} > C\}$$

Theorem 2. *Let assumptions A1 - A12 hold. Then the minimax CUSUM stopping time $T(C)$ is adaptive asymptotically optimal. Besides, for any $\theta_1 \in \Theta_1$ the following equality holds:*

$$\lim_{C\to\infty} \frac{|\ln \alpha(\theta_1, T(C))|}{C} = \max_{\theta_0^*\in\Theta_0^*(\theta_1)} \max_{\tilde\theta_0\in\Theta_0} \min_{\tilde\theta_1\in\Theta_1} t^*(\tilde\theta_0, \tilde\theta_1, \theta_0^*)$$

4 Sketch of the proof of Theorem 2

The proof is based on the following main points.

1. For any point $\vartheta = (\theta_0, \theta_1) \in \Theta$ consider the following stopping time

$$\tau_C(\vartheta) = \inf\left\{ n \geq k+1 : \max_{k+1\leq j\leq n} \sum_{i=j}^{n} \ln \frac{\varphi(\theta_1, z_i|z_{i-k}^{i-1})}{\varphi(\theta_0, z_i|z_{i-k}^{i-1})} > C \right\}$$

Obviously, $\tau_C(\vartheta)$ is the classical CUSUM stopping time for the sequence $\{h(n, \vartheta)\}$.

Then for any $\vartheta = (\theta_0, \theta_1)$, $\tilde\vartheta = (\tilde\theta_0, \tilde\theta_1)$ and any fixed m the following relations hold:

a)

$$\lim_{C\to\infty} \frac{(\tau_C(\vartheta) - m - k)^+}{C} = \mathbf{A}^{-1}(\tilde\theta_1, \vartheta) \qquad \mathbf{P}_{m,\tilde\vartheta} - \text{a.s.}$$

b) the collection (over the parameter C) of random variables $\dfrac{(\tau_C(\vartheta) - m - k)^+}{C}$ is uniformly integrable w.r.t. the measure $\mathbf{P}_{m,\tilde\vartheta}$

c)

$$\lim_{C\to\infty} \frac{\mathbf{E}_{m,\tilde\vartheta}(\tau_C(\vartheta) - m - k)^+}{C} = \mathbf{A}^{-1}(\tilde\theta_1, \vartheta)$$

where $\mathbf{A}(\tilde\theta_1, \vartheta) \stackrel{\text{def}}{=} \mathbf{E}_{1,\tilde\theta_1} h(n, \vartheta)$.

In (Brodsky and Darkhovsky, 2000, p. 229) it was proved (in slightly different notations) that for any fixed $m, \vartheta, \tilde{\vartheta}$ from conditions A3, A9, A10 the property a) follows. It was also proved there that under conditions A3, A9, A10, A11 the property b) holds. Now c) follows from a) and b).

2. For any $\vartheta \in \Theta$ and $\theta_0^* \in \Theta_0^*(\theta_1)$ put

$$\beta_C(\vartheta, \theta_0^*) \stackrel{\text{def}}{=} \sup_n \mathbf{P}_{\infty, \theta_0^*} \left\{ \max_{k+1 \le j \le n} \sum_{i=j}^{n} \ln \frac{\varphi(\theta_1, z_i | z_{i-k}^{i-1})}{\varphi(\theta_0, z_i | z_{i-k}^{i-1})} > C \right\}$$

Then

$$\lim_{C \to \infty} \frac{|\ln \beta_C(\vartheta, \theta_0^*)|}{C} = t^*(\vartheta, \theta_0^*)$$

where $t^*(\vartheta, \theta_0^*)$ was given above. The convergence here is uniform w.r.t. ϑ, θ_0^*.

The proof can be found in (Brodsky and Darkhovsky, 2000, p.261). This proof uses assumptions A8—A12. The uniform convergence is guaranteed because these assumptions are fulfilled uniformly w.r.t. ϑ, θ_0^* .

The points **1** and **2** play crucial role in the proof. The continuance of the proof can be done analogously (Brodsky and Darkhovsky, 2008). As a result we prove that

$$\lim_{C \to \infty} \frac{\sup\limits_{\theta_0^* \in \Theta_0^*(\theta_1)} \mathbf{E}_{m,\vartheta} \left(T(C) - (m+k)\right)^+}{|\ln \alpha\left(\theta_1, T(C)\right)|} = \mathbf{J}^{-1}(\theta_1) \tag{4}$$

From (4) we get the asymptotic optimality of $T(C)$ because $\mathbf{P}_{\infty, \theta_0}\left(T(C) \ge m+k\right) \to 1$ as $C \to \infty$ for any $\theta_0 \in \Theta_0^*(\theta_1)$ and any fixed m.

References

Blum, J. R., Hanson, D. L. and Koopmans, L. H. (1963). On the Strong Law of Large Numbers for a Class of Stochastic Processes. *Z. Wahrscheinlichkeitstheorie*, **2**, 1-11.

Brodsky, B. E. and Darkhovsky, B. S. (2000). *Non-parametric statistical diagnosis: problems and methods.* Kluwer Academic Publishers, The Netherland.

Brodsky, B. and Darkhovsky, B. (2008). Sequential change-point detection for mixing random sequences under composite hypotheses. *Statistical Inference for Stochastic Processses*, **11**, 35-54.