Asymptotically optimal methods of early change-point detection

Boris Brodsky

(Russian Academy of Sciences, Moscow)

3rd International Workshop in Sequential Methodologies

Stanford University, June 14–16, 2011

Plan:

1. Introduction: motivation, examples, problem statement

2. A priori inequalities

- Univariate models

- Multivariate models

3. Methods

- Dependent univariate sequences

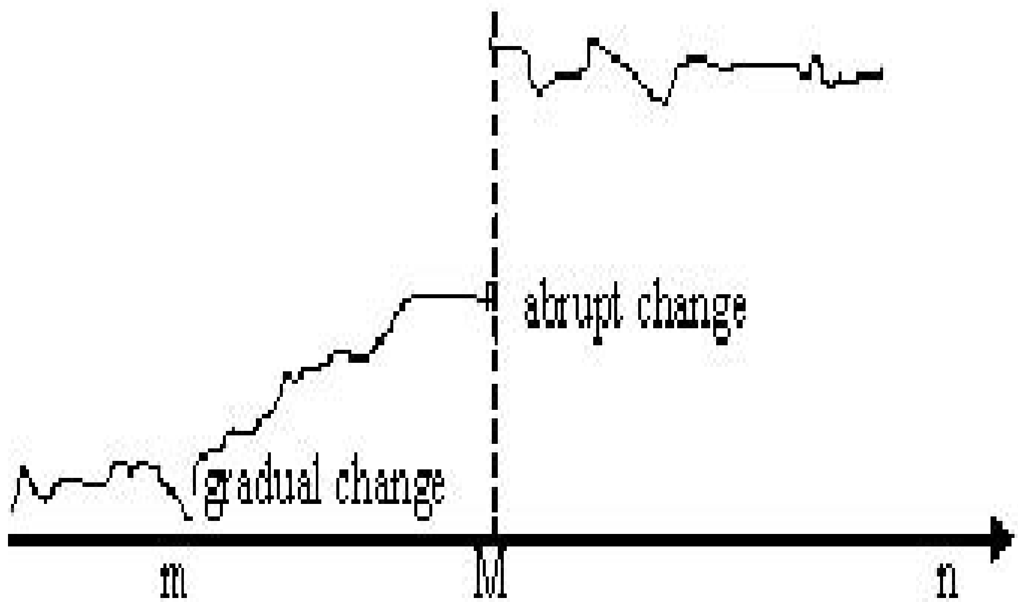
- Multivariate regressions

- State-space models

4. Experiments

1. Introduction

Gradual change-point detection



Examples

1. Ecology: monitoring of air and water pollution
2. Technology: technological breaks
3. Economy: financial crises

Problem

– Detect a change-point m not *post factum* (after M) but *ante factum* ($m \leq \tau \leq M$)

– FAR small

Abrupt changes: *univariate models*

Page (1954)

Girshick and Rubin (1952)

Kolmogorov, Shiryaev (1959)

Shiryaev (1959-1965)

Lorden (1971), Pollack (1985), Moustakides (1986)

Bansal, Papantoni-Kazakos (1983)

Dragalin V.P., Tartakovsky A.G., Veeravalli V. (1999, 2000)

Mei (2006)

Brodsky and Darkhovsky (2000, 2005, 2008)

Multivariate models

Willsky (1976), Willsky and Jones (1976)

Basseville, Benveniste (1983)

Basseville, Nikiforov (1993)

Lai (1995, 1998, 2000)

Fuh (2003, 2004, 2006, 2007)

Gradual changes

Willsky (1976), Willsky and Jones (1976)

Brodsky and Darkhovsky (2000)

2. Asymptotic optimality

2.1. Univariate models; independent observations

$$X = \{x(1), x(2), \ldots\}, \qquad m - \textbf{ a change-point},$$

$$\frac{d}{dz} P\{x(n) \le z\} = \begin{cases} f(z,0), & n \le m \quad \textbf{or } m = \infty \\ f(z, n-m), & n > m, \end{cases}$$

Consider the following decision rule $d_C(n)$ depending on a large parameter C:

$$d_C(n) = \begin{cases} 1, \textbf{stop at time } n \textbf{ and accept } H_1, \\ 0, \textbf{continue under } H_0 \end{cases}$$

Define

$$\alpha_C = \sup_n P_\infty\{d_C(n) = 1\}$$
$$\tau_C = \min\{n : \ d_C(n) = 1\}$$
$$\gamma_C = (\tau_C - m)^+/C$$
$$M_C = \min\{l : \sum_{n=m+l}^{\infty} P_\infty\{\tau_C = n\} \le \alpha_C\}.$$

$$J(n) = E_m(\ln \frac{f(z, n-m)}{f(z,0)}), \qquad j(t) = J(m + [tc]), \quad t \ge 0.$$

Theorem

$$E_m \int_0^{\gamma_C} j(t)dt \ge \frac{|\ln(\alpha_C M_C)|}{C} + O(\frac{1}{C}).$$

Multivariate models; dependent observations

Suppose $Z = (z_1, z_2, \dots)$ **is a sequence of dependent vector-valued observations** $z_n = (z_n^1, \dots, z_n^k)$ **defined on the probability space** $(\Omega, \mathcal{F}, P)$**.**

$$\alpha_C = \sup_n P_\infty\{d_C(n) = 1\}$$
$$\tau_C = \min\{n : \ d_C(n) = 1\}$$
$$\gamma_C = (\tau_C - m)^+/C$$
$$M_C = \min\{l : \sum_{n=m+l}^{\infty} P_\infty\{\tau_C = n\} \leq \alpha_C\}.$$

Theorem

$$E_m \int_0^{\gamma_C} j(t)dt \geq \frac{|\ln(\alpha_C M_C)|}{C} + O(\frac{1}{C}),$$

where

$$J(n) = E_m(\ln \frac{f_n(z_n|z_1 \dots z_{n-1})}{f_0(z_n|z_1 \dots z_{n-1})}),$$
$$j(t) = J(m + [tC]), \quad t \geq 0.$$

Univariate models

$(\Omega, \mathfrak{F}, P)$**,**

A dependent random sequence $X = \{x(1), x(2), \ldots\}$**,**

$$x(n) = a\, I(1 \leq n \leq m) + h(n-m)\, I(n > m) + \xi(n), \qquad n = 1, 2, \ldots$$

where $E\xi(n) = 0$**.**

Assumptions

1) Cramer's condition:

$$\exists H > 0 : \quad E \exp(t\xi(i)) < \infty, \quad |t| < H, \quad \forall i \geq 1;$$

2) ψ**-mixing condition:** $\mathfrak{F}_1^t = \sigma\{\xi(1), \ldots, \xi(t)\},\ \mathfrak{F}_{t+n}^{\infty} = \sigma\{\xi(t+n), \ldots\}$**,**

Let $\mathcal{H}_1$ **and** $\mathcal{H}_2$ **be two** σ**-algebras contained in** $\mathfrak{F}$**.**

$$\psi(\mathcal{H}_1, \mathcal{H}_2) = \sup_{A \in \mathcal{H}_1, B \in \mathcal{H}_2, \mathbf{P}(A)\mathbf{P}(B) \neq 0} \left| \frac{\mathbf{P}(AB)}{\mathbf{P}(A)\mathbf{P}(B)} - 1 \right|$$

$$\psi(n) = \sup_{t \geq 1} \psi(\mathfrak{F}_1^t, \mathfrak{F}_{t+n}^{\infty}) \to 0 \text{ as } n \to \infty.$$

Methods

CUSUM

$$y_n = (y_{n-1} + x(n))^+, \qquad y_0 = 0, \qquad d_N(n) = I(y_n > N).$$

Roberts–Shiryaev

$$R_n = (1 + R_{n-1})e^{x(n)}, \qquad R_0 = 0, \qquad d_N(n) = I(R_n > e^N).$$

"Window-limited"

$$Y_N(n) = N^{-1} \sum_{k=0}^{N-1} g(\frac{k}{N})x(n-k), \qquad n = N, N+1, \ldots$$

$$d_N(n) = I(|Y_N(n)| > C), \qquad \int_0^1 g^2(t)dt = 1.$$

Exponential smoothing

$$Y(n) = (1-\nu)Y(n-1) + \nu x(n), \qquad Y(o) = 0$$

$$d_N(n) = I(|Y(n)| > C), \qquad N = 1/\nu,\ 0 < \nu < 1.$$

Theorem (CUSUM, RSh)

1) $\delta_N = \dfrac{|\ln \alpha_N|}{N} \to \delta^*$ **as** $N \to \infty$**, where** δ^* **is the minimal nonzero root of the equation:**

$$\kappa(t) = \ln \sup_n E \exp(t\xi(n)) = 0;$$

2) $\gamma_N \to \gamma^*$ P_m**-a.s. as** $N \to \infty$**, where**

$$\int_0^{\gamma^*} (h(t) - |a|)dt = 1;$$

3) for $\forall \epsilon > 0$**:**

$$\lim_{N\to\infty} N^{-1} \ln P_m\{|\gamma_N - \gamma^*| > \epsilon\} = -\frac{\epsilon^2}{2\sigma^2} \frac{(h(\gamma^*) - |a|)^2}{\gamma^*}.$$

Theorem (”window-limited)

Suppose the above formulated ψ-mixing and Cramer’s conditions are satisfied and $C < h(T)$. Then

i)

$$\lim_{N\to\infty} N^{-1}|\ln \max_{1\le n\le N} \mathbf{P}_\infty\left(d_N(n)=1\right)| = \frac{C^2}{2\sigma^2},$$

ii)

$$\gamma_N \stackrel{\mathbf{P}_m\ a.s.}{\longrightarrow} \gamma^e_{gm} \quad \textbf{as } N\to\infty,$$

where γ^e_{gm} is the minimal root of the equation $\int_0^{\gamma^e_{gm}} h(t)g(t)dt = C$;

iii)

$$\sqrt{N}(\gamma_N - \gamma^e_{gm}) \stackrel{d}{\to} \eta,$$

where η is the Gaussian random variable with zero mean and the dispersion $\dfrac{\sigma^2}{h^2(\gamma^e_{gm})g^2(\gamma^e_{gm})}$.

Asymptotic optimality

Gaussian sequence $x_1, x_2, \ldots$ **with the dispersion** $\sigma^2 = 1$ **and the trend** $h(t)$, $t \geq 0$ **in the mathematical expectation of observations.**

CUSUM and RSh

$$\int_0^{\gamma^e} \frac{h^2(t)}{2}\, dt \geq \delta^* = 2|a|.$$

$$\int_0^{\gamma_c} (h(t) - |a|)\, dt = 1.$$

Asymptotical optimality only for abrupt changes:

$$h(t) \equiv const = 2|a|.$$

"Window limited" method:

$h(1) > C,\ \delta^*(\cdot) = C^2/2,\ \sigma^2 = 1$:

$$\int\limits_0^{\gamma^e} \frac{h^2(t)}{2} \geq \frac{C^2}{2}.$$

For "window limited" methods:

$$\int\limits_0^{\gamma} h(t)g(t)dt = C,$$

where $\int\limits_0^1 g^2(t)dt = 1,\ 0 < \gamma \leq 1$**.**

Asymptotically optimal method:

$$g(t) = h(t)/C, \qquad C = (\int_0^1 h^2(t)dt)^{1/2}.$$

Multivariate models

Multivariate regression

$$Y(n) = \Pi(n)X(n) + \nu_n, \quad n = 1, 2, \ldots,$$

The $M \times K$ **matrix** Π **changes abruptly at some unknown change-point** m**, i.e.**

$$\Pi(n) = \mathbf{a}I(n \leq m) + \mathbf{b(n)}I(n > m), \quad n = N, N+1, \ldots$$

where $\|\mathbf{a} - \mathbf{b(n)}\| > 0$**.**

This model generalizes many widely used regression models, i.e.

- static and dynamic regression models with multiple predictors

- ARMA (autoregression and moving average) models for time series

- systems of simultaneous regression equations in econometrics

- stochastic dynamical systems with fully observed state variables in control theory.

Assumptions

Suppose that predictors $X(n)$ and noises ν_n are continuously distributed and strictly stationary and the following conditions are satisfied:

1) the vector $X(n) = (x_{1n}, \ldots, x_{Kn})^{'}$ is $\mathcal{F}_{n-1}$ measurable.

2) there exists a continuous matrix function $V(t)$, $t \in [0,1]$ such that for any $0 \leq t_1 \leq t_2 \leq 1$

$$\frac{1}{N} \sum_{j=[t_1 N]}^{[t_2 N]} X(j) X^{'}(j) \rightarrow \int_{t_1}^{t_2} V(t) dt, \quad P - \textbf{a.s. as } N \rightarrow \infty,$$

where $\int_{t_1}^{t_2} V(t)dt$ is the positive definite matrix;

3) the random vector sequence $\{(X(n), \nu_n)\}$ satisfies ψ-mixing and the unified Cramer condition.

4) $\{\nu_n\}$ is a martingale-difference sequence w.r.t. the flow $\{\mathcal{F}_n\}$.

Method

For any $n = N, N+1, \ldots$ **consider** N **last vectors of observations** $Y(i), X(i), i = n-N+1, \ldots, n$**.**

First, consider the $K \times K$ **matrices:**

$$\mathcal{T}^n(1,l) = \sum_{i=1}^{l} X(i+n-N)X^{'}(i+n-N), \quad l = 1, \ldots, N,$$

second, the $K \times M$ **matrices:**

$$z^n(1,l) = \sum_{i=1}^{l} X(i+n-N)Y^{'}(i+n-N), \quad l = 1, \ldots, N,$$

and third, the decision statistic

$$Y_N^n(l) = \frac{1}{N}(z^n(1,l) - \mathcal{T}^n(1,l)(\mathcal{T}^n(1,N))^{-1} z^n(1,N)).$$

where $l = 1, \ldots, N,$ $Y_N^n(N) = 0$ **and by definition,** $Y_N^n(0) = 0$**.**

Fix the number $0 < \beta < 1/2$**. For detection of the change-point** $m > N$**, we define the stopping time**

$$\tau_N = \inf\{n : \max_{[\beta N] \le l \le N} \|Y_N^n(l)\| > C\}$$

where C **is a certain decision threshold,** $\|A\|$ **is the Gilbert norm of the matrix** A**.**

Theorem

Suppose assumptions 1)-2) are satisfied. Then

1) for the 1st type error:

$$P_0\{\max_n \|Z_N(n)\| > C\} \le m_0(C_1) \begin{cases} \exp(-\frac{TN\beta C_1}{4m_0(C_1)}), & C_1 > hT \\ \exp(-\frac{N\beta C_1^2}{4hm_0(C_1)}), & C_1 \le hT, \end{cases}$$

where $C_1 = C/(1+\sqrt{K})$**.**

2) for the 2nd type error, define:

$$S(\theta) = \int_0^{1-\theta} V(\tau)d\tau \cdot I^{-1} \cdot \int_{1-\theta}^{1} V(\tau)(\mathbf{b}(\tau) - \mathbf{a})d\tau$$

$$g(\tilde{\theta}) = \max_\theta g(\theta), \quad g(\theta) = \|S(\theta)\|^2,$$

$$d = (g(\tilde{\theta}) - C)/(1+\sqrt{K}).$$

Then

$$\delta_N \le m_0(d) \begin{cases} \exp(-\frac{TN\beta d}{4m_0(d)}), & d > hT \\ \exp(-\frac{N\beta d^2}{4hm_0(d)}), & d \le hT. \end{cases}$$

3) normed delay time:

$$\gamma_N = \frac{(\tau_N - m)^+}{N} \to \gamma^*, \quad P_m - \text{ **a.s. as** } N \to \infty,$$

where γ^* **is the minimal root of the equation** $g(t) = C$**.**

The proposed method is *asymptotically optimal by the order of the performance measures* **(w.r.t.** $N \to \infty$**).**

State-space models

Model

For $0 \le n \le m$**:**

$$Y(n+1) = H\,X(n+1) + \xi(n+1)$$
$$X(n+1) = \phi(X(n)) + \eta(n+1)$$

and for $n > m$**:**

$$Y(n+1) = D(n)\,X(n+1) + \xi(n+1)$$
$$X(n+1) = \Lambda_n(X(n)) + \eta(n+1),$$

where $\|D(n) - H\| > 0,\ \sup_X \|\Lambda_n(X) - \phi(X)\| > 0$**.**

Stability assumptions

Let $\Lambda_n(X) = Q_n(x)\,\phi(X)$**. Then suppose that**

$$\frac{1}{N}\sum_{i=1}^{[Nt]} \phi(X(i))\,\phi^{'}(X(i)) \to \int\limits_0^t V(\tau)d\tau, \quad P-\textbf{a.s. as } N \to \infty$$

$$\frac{1}{N}\sum_{i=1}^{[Nt]} \phi(X(i))\,\phi^{'}(X(i))\,Q_i^{'}(X(i)) \to \int\limits_0^t U(\tau)d\tau, \quad P-\textbf{a.s. as } N \to \infty$$

Let $I = \int\limits_0^1 V(\tau)d\tau$**. Define**

$$S(\theta) = \int\limits_0^{1-\theta} V(\tau)d\tau \cdot I^{-1} \cdot \int\limits_{1-\theta}^{1} (V(\tau)H^{'} - U(\tau)D^{'}(\tau))\,d\tau,$$

Theorem

Suppose Cramer's and ψ-mixing conditions are satisfied. Then

1) for the 1st type error:

$$P_0\{\max_n \|Z_N(n)\| > C\} \le m_0(C_1) \begin{cases} \exp(-\frac{TN\beta C_1}{4m_0(C_1)}), & C_1 > hT \\ \exp(-\frac{N\beta C_1^2}{4hm_0(C_1)}), & C_1 \le hT, \end{cases}$$

where $C_1 = C/(1+\sqrt{K})$**;**

2) for the 2nd type error: define

$$q = \max_{i,j,u} Ea_{ij}(u),\ a(u) = \Omega(u)X^{'}(u)$$
$$g(\theta) = \|S(\theta)\|^2, \quad g(\tilde{\theta}) = \max_{\theta} g(\theta), \quad d = (g(\tilde{\theta}) - C - q))/(1+\sqrt{K}).$$

Then

$$P_m\{\max_n \|Z_N(n)\| \le C\} \le m_0(d) \begin{cases} \exp(-\frac{TN\beta d}{4m_0(d)}), & d > hT \\ \exp(-\frac{N\beta d^2}{4hm_0(d)}), & d \le hT, \end{cases}$$

3) for the normalized delay time

$$\gamma_N \to \gamma^*, \quad P- \text{ a.s. as } N \to \infty,$$

where γ^* **is the minimal root of the equation** $g(t) = C + q$**.**

The proposed method is *asymptotically optimal by the order of the performance measures* **(w.r.t.** $N \to \infty$**).**

Experiments

Univariate models

E$T \simeq 577$ for all methods:

1) CUSUM: $y_n = (y_{n-1} + x_n)^+;\ a = -0.5$ **and the threshold of detection was equal to 4.5;**

2) RSh: $y_n = (1 + y_{n-1}) \exp(x_n);\ a = -0.5$ **and the threshold of detection was equal to 330;**

3) Exp: $y_n = (1-\nu)y_{n-1} + \nu x_n;\ \nu = 0.02$ **and the threshold of detection was equal to 0.255.**

4) WL: $y_n = N^{-1} \sum_{i=0}^{N-1} g(i/N)\, x(n-i)$; $N = 100$; $C = 0.255$,
$g(i/N) = \frac{H}{T} \frac{i}{N}$.

The length of the "transition period" was equal to $T = 3000$ and the value H for the linear trend model was changed in the interval 0.1 - 1000. In 5000 independent trials of each experiment the average delay time in detection Eτ and the value $\sigma\tau$ were computed. The results are reported in Table 1.

Table 1.

Hmax		0.1	0.5	1	2	5	10
RSh	$E\tau$	**527.2**	**390.3**	**314.0**	**232.3**	**145.8**	**99.4**
	$\sigma\tau$	**447.7**	**284.6**	**204.6**	**135.7**	**72.3**	**43.9**
CUSUM	$E\tau$	**517.2**	**388.9**	**316.8**	**235.9**	**150.3**	**103.3**
	$\sigma\tau$	**457.1**	**281.0**	**223.7**	**143.6**	**79.6**	**46.4**
Exp	$E\tau$	**489.9**	**342.4**	**266.6**	**197.4**	**127.4**	**88.8**
	$\sigma\tau$	**407.6**	**237.2**	**167.5**	**113.5**	**62.4**	**38.1**
WL	$E\tau$	**485.2**	**369.8**	**240.8**	**188.5**	**118.5**	**95.3**
	$\sigma\tau$	**499.5**	**343.3**	**229.8**	**150.1**	**81.5**	**44.6**

Hmax		50	100	200	500	1000
RSh	$E\tau$	**37.9**	**24.5**	**16.3**	**9.6**	**6.6**
	$\sigma\tau$	**12.2**	**7.0**	**3.8**	**1.9**	**1.2**
CUSUM	$E\tau$	**39.3**	**24.9**	**16.4**	**9.3**	**6.3**
	$\sigma\tau$	**12.7**	**7.4**	**4.1**	**2.0**	**1.2**
Exp	$E\tau$	**35.2**	**23.6**	**16.2**	**9.7**	**6.7**
	$\sigma\tau$	**10.8**	**6.6**	**3.8**	**1.9**	**1.1**
WL	$E\tau$	**34.5**	**25.6**	**17.0**	**10.7**	**6.8**
	$\sigma\tau$	**15.2**	**7.4**	**4.5**	**2.1**	**1.8**

Multivariate models

Multivariate regression

The following system of simultaneous equations was considered:

$$
\begin{aligned}
y_i &= c_0 + c_1 y_{i-1} + c_2 z_{i-1} + c_3 x_i + \epsilon_i \\
z_i &= d_0 + d_1 y_i + d_2 x_i + \xi_i \\
x_i &= 0.5 x_{i-1} + \nu_i \\
\epsilon_i &= 0.3 \epsilon_{i-1} + \eta_i,
\end{aligned}
$$

where $\xi_i, \nu_i, \eta_i,\ i = 1, 2, \ldots$ **are independent** $\mathcal{N}(0,1)$ **r.v.'s.**

So $(y_i, z_i)'$ **is the vector of endogenous variables,** x_i **is the exogenous variable, and** $(1, y_{i-1}, z_{i-1}, x_i)'$ **is the vector of predetermined variables of this system.**

The dynamics of this system is characterized by the following vector of coefficients: $\mathbf{u} = [c_0\ c_1\ c_2\ c_3\ d_0\ d_1\ d_2]$**. The initial stationary dynamics is characterized by the coefficients** $[0.1\ 0.5\ 0.3\ 0.7\ 0.2\ 0.4\ 0.6]$**.**

Table 2. Decision bounds of the nonparametric test (SSE model)

N	**20**	**50**	**100**	**200**	**300**	**400**
$p = 0.95$	**0.99**	**0.67**	**0.49**	**0.39**	**0.30**	**0.25**
$p = 0.99$	**1.50**	**0.85**	**0.65**	**0.47**	**0.38**	**0.32**
th	**1.45**	**0.91**	**0.65**	**0.46**	**0.37**	**0.32**

In the following series of experiments the models with changes in the coefficient d_2 were considered. For each sample volume N and the chosen values of the decision threshold th, the estimates of the 1st ('false alarm') and the 2nd type error probabilities were computed, as well as the average delay time in change-point detection in $k = 5000$ independent trials. The results are reported in Table 3.

Table 3. Performance characteristics of the nonparametric test (SSE model, 5000 replications, pr - empirical false alarm rate, w_2 - type 2 error, $E\tau$ - average delay time)

abrupt	N	**20**	**50**	**100**	**200**
th		**1.50**	**0.85**	**0.65**	**0.47**
pr		**0.02**	**0.03**	**0.02**	**0.03**
$d_2 = 0.95$	w_2	**0.09**	**0**	**0**	**0**
	$E\tau$	**3.80**	**1.71**	**1.21**	**1.01**
$d_2 = 0.9$	w_2	**0.19**	**0.02**	**0**	**0**
	$E\tau$	**4.83**	**2.46**	**1.04**	**1.10**
$d_2 = 0.8$	w_2	**0.45**	**0.15**	**0.04**	**0**
	$E\tau$	**6.52**	**9.20**	**13.2**	**11.2**

gradual ($\dot{d}_2 = \Delta d$)		$N = 100$
th		**0.65**
pr		**0.02**
$\Delta d = 10^{-4}$	w_2	**0.23**
	$E\tau$	**97.4**
$\Delta d = 10^{-3}$	w_2	**0.02**
	$E\tau$	**27.8**
$\Delta d = 10^{-2}$	w_2	**0**
	$E\tau$	**5.6**

State-space models

Lai's example: changes in means

Now let us consider the following example of the multivariate state space model:

$$x_{t+1} = Fx_t + (\theta,\ 0)' I_{\{t \geq r\}} + w_t$$
$$y_t = (1,\ 0)x_t + \epsilon_t,$$

where $F = \begin{pmatrix} 0.7 & 0.1 \\ 0 & 0.7 \end{pmatrix}$ **and** w_t, ϵ_t **are independent Gaussian with zero means,** $Var(\epsilon_t) = 1$, $Cov(w_t) = \begin{pmatrix} 0.745 & -0.07 \\ -0.07 & 0.51 \end{pmatrix}$ **and** θ **is an unknown scalar parameter representing the change magnitude.**

This example was considered in Lai and Shan (1999). The threshold c **of the GLR rule was chosen in order to obtain the average time between false alarms equal to 500. Each result in this table is based on 1000 independent trials.**

Table 4. Comparison of GLR and nonparametric test

θ	$GLR\, rule;\ c = 5.35$	$Nonparametric\, test;\ N = 30,\ c = 0.52$
0	**509**	**507**
1.5	**3.28**	**4.23**
1.2	**3.88**	**5.07**
1.0	**4.48**	**7.13**
0.9	**4.89**	**9.19**
0.8	**5.41**	**11.41**
0.7	**6.25**	**15.74**
0.6	**7.65**	**18.70**
0.5	**9.79**	**23.73**
0.4	**13.13**	**31.65**
0.3	**21.41**	**46.56**
0.2	**44.06**	**72.22**
0.1	**144**	**153.28**

Non-Gaussian distributions; changes in coefficients

The baseline model is of the following functional form

$$x_{t+1} = Fx_t + w_t$$
$$y_t = (1,\ 0)x_t + \epsilon_t,$$

where $F = \begin{pmatrix} 0.7 & 0.1 \\ 0 & 0.7 \end{pmatrix}$ **but** w_t **has the multivariate t-distribution with the correlation matrix** $\sigma = \begin{pmatrix} 1.0 & 0.8 \\ 0.8 & 1.0 \end{pmatrix}$ **and three degrees of freedom;** ϵ_t **has the standard uniform d.f. on the segment** $[0; 1]$**.**

At an unknown change-point m **the matrix** F **changes to** $G = \begin{pmatrix} \theta & 0.1 \\ 0 & 0.7 \end{pmatrix}$**, where** $\theta \neq 0.7$**.**

Table 5. Changes in coefficients of state-space models; non-Gaussian distributions $(N = 30,\ c = 2.02,\ \theta_n = \theta_{n-1} + \Delta\theta,\ \theta_0 = 0.7)$

abrupt	θ	**0.7**	**0.8**	**0.9**	**1.0**	**1.1**	**1.2**
	ET	**507**	**378.72**	**104.61**	**29.52**	**16.66**	**11.30**
gradual	$\Delta\theta$	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	**0.5**
	ET	**409.4**	**325.9**	**142.0**	**32.6**	**9.4**	**4.5**

Non-linear state-space model

Consider the following state-space model:

$$y_i = d_1 + d_2 x_i + d_3 x_i^2 + \eta_i$$
$$x_i = \frac{d_4}{d_5 + d_6 e^{-x_{i-1}}} + \xi_i,$$

where $x_i, y_i,\ i = 1, 2, \dots$ **is the state variable and the observed variable, respectively;** η_i, ξ_i **are independent Gaussian random sequences** $N(0,1)$ **and the baseline model is characterized by the following parameters:** $d_1 = 1,\ d_2 = 0.3,\ d_3 = 0.1,\ d_4 = 1,\ d_5 = 1,\ d_6 = 1$**.**

We consider possible changes in coefficients of this model at an unknown change-point m**. In experiments we estimate the false alarm probability (1st type error)** pr_1**, the probability of the 2nd type error** w_2**, and the average delay time in change-point detection** $s_N = E(\tau_N - m | \tau_N > m)$**. Each value was computed as an average in 5000 independent Monte Carlo trials.**

Table 6. Non-linear state-space model

abrupt	N	**100**	**200**	**300**	**500**	**700**
C		**0.33**	**0.23**	**0.21**	**0.19**	**0.17**
pr_1		**0.04**	**0.05**	**0.04**	**0.04**	**0.04**
$d_2 = 0.8$	w_2	**0.44**	**0.11**	**0.07**	**0.05**	**0**
	s_N	**42.2**	**58.9**	**79.1**	**96.8**	**116.1**
$d_4 = 1.3$	w_2	**0.20**	**0.12**	**0.08**	**0.05**	**0**
$d_5 = 0.5$	s_N	**59.6**	**79.4**	**91.0**	**115.7**	**124.2**
$d_3 = 0.4$	w_2	**0.11**	**0.02**	**0.03**	**0.02**	**0**
	s_N	**84.2**	**80.3**	**91.9**	**117.7**	**123.3**

gradual $\dot{d}_2 = \Delta d$		$N = 300$
th		**0.21**
pr		**0.04**
$\Delta d = 10^{-4}$	w_2	**0.04**
	s_N	**56.2**
$\Delta d = 10^{-3}$	w_2	**0**
	s_N	**28.8**
$\Delta d = 10^{-2}$	w_2	**0**
	s_N	**6.3**

Conclusion

1. The a priori informational inequalities for the main performance measures in sequential detection of abrupt and gradual changes for univariate and multivatiate stochastic models are proved.

2. It is usually assumed that statistical characteristics of observations change instantaneously from one stationary level into another at some unknown points. The optimality and asymptotic optimality of CUSUM, GRSh and "window-limited" tests was established only under these assumptions. However, in many practically relevant situations of *gradual changes* in statistical characteristics of data the asymptotic optimality of CUSUM, GRSh and other well-known tests may be violated. In this paper we demonstrate that CUSUM and GRSh tests will be asymptotically optimal in the problem of "early detection" only in the classic situation of an abrupt change from one known density function $f_0(\cdot)$ to another (a priori known) density function $f_1(\cdot)$.

3. The asymptotically optimal methods of early change-point detection in univariate and multivariate stochastic models are proposed.

4. The Monte Carlo tests are performed for the proposed methods of early change-point detection in univariate and multivariate models.

Thank you for attention