Inductive Conformal Martingales for Change-Point Detection

Denis Volkhonskiy, Evgeny Burnaev, Ilia Nouretdinov,
Alexander Gammerman, Vladimir Vovk

Yandex School of Data Analysis
Skolkovo Institute of Science and Technology
Royal Holloway, University of London

dvolkhonskiy@gmail.com

June 19, 2017

Quickest Change-Point Detection

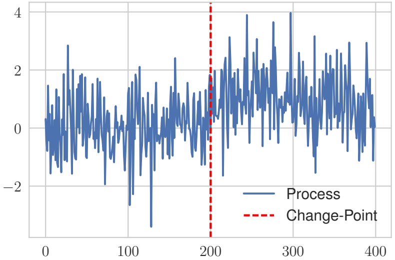
Change-Point Problem

The **change-point detection problem** is formulated as following:

- $z_1, z_2, \ldots, z_n, \ldots$ are independent random variables;
- $z_1, z_2, \ldots, z_{\theta-1}$ are each distributed according to a distribution $f_0(z)$;
- $z_\theta, z_{\theta+1}, \ldots$ are each distributed according to a distribution $f_1(z)$;
- Change-Point (CP) $2 \leq \theta \leq \infty$ is unknown;

The task is to find the **stopping time** τ, such that

- **Probability of False Alarm:** $\mathbb{P}(\tau \leq \theta) \leq \alpha$ for all θ
- **Mean Delay:** $\mathbb{E}(\tau - \theta \mid \tau > \theta) \to \min_\tau$



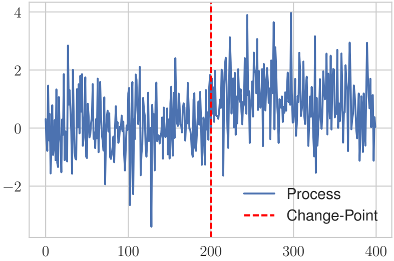
Change-Point Problem

The **change-point detection problem** is formulated as following:

- $z_1, z_2, \ldots, z_n, \ldots$ are independent random variables;
- $z_1, z_2, \ldots, z_{\theta-1}$ are each distributed according to a distribution $f_0(z)$;
- $z_\theta, z_{\theta+1}, \ldots$ are each distributed according to a distribution $f_1(z)$;
- Change-Point (CP) $2 \leq \theta \leq \infty$ is unknown;

The task is to find the **stopping time** τ, such that

- **Probability of False Alarm:** $\mathbb{P}(\tau \leq \theta) \leq \alpha$ for all θ
- **Mean Delay:** $\mathbb{E}(\tau - \theta \mid \tau > \theta) \to \min_{\tau}$



Quickest Change-Point Detection

- **Important applications:** healthcare, security, production, equipment maintenance, Internet of things, traffic routing, etc.
- **Challenges**:
  - Large volumes of time-series data;
  - Complex, intractable or unknown dynamic models;
  - Apparent non-stationary and quasi-periodicity.

Automatic Change-Point detection is critical in today's world where the sheer volume of data makes it impossible to tag change-point manually

EEG for a human

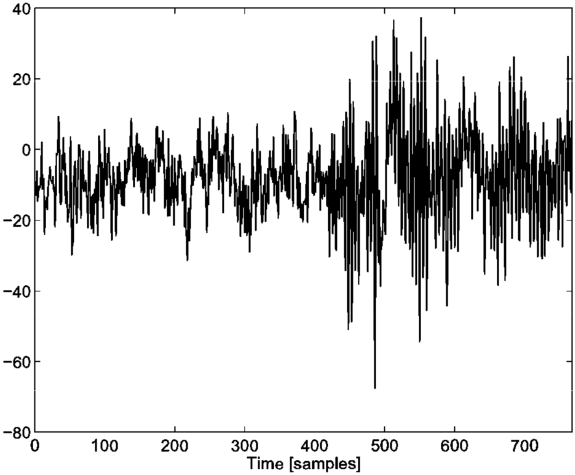


Figure: EEG for a human in a room, where the light is turned of at time 387

Eathquake Prediction

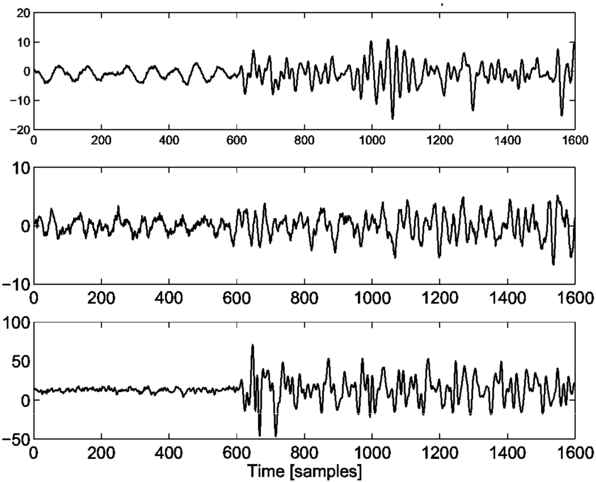


Figure: Seismological signals. Earthquake starts at time 600

Standard Approaches

Standard algorithms for quickest Change-Point detection: Cumulative Sum (CUSUM), Shiryaev-Roberts, Posterior Probability statistics.

- Based on **likelihood ratio** $\frac{L_n^\theta}{L_n}$, where

$$L_n^\theta = \prod_{i=1}^{\theta-1} f_0(z_i) \prod_{i=\theta}^{n} f_1(z_i) \quad (1)$$

  the likelihood of observations $z_1, \ldots, z_n$ when $\theta \in [1, n]$, and by

$$L_n = \prod_{i=1}^{n} f_0(z_i) \quad (2)$$

  the likelihood of observations $z_1, \ldots, z_n$ without Change-Point.
- **Our goal** is to provide distribution-free algorithm for Change-Point detection, that is based on Conformal Martingales.

Conformal Martingales

Conformal Predictors

Conformal martingales:

- **Non-Conformity measure** (the measure of strangeness)
  $\alpha_i = A(z_i, \{z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n\})$
  - Example: distance to the nearest neighbor.

- **P-values**

$$p_n = p(z_n, z_{n-1}, \dots, z_1) = \frac{\#\{i : \alpha_i > \alpha_n\} + U\#\{i : \alpha_i = \alpha_n\}}{n}$$

  where $U \sim Uniform[0, 1]$ independent of $z_1, z_2, \dots$
  - Small p-values $\Rightarrow$ strange objects (other distribution)

Conformal Predictors

Conformal martingales:

- **Non-Conformity measure** (the measure of strangeness)
  $\alpha_i = A(z_i, \{z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n\})$
  - Example: distance to the nearest neighbor.

- **P-values**

$$p_n = p(z_n, z_{n-1}, \ldots, z_1) = \frac{\#\{i : \alpha_i > \alpha_n\} + U\#\{i : \alpha_i = \alpha_n\}}{n}$$

where $U \sim Uniform[0, 1]$ independent of $z_1, z_2, \ldots$

  - Small p-values $\Rightarrow$ strange objects (other distribution)

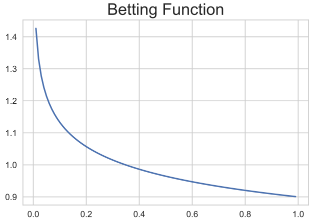
Conformal Martingales

Theorem

If p-values are **not independent and uniformly distributed** *in* $[0, 1]$, *then* $z_1, \ldots, z_n, \ldots$ **don't satisfy the i.i.d.** *assumption. [Vovk et al., 2003]*

- **Conformal Martingale:** $S_n = \prod_{i=1}^{n} g_i(p_i)$, $n = 1, 2, \ldots,$ where $g_i(p_i) = g_i(p_i \mid p_1, \ldots, p_{i-1})$ is what we call a **betting function**.

1. Strange objects (not i.i.d., Change-Point);
2. p-values are not Uniform;
3. penalize with betting function;
4. Martingale grows.



Betting Functions

- **Constant** Betting function;

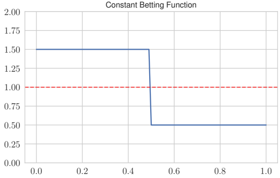
$$g(p) = \begin{cases} 1.5, & \text{if } p \in [0, 0.5), \\ 0.5, & \text{if } p \in [0.5, 1]. \end{cases}$$

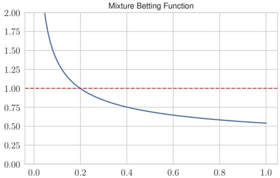
- **Mixture** Betting Function;

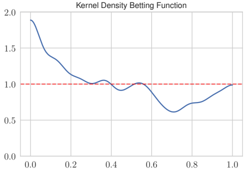
$$g(p) = \int_0^1 \varepsilon p^{\varepsilon - 1} d\varepsilon.$$

- **Kernel** Density Betting Functions;

$$g_n(p_n) = K_{p_{n-L},\dots,p_{n-1}}(p_n).$$







Contribution: Inductive Conformal Martingales for Change-Point detection

Conformal Martingale:

+ Theoretically valid;
− Computationally very inefficient;
− Wasn't designed for quickest change-point detection.

Inductive Conformal Martingales:

- Computation of **non-conformity scores** with fixed training set:

$$\alpha_i = A(z_i, \{z^*_{-(m-1)}, \dots, z^*_0\});$$

- **Increase efficiency** a lot: no need in *Leave-One-Out*-like method;
- Save validity.

Contribution: Inductive Conformal Martingales for Change-Point detection

Conformal Martingale:

+ Theoretically valid;
− Computationally very inefficient;
− Wasn't designed for quickest change-point detection.

Inductive Conformal Martingales:

- Computation of **non-conformity scores** with fixed training set:

$$\alpha_i = A(z_i, \{z^*_{-(m-1)}, \dots, z^*_0\});$$

- **Increase efficiency** a lot: no need in *Leave-One-Out*-like method;
- Save validity.

Validity of Inductive Conformal Martingales

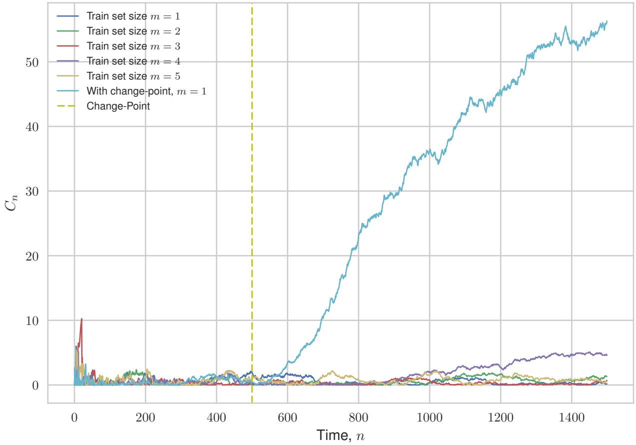


Figure: Validity Test of ICM: case of small train sets

Contribution: adaptation of ICMs for Change-Point detection

The **stopping rule** is $\tau_{\mathrm{CM}}^{S_n} = \inf\{n : S_n \geq h\}$ for some constant h.

Problems:

– It takes a lot of time to reach the change-point;

– Martingale can decrease to $-\infty$.

We propose:

- "Cut" the martingale:

$$C_n = \max\{0, C_{n-1} + \log(g_n(p_n))\}, \tag{3}$$

  where p_n is a p-value, and g_n is a betting function

- Stopping rule: $\tau_{\mathrm{CM}}^{C_n} = \inf\{n : C_n \geq h\}$

\+ This modification performs better in terms of the *mean delay* until Change-Point detection;

Contribution: adaptation of ICMs for Change-Point detection

The **stopping rule** is $\tau_{\mathrm{CM}}^{S_n} = \inf\{n : S_n \geq h\}$ for some constant h.

Problems:

- It takes a lot of time to reach the change-point;
- Martingale can decrease to $-\infty$.

We propose:

- "Cut" the martingale:

$$C_n = \max\{0, C_{n-1} + \log(g_n(p_n))\}, \tag{3}$$

  where p_n is a p-value, and g_n is a betting function
- Stopping rule: $\tau_{\mathrm{CM}}^{C_n} = \inf\{n : C_n \geq h\}$

\+ This modification performs better in terms of the *mean delay* until Change-Point detection;

ICM example

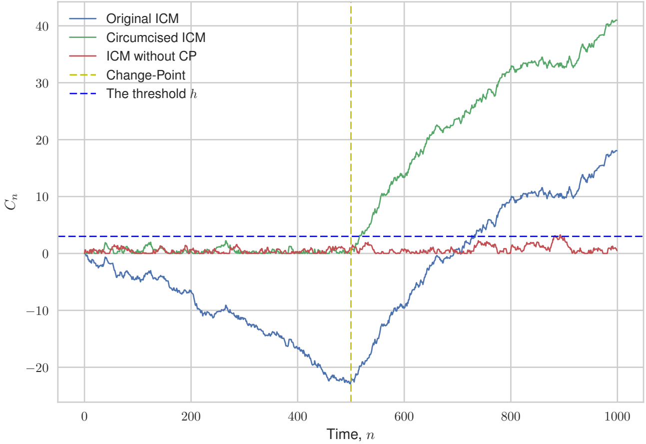
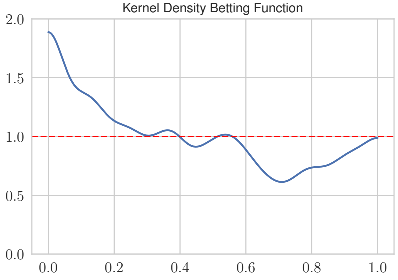


Figure: Example of the ICM in case of data with Change-Point (at $\theta = 500$) and without Change-Point

Contribution: Kernel Betting Function

Kernel Density Betting Function ([Fedorova et al., 2012], estimated from p-values seen before):

- – Computationally inefficient;
- – Need some time to update from uniform distribution after change-point;
- + Theoretically has the best possible growing rate [Fedorova et al., 2012].



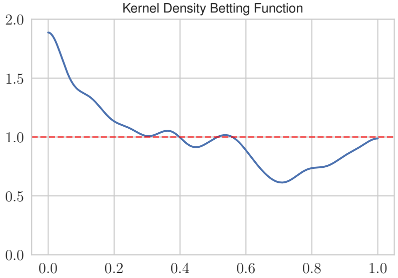
We propose to **precompute** it on p-values for data with an example of a change-point:

- + Computationally efficient (no need to recompute);
- + Faster Change-Point detection.

Contribution: Kernel Betting Function

Kernel Density Betting Function ([Fedorova et al., 2012], estimated from p-values seen before):

- Computationally inefficient;
- Need some time to update from uniform distribution after change-point;
+ Theoretically has the best possible growing rate [Fedorova et al., 2012].



We propose to **precompute** it on p-values for data with an example of a change-point:

+ Computationally efficient (no need to recompute);
+ Faster Change-Point detection.

Experiments

Oracles for Change-Point detection

Classical approach (CUSUM, Shiryaev-Roberst):

- **Optimal** in terms of Mean Delay;
- Need to know the data model.

Conformal Martingales:

- Only i.i.d assumption;
- We expect them to be worse.

Oracles, based on standard Change-Point detection algorithms:

- We assume the distribution class to be known;
- We don't know the parameters;
- Likelihood:

$$\overline{L}_n = \int \prod_{i=1}^{n} f(z_i \mid \mathbf{c}_0) q(\mathbf{c}_0) d\mathbf{c}_0. \tag{4}$$

Classical approach (CUSUM, Shiryaev-Roberst):

- **Optimal** in terms of Mean Delay;
- Need to know the data model.

Conformal Martingales:

- Only i.i.d assumption;
- We expect them to be worse.

Oracles, based on standard Change-Point detection algorithms:

- We assume the distribution class to be known;
- We don't know the parameters;
- Likelihood:

$$\overline{L}_n = \int \prod_{i=1}^{n} f(z_i \mid \mathbf{c}_0) q(\mathbf{c}_0) d\mathbf{c}_0. \quad (4)$$

Experimental Setup

Non-Conformity Measures:

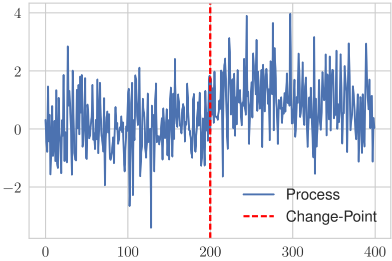
- *k Nearest Neighbors (kNN NCM)* — average distance to k nearest neighbors.
- *Likelihood Ratio (LR NCM)* — the value of likelihood ratio $\frac{f_1(x)}{f_0(x)}$.

Parameters:

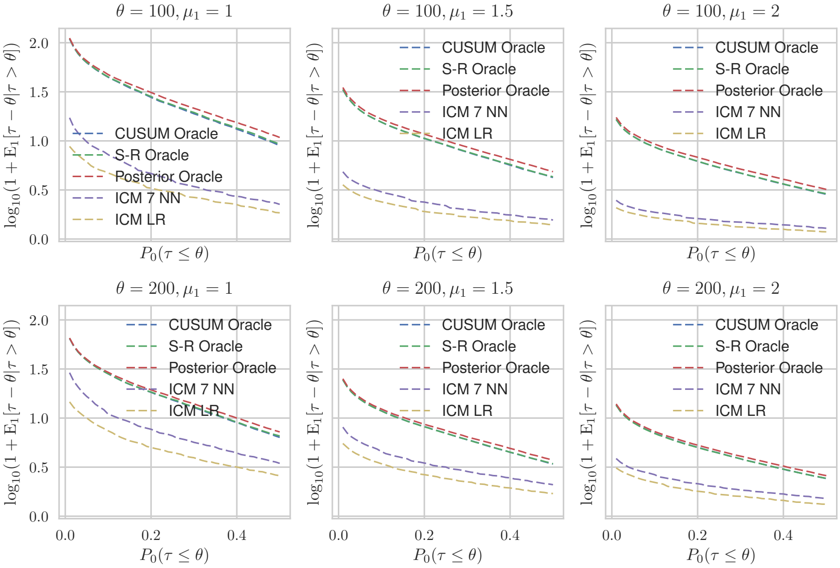
- $f_0 \sim \mathcal{N}(0, 1)$;
- $f_1 \sim \mathcal{N}(m_1, 1)$, $m_1 \in \{1, 1.5, 2\}$;
- $\theta \in [100, 200]$

As a **performance characteristics** we use:

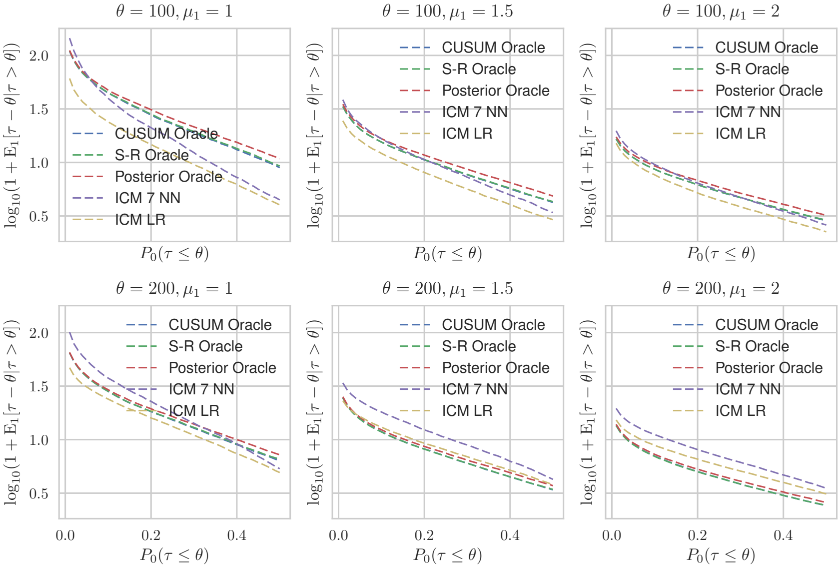
- **Mean delay** until Change-Point detection $\mathbb{E}_1(\tau - \theta \mid \tau > \theta)$,
- **Probability of False Alarm (FA)** $\mathbb{P}_0(\tau \leq \theta)$.



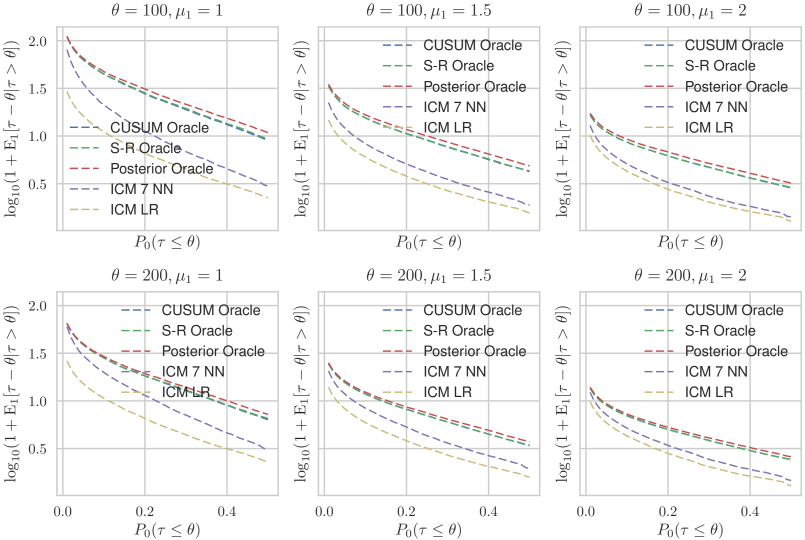
Mixture Betting Function. Comparison with Oracles



Kernel Density Betting Function. Comparison with Oracles



Precomputed Kernel Density Betting Function. Comparison with Oracles



Kernel and Precomputed Kernel Betting Functions

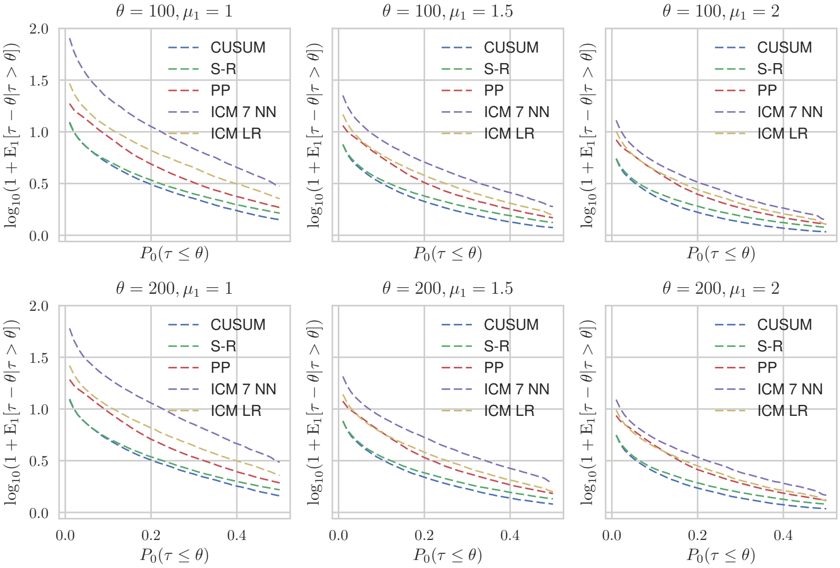
Table: Mean Delay. Kernel Density Betting Function

Param.\Probab. of FA	5%					10%				
	ICM LR	**ICM kNN**	**CUSUM Oracle**	**S-R Oracle**	**Posterior Oracle**	**ICM LR**	**ICM kNN**	**CUSUM Oracle**	**S-R Oracle**	**Posterior Oracle**
$\theta = 200, \mu_1 = 1$	**30.06**	54.14	37.78	37.80	38.73	**22.90**	36.57	27.24	27.24	28.25
$\theta = 200, \mu_1 = 1.5$	15.44	22.02	14.62	**14.52**	15.16	12.08	17.13	10.85	**10.81**	11.36
$\theta = 200, \mu_1 = 2$	10.00	12.81	8.02	**7.98**	8.30	7.83	10.15	6.00	**5.97**	6.28

Table: Mean Delay. Precomputed Kernel Density Betting Function

Param.\Probab. of FA	5%					10%				
	ICM LR	**ICM kNN**	**CUSUM Oracle**	**S-R Oracle**	**Posterior Oracle**	**ICM LR**	**ICM kNN**	**CUSUM Oracle**	**S-R Oracle**	**Posterior Oracle**
$\theta = 200, \mu_1 = 1$	**14.14**	28.70	37.78	37.80	38.73	**9.65**	18.91	27.24	27.24	28.25
$\theta = 200, \mu_1 = 1.5$	**7.24**	10.80	14.62	14.52	15.16	**4.92**	7.39	10.85	10.81	11.36
$\theta = 200, \mu_1 = 2$	**4.90**	6.15	8.02	7.98	8.30	**3.29**	4.18	6.00	5.97	6.28

Comparison with Optimal detectors



Conclusion

- An **adaptation** of Conformal Martingales for change-point detection problem was proposed;
- We demonstrated the **efficiency** of this approach by comparing it with natural oracles, which are likelihood-based change-point detectors;
- We **proposed and compared** several approaches to calculating a **betting function**;
- We also compared Inductive Conformal Martingales with methods that are optimal for known pre- and post-CP distributions, such as CUSUM, Shiryaev-Roberts and Posterior Probability statistics.
- Paper: [Volkhonskiy et al., 2017]

References I

Fedorova, V., Gammerman, A., Nouretdinov, I., and Vovk, V. (2012).
Plug-in martingales for testing exchangeability on-line.
In *Proceedings of the Twenty-Ninth International Conference on Machine Learning (ICML)*.

Volkhonskiy, D., Burnaev, E., Gammerman, A., Nouretdinov, I., and Vovk, V. (2017).
Inductive conformal martingales for change-point detection.
In *Proceedings of the Sixth Symposium on Conformal and Probabilistic Prediction with Application (COPA 2017)*.

Vovk, V., Nouretdinov, I., and Gammerman, A. (2003).
Testing exchangeability on-line.
In *Proceedings of the Twentieth International Conference on Machine Learning (ICML)*, volume 12, pages 768–775.