Multiscale parametric approach for change point detection

Anonymous Author(s)
Affiliation
Address
email

Abstract

This work presents a novel algorithm for change point detection, that can be applied for analysis of data of unknown nature. It is based on likelihood-ratio test statistics, as its behaviour can be described in terms of χ^2-distribution even in case of model misspecification. To discover change point in the quickest way, statistics is calculated in a set of running windows of different scales. Algorithm is self-tuned: critical values are justified by data and calculated with multiplier bootstrap procedure. To make the method more robust for outliers, the concept of change-point patterns is presented.

1 Introduction

The problem of change point detection has a wide range of applications, that varies from life-critical to pure scientific ones. It appears each time one needs to explore a set of random data and make a decision about homogeneity of its structure. In other words, the problem can be stated as two following questions: were there any structural changes in the nature of observed data? At which moments, if so? These and similar questions arise in many areas of theoretical and engineering research. For example, algorithms of change point detection are used for identification and elimination of faults of aeroplane's navigation system, so as to perform better geolocation Nikiforov [2003]. There are many other examples of real-world applications, such as analysis of stock markets Lavielle and Teyssiere [2006] or anomaly detection in computer traffic Tartakovsky et al. [2006], Casas et al. [2010]. This work mainly focuses on the *sequential* or *online* change point detection. In this case the data is aggregated from running random process. Let Y_τ be the observation at the current moment τ, $\tau > 1$. The moment τ is a *change point*, if stochastic properties of observed signal have undergone some changes:

$$\begin{cases} Y_t \backsim I\!P_1 & t < \tau, \\ Y_t \backsim I\!P_2 & t \geq \tau. \end{cases}$$

The goal is to find such structural brakes as soon as possible. This problem can be solved by using likelihood ratio test (LRT). This idea of application LRT for change point detection goes back at least as far as Quandt [1960]. Liu et al. [2008], Zou et al. [2007] investigate LRT for change point detection in nonparametric cases. In general, nonparametric approaches have greater delay in change point detection than their parametric alternatives. Introduction of *parametric assumption*: $I\!P_1, I\!P_2 \in (I\!P(\theta), \theta \in \Theta \subseteq I\!R^p)$ allows to reduce average time of delay. The state-of-the-art review of parametric models based on LRT and its application to economics and bio-informatics are presented by Chen and Gupta [2012]. The paper Gombay [2000] explores how it can be used for sequential change point detection in case $I\!P(\theta)$ is exponential family. Lai [1995] proposes 'window-limited LRT schemes': test statistics is calculated in rolling window. This concept naturally expands to on-line change point detection. Many works are dedicated to asymptotic behaviour of LRT, e.g. Jandhyala and Fotopoulos [1999] obtains lower and upper bounds for distribution of asymptotic maximum likelihood estimator. The work of Kim [1994] provides a very detailed study of its asymptotic behaviour in linear regression models. Similar results for a change in mean of a Gaussian process are in Fotopoulos et al. [2010]. As far as the authors know, the most comprehensive study of the LRT behaviour is done by Fan et al. [2001]. It shows that LRT is asymptotically χ^2 distributed. The idea of the proof is based on the *Wilk's phenomenon* Wilks [1938], Boucheron and Massart [2011]. The aim of the present paper is to described the LRT behaviour in finite-sample case using non-asymptotic Wilks and Fischer theorems Spokoiny [2012]. It is shown that the distribution of LRT is similar to ordinary χ^2 under H_0. Otherwise if the sample is not homogeneous, the systematic drift of LRT appears. Thus, under H_1 the test statistic behaves like non-central χ^2 random value. This drift is referred to as a *change-point pattern*. The result holds for both correct and misspecified parametric models. The cornerstone of the new change point detection procedure is the concept of change-point pattern. The geometry of a pattern depends on a type of switch between distributions the data obeys before and after a change respectively. Three examples are presented at the Fig. 1. The triangle pattern appears in case of an abrupt switch from $\mathbb{P}_{\theta_1^*}$ to $\mathbb{P}_{\theta_2^*}$. A smooth transition between two regimes entails the trapezium change-point pattern. And an inverted triangle pattern appears due to a change in coefficients of linear regression. The control of a change-point pattern instead of a single LRT-value allows to reduce false-alarm rate to zero.

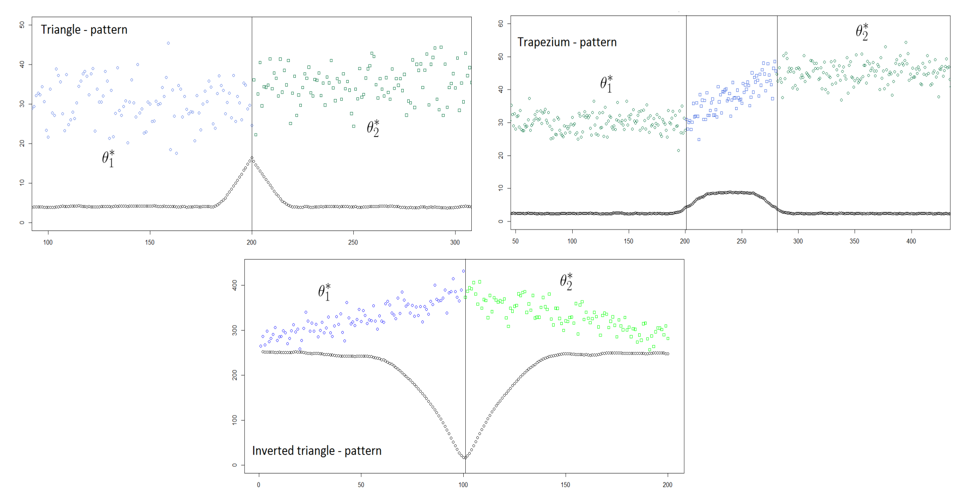

Figure 1: Type of change point and the geometry of change-point pattern

Any procedure of a change point detection requires information about the nature of the data observed after a structural break. The time of information aggregation is referred to as *detection delay*. The proposed algorithm provides optimal *detection delay* in the class of similar methods. The optimality is achieved by introduction of *multiscale* approach. The technique is popular, e.g. Frick et al. [2014], Spokoiny [2009] and allows analysis of the data using different scales simultaneously. The procedure proposed below computes the test statistics in rolling windows of different sizes and controls change-point pattern at each level. The greater number of scales at which a change-point pattern is detected, the more sure algorithm is, that there is a change point.

Under some assumptions on the frequency of change points provided in Section 3, the methods is applied to the *multiple* change point problem. The survey on existing models can be found in Chib [1998]. The last, but not at all the least feature of the proposed algorithm is that critical values for test statistic are computed in a data-driven way. The idea is to use the multiplier bootstrap procedure Chernozhukov et al. [2013]. The work of Spokoiny and Zhilova [2014]shows that it can be used for the construction of confides intervals even in case of a misspecified parametric model. Despite the fact, that theoretical properties of data-driven critical values are beyond the scope of this paper, the procedure of computation is presented in Algorithm 3.

The paper is organized as follows. Section 2 presents the description of the algorithm. Theoretical properties of the procedure are discussed in Section 3. Section 4 compares the new algorithm with existing methods using simulated data sets. It also illustrates the performance of the method on a real data set.

2 Algorithm

This section provides the description of the proposed algorithm. Let $(I\!P(\theta),\ \theta \in \Theta \subseteq \mathbb{R}^p)$ be a local parametric assumption about the nature of data inside a window $(Y_{t-h}, ..., Y_{(t+h-1)})$. The generalised likelihood ratio test is

$$T_h(t) = \sup_{\theta \in \Theta} L(\theta; Y_{t-h}, ..., Y_{t-1}) + \sup_{\theta \in \Theta} L(\theta; Y_t, ..., Y_{t+h-1})$$
$$- \sup_{\theta \in \Theta} \{L(\theta; Y_{t-h}, ..., Y_{t-1}) + L(\theta; Y_t, ..., Y_{t+h-1})\},$$

where $L(\theta; \cdot)$ is a log-likelihood function. To control a change point pattern, the procedure monitors $2h$ values of the LRT simultaneously:

$$\mathbb{T}_h(t) = (\sqrt{2T_h(t-h)}, ..., \sqrt{2T_h(t+h-1)}).$$

The test statistics in hand is a convolution of $\mathbb{T}_h(t)$ with a predefined change-point pattern $P_h \in I\!R^{2h}$.

$$\widehat{\mathbb{T}}_h(t) = \langle \mathbb{T}_h(t), P_h \rangle .$$

Under *online* framework, the algorithm marks a time moment τ as a change point, if test statistics $\widehat{T}_h(\tau + h)$ exceeds critical value $z(h)$ at the moment $\tau + h$:

$$\{\tau : \widehat{\mathbb{T}}_h(\tau - h) > z(h)\}.$$

Under *offline* setting, τ is a change point if

$$\{\tau = \operatorname*{argmax}_{t \in \{1,...,M\}} \sum_{h \in H} w_h \widehat{\mathbb{T}}_h(t), \quad \exists h \in H : \widehat{\mathbb{T}}_h(\tau) > z(h)\},$$

where M is a time moment till which the data is observed, $\{w_h\}_{h \in H}$ – weights for window size preferences.

In both cases the procedure repeats itself simultaneously on different scales $H = \{h_1, ..., h_N\}$. The greater the number k of such scales $h_{i_1}, ..., h_{i_k}$ where τ is marked as a change point, the more sure algorithm is, that τ the *true* change point is.

Algorithm 1, 2 summarises these ideas for sequential case. Here the current moment is supposed to be $\tau + 2h_N - 2$ and a candidate for the change point is τ.

Algorithm 3 presents the procedure of calculation of a critical value z_h for a fixed window size $2h$. Let $\mathbb{Y} = (Y_1, ..., Y_M)$ be a training set. Let weighted likelihood function be a convolution of i.i.d likelihood components and a weight vector $(u_1, \ldots, u_M)$:

$$L^\flat(\theta; Y_1, \ldots, Y_M) = \sum_{m=1}^{M} u_m l(\theta, Y_m), \tag{Lb}$$

where $\{u_m\}_{m=1}^{M}$ are i.i.d. and $I\!E u_m = \operatorname{Var} u_m = 1$. Then bootstrapped generalised likelihood ratio test is

$$T_h^\flat(t) = \sup_{\theta \in \Theta} L^\flat(\theta; Y_{t-h}, \ldots, Y_{t-1}) + \sup_{\theta \in \Theta} L^\flat(\theta; Y_t, \ldots, Y_{t+h-1}) \tag{Tb}$$
$$- \sup_{\theta \in \Theta} \{L^\flat(\theta; Y_{t-h}, \ldots, Y_{t-1}) + L^\flat(\theta + \widehat{\theta}_{12}; Y_t, \ldots, Y_{t+h-1})\},$$

$$\widehat{\theta}_{12} = \operatorname*{argmax}_{\theta} L(\theta; Y_t, \ldots, Y_{t+h-1}) - \operatorname*{argmax}_{\theta} L(\theta; Y_{t-h}, \ldots, Y_{t-1})$$

$Q_h(t) = 0$ – change point signals;
get z(h) by Algorithm 3;
foreach *window position* t **do**
  foreach h **do**
    add $T_h(t)$ to $\mathbb{T}_h$;
    $\widehat{\mathbb{T}}_h = \langle \mathbb{T}_h(t-h), P \rangle$;
    if $\widehat{\mathbb{T}}_h > z(h)$ *and* $Q_{i \le h}(t - 2h : t) = 0$ **then**
      $Q_h(t) = 1$;
    end
  end
  if $\max_h Q_h(t) = 1$ **then**
    t is change point;
  end
end

Algorithm 1: LRTOnline.

S – change points set; w_j – window size weights;
function FindCP($Y_1, \dots, Y_M$):
get z(h) by Algorithm 3;
foreach h **do**
  foreach *window position* t **do**
    compute $T_h(t)$;
  end
  foreach τ **do**
    $\widehat{\mathbb{T}}_h(\tau) = \langle \mathbb{T}_h(\tau), P \rangle$;
  end
end
$\tau = \text{argmax}_\tau \sum_{h \in H} w_h \widehat{\mathbb{T}}_h(\tau)$;
if $\exists h : \mathbb{T}_h(\tau) > z(h)$ **then**
  add τ to S;
  FindCP($Y_1, \dots, Y_\tau$); FindCP($Y_\tau, \dots, Y_M$);
end

Algorithm 2: LRTOffline.

Data: $(Y_1, \dots, Y_M)$, h, P, S – weights generation counts
Result: f_h^b – bootstrap distribution of maximal convolution inside the window
for $s = 1$ **to** S **do**
  generate $u = (u_1, \dots, u_M)$;
  foreach *window position* t **do**
    compute $T_h^b(t)$;
  end
  foreach τ **do**
    $\widehat{\mathbb{T}}_h^b(\tau) = \langle \mathbb{T}(\tau), P \rangle$;
  end
  add $\max_\tau \widehat{\mathbb{T}}^b(\tau)$ to f_h^b;
end
Data: $H = (h_1, \dots, h_N)$, f_h^b, α – confidence
Result: critical values $z(h)$
Multiplicity correction:
for *s = 1* **to** S **do**
  generate $u = (u_1, \dots, u_M)$;
  add $\min_h$ p-value($\max_\tau \widehat{\mathbb{T}}_h^b(\tau), f_h^b$) to empirical distribution $I\!P_f$
end
find α' from condition $I\!P_f(x < \alpha - \alpha') = \alpha$;
foreach h *in* H **do**
  $z(h) = \text{quantile}(f_h^b, \alpha - \alpha')$;
end

Algorithm 3: Critical values calibration

3 Theoretical results

3.1 LRT statistic

This section presents main results that describe theoretical properties of the likelihood-ratio statistics (LRT). They are essential for the proposed algorithm of change point detection. Further assume that log-likelihood function $L(\theta) = L(Y, \theta)$ has rather precise approximation by its quadratic part in local region $\Theta_0(r)$ of θ^*, $\Theta_0(r) \subseteq I\!R^p$, where

$$\theta^* = \operatorname*{argmax}_\theta I\!E L(\theta), \quad \widehat{\theta} = \operatorname*{argmax}_\theta L(\theta)$$

and $\Theta_0(r) = \{\|D(\theta - \theta^*)\| < r\}$. Spokoiny [2012] provides required conditions for justified quadratic approximation and parameter concentration in the local region. Approximation error involves the next variables for its estimation:

$$\alpha(\theta, \theta_0) = L(\theta) - L(\theta_0) - (\theta - \theta_0)^T \nabla L(\theta_0) + \frac{1}{2}\|D(\theta - \theta_0)\|^2,$$

$$\chi(\theta, \theta_0) = D^{-1} \nabla \alpha(\theta, \theta_0) = D^{-1}(\nabla L(\theta) - \nabla L(\theta_0)) + D(\theta - \theta_0).$$

Let in region $\Theta_0(r)$ with probability $1 - e^{-x}$:

$$\frac{|\alpha(\theta, \theta^*)|}{\|D(\theta - \theta^*)\|} \le \diamondsuit(r, x), \quad \|\chi(\theta, \theta^*)\| \le \diamondsuit(r, x). \tag{A}$$

Firstly, to provide a simple non-strict explanation of what kind of distribution the main statistic T_h is supposed to have, review T_h as

$$T_h = L(\widehat{\theta}) - L(\widehat{\theta}_{H_0}), \quad L(\theta) = L_1(\theta_1) + L_2(\theta_2),$$

$$L_1 = L(Y_1, \ldots, Y_h), \; L_2 = L(Y_h, \ldots, Y_{2h}),$$

where $\widehat{\theta}_{H_0}$ is argmax of L under condition $H_0 : \theta_1^* = \theta_2^*$. Then due to quadratic approximation T_h corresponds to Tailor equation with point $\widehat{\theta}$:

$$T_h \approx \frac{1}{2} \| D(\widehat{\theta} - \widehat{\theta}_{H_0}) \|^2.$$

If $\widehat{\theta}$ and $\widehat{\theta}_{H_0}$ tend to be Normal and H_0 is true then their difference are close to a centered Normal variable. If H_0 is false – the Normal variable will have mean that is equal to $D(\theta^* - \theta^*_{H_0})$.

The next theorem presents generalized result for non-quadratic model.

Theorem 1 *Assume condition (L*) and quadratic Laplace approximation (A) of L_1 and L_2 are fulfilled with probability $1 - 2e^{-x}$, additionally with probability $1 - 2e^{-x}$*

$$\|\xi_i\| \le z(x), \quad z^2(x) = \max_i p_{B_i} + 6\lambda_{B_i} x,$$

$$B_i = D_i^{-1} \operatorname{Var}(\nabla L_i(\theta^*)) D_i^{-1}, \quad p_B = \operatorname{tr}(B), \quad \lambda_B = \lambda_{\max}(B). \tag{B}$$

Then in the local region with probability $1 - 8e^{-x}$

$$2T_h = \|\xi_{12} + \theta^*_{12}\|^2 + O(\{r + z(x)\} \diamondsuit(r, x)),$$

where

$$\xi_{12} = \varSigma(D_2^{-1}\xi_2 - D_1^{-1}\xi_1), \quad \theta^*_{12} = \varSigma(\theta_2^* - \theta_1^*),$$

$$\varSigma^2 = D_2^2 D^{-2} D_1^2 D^{-2} D_2^2 + D_1^2 D^{-2} D_2^2 D^{-2} D_1^2 = D_1^2 D^{-2} D_2^2 \approx \frac{1}{4} D^2. \tag{S}$$

Remark 1 *In increasing sample size $n \to \infty$ the stochastic component tends to Normal distribution:*

$$\xi_{12} \to \mathcal{N}(0, B_1 + B_2).$$

Remark 2 *For the condition $\widehat{\theta} \in \varTheta_1(r) \cap \varTheta_2(r)$ the restriction of the parameter θ^* variability is required*

$$\|D(\theta_1^* - \theta_2^*)\| \le r. \tag{L*}$$

Proof of a similar statement (theorem 1) for statistic $\sqrt{2T_h}$ one could get from condition (A). With probability $1 - 2e^{-x}$

$$\begin{aligned}\left| T_h(\widehat{\theta}_1, \widehat{\theta}_2) - \frac{1}{2}\|\varSigma(\widehat{\theta}_2 - \widehat{\theta}_1)\|^2 \right| &\le 2\|D_1(\widehat{\theta}_1 - \widehat{\theta})\| \diamondsuit(r,x) + 2\|D_2(\widehat{\theta}_2 - \widehat{\theta})\| \diamondsuit(r,x) \\ &\le 4\|\varSigma(\widehat{\theta}_2 - \widehat{\theta}_1)\| \diamondsuit(r,x).\end{aligned}$$

Inequality $|a - b| \le |a^2 - b^2|/b, \; b > 0$ converts the previous term to

$$\left| \sqrt{2T_h(\widehat{\theta}_1, \widehat{\theta}_2)} - \|\varSigma(\widehat{\theta}_2 - \widehat{\theta}_1)\| \right| \le 8\diamondsuit(r,x).$$

Replacement $(\widehat{\theta}_1, \widehat{\theta}_2)$ with $(D_1^{-1}\xi_1 + \theta_1^*, \; D_2^{-1}\xi_2 + \theta_2^*)$ results in

$$\left| \|\varSigma(\widehat{\theta}_2 - \widehat{\theta}_1)\| - \|\xi_{12} + \theta^*_{12}\| \right|$$

$$\le \|\varSigma(\widehat{\theta}_1 - \theta_1^*) - \varSigma D_1^{-1}\xi_1\| + \|\varSigma(\widehat{\theta}_2 - \theta_2^*) - \varSigma D_2^{-1}\xi_2\| \le 2\diamondsuit(r,x).$$

The next theorem summarizes the statements above.

Theorem 2 *Assume condition (L*) and quadratic Laplace approximation (A) with probability* $1 - 2e^{-x}$ *are fulfilled. Then with probability* $1 - 4e^{-x}$ *in the local region* $\Theta_1(r) \cap \Theta_2(r)$ *took place*

$$\left| \sqrt{2T_h} - \|\xi_{12} + \theta^*_{12}\| \right| \leq 10\diamondsuit(r, x).$$

where ξ_{12} *and* θ^*_{12} *are defined in theorem 1.*

Remark 3 *The constant near* $\diamondsuit(r, x)$ *could be decreased, expanding series of* $L_1(\theta)$*,* $L_2(\theta)$ *and* $L(\theta)$ *in the local regions around* θ^*_1*,* θ^*_2 *and* θ^* *instead of MLE values:*

$$\left| 2T_h - \|\xi_{12} + \theta^*_{12}\|^2 \right| \leq (4\diamondsuit(r, x)r + 2\delta(r)r^2).$$

Remark 4 *Weighted LRT statistic (Tb) has similar approximation:*

$$2T^\flat_h \approx \|D(\widehat{\theta}^b - \widehat{\theta}^b_{H_0})\|^2 = \|\xi^\flat_{12}\|^2.$$

where $\widehat{\theta}^\flat_{H_0}$ *is argmax of* $L^\flat$ *under condition* $H_0 : \widehat{\theta}_2 - \widehat{\theta}_1 = \widehat{\theta}_{12}$*, which is true. That's why the mean of difference* $(\widehat{\theta}^b - \widehat{\theta}^b_{H_0})$ *is zero.*

4 Experiments

4.1 Experiments with synthetic data

This section presents results of the comparison of the proposed algorithm of change point detection (referred as *LRTOnline* or *LRTOffline*) with two other methods: *Bayesian online changepoint detection (BOCPD)* Adams and MacKay [2007] and *cpt.meanvar(PELT,...)* (RMeanVar) from RPa [2014]. The first method is constructed for online inference, but so far as it returns CP location with each CP signal, it is also applicable for offline testing scenario. The idea of this method is predictive filtering: its forecasts a new data point using only the information have been observed already, where the distribution family is fixed (Normal for the tests in this paper). Bayesian inference calculates the length of the observed data (from the last CP). The second algorithm also uses preliminary specified model. Its design focuses into finding multiple changes in mean and variance in Normally (another distributions also supported) distributed data. The returned set of change points is the result of sequential testing H_0 (existing number of change points) against H_1 (one extra change point) applying the likelihood ratio statistic of the whole data coupled with the penalty for CP count. Originally the method has offline change point detection interface, but one could adapt it for online case by buffering incoming data elements and clearing the buffer when at least one CP have been observed in the buffered data. In total, each of these two algorithms has modification in the way that allows one to use it in both online and offline testing mode.

LRTOffline configuration:
window sizes $(h_1, \ldots, h_W) = (10, 20, 40, 70)$; confidence for the upper bound of convolution with pattern $= 0.1$; window weights $(u_1, \ldots, u_W) = (1.0, 2.0, 0.5, 0.2)$.

LRTOnline configuration:
window sizes $(h_1, \ldots, h_W) = (30, 50, 70)$; confidence $= 0.1$.

Quality of measurements uses three following metrics: Normalised Mutual Information (NMI), Delay (average time interval in which CP have been detected after it had taken place), Precision and Recall. The next equation defines NMI measure of two partitions (X, Y) of time range by change points

$$\text{NMI}(X, Y) = 2\frac{H(X) + H(Y) - H(X, Y)}{H(X) + H(Y)}.$$

Higher NMI values (they are in $[0, 1]$) correspond to better quality. Quality comparison in offline case apply NMI measure, while for online mode involves Delay, Precision and Recall.

Synthetic test data have been generated for different values of difference norm of the data distribution parameter. Such values are denoted as *delta*. Each delta corresponds to 10 sampled data sequences over which one compute measure average. In online mode each data sequence could have one or none change points, in offline mode – two, one or none change points.

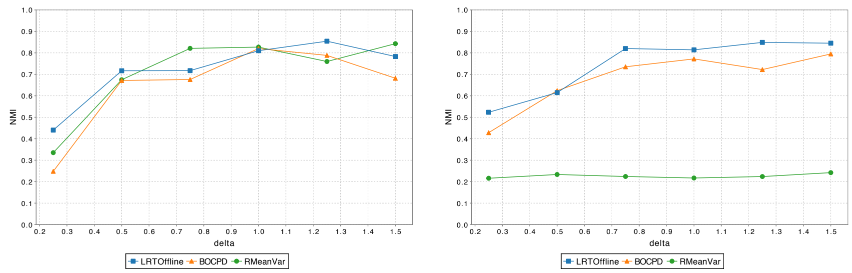

Figure 2: Offline mode. First data: $\mathcal{N}(\theta(1), \theta(2))$, second data: $\mathrm{Po}(\theta)$, delta = $\|\theta^*_{12}\|$, data size = 340, PA for all methods is $\mathcal{N}(\theta(1), \theta(2))$, NMI – Normalized Mutual Information between predicted and reference partitions of time interval with change points, tests per delta = 10, change point per test = $\{0, 1, 2\}$.

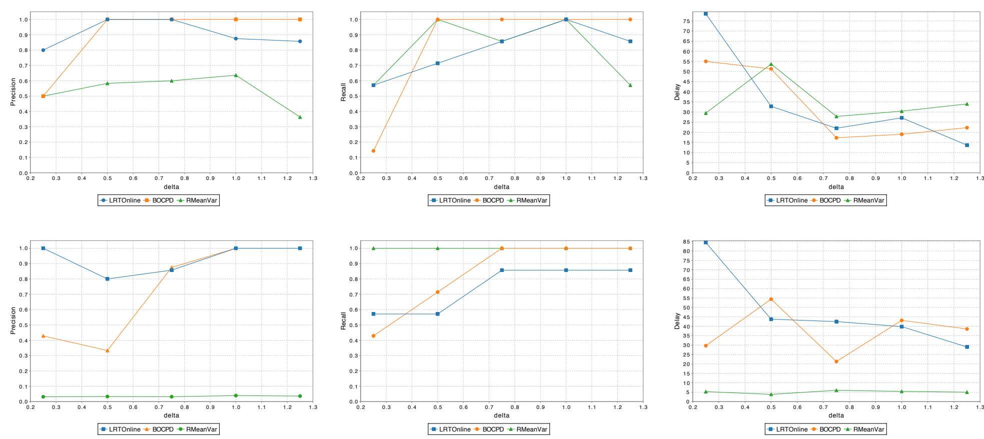

Figure 3: Online mode. First row data: $\mathcal{N}(\theta(1), \theta(2))$, second row data: $\mathrm{Po}(\theta)$, delta = $\|\theta^*_{12}\|$, data size = 340, PA for all methods is $\mathcal{N}(\theta(1), \theta(2))$, tests per delta = 10, change point per test = $\{0, 1\}$.

In the offline tests with Normal data all the methods achieves similar NMI scores, nonetheless LRTOffline is more stable for decreasing strength of CP (delta). In the tests with Poisson data (misspecification) RMeanVar has relatively low quality. The online tests characterizes the proposed method (LRTOnline) as more stable along different delta values what is accomplished by multiscale heuristic.

The experiments reveal following meaningful properties of the proposed method configuration:

1. Quality is sensitive to selection of interval for upper bound calibration of convolution in offline mode. For example in data $\mathcal{N}(0, 1)$.sample(100) $\cup\, \mathcal{N}(1, 2)$.sample(100) is preferable to use only slice of 0 to 100 elements for calibration, because of lower $\mathrm{Var}\,\xi_{12}$. Generally one should find change points in $tr(B_1 + B_2)$ according to remark 1 from Section 3 and run calibration in the range with the lowest values of $tr(B_1 + B_2)$. This improvement additional requires approximation of the convolution maximum in larger data ranges.
2. It is influenced to find out the minimal h sufficient for bootstrap usage. Small h improves Delay but makes unable to keep high level of Precision and Recall in online mode.

4.2 Experiments with real data

Here data from 1972-75 Dow Jones Returns Adams and MacKay [2007] describes several major events with potential macroeconomic effects (the most significant among them are the Watergate affair and the OPEC oil embargo). Convolutions plot with its upper bounds on this dataset appeared to be a nice illustration of multiscale search importance: CP near $t = 325$ is better perceptible when window size is equal to 30 and CP near $t = 600$ has more perceptible convolution when window size is equal to 70. Two plots presented below includes convolutions with Static and Fitted Patterns, where one could remark better separability of convolution peaks in fitted case.

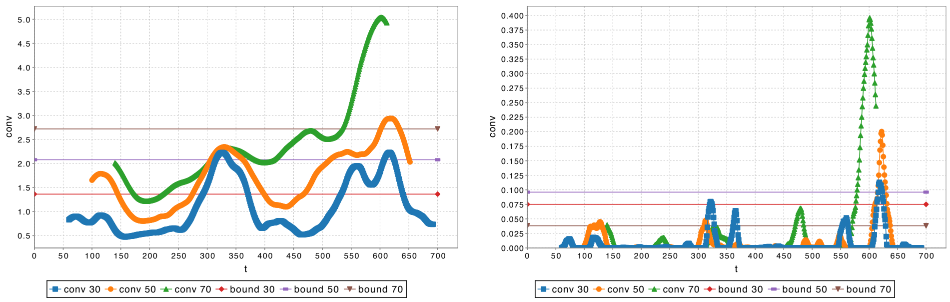

Figure 4: Data: daily returns of the Dow Jones Industrial Average from July 3, 1972 to June 30, 1975. Left plot – convolution with static triangle pattern; right plot – convolution with fitted triangle pattern. The time axis is in business days, conv 30 (50, 70) corresponds to pattern with window size 30 (50, 70), bound 30 (50, 70) corresponds to convolution upper bound. Three reference CP are presented: the conviction of G. Gordon Liddy and James W. McCord, Jr. on January 30, 1973 ($t = 142$); the beginning of the OPEC embargo against the United States on October 19, 1973 ($t = 325$); the resignation of President Nixon on August 9, 1974 ($t = 548$).

4.3 Sources

Demo of the LRTOnline method is available by link
https://localcpdetector.shinyapps.io/localCP

Scala project with LRTOffline and LRTOnline methods could be cloned from
https://github.com/nazarblch/cpd
which also includes testing system for abrupt change points detection applications and generated data used in the experiments.

5 Conclusion

This paper presented new change point detection method that works in offline and online modes. The method accounts properties of LRT statistic, which has shifted χ^2-distribution. Bootstrap technique appeared to be rather effective for LRT critical values calibration. Experiments and quality measurements confirm stability of the proposed algorithm in possibility to detect change points with different significance. The introduced concept of patterns allows to reveal different types of change point and filter regions with outliers.

References

I. Nikiforov. A lower bound for the detection/isolation delay in a class of sequential tests. *IEEE Trans. Inf. Theory*, 49:3037–3047, 2003.

M. Lavielle and G. Teyssiere. Detection of multiple change–points in multivariate time series. *Lithuanian Math. J.*, 46:287–306, 2006.

A. Tartakovsky, B. Rozovskii, R. Blažek, and H. Kim. Detection of intrusions in information systems by sequential change-point methods. *Stat. Methodology*, 3:252—340, 2006.

Pedro Casas, Sandrine Vaton, Lionel Fillatre, and Igor Nikiforov. Optimal volume anomaly detection and isolation in large-scale ip networks using coarse-grained measurements. *Computer Networks*, 54(11):1750–1766, 2010.

Richard E Quandt. Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American statistical Association*, 55(290):324–330, 1960.

Yukun Liu, Changliang Zou, and Runchu Zhang. Empirical likelihood ratio test for a change-point in linear regression model. *Communications in Statistics—Theory and Methods*, 37(16):2551–2563, 2008.

Changliang Zou, Yukun Liu, Peng Qin, and Zhaojun Wang. Empirical likelihood ratio test for the change-point problem. *Statistics & probability letters*, 77(4):374–382, 2007.

J. Chen and A.K. Gupta. *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*. Springer, 2012. ISBN 0817648003. URL `http://www.amazon.com/Parametric-Statistical-Change-Point-Analysis/dp/0817648003`.

Edit Gombay. Sequential change-point detection with likelihood ratios. *Statistics & probability letters*, 49(2):195–204, 2000.

Tze Leung Lai. Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 613–658, 1995.

BK Jandhyala and Stergios B Fotopoulos. Capturing the distributional behaviour of the maximum likelihood estimator of a changepoint. *Biometrika*, 86(1):129–140, 1999.

Hyune-Ju Kim. Tests for a change-point in linear regression. *Lecture Notes-Monograph Series*, pages 170–176, 1994.

Stergios B Fotopoulos, Venkata K Jandhyala, and Elena Khapalova. Exact asymptotic distribution of change-point mle for change in the mean of gaussian sequences. *The Annals of Applied Statistics*, pages 1081–1104, 2010.

Jianqing Fan, Chunming Zhang, and Jian Zhang. Generalized likelihood ratio statistics and wilks phenomenon. *Annals of Statistics*, pages 153–193, 2001.

Samuel S Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.

Stéphane Boucheron and Pascal Massart. A high-dimensional wilks phenomenon. *Probability theory and related fields*, 150(3-4):405–433, 2011.

V. Spokoiny. Penalized maximum likelihood estimation and effective dimension. *eprint arXiv:1205.0498*, 2012.

Klaus Frick, Axel Munk, and Hannes Sieling. Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):495–580, 2014.

V. Spokoiny. Multiscale local change point detection with applications to value-at-risk. *Ann. of Stat.*, 2009.

Siddhartha Chib. Estimation and comparison of multiple change-point models. *Journal of econometrics*, 86(2):221–241, 1998.

V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. of Stat.*, 41:2786—-2819, 2013.

V. Spokoiny and M. Zhilova. Bootstrap confidence sets under a model misspecification. *Preprint no. 1992, WIAS*, 2014.

R.P Adams and D.J.C. MacKay. Bayesian online changepoint detection, 2007. URL `http://hips.seas.harvard.edu/content/bayesian-online-changepoint-detection`.

changepoint: An r package for changepoint analysis, 2014. URL `http://cran.r-project.org/web/packages/changepoint`.