Restarted Bayesian Online Change-point Detector achieves Optimal Detection Delay

Reda ALAMI

Joint work with Odalric Maillard and Raphael Féraud.

reda.alami@total.com

Presented at ICML 2020

Overview

- A *pruning* version of the Bayesian Online Change-point Detector.

Overview

- A *pruning* version of the Bayesian Online Change-point Detector.
- High probability guarantees in term of:

Overview

- A *pruning* version of the Bayesian Online Change-point Detector.
- High probability guarantees in term of:
  - **False alarm rate.**

Overview

- A *pruning* version of the Bayesian Online Change-point Detector.
- High probability guarantees in term of:
  - **False alarm rate.**
  - **Detection delay**.

Overview

- A *pruning* version of the Bayesian Online Change-point Detector.
- High probability guarantees in term of:
  - **False alarm rate.**
  - **Detection delay**.
- The detection delay is asymptotically optimal

Overview

- A *pruning* version of the Bayesian Online Change-point Detector.
- High probability guarantees in term of:
  - **False alarm rate.**
  - **Detection delay**.
- The detection delay is asymptotically optimal (reaching the existing lower bound [Lai and Xing, 2010]).

Overview

- A *pruning* version of the Bayesian Online Change-point Detector.
- High probability guarantees in term of:
  - **False alarm rate.**
  - **Detection delay**.
- The detection delay is asymptotically optimal (reaching the existing lower bound [Lai and Xing, 2010]).
- Empirical comparisons with the original BOCPD [Fearnhead and Liu, 2007]

Overview

- A *pruning* version of the Bayesian Online Change-point Detector.
- High probability guarantees in term of:
  - **False alarm rate.**
  - **Detection delay**.
- The detection delay is asymptotically optimal (reaching the existing lower bound [Lai and Xing, 2010]).
- Empirical comparisons with the original BOCPD [Fearnhead and Liu, 2007] and the Improved Generalized Likelihood Ratio test [Maillard, 2019].

Setting & Notations

Setting & Notations

- $\mathcal{B}(\mu_t)$: Bernoulli distribution of mean $\mu_t \in [0, 1]$.

Setting & Notations

- $\mathcal{B}\left(\mu_t\right)$: Bernoulli distribution of mean $\mu_t \in [0, 1]$.
- Piece-wise stationary process: $\forall c \in [1, C], \forall t \in \mathcal{T}_c = [\tau_c, \tau_{c+1})\ \mu_t = \theta_c$

Setting & Notations

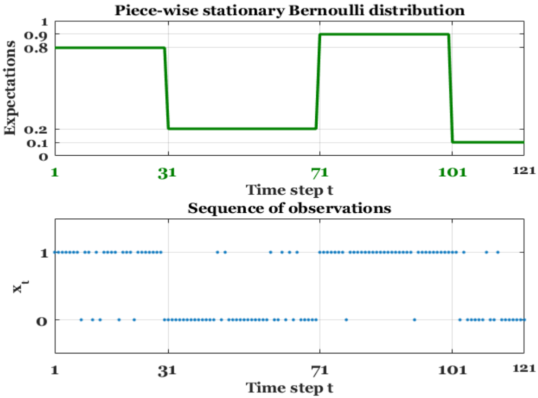
- $\mathcal{B}(\mu_t)$: Bernoulli distribution of mean $\mu_t \in [0, 1]$.
- Piece-wise stationary process: $\forall c \in [1, C], \forall t \in \mathcal{T}_c = [\tau_c, \tau_{c+1})\ \mu_t = \theta_c$
- Sequence of observations: $\mathbf{x}_{s:t} = (x_s, ...x_t)$.

Setting & Notations

- $\mathcal{B}(\mu_t)$: Bernoulli distribution of mean $\mu_t \in [0, 1]$.
- Piece-wise stationary process: $\forall c \in [1, C], \forall t \in \mathcal{T}_c = [\tau_c, \tau_{c+1})\ \mu_t = \theta_c$
- Sequence of observations: $\mathbf{x}_{s:t} = (x_s, ...x_t)$.
- Length: $n_{s:t} = t - s + 1$.

Setting & Notations

- $\mathcal{B}(\mu_t)$: Bernoulli distribution of mean $\mu_t \in [0, 1]$.
- Piece-wise stationary process: $\forall c \in [1, C], \forall t \in \mathcal{T}_c = [\tau_c, \tau_{c+1})\ \mu_t = \theta_c$
- Sequence of observations: $\mathbf{x}_{s:t} = (x_s, ...x_t)$.
- Length: $n_{s:t} = t - s + 1$.



Bayesian Online Change-point Detector

Runlength inference

Runlength inference

Bayesian Online Change-point Detector

Runlength inference

Runlength inference

Runlength r_t: number of time steps since the last change-point.

$$\forall r_t \in [0, t-1] \qquad \underbrace{p\left(r_t | \mathbf{x}_{1:t}\right)}_{\text{Runlength distribution at } t} \propto \sum_{r_{t-1} \in [0, t-2]} \underbrace{p\left(r_t | r_{t-1}\right)}_{\text{hazard}} \underbrace{p\left(x_t | r_{t-1}, \mathbf{x}_{1:t-1}\right)}_{\text{UPM}} p\left(r_{t-1} | \mathbf{x}_{1:t-1}\right)$$

Bayesian Online Change-point Detector

Runlength inference

Runlength r_t: number of time steps since the last change-point.

$$\forall r_t \in [0, t-1] \qquad \underbrace{p\left(r_t | \mathbf{x}_{1:t}\right)}_{\text{Runlength distribution at } t} \propto \sum_{r_{t-1} \in [0, t-2]} \underbrace{p\left(r_t | r_{t-1}\right)}_{\text{hazard}} \underbrace{p\left(x_t | r_{t-1}, \mathbf{x}_{1:t-1}\right)}_{\text{UPM}} p\left(r_{t-1} | \mathbf{x}_{1:t-1}\right)$$

Constant hazard rate assumption ($h \in (0, 1)$) (geometric inter-arrival time of change-point):

Bayesian Online Change-point Detector

Runlength inference

Runlength r_t: number of time steps since the last change-point.

$$\forall r_t \in [0, t-1] \qquad \underbrace{p\left(r_t | \mathbf{x}_{1:t}\right)}_{\text{Runlength distribution at } t} \propto \sum_{r_{t-1} \in [0, t-2]} \underbrace{p\left(r_t | r_{t-1}\right)}_{\text{hazard}} \underbrace{p\left(x_t | r_{t-1}, \mathbf{x}_{1:t-1}\right)}_{\text{UPM}} p\left(r_{t-1} | \mathbf{x}_{1:t-1}\right)$$

Constant hazard rate assumption ($h \in (0,1)$) (geometric inter-arrival time of change-point):

$$\begin{cases} p(r_t = r_{t-1} + 1 | \mathbf{x}_{1:t}) & \propto (1-h)\, p(x_t | r_{t-1}, \mathbf{x}_{1:t-1}) p(r_{t-1} | \mathbf{x}_{1:t-1}) \\ p(r_t = 0 | \mathbf{x}_{1:t}) & \propto h \sum_{r_{t-1}} p(x_t | r_{t-1}, \mathbf{x}_{1:t-1}) p(r_{t-1} | \mathbf{x}_{1:t-1}) \end{cases}$$

Bayesian Online Change-point Detector

Runlength inference

Runlength r_t: number of time steps since the last change-point.

$$\forall r_t \in [0, t-1] \qquad \underbrace{p\left(r_t | \mathbf{x}_{1:t}\right)}_{\text{Runlength distribution at } t} \propto \sum_{r_{t-1} \in [0, t-2]} \underbrace{p\left(r_t | r_{t-1}\right)}_{\text{hazard}} \underbrace{p\left(x_t | r_{t-1}, \mathbf{x}_{1:t-1}\right)}_{\text{UPM}} p\left(r_{t-1} | \mathbf{x}_{1:t-1}\right)$$

Constant hazard rate assumption ($h \in (0, 1)$) (geometric inter-arrival time of change-point):

$$\begin{cases} p(r_t = r_{t-1} + 1 | \mathbf{x}_{1:t}) & \propto (1-h)\, p(x_t | r_{t-1}, \mathbf{x}_{1:t-1}) p(r_{t-1} | \mathbf{x}_{1:t-1}) \\ p(r_t = 0 | \mathbf{x}_{1:t}) & \propto h \sum_{r_{t-1}} p(x_t | r_{t-1}, \mathbf{x}_{1:t-1}) p(r_{t-1} | \mathbf{x}_{1:t-1}) \end{cases}$$

$p(x_t | r_{t-1}, \mathbf{x}_{1:t-1})$ is computed via the Laplace predictor as MLE:

$$\mathtt{Lp}\left(x_{t+1} | \mathbf{x}_{s:t}\right) := \begin{cases} \frac{\sum_{i=s}^{t} x_i + 1}{n_{s:t} + 2} & \text{if } x_{t+1} = 1 \\ \frac{\sum_{i=s}^{t} (1 - x_i) + 1}{n_{s:t} + 2} & \text{if } x_{t+1} = 0 \end{cases}$$

Bayesian Online Change-point Detector

Forecaster Learning

Forecaster Learning

Instead of runlength $r_t \in [0, t-1]$, use the forecaster notion.
Forecaster weight: $\forall s \in [1, t] \;\; v_{s,t} := p(r_t = t - s | \mathbf{x}_{s:t})$

Bayesian Online Change-point Detector

Forecaster Learning

Forecaster Learning

Instead of runlength $r_t \in [0, t-1]$, use the forecaster notion.
Forecaster weight: $\forall s \in [1, t]\;\; v_{s,t} := p(r_t = t - s | \mathbf{x}_{s:t})$

$$v_{s,t} = \begin{cases} (1-h)\exp\left(-l_{s,t}\right) v_{s,t-1} & \forall s < t, \\ h \sum_{i=1}^{t-1} \exp\left(-l_{i,t}\right) v_{i,t-1} & s = t\,. \end{cases}$$

Bayesian Online Change-point Detector

Forecaster Learning

Forecaster Learning

Instead of runlength $r_t \in [0, t-1]$, use the forecaster notion.
Forecaster weight: $\forall s \in [1, t]\ \ v_{s,t} := p(r_t = t - s | \mathbf{x}_{s:t})$

$$v_{s,t} = \begin{cases} (1-h)\exp\left(-l_{s,t}\right) v_{s,t-1} & \forall s < t, \\ h \sum_{i=1}^{t-1} \exp\left(-l_{i,t}\right) v_{i,t-1} & s = t\,. \end{cases}$$

Instantaneous loss: $l_{s,t} := -\log \mathtt{Lp}\left(x_t | \mathbf{x}_{s':t-1}\right)$.

Bayesian Online Change-point Detector

Forecaster Learning

Forecaster Learning

Instead of runlength $r_t \in [0, t-1]$, use the forecaster notion.
Forecaster weight: $\forall s \in [1, t]\ \ v_{s,t} := p(r_t = t - s | \mathbf{x}_{s:t})$

$$v_{s,t} = \begin{cases} (1-h)\exp\left(-l_{s,t}\right) v_{s,t-1} & \forall s < t, \\ h \sum_{i=1}^{t-1} \exp\left(-l_{i,t}\right) v_{i,t-1} & s = t\,. \end{cases} \qquad v_{s,t} = \begin{cases} (1-h)^{n_{s:t}}\, h^{\mathbb{I}\{s \neq 1\}} \exp\left(-\widehat{L}_{s:t}\right) V_s & \forall s < t, \\ hV_t & s = t. \end{cases}$$

Instantaneous loss: $l_{s,t} := -\log \texttt{Lp}\left(x_t | \mathbf{x}_{s':t-1}\right)$.

Bayesian Online Change-point Detector

Forecaster Learning

Forecaster Learning

Instead of runlength $r_t \in [0, t-1]$, use the forecaster notion.
Forecaster weight: $\forall s \in [1,t]\;\; v_{s,t} := p(r_t = t-s|\mathbf{x}_{s:t})$

$$v_{s,t} = \begin{cases} (1-h)\exp\left(-l_{s,t}\right) v_{s,t-1} & \forall s < t, \\ h\sum_{i=1}^{t-1} \exp\left(-l_{i,t}\right) v_{i,t-1} & s = t\,. \end{cases} \qquad v_{s,t} = \begin{cases} (1-h)^{n_{s:t}}\, h^{\mathbb{I}\{s\neq 1\}} \exp\left(-\widehat{L}_{s:t}\right) V_s & \forall s < t, \\ hV_t & s = t. \end{cases}$$

Instantaneous loss: $l_{s,t} := -\log \mathtt{Lp}\left(x_t|\mathbf{x}_{s':t-1}\right)$.
$\widehat{L}_{s:t} := \sum_{s'=s}^{t} l_{s,t}$: cumulative loss and $V_t = \sum_{s=1}^{t} v_{s,t}$

Main difficulty to provide the theoretical guarantees

Lemma (Computing the initial weight V_t)

Main difficulty to provide the theoretical guarantees

Lemma (Computing the initial weight V_t)

$$V_t = (1-h)^{t-2} \sum_{k=1}^{t-1} \left(\frac{h}{1-h} \right)^{k-1} \tilde{V}_{k:t},$$

.

Main difficulty to provide the theoretical guarantees

Lemma (Computing the initial weight V_t)

$$V_t = (1-h)^{t-2} \sum_{k=1}^{t-1} \left(\frac{h}{1-h}\right)^{k-1} \tilde{V}_{k:t}, \textit{ where:}$$

$$\tilde{V}_{k:t} = \sum_{i_1=1}^{t-k} \sum_{i_2=i_1+1}^{t-(k-1)} \cdots \sum_{i_{k-1}=i_{k-2}+1}^{t-2} \exp\left(-\widehat{L}_{1:i_1}\right) \times \prod_{j=1}^{k-2} \exp\left(-\widehat{L}_{i_j+1:i_{j+1}}\right) \times \exp\left(-\widehat{L}_{i_{k-1}+1:t-1}\right),$$

.

Main difficulty to provide the theoretical guarantees

Lemma (Computing the initial weight V_t)

$$V_t = (1-h)^{t-2} \sum_{k=1}^{t-1} \left(\frac{h}{1-h}\right)^{k-1} \tilde{V}_{k:t}, \textit{ where:}$$

$$\tilde{V}_{k:t} = \sum_{i_1=1}^{t-k} \sum_{i_2=i_1+1}^{t-(k-1)} \cdots \sum_{i_{k-1}=i_{k-2}+1}^{t-2} \exp\left(-\widehat{L}_{1:i_1}\right) \times \prod_{j=1}^{k-2} \exp\left(-\widehat{L}_{i_j+1:i_{j+1}}\right) \times \exp\left(-\widehat{L}_{i_{k-1}+1:t-1}\right),$$

$$\textit{with: } \sum_{i_1=1}^{t-k} \sum_{i_2=i_1+1}^{t-(k-1)} \cdots \sum_{i_{k-1}=i_{k-2}+1}^{t-2} 1 = \binom{t-2}{k-1} .$$

Main difficulty to provide the theoretical guarantees

Lemma (Computing the initial weight V_t)

$$V_t = (1-h)^{t-2} \sum_{k=1}^{t-1} \left(\frac{h}{1-h}\right)^{k-1} \tilde{V}_{k:t}, \textit{ where:}$$

$$\tilde{V}_{k:t} = \sum_{i_1=1}^{t-k} \sum_{i_2=i_1+1}^{t-(k-1)} \cdots \sum_{i_{k-1}=i_{k-2}+1}^{t-2} \exp\left(-\widehat{L}_{1:i_1}\right) \times \prod_{j=1}^{k-2} \exp\left(-\widehat{L}_{i_j+1:i_{j+1}}\right) \times \exp\left(-\widehat{L}_{i_{k-1}+1:t-1}\right),$$

$$\textit{with: } \sum_{i_1=1}^{t-k} \sum_{i_2=i_1+1}^{t-(k-1)} \cdots \sum_{i_{k-1}=i_{k-2}+1}^{t-2} 1 = \binom{t-2}{k-1} \textit{ and } \widehat{L}_{s:t} := \sum_{s'=s}^{t} l_{s,t}.$$

Main difficulty to provide the theoretical guarantees

Lemma (Computing the initial weight V_t)

$$V_t = (1-h)^{t-2} \sum_{k=1}^{t-1} \left(\frac{h}{1-h} \right)^{k-1} \tilde{V}_{k:t}, \textit{ where:}$$

$$\tilde{V}_{k:t} = \sum_{i_1=1}^{t-k} \sum_{i_2=i_1+1}^{t-(k-1)} \dots \sum_{i_{k-1}=i_{k-2}+1}^{t-2} \exp\left(-\widehat{L}_{1:i_1}\right) \times \prod_{j=1}^{k-2} \exp\left(-\widehat{L}_{i_j+1:i_{j+1}}\right) \times \exp\left(-\widehat{L}_{i_{k-1}+1:t-1}\right),$$

$$\textit{with: } \sum_{i_1=1}^{t-k} \sum_{i_2=i_1+1}^{t-(k-1)} \dots \sum_{i_{k-1}=i_{k-2}+1}^{t-2} 1 = \binom{t-2}{k-1} \textit{ and } \widehat{L}_{s:t} := \sum_{s'=s}^{t} l_{s,t}.$$

Combinatorial number of cumulative losses: very difficult to use classical concentrations.

From BOCPD to Restarted-BOCPD

From BOCPD to Restarted-BOCPD

Some modifications of BOCPD to tackle the theoretical difficulty.

From BOCPD to Restarted-BOCPD

Some modifications of BOCPD to tackle the theoretical difficulty.

- Restart time $r >= 0$ (updated for each time a change-point is raised).

From BOCPD to Restarted-BOCPD

Some modifications of BOCPD to tackle the theoretical difficulty.

- Restart time $r >= 0$ (updated for each time a change-point is raised).
- Initial weight function: $\mathcal{V}_{r:t-1} := \exp\left(-\widehat{L}_{r:t-1}\right)$ instead of V_t.

From BOCPD to Restarted-BOCPD

Some modifications of BOCPD to tackle the theoretical difficulty.

- Restart time $r >= 0$ (updated for each time a change-point is raised).
- Initial weight function: $\mathcal{V}_{r:t-1} := \exp\left(-\widehat{L}_{r:t-1}\right)$ instead of V_t.
- Hyper-parameter $\eta_{r,s,t}$ instead of the hazard rate $h \in (0,1)$.

From BOCPD to Restarted-BOCPD

Some modifications of BOCPD to tackle the theoretical difficulty.

- Restart time $r >= 0$ (updated for each time a change-point is raised).
- Initial weight function: $\mathcal{V}_{r:t-1} := \exp\left(-\widehat{L}_{r:t-1}\right)$ instead of V_t.
- Hyper-parameter $\eta_{r,s,t}$ instead of the hazard rate $h \in (0,1)$.
- Restart criterion: $\texttt{Restart}_{r:t} = \mathbb{I}\left\{\exists s \in (r,t] : \vartheta_{r,s,t} > \vartheta_{r,r,t}\right\}$.

From BOCPD to Restarted-BOCPD

Some modifications of BOCPD to tackle the theoretical difficulty.

- Restart time $r >= 0$ (updated for each time a change-point is raised).
- Initial weight function: $\mathcal{V}_{r:t-1} := \exp\left(-\widehat{L}_{r:t-1}\right)$ instead of V_t.
- Hyper-parameter $\eta_{r,s,t}$ instead of the hazard rate $h \in (0,1)$.
- Restart criterion: $\texttt{Restart}_{r:t} = \mathbb{I}\left\{\exists s \in (r,t] : \vartheta_{r,s,t} > \vartheta_{r,r,t}\right\}$.

R-BOCPD update rule

For some starting time r:

$$\vartheta_{r,s,t} \leftarrow \begin{cases} \frac{\eta_{r,s,t}}{\eta_{r,s,t-1}} \exp\left(-l_{s,t}\right) \vartheta_{r,s,t-1} & \forall s < t, \\ \eta_{r,t,t} \times \mathcal{V}_{r:t-1} & s = t. \end{cases}$$

From BOCPD to Restarted-BOCPD

Some modifications of BOCPD to tackle the theoretical difficulty.

- Restart time $r >= 0$ (updated for each time a change-point is raised).
- Initial weight function: $\mathcal{V}_{r:t-1} := \exp\left(-\widehat{L}_{r:t-1}\right)$ instead of V_t.
- Hyper-parameter $\eta_{r,s,t}$ instead of the hazard rate $h \in (0,1)$.
- Restart criterion: $\texttt{Restart}_{r:t} = \mathbb{I}\left\{\exists s \in (r,t] : \vartheta_{r,s,t} > \vartheta_{r,r,t}\right\}$.

R-BOCPD update rule

For some starting time r:

$$\vartheta_{r,s,t} \leftarrow \begin{cases} \frac{\eta_{r,s,t}}{\eta_{r,s,t-1}} \exp\left(-l_{s,t}\right) \vartheta_{r,s,t-1} & \forall s < t, \\ \eta_{r,t,t} \times \mathcal{V}_{r:t-1} & s = t. \end{cases}$$

Recall BOCPD update rule

$$v_{s,t} \leftarrow \begin{cases} (1-h) \exp\left(-l_{s,t}\right) v_{s,t-1} & \forall s < t, \\ h \times V_t & s = t. \end{cases}$$

Analysis of R-BOCPD

False alarm control

Theorem: False alarm rate control

Assume that $(x_r, ...x_t) \sim \mathcal{B}(\theta)$. Let: $\alpha > 1$. If:

$$\forall t \in [r, \tau), s \in (r, t] : \eta_{r,s,t} < \frac{\sqrt{n_{r:s-1} \times n_{s:t}}}{10 n_{r:t+1}} \left(\frac{\log(4\alpha + 2)\delta^2}{4 n_{r:t} \log((\alpha+3)\, n_{r:t})} \right)^{\alpha}$$

then, with probability higher than $1 - \delta$, no false alarm occurs on the interval $[r, \tau)$:

$$\forall \delta \in (0,1) \quad \mathbb{P}_\theta \left\{ \exists\, t \in [r, \tau) : \texttt{Restart}_{r:t} = 1 \right\} \leqslant \delta.$$

For $\alpha \approx 1$, $\eta_{r,s,t} = O\left(\frac{1}{t-r+1}\right)$

Analysis of R-BOCPD

Detection delay control

Theorem: Detection delay control

Let $(x_r, \dots x_{\tau-1}) \sim \mathcal{B}(\theta_1)$, $(x_\tau, \dots x_t) \sim \mathcal{B}(\theta_2)$ and $\Delta = |\theta_1 - \theta_2|$: the change-point gap.
Then, let: $f_{r,s,t} = \log n_{r:s} + \log n_{s:t+1} - \frac{1}{2}\log n_{r:t} + \frac{9}{8}$.
If $\eta_{r,s,t} > \exp\left(-2n_{r,s-1}\left(\Delta_{r,s,t} - \mathcal{C}_{r,s,t,\delta}\right)^2 + f_{r,s,t}\right)$, then, the change-point τ is detected (with a probability at least $1-\delta$) with a delay not exceeding $\mathfrak{D}_{\Delta,r,\tau}$, such that:

$$\mathfrak{D}_{\Delta,r,\tau} = \min\left\{ d \in \mathbb{N}^\star : d > \frac{\left(1 - \frac{\mathcal{C}_{r,\tau,d+\tau-1,\delta}}{\Delta}\right)^{-2}}{2\Delta^2} \times \frac{-\log \eta_{r,\tau,d+\tau-1} + f_{r,\tau,d+\tau-1}}{1 + \frac{\log \eta_{r,\tau,d+\tau-1} - f_{r,\tau,d+\tau-1}}{2n_{r,\tau-1}\left(\Delta - \mathcal{C}_{r,\tau,d+\tau-1,\delta}\right)^2}} \right\},$$

$$\text{with: } \mathcal{C}_{r,s,t,\delta} = \frac{\sqrt{2}}{2}\left(\sqrt{\frac{1+\frac{1}{n_{r:s-1}}}{n_{r:s-1}} \log\left(\frac{2\sqrt{n_{r:s}}}{\delta}\right)} + \sqrt{\frac{1+\frac{1}{n_{s:t}}}{n_{s:t}} \log\left(\frac{2n_{r:t}\sqrt{n_{s:t}+1}\log^2(n_{r:t})}{\log(2)\delta}\right)}\right).$$

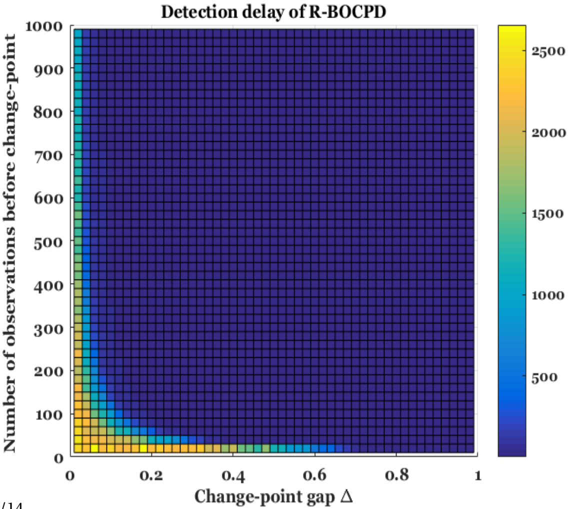
$\eta_{r,s,t} = \Omega\left(\exp(-n_{r,s,t})\right)$ and $\mathcal{C}_{r,s,t,\delta} = O\left(\sqrt{\log\left(n_{r:s}/\delta\right)/n_{r:s-1}} + \sqrt{\log\left(n_{s:t+1}/\delta\right)/n_{s:t}}\right)$

Analysis of R-BOCPD

Asymptotic analysis of the Detection delay

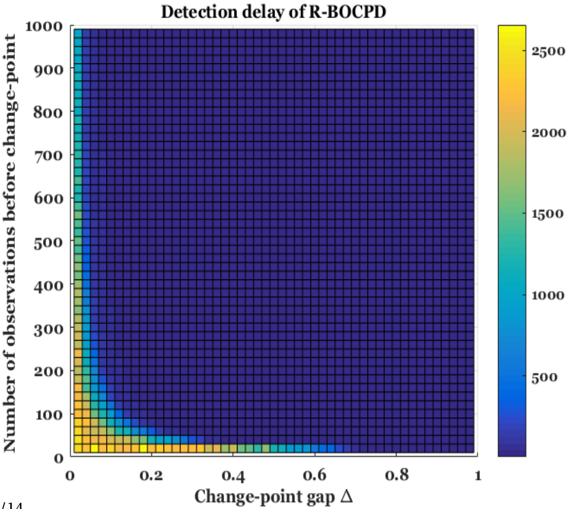
Analysis of R-BOCPD

Asymptotic analysis of the Detection delay



Analysis of R-BOCPD

Asymptotic analysis of the Detection delay



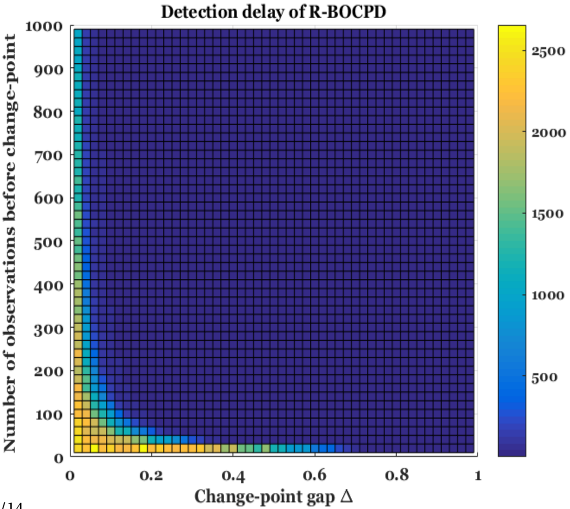
Asymptotic Analysis

if $\eta_{r,s,t} = \frac{1}{t-r+1}$, then in the asymptotic regime:

$$\mathfrak{D}_{|\theta_2-\theta_1|,r,\tau} \underset{\tau\to\infty}{\to} \frac{o\left(\log\frac{1}{\delta}\right)}{2\left|\theta_2-\theta_1\right|^2}$$

Analysis of R-BOCPD

Asymptotic analysis of the Detection delay



Asymptotic Analysis

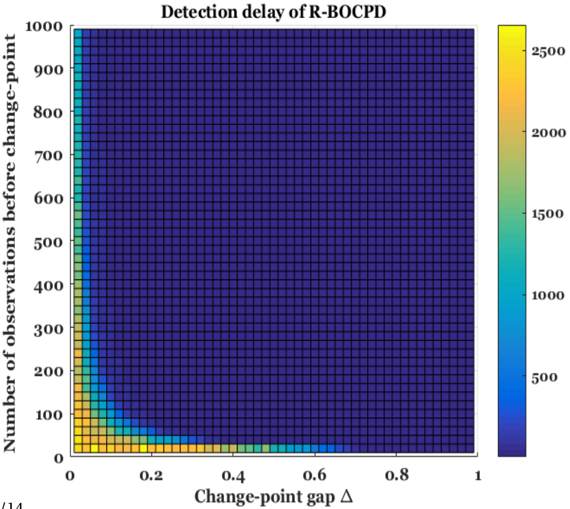
if $\eta_{r,s,t} = \frac{1}{t-r+1}$, then in the asymptotic regime:

$$\mathfrak{D}_{|\theta_2-\theta_1|,r,\tau} \underset{\tau\to\infty}{\to} \frac{o\left(\log\frac{1}{\delta}\right)}{2\left|\theta_2-\theta_1\right|^2}$$

$$= O\left(\frac{o\left(\log\frac{1}{\delta}\right)}{\boldsymbol{KL}\left(\theta_2,\theta_1\right)}\right)$$

Analysis of R-BOCPD

Asymptotic analysis of the Detection delay



Asymptotic Analysis

if $\eta_{r,s,t} = \frac{1}{t-r+1}$, then in the asymptotic regime:

$$\mathfrak{D}_{|\theta_2-\theta_1|,r,\tau} \underset{\tau\to\infty}{\to} \frac{o\left(\log\frac{1}{\delta}\right)}{2\left|\theta_2-\theta_1\right|^2} = O\left(\frac{o\left(\log\frac{1}{\delta}\right)}{\boldsymbol{KL}\left(\theta_2,\theta_1\right)}\right)$$

Existing lower bound [Lai and Xing, 2010].

Empirical comparisons

Comparison with the original BOCPD: Benchmark 1

Empirical comparisons

Comparison with the original BOCPD: Benchmark 1

Benchmark 1: Highlighting the use of the function $\mathcal{V}_{r:t-1}$ instead of V_t

Empirical comparisons

Comparison with the original BOCPD: Benchmark 1

Benchmark 1: Highlighting the use of the function $\mathcal{V}_{r:t-1}$ instead of V_t

- Generate 2500 trajectories (sequences) of length $T = 5000$.

Empirical comparisons

Comparison with the original BOCPD: Benchmark 1

Benchmark 1: Highlighting the use of the function $\mathcal{V}_{r:t-1}$ instead of V_t

- Generate 2500 trajectories (sequences) of length $T = 5000$.
- Vary the number of observation before the change in $[10, 1000]$.

Empirical comparisons

Comparison with the original BOCPD: Benchmark 1

Benchmark 1: Highlighting the use of the function $\mathcal{V}_{r:t-1}$ instead of V_t

- Generate 2500 trajectories (sequences) of length $T = 5000$.
- Vary the number of observation before the change in $[10, 1000]$.
- Vary the change-point gap Δ in $[0.01, 1]$.

Empirical comparisons

Comparison with the original BOCPD: Benchmark 1

Benchmark 1: Highlighting the use of the function $\mathcal{V}_{r:t-1}$ instead of V_t

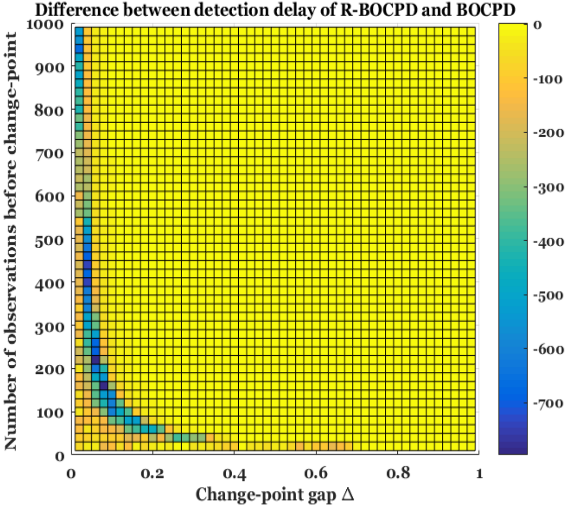
- Generate 2500 trajectories (sequences) of length $T = 5000$.
- Vary the number of observation before the change in $[10, 1000]$.
- Vary the change-point gap Δ in $[0.01, 1]$.
- Plot detection delays differences between R-BOCPD and BOCPD.

Empirical comparisons

Comparison with the original BOCPD: Benchmark 1

Benchmark 1: Highlighting the use of the function $\mathcal{V}_{r:t-1}$ instead of V_t

- Generate 2500 trajectories (sequences) of length $T = 5000$.
- Vary the number of observation before the change in $[10, 1000]$.
- Vary the change-point gap Δ in $[0.01, 1]$.
- Plot detection delays differences between R-BOCPD and BOCPD.



Empirical comparisons

Comparison with the original BOCPD: Benchmark 2

Empirical comparisons

Comparison with the original BOCPD: Benchmark 2

Benchmark 2: Highlighting the use of the restart procedure $\texttt{Restart}_{r:t}$

Empirical comparisons

Comparison with the original BOCPD: Benchmark 2

Benchmark 2: Highlighting the use of the restart procedure $\texttt{Restart}_{r:t}$

- Piece-wise stationary Bernoulli process $\tau_1 = 1, \tau_2 = 301, \tau_3 = 701, \tau_4 = 1051.$

Empirical comparisons

Comparison with the original BOCPD: Benchmark 2

Benchmark 2: Highlighting the use of the restart procedure $\texttt{Restart}_{r:t}$

- Piece-wise stationary Bernoulli process $\tau_1 = 1, \tau_2 = 301, \tau_3 = 701, \tau_4 = 1051.$
- Run R-BOCPD and BOCPD.

Empirical comparisons

Comparison with the original BOCPD: Benchmark 2

Benchmark 2: Highlighting the use of the restart procedure $\texttt{Restart}_{r:t}$

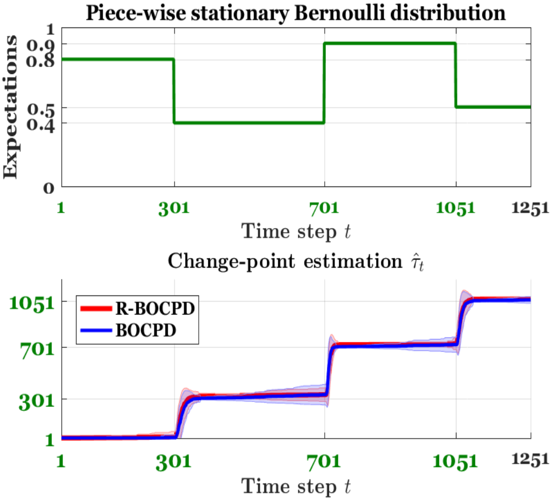
- Piece-wise stationary Bernoulli process $\tau_1 = 1, \tau_2 = 301, \tau_3 = 701, \tau_4 = 1051$.
- Run R-BOCPD and BOCPD.
- Plot the change-point estimation $\widehat{\tau}_t$ for both R-BOCPD and BOCPD.

Empirical comparisons

Comparison with the original BOCPD: Benchmark 2

Benchmark 2: Highlighting the use of the restart procedure $\texttt{Restart}_{r:t}$

- Piece-wise stationary Bernoulli process $\tau_1 = 1, \tau_2 = 301, \tau_3 = 701, \tau_4 = 1051$.
- Run R-BOCPD and BOCPD.
- Plot the change-point estimation $\widehat{\tau}_t$ for both R-BOCPD and BOCPD.



Empirical comparisons

Comparison with the Improved GLR [Maillard, 2019]

Empirical comparisons

Comparison with the Improved GLR [Maillard, 2019]

Improved GLR final formulation

$$\textsc{ImpGLR}_\delta\left(y_1, \dots, y_t\right) = \mathbb{I}\left\{\exists s \in [1, t) : \left| \frac{1}{s} \sum_{i=1}^{s} y_i - \frac{1}{t-s} \sum_{i=s+1}^{t} y_i \right| \geqslant \mathscr{C}_{\delta,s,t} \right\}$$

Empirical comparisons

Comparison with the Improved GLR [Maillard, 2019]

Improved GLR final formulation

$$\textsc{ImpGLR}_\delta\left(y_1, \ldots, y_t\right) = \mathbb{I}\left\{\exists s \in [1,t) : \left|\frac{1}{s}\sum_{i=1}^{s} y_i - \frac{1}{t-s}\sum_{i=s+1}^{t} y_i\right| \geqslant \mathscr{C}_{\delta,s,t}\right\}$$

$$\mathscr{C}_{\delta,s,t} = \frac{\sqrt{2}}{2}\left(\sqrt{\frac{1}{s} + \frac{1}{s^2}\log\left(\frac{2\sqrt{s+1}}{\delta}\right)} + \sqrt{\frac{1}{t-s} + \frac{1}{(t-s)^2}\log\left(\frac{2t\sqrt{t-s+1}\log^2(t)}{\log(2)\delta}\right)}\right)$$

Empirical comparisons

Comparison with the Improved GLR [Maillard, 2019]

Empirical comparisons

Comparison with the Improved GLR [Maillard, 2019]

Benchmark :

Empirical comparisons

Comparison with the Improved GLR [Maillard, 2019]

Benchmark :

- Generate 2500 trajectories (sequences) of length $T = 2500$.

Empirical comparisons

Comparison with the Improved GLR [Maillard, 2019]

Benchmark :

- Generate 2500 trajectories (sequences) of length $T = 2500$.
- Vary the number of observation before the change in $[10, 500]$.

Empirical comparisons

Comparison with the Improved GLR [Maillard, 2019]

Benchmark :

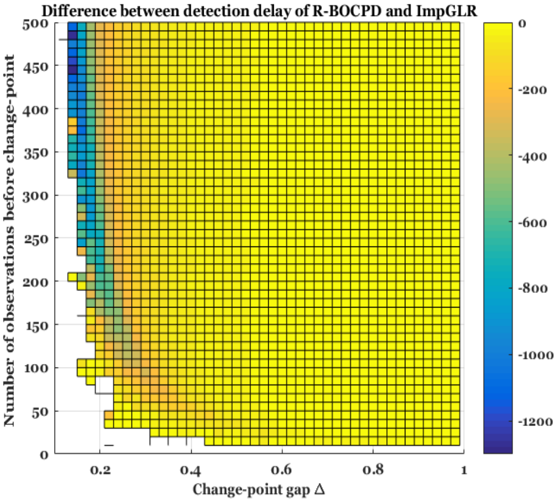
- Generate 2500 trajectories (sequences) of length $T = 2500$.
- Vary the number of observation before the change in $[10, 500]$.
- Vary the change-point gap $\Delta \in [0.01, 1]$.
- Plot the difference of detection delays between R-BOCPD and Improved GLR.

Empirical comparisons

Comparison with the Improved GLR [Maillard, 2019]

Benchmark :

- Generate 2500 trajectories (sequences) of length $T = 2500$.
- Vary the number of observation before the change in $[10, 500]$.
- Vary the change-point gap $\Delta \in [0.01, 1]$.
- Plot the difference of detection delays between R-BOCPD and Improved GLR.



Fearnhead, P. and Liu, Z. (2007).
On-line inference for multiple changepoint problems.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69(4):589–605.

Lai, T. L. and Xing, H. (2010).
Sequential change-point detection when the pre-and post-change parameters are unknown.
Sequential analysis, 29(2):162–175.

Maillard, O.-A. (2019).
Sequential change-point detection: Laplace concentration of scan statistics and non-asymptotic delay bounds.
In *Algorithmic Learning Theory*, pages 610–632.