Mean Shift Detection for State Space Models

Julia Kuhn[•,⋆], Michel Mandjes[⋆], Thomas Taimre[•]

[•] *The University of Queensland,*
[⋆] *University of Amsterdam*

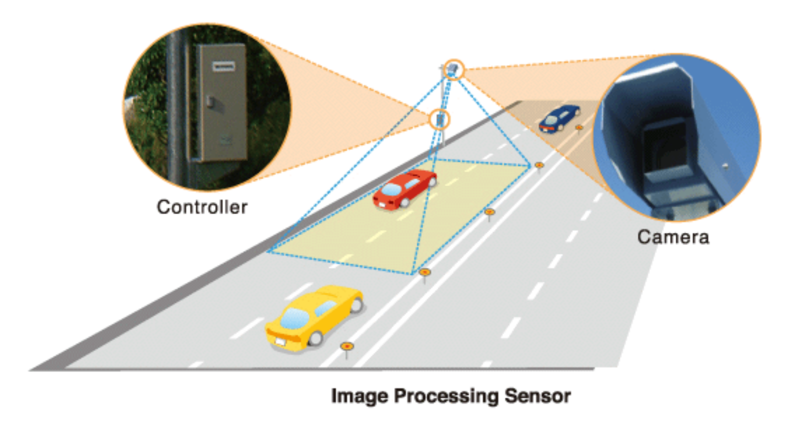
3 November 2015

A simple state space model

$$X_{t+1} = AX_t + Y_t$$

$$V_t = BX_t + Z_t$$

Application to traffic modelling



Source: Sumitomo Electric


D.U.I. CHECKPOINT AHEAD

State space model with shift

$$X_{t+1} = AX_t + Y_t + M\,\mathbb{1}_{\{t \geq k\}}$$

$$V_t = BX_t + Z_t + N\,\mathbb{1}_{\{t \geq k\}}$$

We need to decide whether a change has occurred or not:

H_0 : No change has occurred.

H_1 : There is a change point k with $1 \leq k \leq m$.

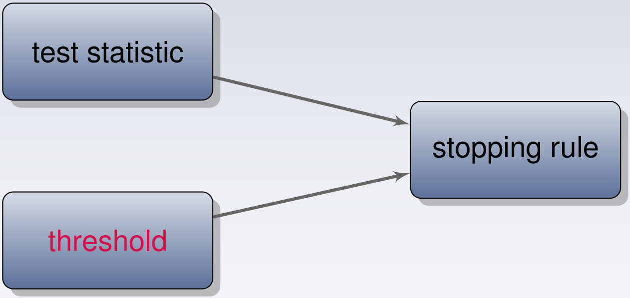
We need to decide whether a change has occurred or not:

H_0 : No change has occurred.

H_1 : There is a change point k with $1 \leq k \leq m$.

Note that $H_1 = \bigcup_k H_1(k)$.

How do we do that?



CUSUM method

Test statistic: $\mathrm{LLR}_{k,m} = \sum_{t=k}^{m} \ell(V_t)$

CUSUM method

Test statistic: $\mathrm{LLR}_{k,m} = \sum_{t=k}^{m} \ell(V_t)$

Stopping rule: $\tau = \inf\{m : \max_k \mathrm{LLR}_{k,m} > b\}$

CUSUM method

Test statistic: $\mathrm{LLR}_{k,m} = \sum_{t=k}^{m} \ell(V_t)$

Stopping rule: $\tau = \inf\{m : \max_k \mathrm{LLR}_{k,m} > b\}$

Performance criterion: $\mathbb{E}_0 \tau \geq \kappa$

CUSUM method

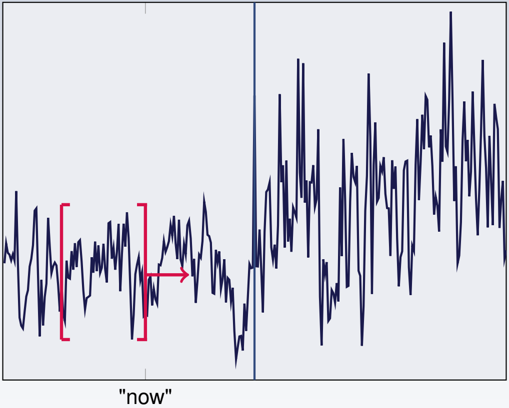
Test statistic: $\mathrm{LLR}_{k,m} = \sum_{t=k}^{m} \ell(V_t)$

Stopping rule: $\tau = \inf\{m : \max_k \mathrm{LLR}_{k,m} > b\}$

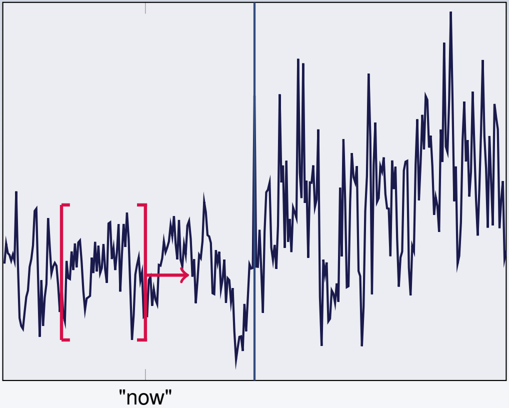
Performance criterion: $\mathbb{E}_0 \tau \geq \kappa$

Threshold: Asymptotics (for i.i.d.), recursive integral equations, numerical methods

Window-limited testing procedure

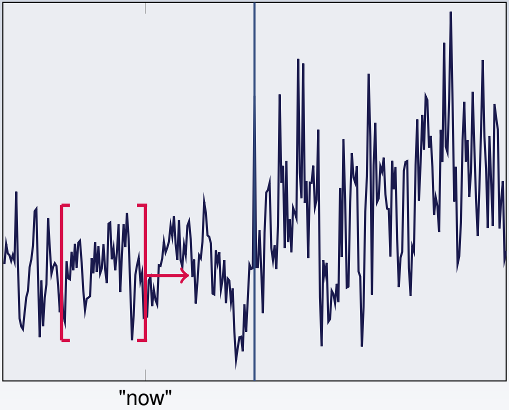


Window-limited testing procedure



$$\tau = \inf\left\{m : \max_{\{1\le k\le m\}} \mathrm{LLR}_{k,m} > b\right\}$$

Window-limited testing procedure



$$\tau_{WL} = \inf\left\{m : \max_{\{m-n\leq k\leq m\}} \mathrm{LLR}_{k,m} > b\right\}$$

Our (stronger) performance criterion

Choose the threshold b in such a way that the probability of raising an alarm for the current window is kept at a given low level:

$$\mathbb{P}_0\left(\max_{k\in\{1,\dots,n\}} \mathrm{LLR}_{k,n} > b\right) = \alpha$$

False Alarm Control for Change Point Detection: Beyond Average Run Length

J. Kuhn[•,⋆], M. Mandjes[•], T. Taimre[⋆]

December 2, 2015

[•] Korteweg-de Vries Institute for Mathematics, University of Amsterdam, the Netherlands.
[⋆] School of Mathematics and Physics, The University of Queensland, Australia.
E-mail: j.kuhn@uva.nl

Abstract

A popular method for detecting changes in the probability distribution of a sequence of observations is CUSUM, which proceeds by sequentially evaluating a log-likelihood ratio test statistic and comparing it to a predefined threshold; a change point is detected as soon as the threshold is exceeded. It is desirable to choose the threshold in such a way that the number of false detections is kept to a specified level while on the other hand ensuring a quick detection if a change has occurred. In this paper we analyse the distribution of the CUSUM stopping time when observations may be correlated, with the aim of devising simple yet effective methods for selecting the threshold. In addition to the standard CUSUM procedure we consider window-limited testing where only the n most recent observations are considered at each time point. Traditionally, the number of false alarms is measured by the average run length – the expected time until the first false alarm. However, this is a reasonable criterion only when the expectation is finite. We thus propose an alternative criterion that ensures a large average run length and is more generally applicable. We prove that CUSUM is asymptotically optimal under this criterion, and investigate methods for selecting the threshold such that it is approximately achieved. Apart from the above, we note that the average run length criterion does not allow one to restrict the variability of false alarms, which we argue can be crucial. Therefore, we make a case for a stronger false alarm criterion, and show how it is related to the average run length. To illustrate the procedures and evaluate their performance, we provide numerical examples featuring a multidimensional state space model.

Our (stronger) performance criterion

Choose the threshold b in such a way that the probability of raising an alarm for the current window is kept at a given low level:

$$\mathbb{P}_0\left(\max_{k\in\{1,\ldots,n\}} \mathrm{LLR}_{k,n} > b\right) = \alpha$$

Gaussian approximation

Recall that LLRs are partial sums: $\mathrm{LLR}_{k,n} \sum_{t=k}^{n} \ell(V_t)$.

$\Rightarrow$ Can apply Donsker's theorem

Gaussian approximation

Recall that LLRs are partial sums: $\mathrm{LLR}_{k,n} \sum_{t=k}^{n} \ell(V_t)$.

$\Rightarrow$ Can apply Donsker's theorem

$$\begin{aligned}
&\mathbb{P}_0 \left(\max_{1 \le k \le n} \mathrm{LLR}_{k,n} > b \right) \\
&\quad \approx \mathbb{P}_0 \left(\max_{t \in [0,n]} \sigma B_t + \mu\, t \ge b \right) \\
&\quad = 1 - \Phi\left(\frac{b - \mu\, n}{\sigma \sqrt{n}} \right) + e^{\frac{2b\mu}{\sigma^2}}\, \Phi\left(\frac{-b - \mu\, n}{\sigma \sqrt{n}} \right)
\end{aligned}$$

Back to the model

$$X_{t+1} = AX_t + Y_t + M\,\mathbb{1}_{\{t \geq k\}}\,, \quad V_t = BX_t + Z_t + N\,\mathbb{1}_{\{t \geq k\}}$$

Back to the model

$$X_{t+1} = AX_t + Y_t + M\,\mathbb{1}_{\{t \geq k\}}\,, \quad V_t = BX_t + Z_t + N\,\mathbb{1}_{\{t \geq k\}}$$

From Kalman filtering:

$$\varepsilon_t := V_t - B\,\widehat{X}_t\,,$$

where $\widehat{X}_t = \mathbb{E}[X_t \mid V_1, \ldots, V_{t-1}]$ is the minimum variance estimator for the hidden state X_t.

Back to the model

$$X_{t+1} = AX_t + Y_t + M\,\mathbb{1}_{\{t \geq k\}}\,, \quad V_t = BX_t + Z_t + N\,\mathbb{1}_{\{t \geq k\}}$$

From Kalman filtering:

$$\varepsilon_t := V_t - B\,\widehat{X}_t\,,$$

where $\widehat{X}_t = \mathbb{E}[X_t \mid V_1, \ldots, V_{t-1}]$ is the minimum variance estimator for the hidden state X_t.

The innovations ε_t are independent.

Testing the innovations

The persistent change in mean in X_t and V_t results in a dynamic change in the innovations; namely, the shift in mean on ε_t is

$$\rho(t,k) = B\left[\psi(t,k) - A\zeta(t-1,k)\right] + N,$$

where $\psi(t,k) = A\psi(t-1,k) + M$, $\zeta(t,k) = A\zeta(t-1,k) + K_t\,\rho(t,k)$, with initial conditions $\psi(k,k) = 0$, $\zeta(k-1,k) = 0$.

Testing the innovations

Thus, we are interested in testing whether there is a change point at some $k \in \{1, \ldots, n\}$, with $t \geq k$:

$$H_0\colon\ \varepsilon_t \sim \mathcal{N}(0,\, \Omega_t)$$

$$H_1\colon\ \bigcup_{k=1}^{n} \Big[H_1(k) : \ \varepsilon_t \sim \mathcal{N}\big(\rho(t,k),\, \Omega_t\big) \Big]$$

Testing the innovations

Thus, we are interested in testing whether there is a change point at some $k \in \{1, \ldots, n\}$, with $t \geq k$:

$$H_0\colon\ \varepsilon_t \sim \mathcal{N}(0,\ \underset{\substack{\downarrow \\ \Omega}}{\Omega_t})$$

$$H_1\colon\ \bigcup_{k=1}^{n} \left[H_1(k) : \ \varepsilon_t \sim \mathcal{N}\big(\rho(\underset{\substack{\downarrow \\ \rho}}{t}, k),\ \underset{\substack{\downarrow \\ \Omega}}{\Omega_t}\big) \right]$$

Testing the innovations

For large $n - k$ we have

$$\mathsf{LLR}_{k,n} \approx \sum_{t=k}^{n} \rho' \Omega^{-1} \varepsilon_t - \frac{1}{2} \rho' \Omega^{-1} \rho \,.$$

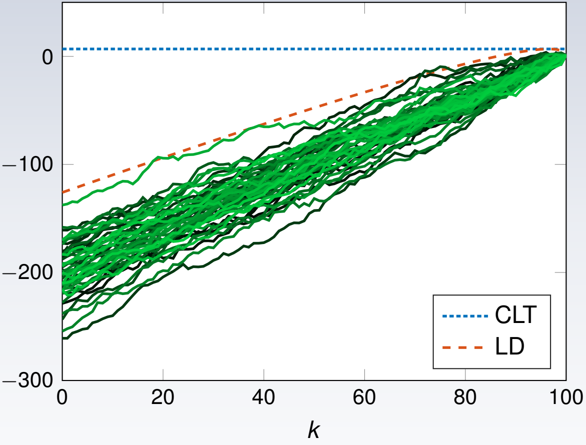
The mean and variance of the asymptotic likelihood increments (under H_0) are

$$\mu = -\frac{1}{2} \rho' \Omega^{-1} \rho \,, \quad \sigma^2 = \rho' \Omega^{-1} \rho \,.$$

Gaussian approximation

$$\begin{aligned}
&\mathbb{P}_0\left(\max_{1\le k\le n} \mathrm{LLR}_{k,n} > b\right) \\
&\quad\approx\ \mathbb{P}_0\left(\max_{t\in[0,n]} \sigma B_t + \mu\, t \ge b\right) \\
&\quad=\ 1-\Phi\left(\frac{b-\mu\, n}{\sigma\sqrt{n}}\right)+e^{\frac{2b\mu}{\sigma^2}}\,\Phi\left(\frac{-b-\mu\, n}{\sigma\sqrt{n}}\right)
\end{aligned}$$

Should b really be constant?



Large deviation approximation

Gärtner-Ellis theorem:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_0 \left(\frac{1}{n} \mathrm{LLR}_{k,n} > b(k) \right) = \mathcal{I}\big(b(k)\big) ,$$

where

$$\mathcal{I}\big(b(k)\big) := \sup_{\lambda \in \mathbb{R}} \left(\lambda b(k) - (1 - \beta_k)\, \Lambda(\lambda) \right)$$

is the Fenchel-Legendre transform of the limiting log-moment generating function of the LLR:

$$\Lambda(\lambda) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}_0 \left(e^{\lambda\, \mathrm{LLR}_{k,n}} \right) .$$

Large deviation approximation

Gärtner-Ellis theorem:

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}_0\left(\frac{1}{n}\mathrm{LLR}_{k,n}>b(k)\right)=\mathcal{I}\big(b(k)\big)\,,$$

where

$$\mathcal{I}\big(b(k)\big):=\sup_{\lambda\in\mathbb{R}}\left(\lambda b(k)-(1-\beta_k)\,\Lambda(\lambda)\right)$$

is the Fenchel-Legendre transform of the limiting log-moment generating function of the LLR:

$$\Lambda(\lambda):=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}_0\left(e^{\lambda\,\mathrm{LLR}_{k,n}}\right)\,.$$

Large deviation approximation

Gärtner-Ellis theorem:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_0 \left(\frac{1}{n} \mathrm{LLR}_{k,n} > b(k) \right) = \mathcal{I}\big(b(k)\big)\,,$$

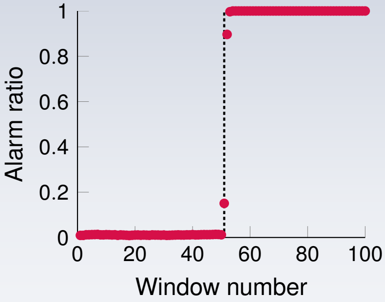
suggests to pick the threshold **function** b such that it satisfies

$$\alpha = \exp\Big(-n\mathcal{I}\big(b(k)\big)\Big)\,,$$

where we can compute

$$\mathcal{I}\big(b(k)\big) = \sup_{\lambda} \left\{ \lambda b(k) + (1-\beta_k)\frac{\lambda}{2}(1-\lambda)\rho'\Omega^{-1}\rho \right\}\,.$$

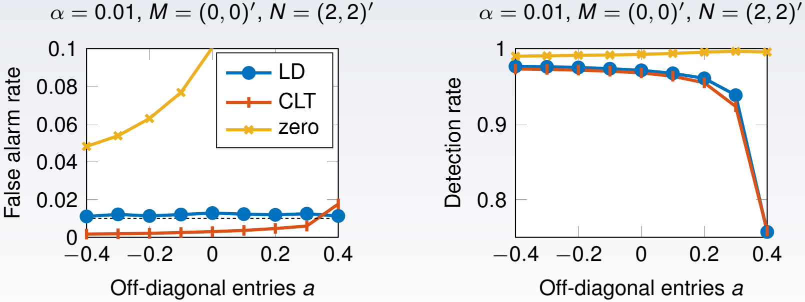
Numerics

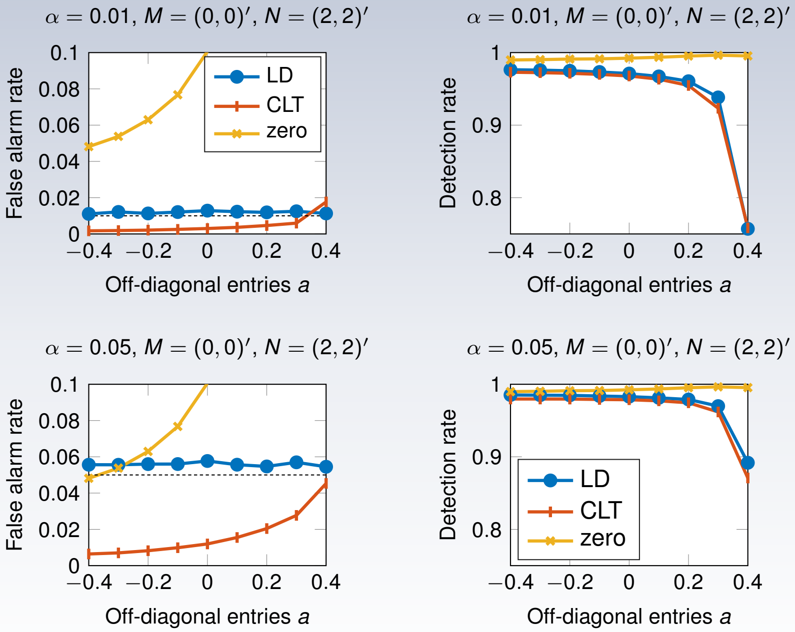


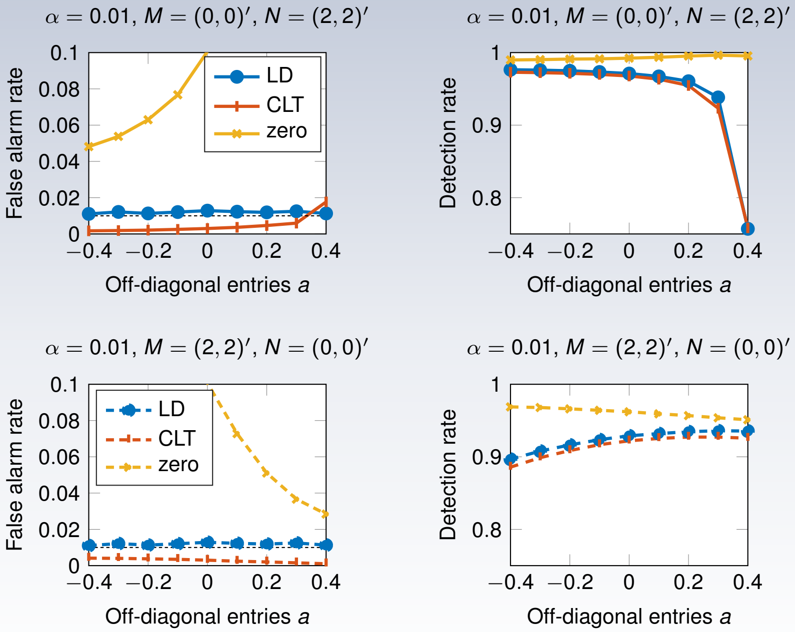
Alarm ratios, obtained with the LD threshold for $A = B = 0.5\, I_2$, $M = N = (2, 2)'$, $\alpha = 0.01$.

$$X_{t+1} = \begin{pmatrix} 0.5 & \boldsymbol{a} \\ \boldsymbol{a} & 0.5 \end{pmatrix} X_t + Y_t + \begin{pmatrix} 0 \\ 0 \end{pmatrix} \mathbb{1}_{\{t \geq k\}}$$

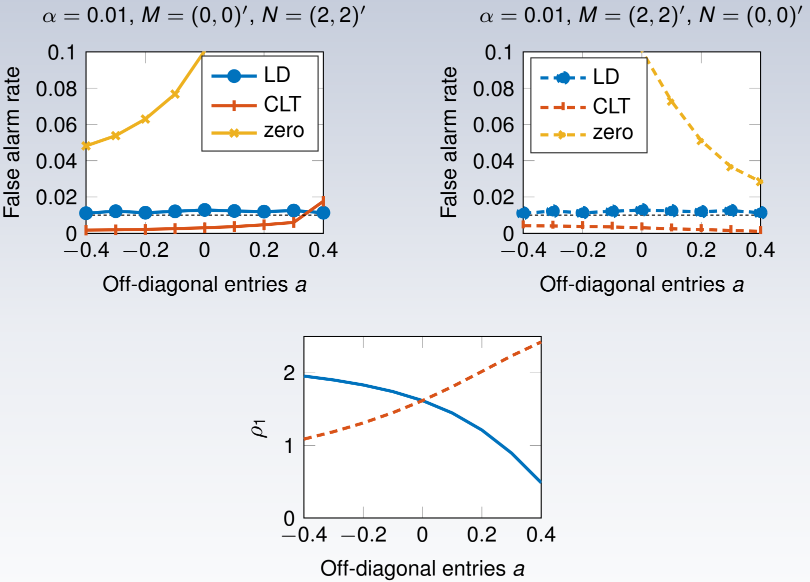
$$V_t = BX_t + Z_t + \begin{pmatrix} 2 \\ 2 \end{pmatrix} \mathbb{1}_{\{t \geq k\}}$$




$\alpha = 0.01$, $M = (0,0)'$, $N = (2,2)'$
False alarm rate
0.1
0.08
0.06
0.04
0.02
0
−0.4
−0.2
0
0.2
0.4
Off-diagonal entries a
LD
CLT
zero
$\alpha = 0.01$, $M = (0,0)'$, $N = (2,2)'$
Detection rate
1
0.9
0.8
−0.4
−0.2
0
0.2
0.4
Off-diagonal entries a
$\alpha = 0.05$, $M = (0,0)'$, $N = (2,2)'$
False alarm rate
0.1
0.08
0.06
0.04
0.02
0
−0.4
−0.2
0
0.2
0.4
Off-diagonal entries a
$\alpha = 0.05$, $M = (0,0)'$, $N = (2,2)'$
Detection rate
1
0.9
0.8
LD
CLT
zero
−0.4
−0.2
0
0.2
0.4
Off-diagonal entries a



$\alpha = 0.01$, $M = (0,0)'$, $N = (2,2)'$
False alarm rate
0.1
0.08
0.06
0.04
0.02
0
−0.4
−0.2
0
0.2
0.4
Off-diagonal entries a
LD
CLT
zero
$\alpha = 0.01$, $M = (0,0)'$, $N = (2,2)'$
Detection rate
1
0.9
0.8
Off-diagonal entries a
$\alpha = 0.01$, $M = (2,2)'$, $N = (0,0)'$
False alarm rate
Off-diagonal entries a
LD
CLT
zero
$\alpha = 0.01$, $M = (2,2)'$, $N = (0,0)'$
Detection rate
Off-diagonal entries a


$\alpha = 0.01$, $M = (0,0)'$, $N = (2,2)'$
$\alpha = 0.01$, $M = (2,2)'$, $N = (0,0)'$
False alarm rate
Off-diagonal entries a
LD
CLT
zero
ρ_1

Conclusions

- We can test the independent sequence of innovations.
- Using approximate LLRs works well.
- LD threshold function is better than CLT threshold.

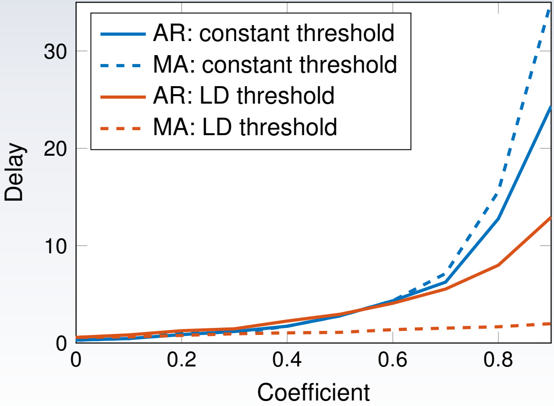
Conclusions

- We can test the independent sequence of innovations.
- Using approximate LLRs works well.
- LD threshold function is better than CLT threshold.
- Generally, we should choose a threshold function.
- Not from this paper: We should replace $\mathbb{E}_0\tau \geq \kappa$ by false alarm probability criterion.

Thank you!

Example: ARMA model

$$\begin{aligned} \text{AR:} \quad & X_t = \varphi X_{t-1} + \varepsilon_t \\ \text{MA:} \quad & X_t = \vartheta\, \varepsilon_{t-1} + \varepsilon_t \end{aligned}$$



Large deviation approximation

Gärtner-Ellis theorem:

$$\frac{1-\beta_k}{n(1-\beta_k)} \log \mathbb{P}_0 \left(\frac{1}{n(1-\beta_k)} \mathrm{LLR}_{k,n} > \frac{b(k)}{1-\beta_k} \right) \to \mathcal{I}\big(b(k)\big) \,,$$

where

$$\mathcal{I}\big(b(k)\big) := \sup_{\lambda \in \mathbb{R}} \left(\lambda b(k) - (1-\beta_k)\, \Lambda(\lambda) \right)$$

is the Fenchel-Legendre transform of the limiting log-moment generating function of the LLR:

$$\Lambda(\lambda) := \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}_0 \left(e^{\lambda \, \mathrm{LLR}_{k,n}} \right) \,.$$