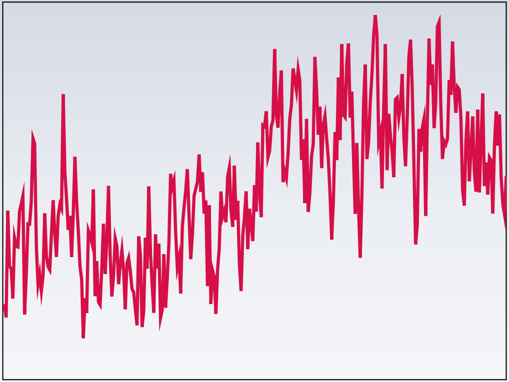
False Alarm Control for Window-Limited Change Detection

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

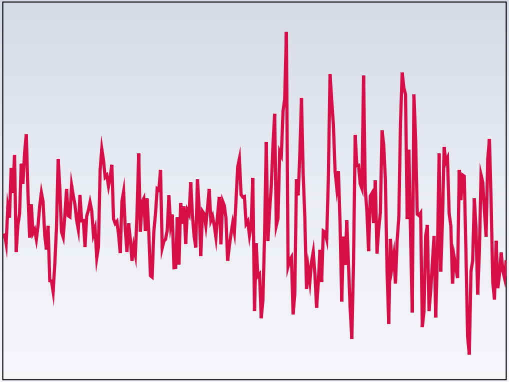
[•] *The University of Queensland,*
[⋆] *University of Amsterdam*

30 September 2015

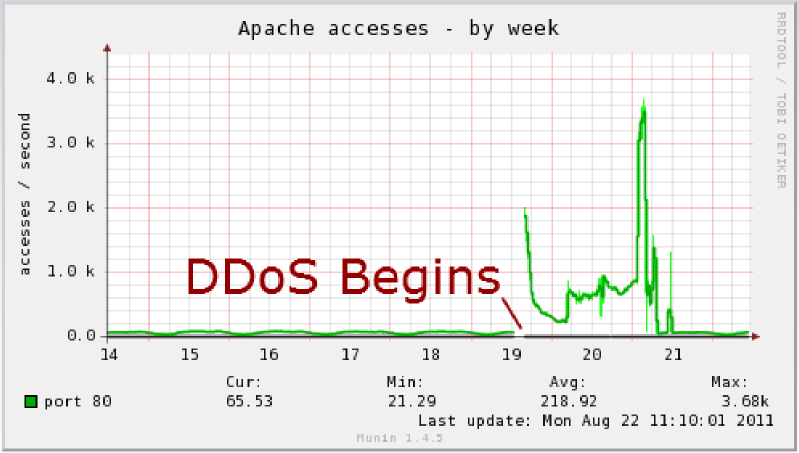
Change in mean



Change in variance



Why is that interesting?



Outline

Change Detection Procedures

Unconditional Local False Alarm Probability

Maximum Local False Alarm Probability

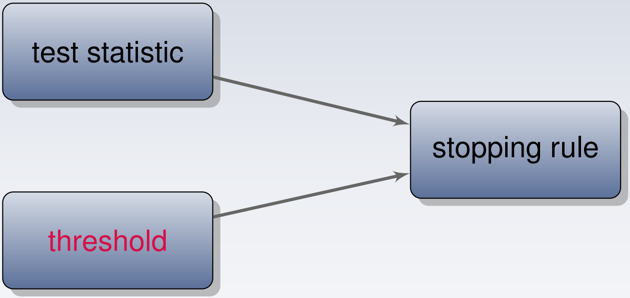
Window Limited False Alarm Probability

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 $1 \leq k \leq m$.

How do we do that?



Traditional approach

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

Traditional approach

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

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

Traditional approach

Test statistic:	$\mathrm{LLR}_{k,m} = \sum_{t=k}^{m} \ell(X_t)$	(CUSUM)
Stopping rule:	$\tau = \inf\{m : \max_k \mathrm{LLR}_{k,m} > b\}$	
Performance criterion:	Average run length: $\mathbb{E}_0\tau$	

Traditional approach

Test statistic:	$\mathrm{LLR}_{k,m} = \sum_{t=k}^{m} \ell(X_t)$ (CUSUM)
Stopping rule:	$\tau = \inf\{m : \max_k \mathrm{LLR}_{k,m} > b\}$
Performance criterion:	Average run length: $\mathbb{E}_0 \tau$
Threshold:	Asymptotics (for i.i.d.), recursive integral equations, numerical methods

Optimality

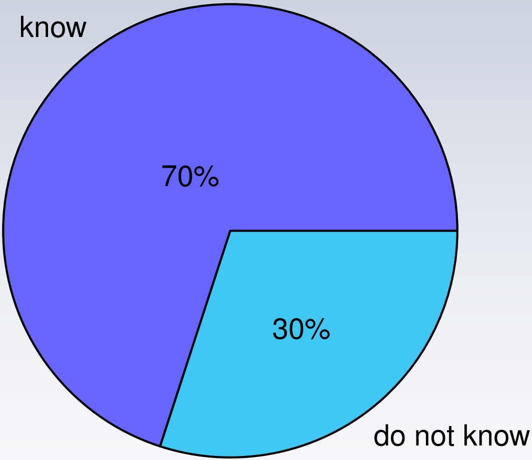
Minimises detection delay among all procedures that satisfy

- $\mathbb{E}_0\tau = \kappa$ (strict optimality)

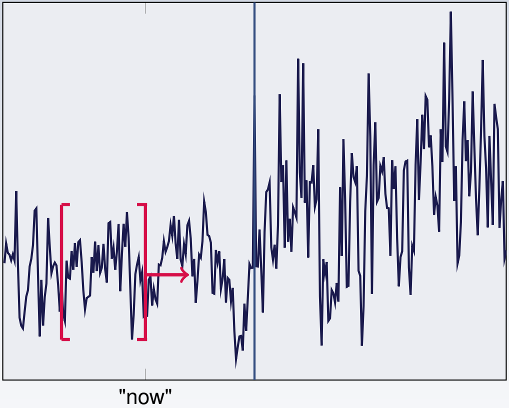
Optimality

Minimises detection delay among all procedures that satisfy

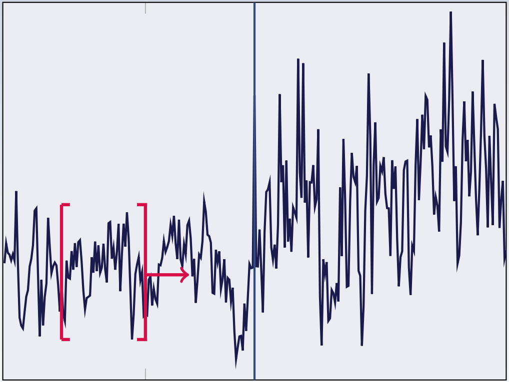
- $\mathbb{E}_0\tau = \kappa$ (strict optimality)
- $\mathbb{E}_0\tau \geq \kappa$ as $\kappa \to \infty$ (asymptotic optimality)


know
70%
30%
do not know

Procedure

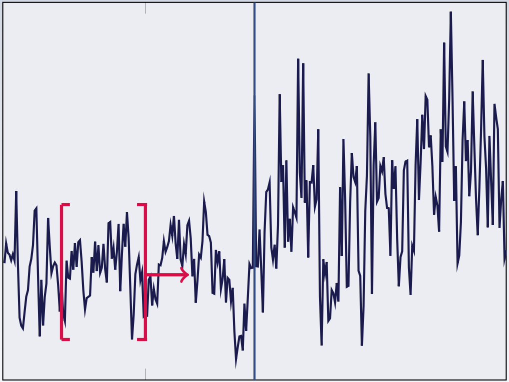


Procedure



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

Procedure



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

Alternative performance criteria

WL FA probability: $\mathbb{P}_0(\tau_{WL} \leq N)$

Alternative performance criteria

WL FA probability: $\mathbb{P}_0(\tau_{WL} \leq N)$

Max. local FA probability: $\sup_{m \geq 0} \mathbb{P}_0(\tau \leq m + n \,|\, \tau \geq m)$

Alternative performance criteria

WL FA probability:	$\mathbb{P}_0(\tau_{WL} \leq N)$
Max. local FA probability:	$\sup_{m\geq 0} \mathbb{P}_0(\tau \leq m+n \mid \tau \geq m)$
Unconditional local FA prob.:	$\sup_{m\geq 0} \mathbb{P}_0(m \leq \tau \leq m+n)$

Alternative performance criteria

WL FA probability:	$\mathbb{P}_0(\tau_{WL} \leq N)$
Max. local FA probability:	$\sup_{m\geq 0} \mathbb{P}_0(\tau \leq m+n \mid \tau \geq m)$
Unconditional local FA prob.:	$\sup_{m\geq 0} \mathbb{P}_0(m \leq \tau \leq m+n)$

Asymptotics

$$\mathrm{ULFA}_n := \sup_{m \geq 0} \mathbb{P}_0(m \leq \tau \leq m + n)$$

- CUSUM is asymptotically optimal for small α among procedures that satisfy $\mathrm{ULFA}_n \leq \alpha$ (Lai, 1998).

Asymptotics

$$\mathrm{ULFA}_n := \sup_{m \geq 0} \mathbb{P}_0(m \leq \tau \leq m + n)$$

- CUSUM is asymptotically optimal for small α among procedures that satisfy $\mathrm{ULFA}_n \leq \alpha$ (Lai, 1998).

- **But:** For large thresholds (small α) this is decreasing in m (Tartakovsky, 2005).

Alternative performance criteria

WL FA probability:	$\mathbb{P}_0(\tau_{WL} \leq N)$
Local FA probability:	$\sup_{m \geq 0} \mathbb{P}_0(\tau \leq m + n \,\vert\, \tau \geq m)$
Unconditional local FA prob.:	$\sup_{m \geq 0} \mathbb{P}_0(m \leq \tau \leq m + n)$

Maximum local FA

$$\begin{aligned}\mathrm{MLFA}_n &:= \sup_{m\geq 0} \mathbb{P}_0\left(\tau \leq m+n \,|\, \tau > m\right)\\ &= \sup_{m\geq 0} \frac{\mathbb{P}_0(m < \tau \leq m+n)}{\mathbb{P}_0(\tau > m)}\end{aligned}$$

Maximum local FA

$$\begin{aligned} \mathrm{MLFA}_n &:= \sup_{m \geq 0} \mathbb{P}_0\left(\tau \leq m+n \,|\, \tau > m\right) \\ &= \sup_{m \geq 0} \frac{\mathbb{P}_0(m < \tau \leq m+n)}{\mathbb{P}_0(\tau > m)} \end{aligned}$$

Again, CUSUM turns out to be optimal for large b among all procedures that satisfy $\mathrm{MLFA}_n \leq \alpha$ (Tartakovsky, 2005).

Maximum local FA

$$\begin{aligned}\mathrm{MLFA}_n &:= \sup_{m \geq 0} \mathbb{P}_0\left(\tau \leq m + n \,|\, \tau > m\right) \\ &= \sup_{m \geq 0} \frac{\mathbb{P}_0(m < \tau \leq m + n)}{\mathbb{P}_0(\tau > m)}\end{aligned}$$

Again, CUSUM turns out to be optimal for large b among all procedures that satisfy $\mathrm{MLFA}_n \leq \alpha$ (Tartakovsky, 2005).

Numerically it seems that the value for m that maximizes $\mathbb{P}_0\left(\tau \leq m + n \,|\, \tau > m\right)$ is....

Maximum local FA

$$\begin{aligned}\mathrm{MLFA}_n &:= \sup_{m\geq 0} \mathbb{P}_0\left(\tau \leq m+n \,|\, \tau > m\right) \\ &= \sup_{m\geq 0} \frac{\mathbb{P}_0(m < \tau \leq m+n)}{\mathbb{P}_0(\tau > m)}\end{aligned}$$

Again, CUSUM turns out to be optimal for large b among all procedures that satisfy $\mathrm{MLFA}_n \leq \alpha$ (Tartakovsky, 2005).

Numerically it seems that the value for m that maximizes $\mathbb{P}_0\left(\tau \leq m+n \,|\, \tau > m\right)$ is.... zero!

Alternative performance criteria

WL FA probability:	$\mathbb{P}_0(\tau_{WL} \leq N)$
Local FA probability:	$\sup_{m \geq 0} \mathbb{P}_0(\tau \leq m + n \mid \tau \geq m)$
Unconditional local FA prob.:	$\sup_{m \geq 0} \mathbb{P}_0(m \leq \tau \leq m + n)$

Asymptotic optimality?

If $\mathbb{P}_0\big(\tau_{WL} \leq N\big) \leq \alpha$, then:

(i) The ARL condition $\mathbb{E}_0\tau \geq \kappa$ is satisfied for $\kappa = (n+1)(1-\alpha)$.

(ii) The ARL condition $\mathbb{E}_0\tau_{WL} \geq \kappa_{WL}$ is satisfied for $\kappa_{WL} = \max\{N(1-\alpha), \kappa\}$.

How we may want to think about it

Define

$$S_{k:n}(m) := \sum_{i=k+m}^{n+m} \ell(X_i)\,.$$

Then

$$\mathbb{P}_0\left(\tau_{WL} \le N\right) = \mathbb{P}_0\big(\boldsymbol{M}_N > b\big)\,,$$

where

$$\boldsymbol{M}_N := \begin{pmatrix} \max\left\{S_{1:n}(0), \ldots, S_{1:n}(N)\right\} \\ \vdots \\ \max\left\{S_{n:n}(0), \ldots, S_{n:n}(N)\right\} \end{pmatrix}.$$

Get a grip on $\boldsymbol{M}_n$

For many popular CP test statistics we can write

$$\boldsymbol{Y}_m = C\boldsymbol{Y}_{m-1} + \vartheta \mathbf{1} \ell(X_{m+n}),$$

where $\boldsymbol{Y}_m := \big(S_{1:n}(m), \ldots, S_{n:n}(m)\big)'$.

Get a grip on $\boldsymbol{M}_n$

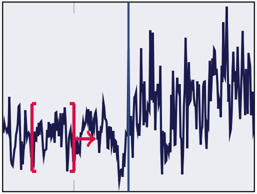
For many popular CP test statistics we can write

$$\boldsymbol{Y}_m = C\boldsymbol{Y}_{m-1} + \vartheta\mathbf{1}\ell(X_{m+n}),$$

where $\boldsymbol{Y}_m := \big(S_{1:n}(m), \ldots, S_{n:n}(m)\big)'$.

Example: CUSUM

$$S_{k:n}(m) = S_{k:n}(m-1) - \ell(X_{k+m-1}) + \ell(X_{n+m})$$



Get a grip on $\boldsymbol{M}_n$

For many popular CP test statistics we can write

$$\boldsymbol{Y}_m = C\boldsymbol{Y}_{m-1} + \vartheta \mathbf{1}\ell(X_{m+n}),$$

where $\boldsymbol{Y}_m := \big(S_{1:n}(m), \ldots, S_{n:n}(m)\big)'$.

Example: CUSUM

$$\boldsymbol{Y}_m = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \boldsymbol{Y}_{m-1} + \mathbf{1}\ell(X_{m+n})$$

Other examples:

Moving average, EWMA, non-parametric CUSUM

Ugly integral equation

Let, for fixed $\boldsymbol{x} \in \mathbb{R}^n$,

$$Q_m(\boldsymbol{x}) := \mathbb{P}\left(\boldsymbol{M}_m \le \boldsymbol{x}\right)$$

for $m \ge 0$.

Proposition

We have $Q_m(\boldsymbol{x}) = \mathcal{K}\, Q_{m-1}(\boldsymbol{x})$ for $m \ge 0$, where

$$\mathcal{K}\, h(\boldsymbol{x}) = \int_{\mathbb{R}} F\left(\frac{1}{\vartheta} \min_{i=1,\dots,n} \left\{\left(\boldsymbol{x} - \Psi(\boldsymbol{z})\right)_i\right\}\right) \mathrm{d}h(\boldsymbol{z})\,.$$

Numerical solution: Withers and Nadarajah (2014)

Extreme value distribution

Theorem (Amram, 1985)

For standard stationary Gaussian random vectors $\boldsymbol{Y}_m$ that are "not too dependent", and large m:

$$\mathbb{P}(\boldsymbol{M}_{m,i} \leq a_m \boldsymbol{x}_i + c_m,\ i = 1, \ldots, n) \approx \prod_{i=1}^{n} \exp\left(-\exp(-\boldsymbol{x}_i)\right),$$

(we know what a_m and c_m are).

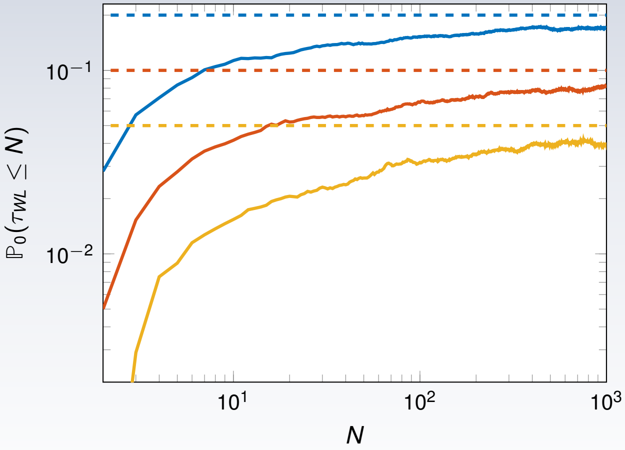
Set the threshold to be the constant

$$b_N := -a_N \log\left(-\frac{1}{n}\log(1-\alpha)\right) + c_N\,,$$

for fixed N and n.

Example

Dashed lines: $\alpha \in \{0.05, 0.1, 0.2\}$



Gaussian approximation

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

Example: State space model

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

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

Large deviation approximation

The Gärtner Ellis theorem suggests to pick the threshold **function** b such that it satisfies

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

for all $0 \leq \beta \leq 1$ such that $n\beta + 1$ integer, where we can compute

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

Large deviation approximation

The Gärtner Ellis theorem suggests to pick the threshold **function** b such that it satisfies

$$\alpha = \exp\left(-n\mathcal{I}\big(b(\beta)\big)\right)$$

for all $0 \leq \beta \leq 1$ such that $n\beta + 1$ integer, where we can compute

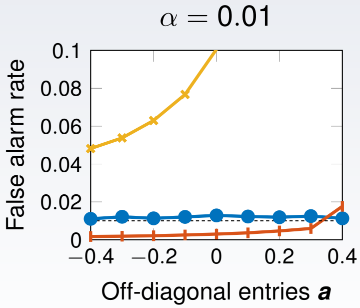
$$\mathcal{I}\big(b(\beta)\big) = \sup_{\lambda}\left\{\lambda b(\beta) + (1-\beta)\frac{\lambda}{2}(1-\lambda)\rho'\Omega^{-1}\rho\right\}.$$

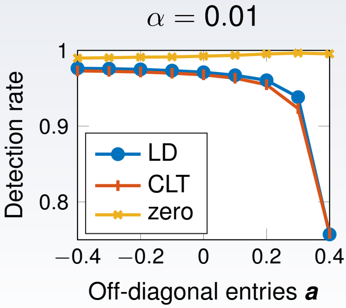
$=>$ Simple closed form expression for $b(\beta)$.

Example: State space model

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

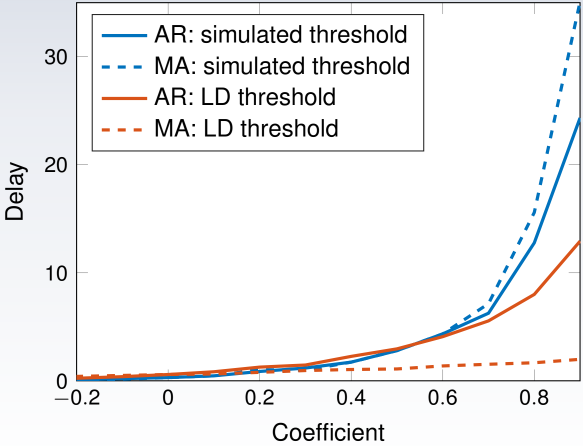
$$V_t = BX_t + Z_t + \begin{pmatrix} 2 \\ 2 \end{pmatrix} \mathbb{1}_{\{t \geq k\}}$$

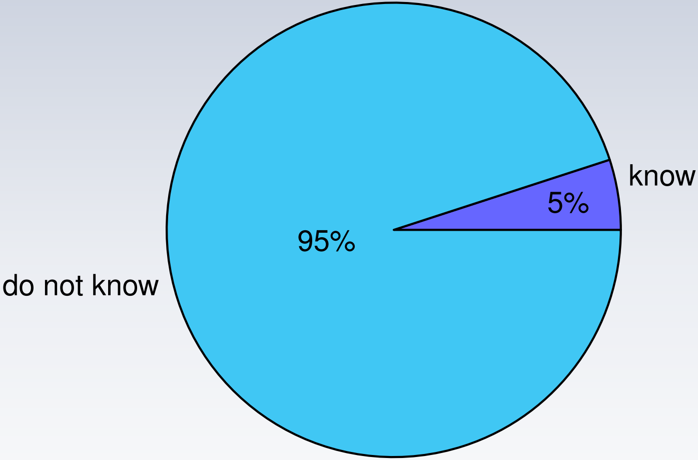




Example: ARMA model

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




know
5%
95%
do not know

Some References

- W. Ellens, J. K., M. M. and P. Zuraniewski Changepoint detection in serially correlated traffic streams. *In submission*.
- J. K., M. M. and T. T. (2015). Mean shift detection in state space models. *MODSIM 2015, Conference Proceedings*.
- L. Lai (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Trans. Information Theory*.
- A. Tartakovsky (2005). Asymptotic performance of a multichart CUSUM test under false alarm probability constraint. *IEEE Decision and Control, Conference Proceedings*.
- C. Withers and S. Nadarajah (2014). The distribution of the maximum of the multivariate AR(p) and multivariate MA(p) processes. *Statistics and Probability Letters*.