Sequential detection of a first-entry-to-a-set (aka Sequential Change Detection)

George V. Moustakides
University of Patras, Greece
& Rutgers University, USA

Outline

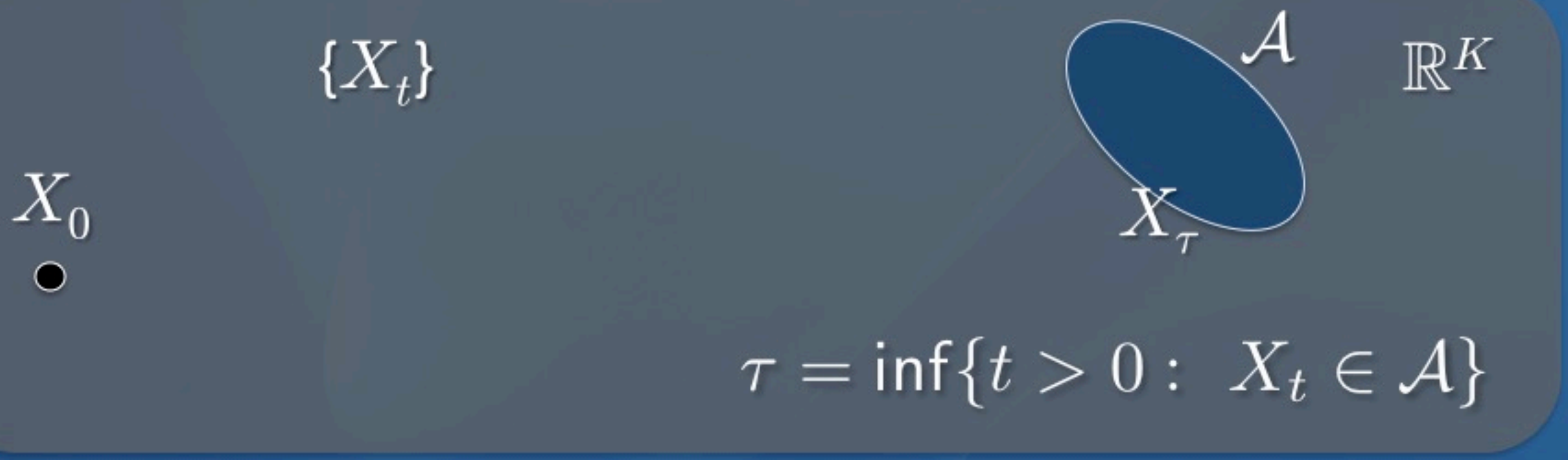
- Detecting the first-entry-to-a-set.
  Connection to Sequential Change Detection
    - Immediate detection
    - Delayed detection
- Known and Unknown entry set
- Existing and recent optimality results
- Under investigation ("Big Data" challenges)

*Collaborative research project funded by the US **NSF***
***UIUC**: V. Veeravalli, G. Fellouris. **Rutgers Univ**: G. Moustakides*

First entry to a set (of a process $\{X_t\}$)

A random vector process $\{X_t\}$ evolves in time in $\mathbb{R}^K$

$\mathcal{A}$ is a subset in $\mathbb{R}^K$



τ is random, depends on realization of $\{X_t\}$.

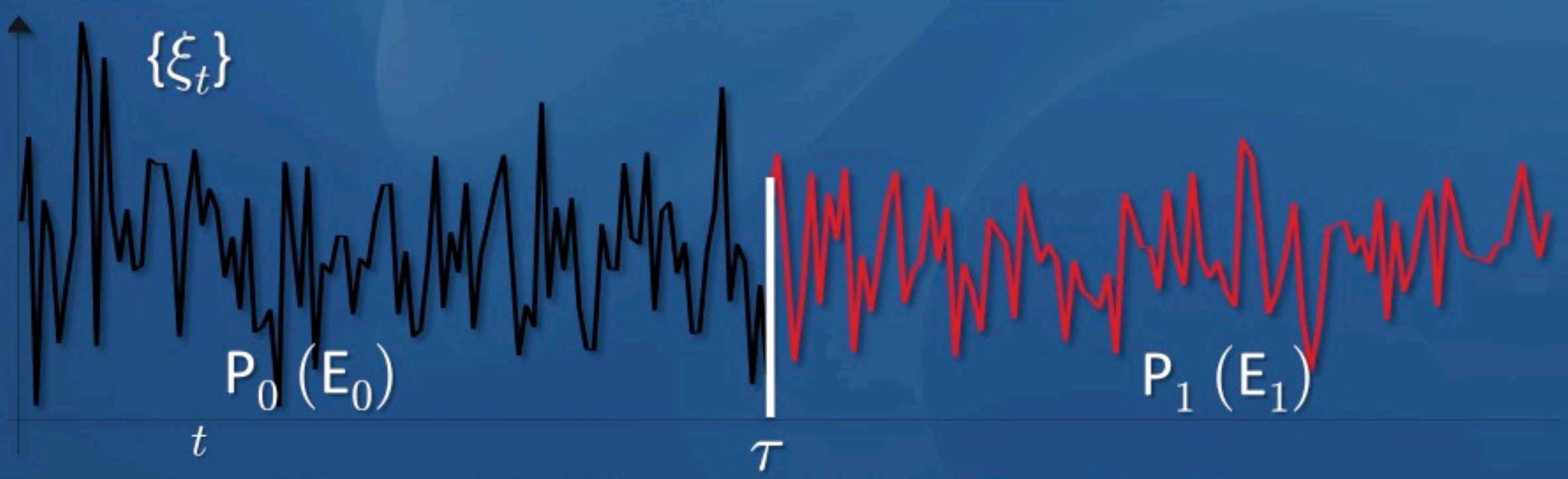
We would like to detect it.

If $\{X_t\}$ observable and $\mathcal{A}$ known, problem is trivial.

If $\{X_t\}$ (partially) hidden, problem becomes challenging.

Instead of $\{X_t\}$ we observe process $\{\xi_t\}$





Sequential Change Detection Problem

- First-entry: Model for change-imposing mechanism.
- Unifies various versions of sequential change detection.
- Deeper understanding of change-imposing mechanism may lead to more efficient detectors.

Detection of τ

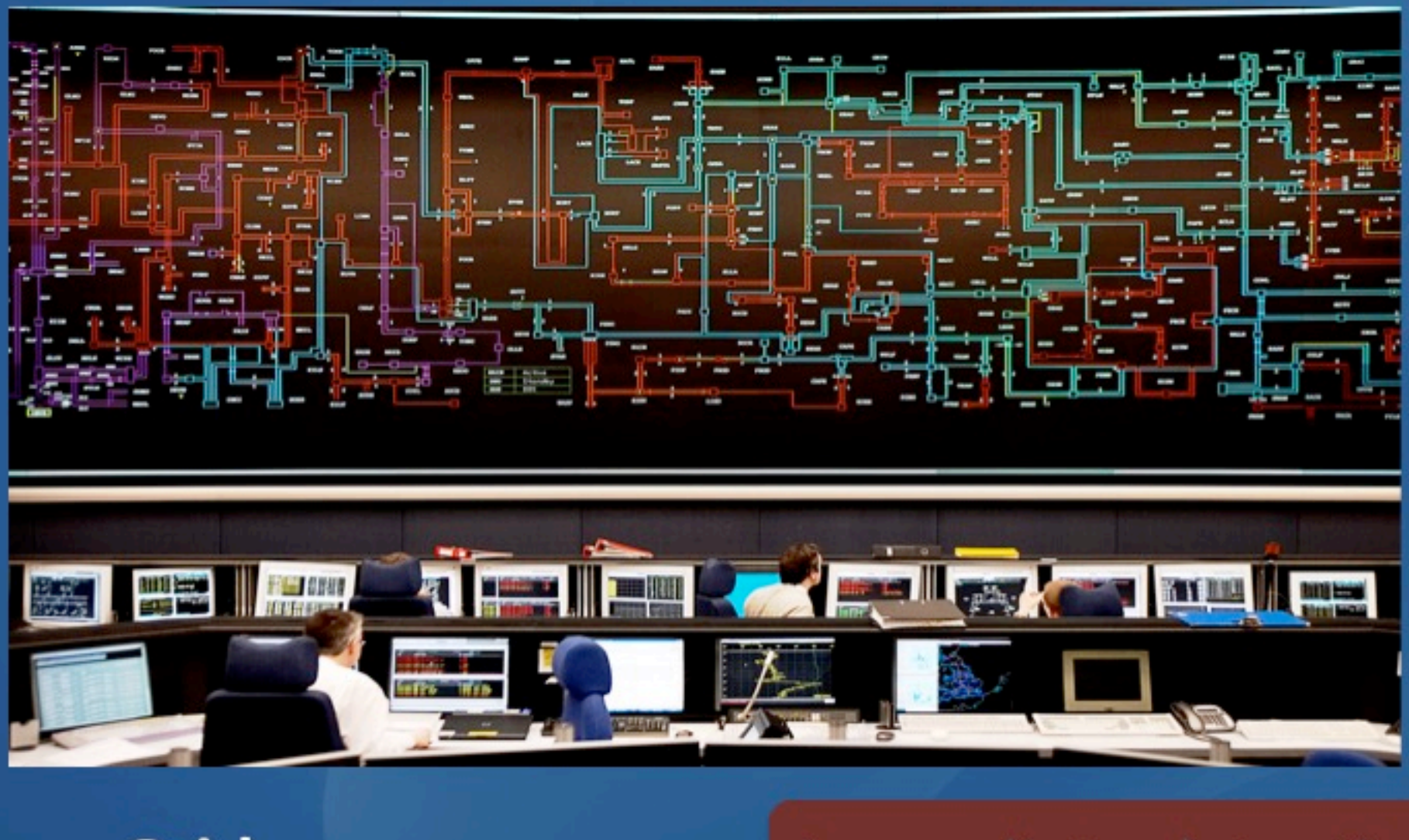
Using observations $\{\xi_t\}$, we like to decide about time τ of first entry of $\{X_t\}$ into $\mathcal{A}$.

T the time we decide that τ has occurred.

T is random, depends on the realization of the observations $\{\xi_t\}$.

Not necessarily!

τ depends on $\{X_t\}$ | $T = \tau$? | T depends on $\{\xi_t\}$



Power Grid:

Immediate detection $T = \tau$

X_t: Energy at major points in the grid.

$\xi_t = X_t + W_t$ noisy measurements.

$\mathcal{A}$: If $X_t \in \mathcal{A}$ then, after short time major blackout.

$\mathcal{A}$ is known



Structure health monitoring: Delayed detection $T \geq \tau$

X_t: Vibrations at every point of the structure (state)

$\xi_t = \mathbf{A}X_t + W_t$ Low dimensional noisy measurements

$\mathcal{A}$: If $X_t \in \mathcal{A}$ then cracks (**change** in the structure)

$\mathcal{A}$ known or unknown.







Independent $\{X_t\}$ and $\{\xi_t\}$:

X_t: Time

ξ_t: Noisy measurements

Independent

$\mathcal{A}$: {12} (when time hits 12 → change)

Quality monitoring of manufacturing process (1930's)
Biomedical Engineering
Electronic Communications
Econometrics
Seismology
Speech & Image Processing (segmentation)
Vibration monitoring (Structural health monitoring)
Security monitoring (fraud detection)
Spectrum monitoring
Scene monitoring
Network monitoring (router failures, **anomaly** detection)
Epidemic detection .....

CUSUM: 2,880 hits in 2015. Google Scholar

We monitor **sequentially** a process $\{\xi_t\}$

At every time t consult available data: ξ_1 ,..., ξ_t , ξ_{t+1}

- ~~**Decide:** first entry did not take place before or at t~~
  Continue sampling
- ~~**Decide:** first entry took place before or at t~~
  Stop sampling! (done with detection)

Sequential Detector ←→ Stopping Rule → T

Examples

$$T = \inf\{t > 0 : \omega(\xi_t) \leq \nu\}$$

$$T = \inf\{t > 0 : \xi_1 + \cdots + \xi_t \geq \nu\}$$

Very rich class

Performance metrics - known entry set

Immediate Detection

$$\mathcal{J}(T) = \mathsf{P}_0(T = \tau | T \geq \tau) \quad (T < \tau : \text{False alarm})$$

$$\sup_T \mathcal{J}(T) = \sup_T \mathsf{P}_0(T = \tau | T \geq \tau)$$

$$\text{subject to:} \quad \mathsf{P}_0(T < \tau) \leq \alpha$$

Delayed Detection

$$\mathcal{J}(T) = \mathsf{E}_1[T - \tau | T \geq \tau]$$

$$\inf_T \mathcal{J}(T) = \inf_T \mathsf{E}_1[T - \tau | T \geq \tau]$$

$$\text{subject to:} \quad \mathsf{P}_0(T < \tau) \leq \alpha$$

Immediate detection - known entry set

$$\sup_T \mathcal{J}(T) = \sup_T \mathsf{P}_0(T = \tau | T \geq \tau)$$

$$\text{subject to:} \quad \mathsf{P}_0(T < \tau) \leq \alpha$$

Optimal Stopping Theory

Pair process $\{(X_t\,, \xi_t)\}$ is i.i.d. before ~~and after~~ τ with joint pdfs f_0, f_1.

Our group (2016): Define the statistic

$$\pi_t = \mathsf{P}_0(X_t \in \mathcal{A} | \xi_t)$$

and stopping rule

$$T_o = \inf\{t > 0 : \pi_t \geq \nu\}, \quad \nu \in (0, 1)$$

Threshold selected to satisfy constraint with equality.

Delayed detection - known entry set

$$\inf_T \mathcal{J}(T) = \inf_T \mathsf{E}_1[T-\tau | T \geq \tau]$$

$$\text{subject to:} \quad \mathsf{P}_0(T < \tau) \leq \alpha$$

Pair process $\{(X_t\,, \xi_t)\}$ is i.i.d. before and after τ with joint pdfs f_0,f_1.

Our group (2016): The optimum test

$$S_t = S_{t-1}\frac{f_1(\xi_t)}{(1-\pi_t)f_0(\xi_t)} + \frac{\pi_t}{1-\pi_t}$$

$$\pi_t = \mathsf{P}_0(X_t \in \mathcal{A}|\xi_t)$$

$$T_o = \inf\{t > 0 : S_t \geq \nu\}$$

Threshold selected to satisfy constraint with equality.

Shiryaev test

Pair process $\{(X_t\,,\xi_t)\}$ is i.i.d. before and after τ with joint pdfs f_0, f_1.

If additionally we assume $\{X_t\}$ and $\{\xi_t\}$ independent then we recover the **Shiryaev test** (1963).

We note that:

$$\pi_t = \mathsf{P}_0(X_t \in \mathcal{A}|\xi_t) = \mathsf{P}_0(X_t \in \mathcal{A}) = \pi$$

The previous test becomes equivalent to:

$$S_t = (S_{t-1}+1)\frac{f_1(\xi_t)}{(1-\pi)f_0(\xi_t)}$$

$$T_{\mathsf{S}} = \inf\{t>0 : S_t \geq \nu\}$$

Oldest test of Sequential Change Detection

Unknown entry set

- What if entry set $\mathcal{A}$ is unknown?
- Bridge example: We do not know which set the state must enter, in order to produce a crack.
- Can we still detect the first-entry to an unknown set? Equivalently detect the change-time τ that inflicts change in statistics?

Focus on
change of the statistics

Performance metrics - unknown entry set

Immediate Detection

Worst-case analysis

$$\mathcal{J}(T) = \inf_{\tau} \mathsf{P}_1(T = \tau + 1 | T \geq \tau)$$

$$\sup_{T} \mathcal{J}(T) = \sup_{T} \inf_{\tau} \mathsf{P}_1(T = \tau + 1 | T \geq \tau)$$

$$\text{subject to : } \mathsf{E}_0[T] \geq \gamma$$

Delayed Detection

$$\mathcal{J}(T) = \sup_{\tau} \mathsf{E}_1[T - \tau | T \geq \tau]$$

$$\inf_{T} \mathcal{J}(T) = \inf_{T} \sup_{\tau} \mathsf{E}_1[T - \tau | T \geq \tau]$$

$$\text{subject to : } \mathsf{E}_0[T] \geq \gamma$$

Delayed detection - unknown entry set

$\{X_t\}$ and $\{\xi_t\}$ independent processes

$$\mathcal{J}_{\mathsf{P}}(T) = \sup_{\tau} \mathsf{E}_1[T - \tau | T \geq \tau]$$

We can show:

$$\mathcal{J}_{\mathsf{P}}(T) = \sup_{t \geq 0} \mathsf{E}_1[T - t | T \geq t] \quad \text{Pollak (1985)}$$

$$\inf_{T} J_{\mathsf{P}}(T) \;\; \text{subject to} : \mathsf{E}_0[T] \geq \gamma$$

Discrete time: i.i.d. data before and after the change with pdfs f_0, f_1.

Shiryaev-Roberts-Pollak test

Compute recursively the following statistic:

$$S_t = (S_{t-1} + 1)\frac{f_1(\xi_t)}{f_0(\xi_t)} \quad T_{\mathsf{P}} = \inf\{t > 0 : S_t \geq \nu\}$$

Pollak (1985): If S_0 specially designed, then

$$[J_{\mathsf{P}}(T_{\mathsf{P}}) - \inf_T J_{\mathsf{P}}(T)] \to 0; \quad \text{as } \gamma \to \infty$$

Exact optimality? Tartakovsky (2012) counterexample.

Continuous-time? Time variation? Dependence? Multiple pre- and/or post-change possibilities?

$\{X_t\}$ and $\{\xi_t\}$ dependent processes

$$\mathcal{J}_{\mathsf{L}}(T) = \sup_{\tau} \mathsf{E}_1[T - \tau | T \geq \tau]$$

We can show:

$$\mathcal{J}_{\mathsf{L}}(T) = \sup_{t \geq 0} \sup_{\xi_1, \ldots, \xi_t} \mathsf{E}_1[(T-t)^+ | \xi_1, \ldots, \xi_t]$$

Lorden (1971)

$$\inf_T J_{\mathsf{L}}(T) \quad \text{subject to : } \mathsf{E}_0[T] \geq \gamma$$

Discrete time: i.i.d. data before and after the change with pdfs f_0, f_1.

CUSUM test

$$S_t = \max\{S_{t-1}, 1\} \frac{f_1(\xi_t)}{f_0(\xi_t)}$$

$$= \{S_{t-1} + (1 - S_{t-1})^+\} \frac{f_1(\xi_t)}{f_0(\xi_t)}$$

$$T_C = \inf\{t > 0 : S_t \geq \nu\}$$

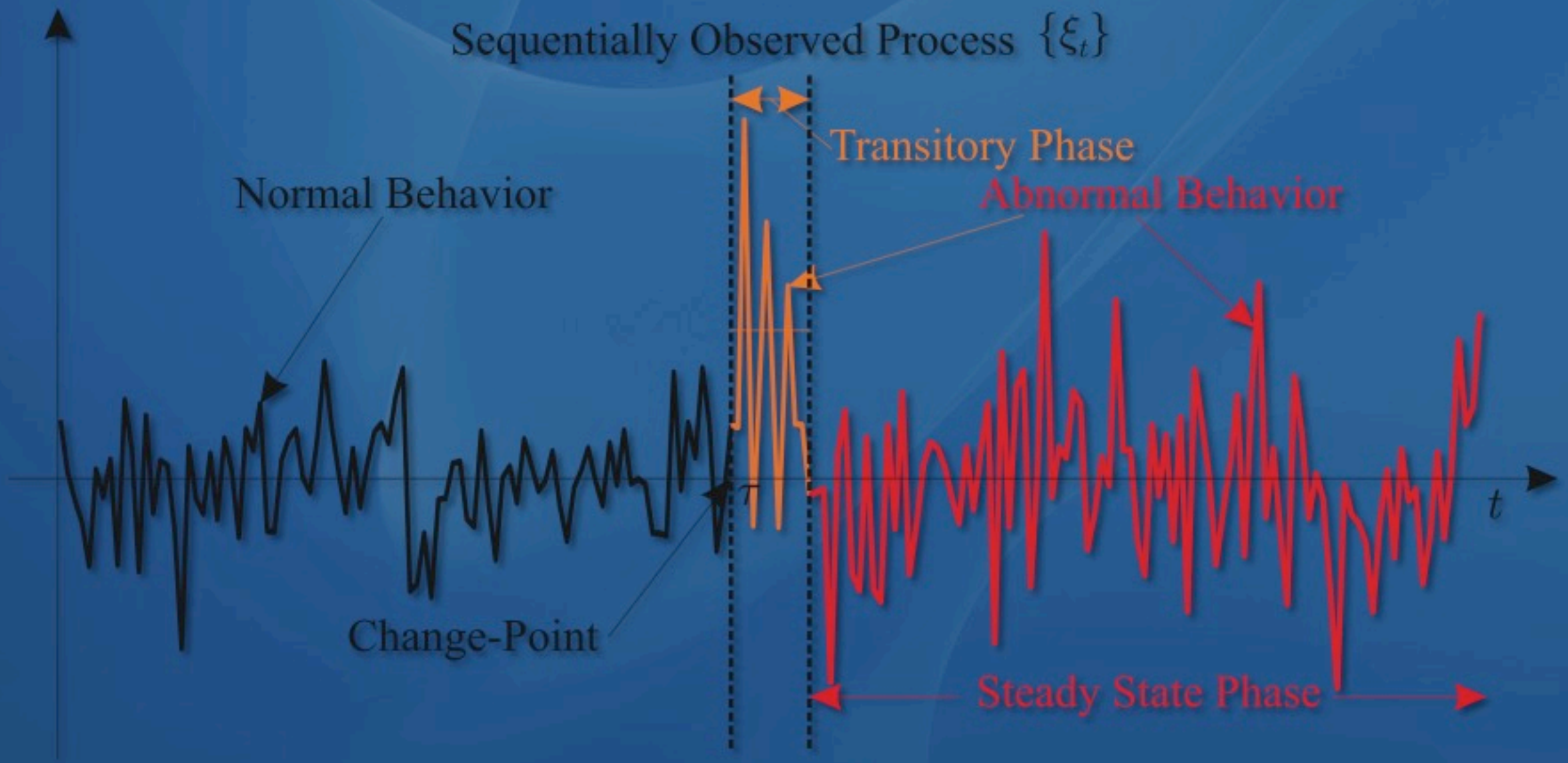
Discrete time: Lorden (1971) asymptotic optimality.
Moustakides (1986) exact optimality.

Continuous time: Shiryaev (1996), Beibel (1996)
Brownian Motion with constant drift before and after.
Moustakides (2004) strict optimality for Ito processes.

Dependence? Multiple pre- and/or post-change possibilities? ***Time variation?***

Our group: Time variation

$f_0(\xi_1),\ldots,f_0(\xi_\tau),f_1(\xi_{\tau+1}),f_1(\xi_{\tau+2}),\ldots$ Abrupt change



$f_0(\xi_1),\ldots,f_0(\xi_\tau),g_1(\xi_{\tau+1}),f_1(\xi_{\tau+2}),\ldots$ Transitory change

Power networks: Power line outage detection

$$\mathcal{J}_{\mathsf{L}}(T) = \sup_{t\geq 0} \sup_{\xi_1,\ldots,\xi_t} \mathsf{E}_1[(T-t)^+ | \xi_1,\ldots,\xi_t] \quad \text{Lorden}$$

$$\inf_T J_{\mathsf{L}}(T) \quad \text{subject to} : \mathsf{E}_0[T] \geq \gamma$$

$$S_t = S_{t-1}\frac{f_1(\xi_t)}{f_0(\xi_t)} + \varphi(S_{t-1})\frac{g_1(\xi_t)}{f_0(\xi_t)}$$

$$T_{\mathsf{C}} = \inf\{t > 0 : S_t \geq \nu\}$$

Moustakides & Veeravalli (2016): for specially designed function $\varphi(S)$ we have optimality.

Classical case $g_1 = f_1$ requires $\varphi(S) = (1 - S)^+$ (CUSUM)

Immediate detection - unknown entry set

$$\mathcal{J}(T) = \mathsf{P}_1(T = \tau + 1 | T \geq \tau)$$ Shiryaev like

$$\mathcal{J}_{\mathsf{P}}(T) = \inf_{t \geq 0} \mathsf{P}_1(T = t + 1 | T \geq t)$$ Pollak like

$$\mathcal{J}_{\mathsf{L}}(T) = \inf_{t \geq 0} \inf_{\xi_1, \ldots, \xi_t} \mathsf{P}_1(T = t + 1 | \xi_1, \ldots, \xi_t)$$

Lorden like

$$T_{\mathsf{Sh}} = \inf\left\{ t > 0 : \frac{f_1(\xi_t)}{f_0(\xi_t)} \geq \nu \right\}$$ Shewhart test (1931)

Optimality: Bojdecki (1979); Pollak and Krieger (2013); Moustakides (2014).

Time variation? Multiple pre- and/or ***post-change possibilities? Dependence?***

Recent result: Dependence

$$\mathcal{J}_{\mathsf{L}}(T) = \inf_{t\geq 0} \inf_{\xi_1,\ldots,\xi_t} \mathsf{P}_1(T = t+1 | \xi_1, \ldots, \xi_t)$$

Lorden like

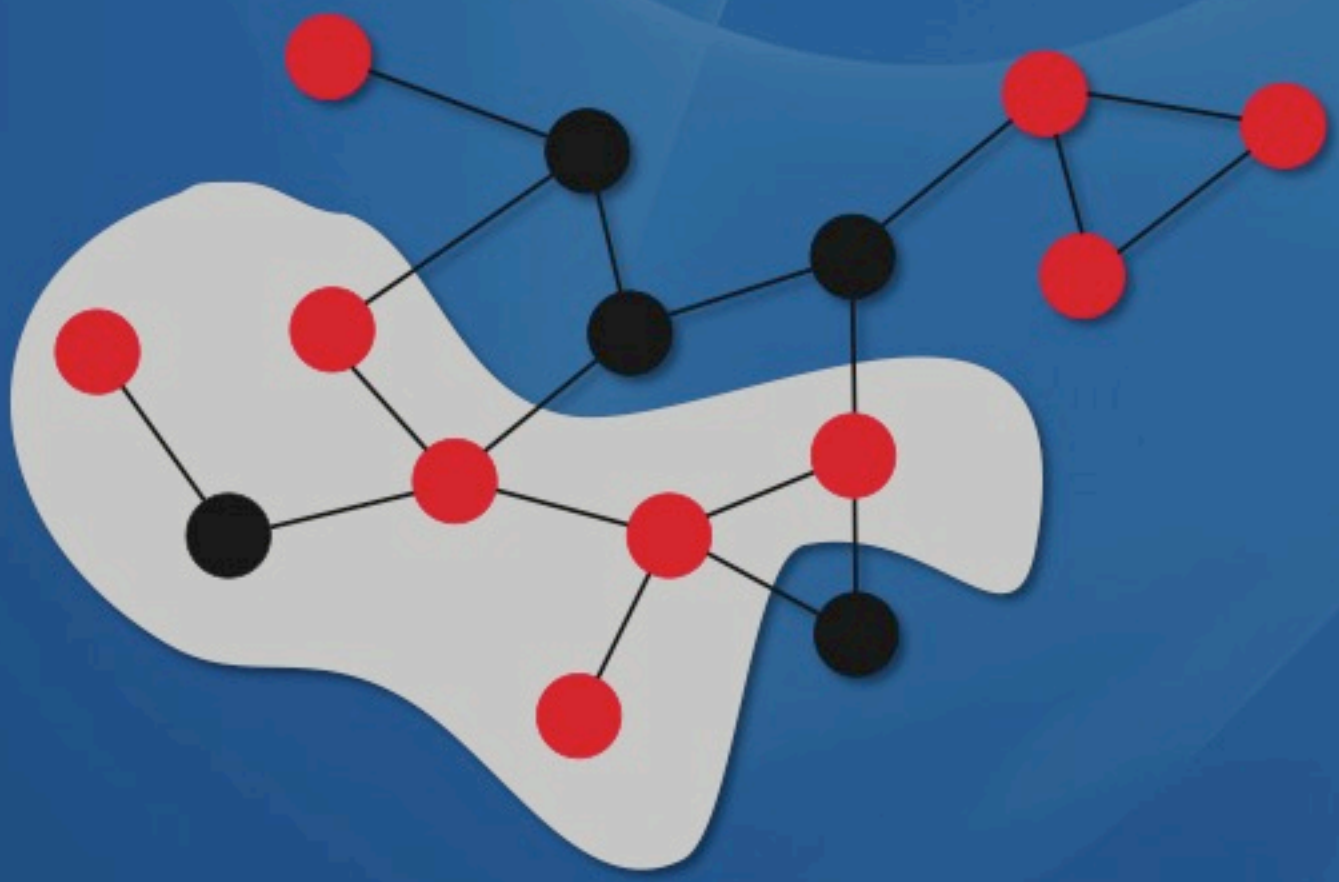
$$\sup_T \mathcal{J}_{\mathsf{L}}(T) \quad \text{subject to : } \mathsf{E}_0[T] \geq \gamma$$

Markovian pre- and post-change pdfs for observations $\{\xi_t\}$

$$T_{\mathsf{Sh}} = \inf\left\{ t > 0 : c(\xi_{t-1}) \frac{f_1(\xi_t|\xi_{t-1})}{f_0(\xi_t|\xi_{t-1})} \geq \nu(\xi_t) \right\}$$

Moustakides (2015): With properly designed functions $c(\xi)$ and $\nu(\xi)$ we solve the constrained optimization.

Under investigation

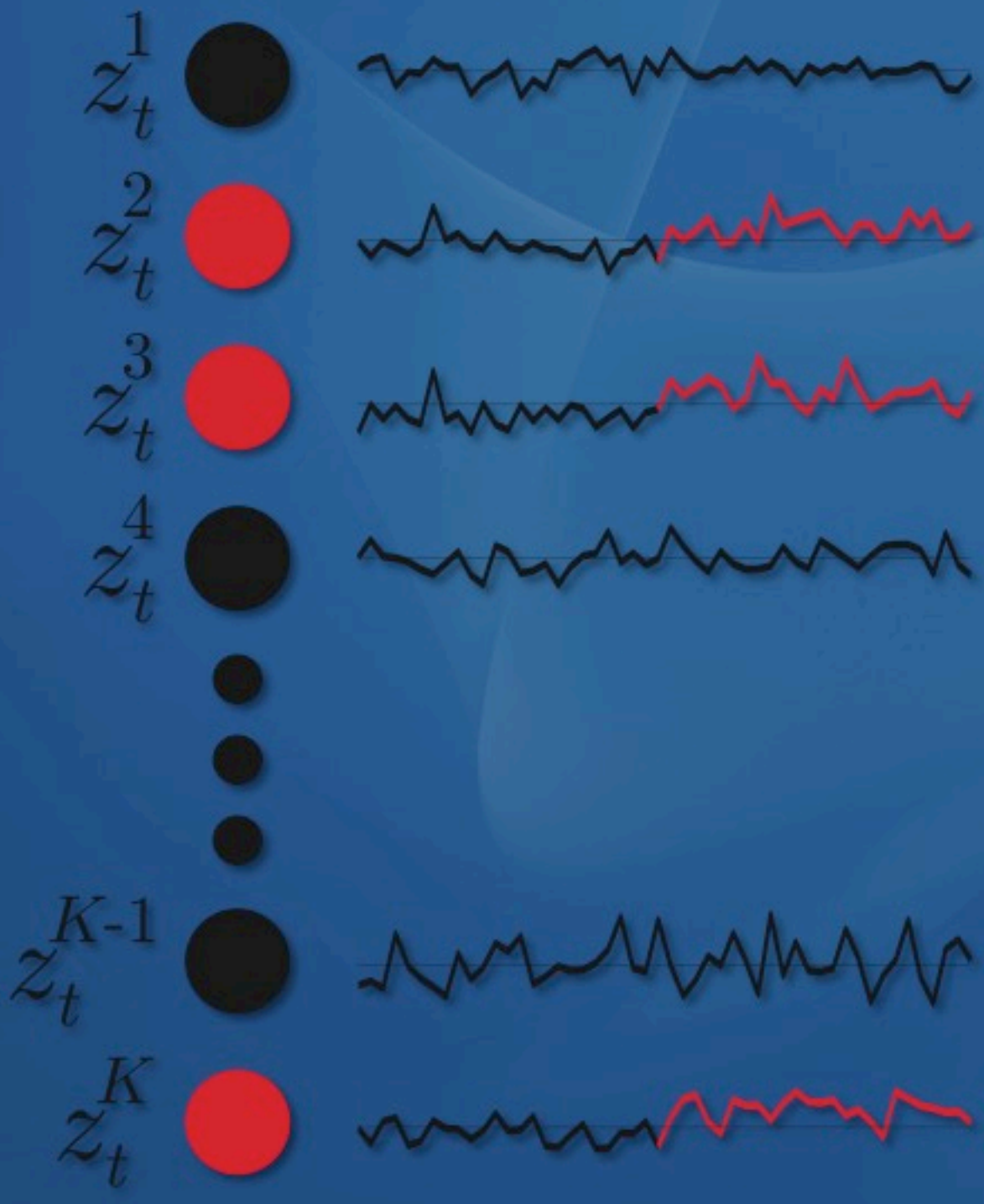


In a graph each node generates a process that is observed.

At time τ there is a change in statistics at a subgraph.

Detect change and isolate changed nodes.

On the subgraphs of interest we impose sparsity or connectivity constraints. For example when there is a unique path between two affected nodes, intermediate nodes must also be affected.



Z_t (Very) long vector.

We are interested in the problem of dimensionality reduction:

$$\xi_t = \mathbf{A} Z_t$$

$\mathbf{A}$ of dimensions $M \times K$ with $M \ll K$

We want to design matrix $\mathbf{A}$ so that with $\{\xi_t\}$ we detect efficiently sparse changes in $\{Z_t\}$.

Matrix $\mathbf{A}$ can have random elements (linear sketches).