Quickest Detection of Changes

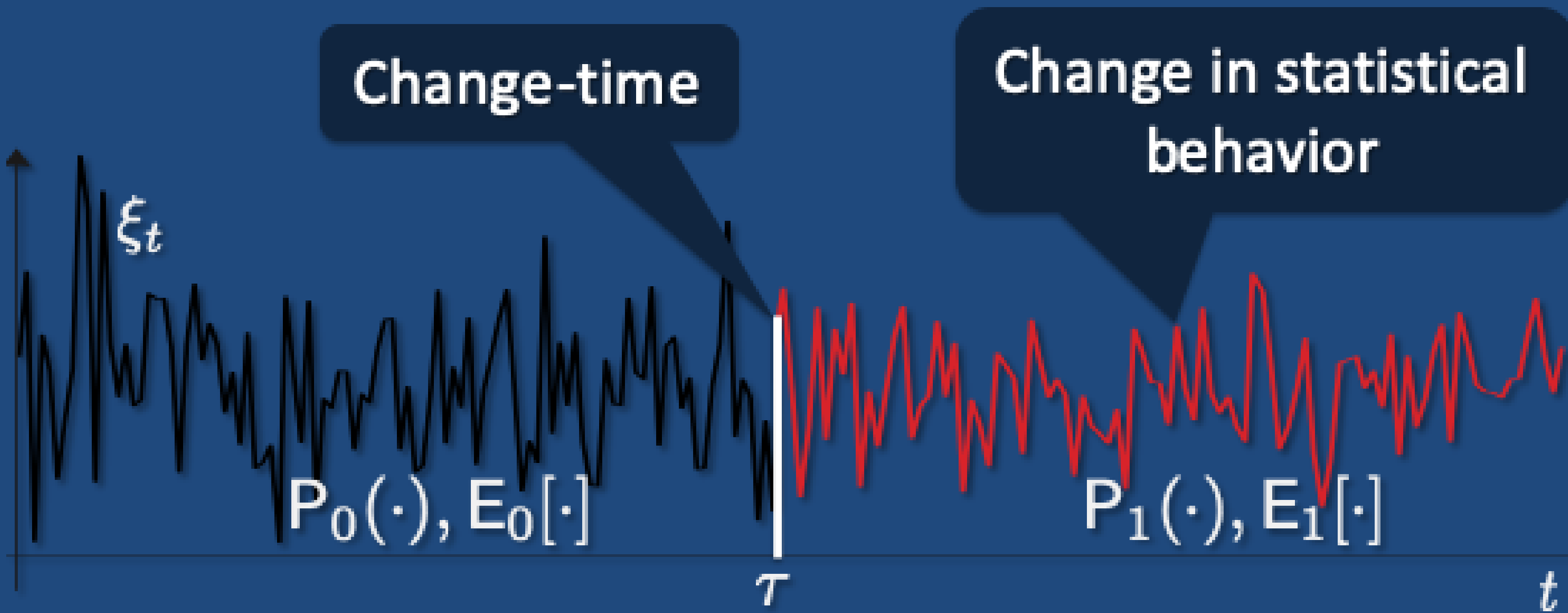
Formulations and Optimality Results

George V. Moustakides

University of Patras, Greece

Outline

- Problem definition
- Model for change mechanism
- Performance metrics
- Optimum detectors
- Decentralized formulations
- Data driven version



Detect change as soon as possible

Data become available sequentially: at each time t obtain new sample ξ_t

Detector: Every instant t consult available data $\xi_1, \ldots, \xi_t$ decide if a change took place or not

1 ξ_1 Was there a change? **NO**

2 ξ_1, ξ_2 Was there a change? **NO**

Take more data (continue sampling)

$\vdots$

$\vdots$

Stop (sampling)

T $\xi_1, \xi_2, \ldots, \xi_T$ Was there a change? **YES**

Equivalently

t $\xi_1, \xi_2, \ldots, \xi_t$

Continue or Stop sampling

Stopping rule

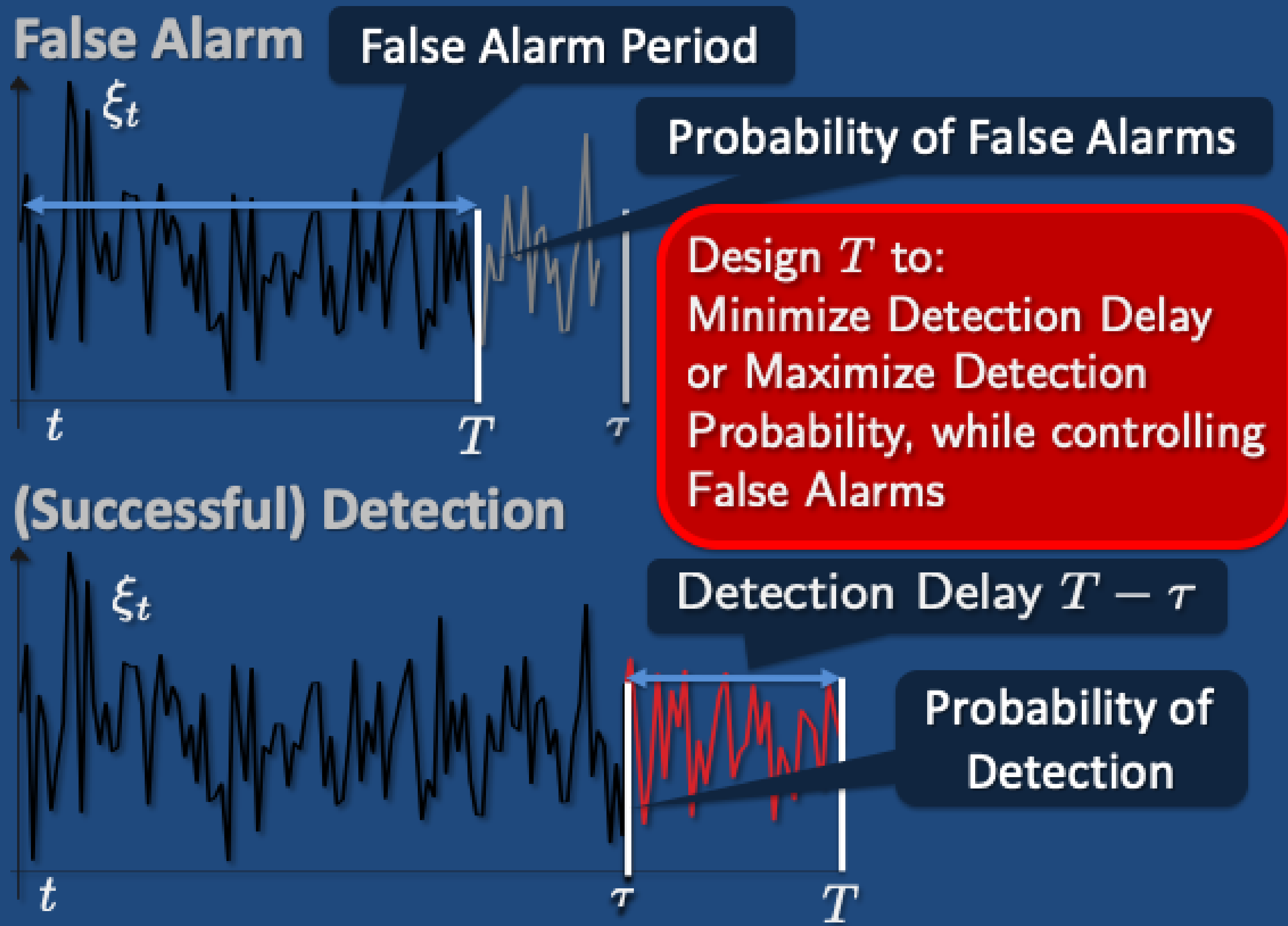
Stop

$$G_t(\xi_1, \ldots, \xi_t) \gtrless \nu$$

Continue

Stopping time

$$T = \inf_t \{G_t(\xi_1, \ldots, \xi_t) \geq \nu\}$$


False Alarm
False Alarm Period
Probability of False Alarms
ξ_t
t
T
τ
Design T to:
Minimize Detection Delay or Maximize Detection Probability, while controlling False Alarms
(Successful) Detection
Detection Delay $T - \tau$
Probability of Detection
ξ_t
t
τ
T

Quality monitoring of manufacturing process

Production line → Continuous Measurements → Quality Assessment

Medical Applications

Epidemic Detection

Disease rate measurements → Increase in rate? → Epidemic outbreak?

Early Detection of Epilepsy Episode

EEG Wearables → Divergence from normal → Episode?

Financial Applications

- Structural Change-detection in Exchange Rates
- Portfolio Monitoring

Electronic Communications

Seismology

Speech & Image Processing (segmentation)

Vibration monitoring (Structural health monitoring)

Security monitoring (fraud detection)

Spectrum monitoring

Scene monitoring

Network monitoring (router failures, attack detection)

⋮

τ

Stopping time

Dependent $\{X_t\}$, $\{\xi_t\}$

X_t: Vibrations at points of the structure

$\xi_t = AX_t + W_t$: Low dimensional noisy observations

If $\|X_t\|^2 \geq \nu$, then crack (change in structure)

Mechanism consults $X_1, \ldots, X_t$ and decides to apply change (**stop** nominal behavior) or not

Independent $\{X_t\}$, $\{\xi_t\}$

X_t: Coordinates of ball

ξ_t: Noisy sensor observations

If X_t inside the volume under the goal, “apply” change

At some point in time:
Attack!!!

Attacker no access to observations…

$\{X_t\}, \{\xi_t\}$ independent

Change mechanism
Consults $\{X_t\}$ and controls stopping time τ

Scientist
Consults $\{\xi_t\}$ and controls stopping time T

Design T to detect τ asap

Detection Delay

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

Hard Limited Delay

$$\mathcal{P}(T) = \mathsf{P}_1(T \leq \tau + M | T > \tau)$$

If stopping rule used by change mechanism unknown computation of $\mathcal{J}(T), \mathcal{P}(T)$ is not possible

We follow a **worst-case analysis**

Detection Delay

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

$$= \sup_{t>0} \sup_{\xi_1,\ldots,\xi_t} \mathsf{E}_1[T-t|T>t,\xi_1,\ldots,\xi_t]$$

Lorden (1971)

$$= \sup_{t>0} \mathsf{E}_1[T-t|T>t] \text{ if } \{X_t\},\{\xi_t\} \text{ independent}$$

Pollak (1985)

Hard Limited Delay

$$\mathcal{P}(T) = \inf_{\tau} \mathsf{P}_1(T \le \tau + M | T>\tau)$$

$$= \inf_{t>0} \inf_{\xi_1,\ldots,\xi_t} \mathsf{P}_1(T \le t+M|T>t,\xi_1,\ldots,\xi_t)$$

$$= \inf_{t>0} \mathsf{P}_1(T \le t+M|T>t) \text{ if } \{X_t\},\{\xi_t\} \text{ independent}$$

Detection Delay

$$\inf_T \mathcal{J}(T) =$$

$$\inf_T \sup_{t>0} \sup_{\xi_1,\ldots,\xi_t} \mathsf{E}_1[T - t | T > t, \xi_1, \ldots, \xi_t]$$

$$\text{subject to :} \quad \mathsf{E}_0[T] \geq \gamma > 1$$

Hard Limited Delay

$$\sup_T \mathcal{P}(T) =$$

$$\sup_T \inf_{t>0} \inf_{\xi_1,\ldots,\xi_t} \mathsf{P}_1(T \leq t + M | T > t, \xi_1, \ldots, \xi_t)$$

$$\text{subject to :} \quad \mathsf{E}_0[T] \geq \gamma > 1$$

Shiryaev 1963

$\{\xi_t\}$ are i.i.d. before and after the change with corresponding pdfs $\mathrm{f}_0(\xi)$, $\mathrm{f}_1(\xi)$

τ is independent from $\{\xi_t\}$ and follows a geometric distribution $\mathrm{P}(\tau = t) = p(1-p)^t, t = 0, 1, \ldots$

$$S_t = (1 + S_{t-1}) \frac{\mathrm{f}_1(\xi_t)}{(1-p)\mathrm{f}_0(\xi_t)}$$

$$T_{\mathrm{S}} = \inf_{t>0} \{S_t \geq \nu\}$$

Threshold to satisfy false alarm constraint with equality

Lorden 1971

$\{\xi_t\}$ are i.i.d. before and after the change with corresponding pdfs $\mathsf{f}_0(\xi)$, $\mathsf{f}_1(\xi)$

τ and $\{\xi_t\}$ are dependent

$$\inf_T \mathcal{J}(T) = \inf_T \sup_{t>0} \sup_{\xi_1,\ldots,\xi_t} \mathsf{E}_1[T-t|T>t,\xi_1,\ldots,\xi_t]$$

$$\text{subject to}: \quad \mathsf{E}_0[T] \geq \gamma > 1$$

$$S_t = \max\{S_{t-1},0\} + \log\frac{\mathsf{f}_1(\xi_t)}{\mathsf{f}_0(\xi_t)}$$

$$T_{\mathrm{C}} = \inf_{t>0}\{S_t \geq \nu\}$$

Threshold to satisfy false alarm constraint with equality

CUSUM test

CUSUM test

$$S_t = \max\{S_{t-1}, 0\} + \log \frac{f_1(\xi_t)}{f_0(\xi_t)}$$

Known since 1954 as the **Page test**

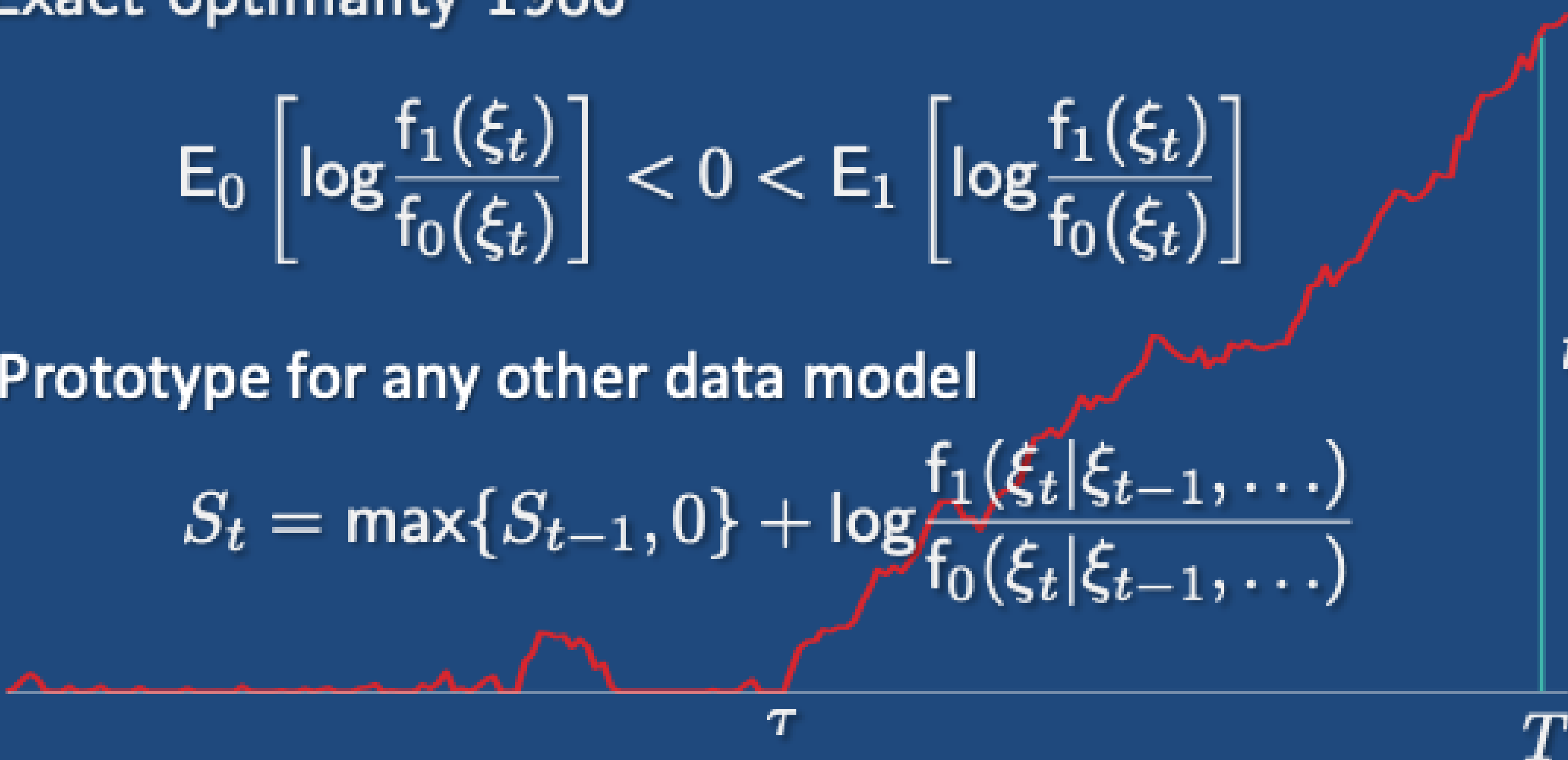
Asymptotic optimality $(\gamma \to \infty)$ 1971

Exact optimality 1986

$$E_0\left[\log \frac{f_1(\xi_t)}{f_0(\xi_t)}\right] < 0 < E_1\left[\log \frac{f_1(\xi_t)}{f_0(\xi_t)}\right]$$

Prototype for any other data model

$$S_t = \max\{S_{t-1}, 0\} + \log \frac{f_1(\xi_t|\xi_{t-1}, \ldots)}{f_0(\xi_t|\xi_{t-1}, \ldots)}$$



$\{\xi_t\}$ Markov before and after the change with corresponding pdfs $f_0(\xi_t|\xi_{t-1})$, $f_1(\xi_t|\xi_{t-1})$

Conjecture

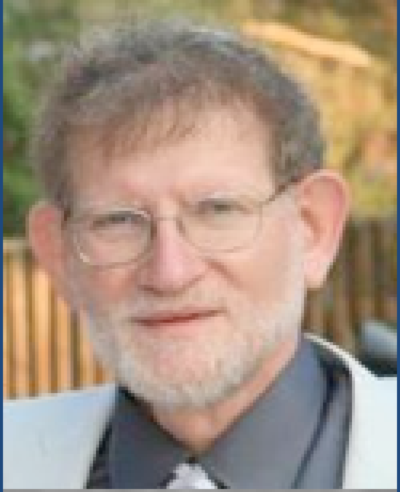
$$S_t = \max\{S_{t-1}, \phi(\xi_{t-1})\} + \log \frac{f_1(\xi_t|\xi_{t-1})}{f_0(\xi_t|\xi_{t-1})}$$

$$T_C = \inf_t\{S_t \geq \nu(\xi_t)\}$$

Functions $\phi(\xi), \nu(\xi)$ are solution to a system of integral equations. Computed either numerically or asymptotically $\gamma \to \infty$

As $\gamma \to \infty$ we have $\phi(\xi) \to 0$ and $\nu(\xi) \to \nu$

Pollak 1985



$\{\xi_t\}$ are i.i.d. before and after the change with corresponding pdfs $f_0(\xi)$, $f_1(\xi)$

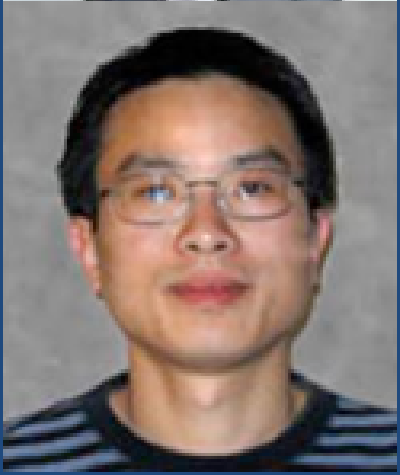
τ and $\{\xi_t\}$ are independent

$$\inf_T \mathcal{J}(T) = \inf_T \sup_{t>0} \mathsf{E}_1[T - t | T > t]$$

$$\text{subject to}: \ \mathsf{E}_0[T] \geq \gamma > 1$$

Shiryaev-Roberts-Pollak test

Mei 2006



$$S_t = (S_{t-1} + 1)\frac{f_1(\xi_t)}{f_0(\xi_t)} \quad T_{\mathrm{SRP}} = \inf_t \{S_t \geq \nu\}$$

If S_0 specially designed T_{SRP} asymptotically $(\gamma \to \infty)$ optimum.

Tartakovsky 2010

Exact optimality? 1997-2006

Tartakovsky 2019

τ and $\{\xi_t\}$ are independent

$$\inf_T \mathcal{J}(T) = \inf_T \sup_{t>0} \mathsf{E}_1[T - t | T > t]$$

$$\text{subject to}: \ \mathsf{E}_0[T] \geq \gamma > 1$$

Multiple post-change possibilities. Data after change i.i.d. with pdf $\mathsf{f}_1(\xi), \ldots, \mathsf{f}_k(\xi)$

$$S_t^i = (S_{t-1}^i + 1)\frac{\mathsf{f}_i(\xi_t)}{\mathsf{f}_0(\xi_t)}, \ i = 1, \ldots, k$$

$$T_{\mathrm{T}} = \inf_{t>0}\{S_t^1 + \cdots + S_t^k \geq \nu\}$$

T_{T} asymptotically $(\gamma \to \infty)$ optimum

Hard Limited Delay

τ and $\{\xi_t\}$ are dependent

$$\sup_T \mathcal{P}(T) = \sup_T \inf_{t>0} \inf_{\xi_1,\dots,\xi_t} \mathsf{P}_1(T \le t + M | T > t, \xi_1, \dots, \xi_t)$$

$$\text{subject to}: \ \mathsf{E}_0[T] \ge \gamma > 1$$

τ and $\{\xi_t\}$ are independent

$$\sup_T \mathcal{P}(T) = \sup_T \inf_{t>0} \mathsf{P}_1(T \le t + M | T > t)$$

$$\text{subject to}: \ \mathsf{E}_0[T] \ge \gamma > 1$$

If $\{\xi_t\}$ are i.i.d. before and after the change with pdfs $\mathsf{f}_0(\xi)$, $\mathsf{f}_1(\xi)$, and interested in $M = 1$ (detect immediately)

$$T_{\mathrm{Sh}} = \inf_{t>0} \left\{ \frac{\mathsf{f}_1(\xi_t)}{\mathsf{f}_0(\xi_t)} \ge \nu \right\}$$

Shewhart test (1931)

Markovian pre- and post-change pdfs for observations

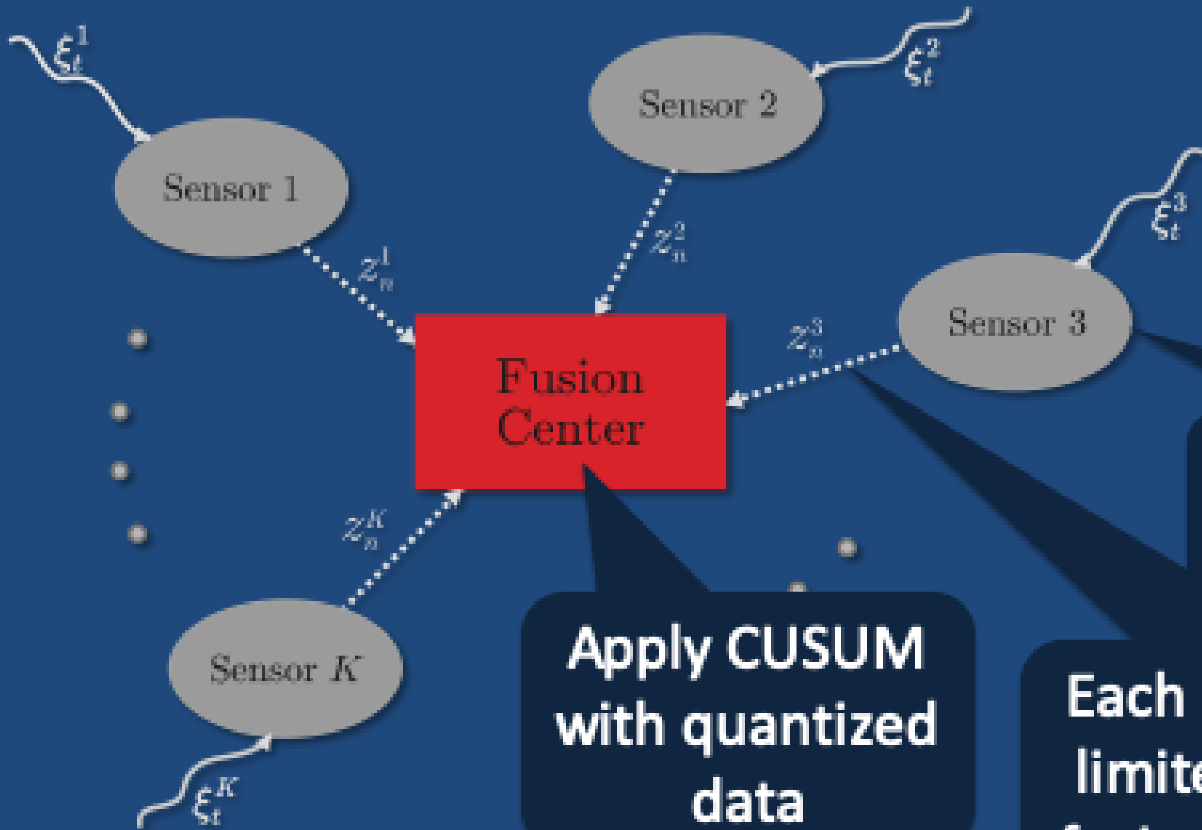
$$\sup_T \mathcal{P}(T) = \sup_T \inf_{t>0} \inf_{\xi_1,\ldots,\xi_t} \mathsf{P}_1(T = t+1 | T > t, \xi_1, \ldots, \xi_t)$$

$$\text{subject to}: \quad \mathsf{E}_0[T] \geq \gamma > 1$$

$$T_{\mathrm{Sh}} = \inf_{t>0}\left\{ c(\xi_{t-1}) \frac{\mathsf{f}_1(\xi_t|\xi_{t-1})}{\mathsf{f}_0(\xi_t|\xi_{t-1})} \geq \nu(\xi_t) \right\}$$

2015: Functions $c(\xi), \nu(\xi)$ satisfy system of integral equation. Can be computed numerically or asymptotically ($\gamma \to \infty$). Simple solution for conditionally Gaussian: $\xi_t = \alpha(\xi_{t-1}) + w_t$ where $\{w_t\}$ i.i.d. $\mathcal{N}(0,1)$

Veeravalli 2001



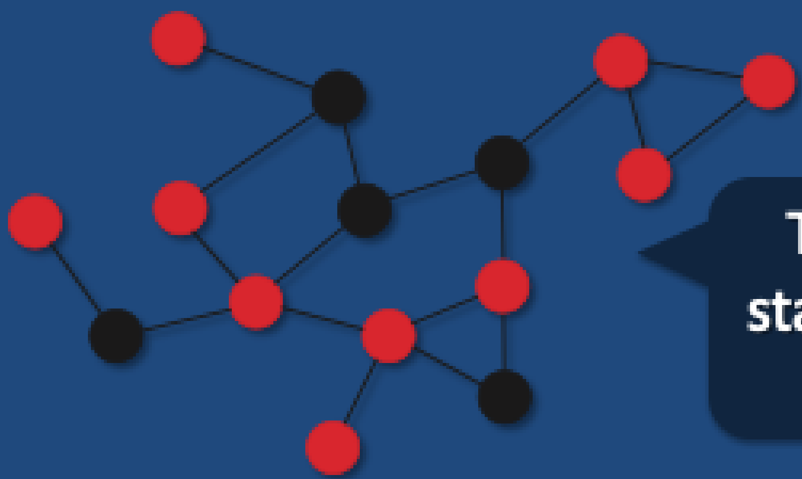
Simultaneous change in all sensors

Apply CUSUM with quantized data

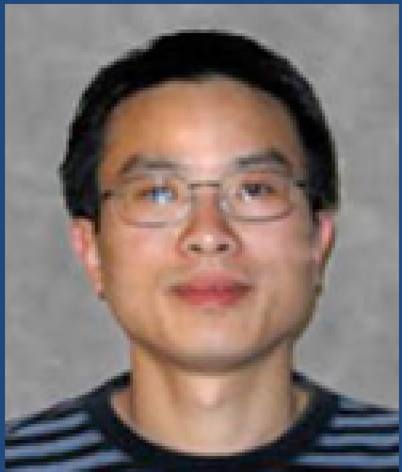
Each sensor must send limited information to fusion center (i.e. 1 bit)

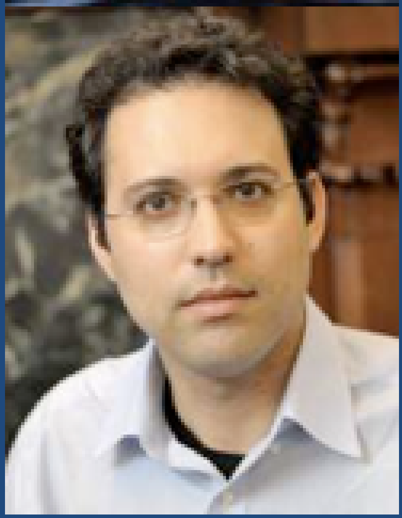
Veeravalli 2001: Apply optimum quantization on each likelihood ratio $\frac{f_1^i(\xi_t)}{f_0^i(\xi_t)}$

Mei 2011 - Fellouris 2016



There is a change in statistical behavior of a subgraph.





$$S_t^i = \max\{S_{t-1}^i, 0\} + \log \frac{f_1^i(\xi_t)}{f_0^i(\xi_t)}$$

$$T = \inf_{t>0} \{\max_i S_t^i \geq \nu\}$$

$$T = \inf_{t>0} \{S_t^1 + \cdots + S_t^K \geq \nu\}$$

$$T = \inf_{t>0} \{\textstyle\sum_{i=1}^K \log\left(1 - \pi + \pi e^{S_t^i}\right) \geq \nu\}$$

What if $f_0(\xi), f_1(\xi)$ NOT known

Use data to estimate $f_0(\xi), f_1(\xi)$

$\{\xi_1^0, \ldots, \xi_N^0\}$ pre-change (training) data

$\{\xi_1^1, \ldots, \xi_N^1\}$ post-change (training) data

Density estimation, not very efficient

$$S_t = \max\{S_{t-1}, 0\} + \log \frac{f_1(\xi_t)}{f_0(\xi_t)}$$

Can we train a NN to estimate the $\log \frac{f_1(\xi_t)}{f_0(\xi_t)}$ using the training data ?

Need proper optimization problem!!!

Consider scalar functions $\phi(z), \psi(z)$ and scalar function $u(\xi)$. Define average cost:

$$\mathcal{J}(u) = \mathsf{E}_0\big[\phi\big(u(\xi)\big)\big] + \mathsf{E}_1\big[\psi\big(u(\xi)\big)\big]$$

If we want $\min_u \mathcal{J}(u)$ to yield

$$u(\xi) = \log \frac{\mathsf{f}_1(\xi)}{\mathsf{f}_0(\xi)}$$

it is necessary and sufficient

$$\psi'(u) = -e^{-u}\phi'(u)$$

$$\mathcal{J}(u) = \mathsf{E}_0\big[\phi\big(u(\xi)\big)\big] + \mathsf{E}_1\big[\psi\big(u(\xi)\big)\big]$$

Let $u(\xi, \theta)$ be the output of a NN with input ξ then

$$\hat{\mathcal{J}}(\theta) = \tfrac{1}{N}\sum_{l=1}^{N}\phi\big(u(\xi_l^0, \theta)\big) + \tfrac{1}{N}\sum_{l=1}^{N}\psi\big(u(\xi_l^1, \theta)\big)$$

Optimize over θ (training)

$$\theta_t = \theta_{t-1} - \mu\left\{\nabla_\theta\phi\big(u(\xi_t^0, \theta_{t-1})\big) + \nabla_\theta\psi\big(u(\xi_t^1, \theta_{t-1})\big)\right\} \to \theta_*$$

$$S_t = \max\{S_{t-1}, 0\} + u(\xi_t, \theta_*) \qquad T_{\mathrm{C}} = \inf_{t>0}\{S_t \geq \nu\}$$

$$\xi_t = \begin{cases} \text{Before change : i.i.d. Gaussian} \\ \text{After change : } \operatorname{sign}(\xi_{t-1})\sqrt{|\xi_{t-1}|} + w_t. \end{cases}$$

$$\log \frac{f_1(\xi_t|\xi_{t-1})}{f_0(\xi_t|\xi_{t-1})} = \underbrace{\log \frac{f_1(\xi_t, \xi_{t-1})}{f_0(\xi_t, \xi_{t-1})}}^{2 \times 50 \times 1} - \underbrace{\log \frac{f_1(\xi_{t-1})}{f_0(\xi_{t-1})}}^{1 \times 20 \times 1}$$