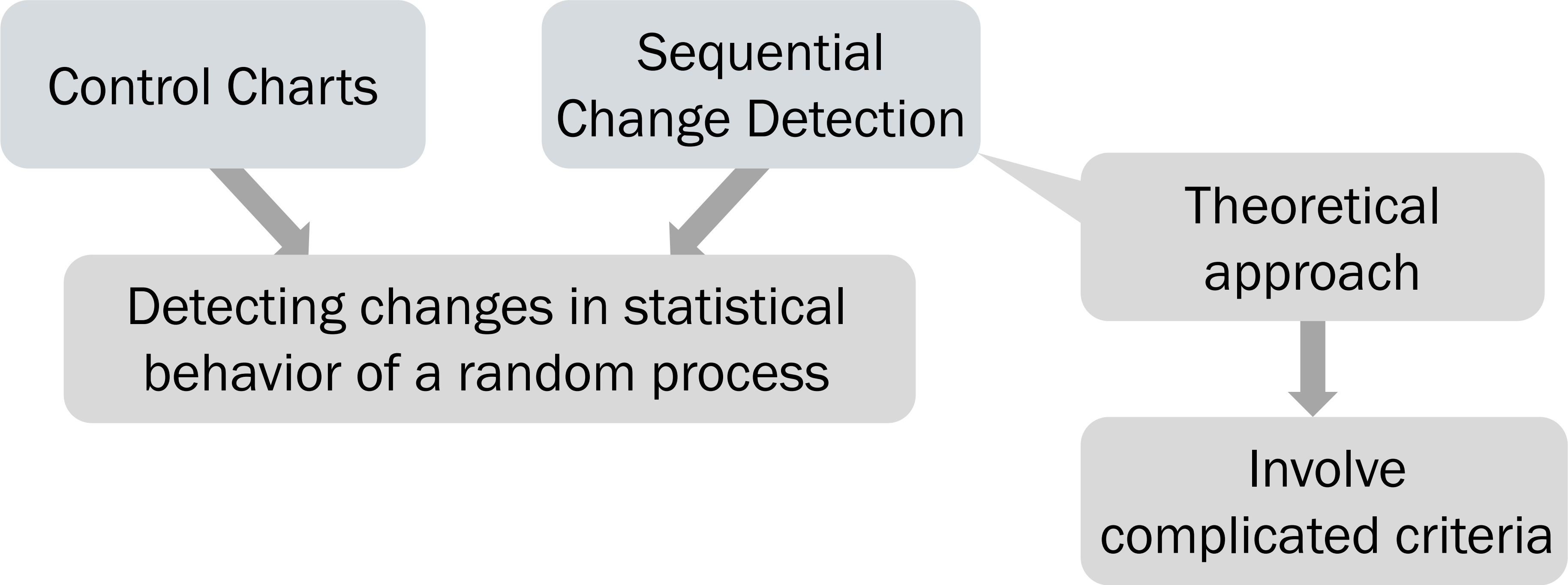
Designing Optimum Tests for Sequential Detection of Changes

George Moustakides

Electrical and Computer Engineering
University of Patras, GREECE

Preview



$$\sup_{t \geq 0} \mathbb{E}_1 \left[T - t \,|\, T > t \right] \quad \text{(1983, Pollak)}$$

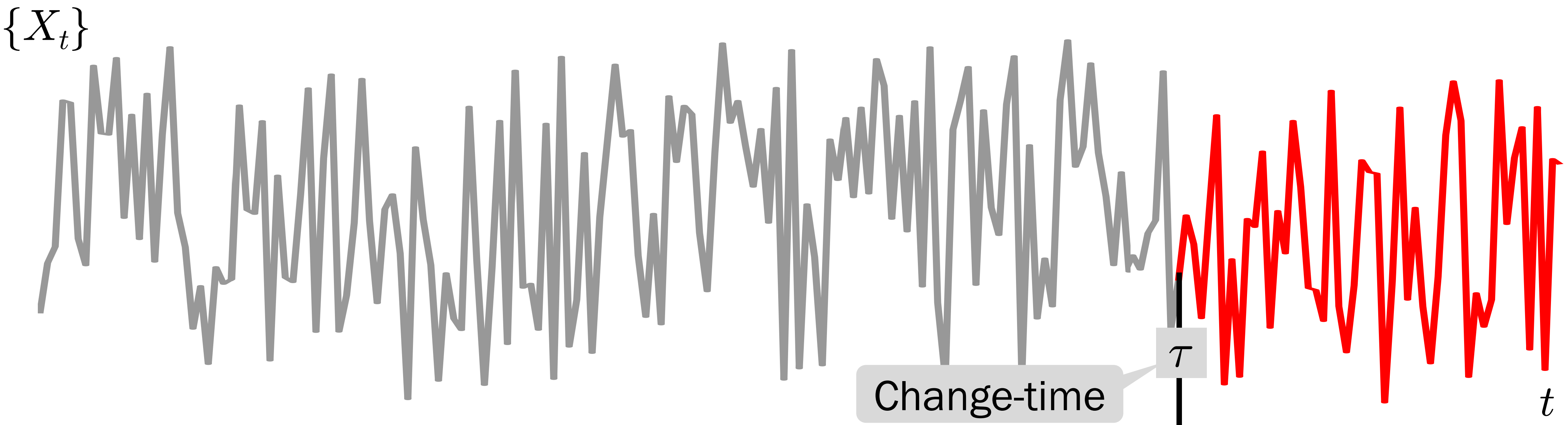
$$\sup_{t \geq 0} \sup_{X_1, \ldots, X_t} \mathbb{E}_1 \left[T - t \,|\, T > t, X_1, \ldots, X_t \right] \quad \text{(1971, Lorden)}$$

Outline

- Sequential Detection of Changes - Problem Definition
- Detection Strategies
- Understanding Changes
- Performance Criteria and Optimum Tests
  - Pollak Criterion and the SRP Test
  - Lorden Criterion and the CUSUM Test
  - Maximal Detection Probability and the Shewhart Test
- Advanced Versions
  - Unknown Parameters
  - Data-Driven

Problem Definition

Observe sequentially a random process $\{X_t\}$ evolving in time



Control Charts	In Control	Out of Control
Sequential Change Detection	Pre-Change Nominal	Post-Change Alternative

Detect change in statistical behavior as soon as possible

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

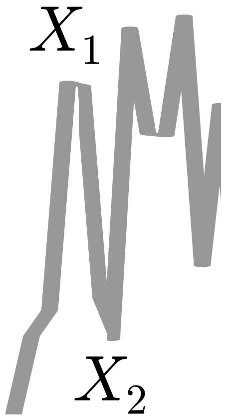
Network Monitoring (router failures, attack detection)

⋮

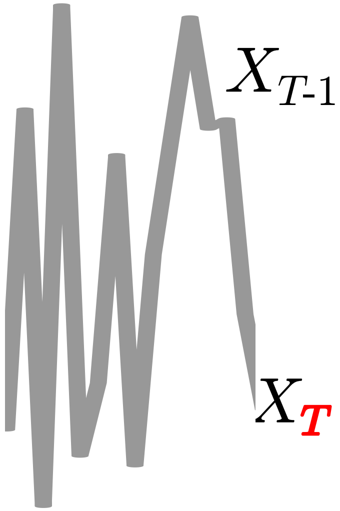
Detection Strategies

Sequential Test: At each time t use available observations $\{X_1,\ldots,X_t\}$ to decide whether a change has occurred at t or before (no future information)

Common tests: Declare change if $S_t(X_1,\ldots,X_t) \geq \nu_t$



...



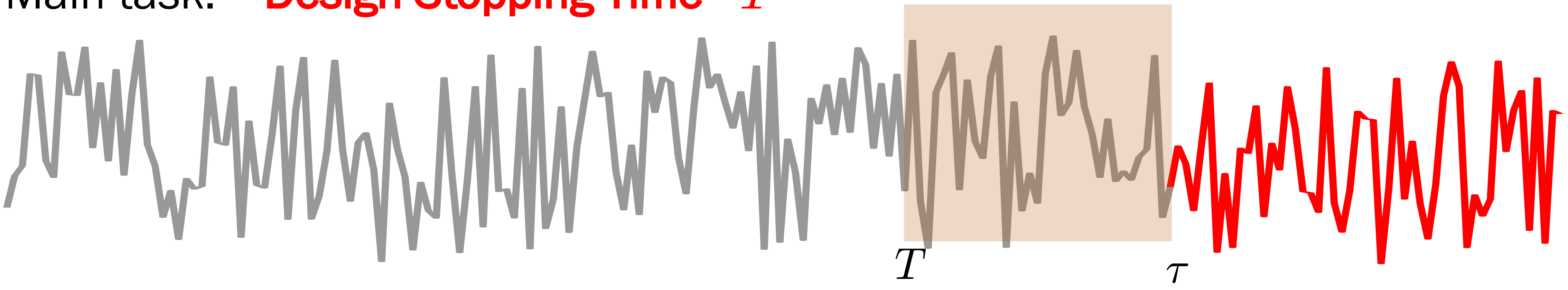
Sequential Test

equivalent to

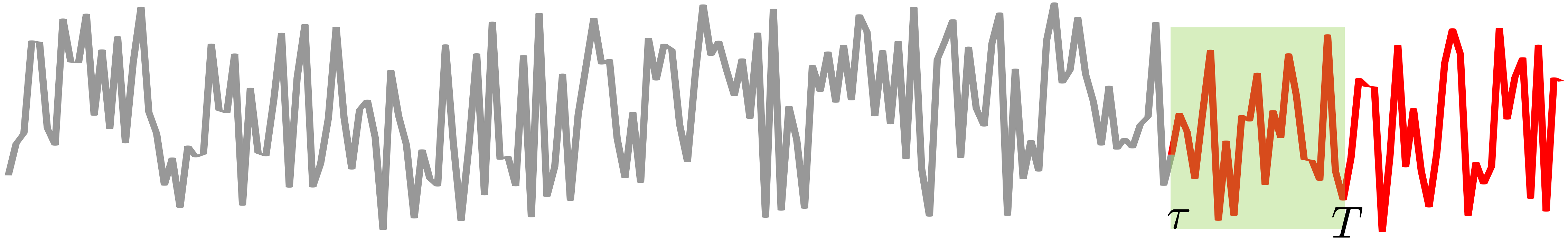
Stopping Time

1	X_1	Was there a change?	**No**,	take one more sample
2	X_1,X_2	Was there a change?	**No**,	take one more sample
	$\vdots$			
T-1	$X_1,X_2,\ldots,X_{T\text{-}1}$	Was there a change?	**No**,	take one more sample
T	$X_1,X_2,\ldots,X_T$	Was there a change?	**Yes**,	**Stop sampling** (stopping time)

Main task: **Design Stopping Time** T

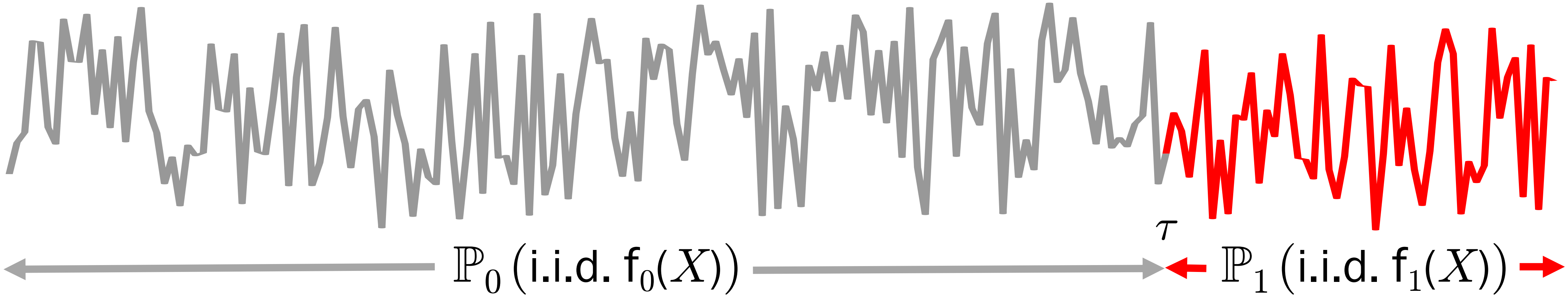


Premature Stopping $\Rightarrow$ False Alarm (infrequent)

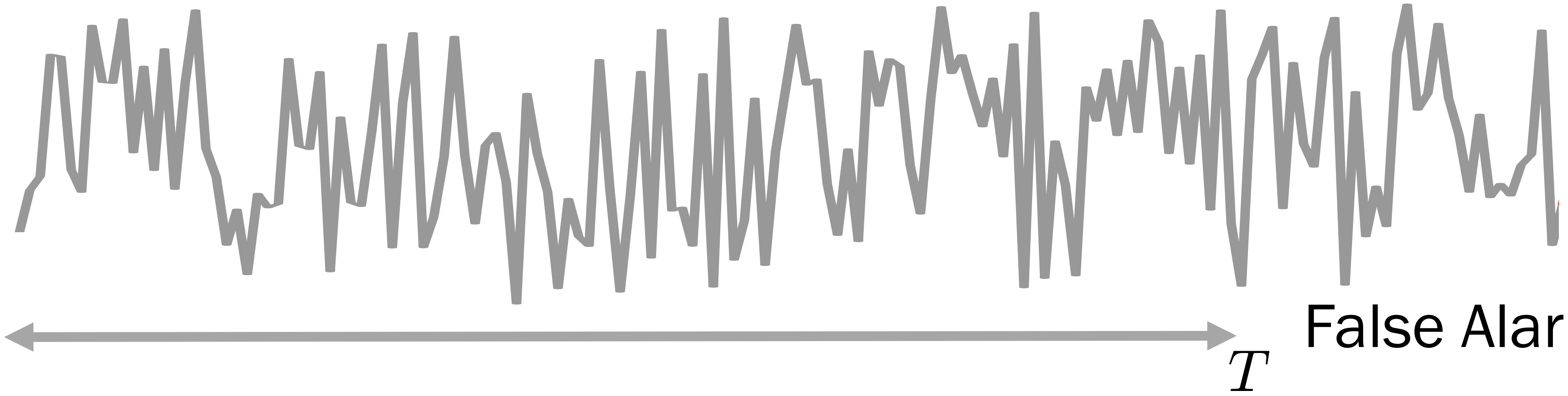


Successful Detection $\Rightarrow$ $T - \tau$ Detection Delay (short)

For the theoretical design of T need to quantify both



False Alarm

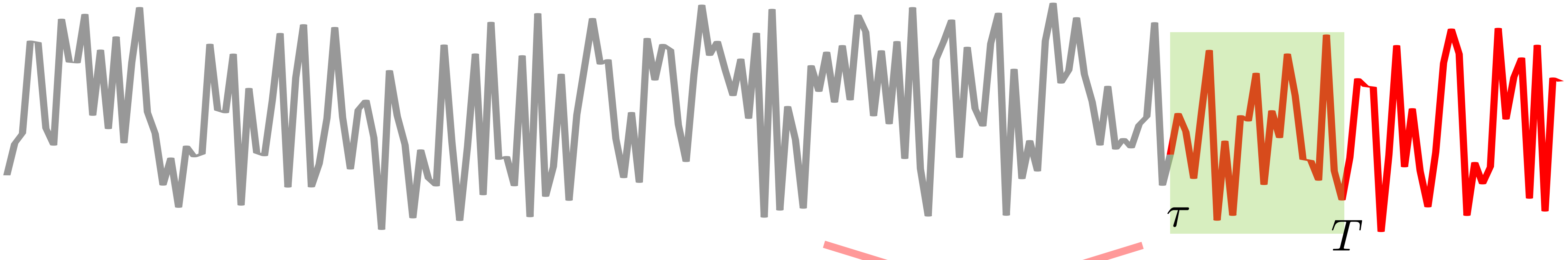


False Alarm Period T

Average False Alarm Period (large): $\mathbb{E}_0[T] \geq \gamma$

Computable

Detection Delay



Average Detection Delay: ~~$\mathbb{E}_1\big[\max\{T-\tau,0\}\big]$~~

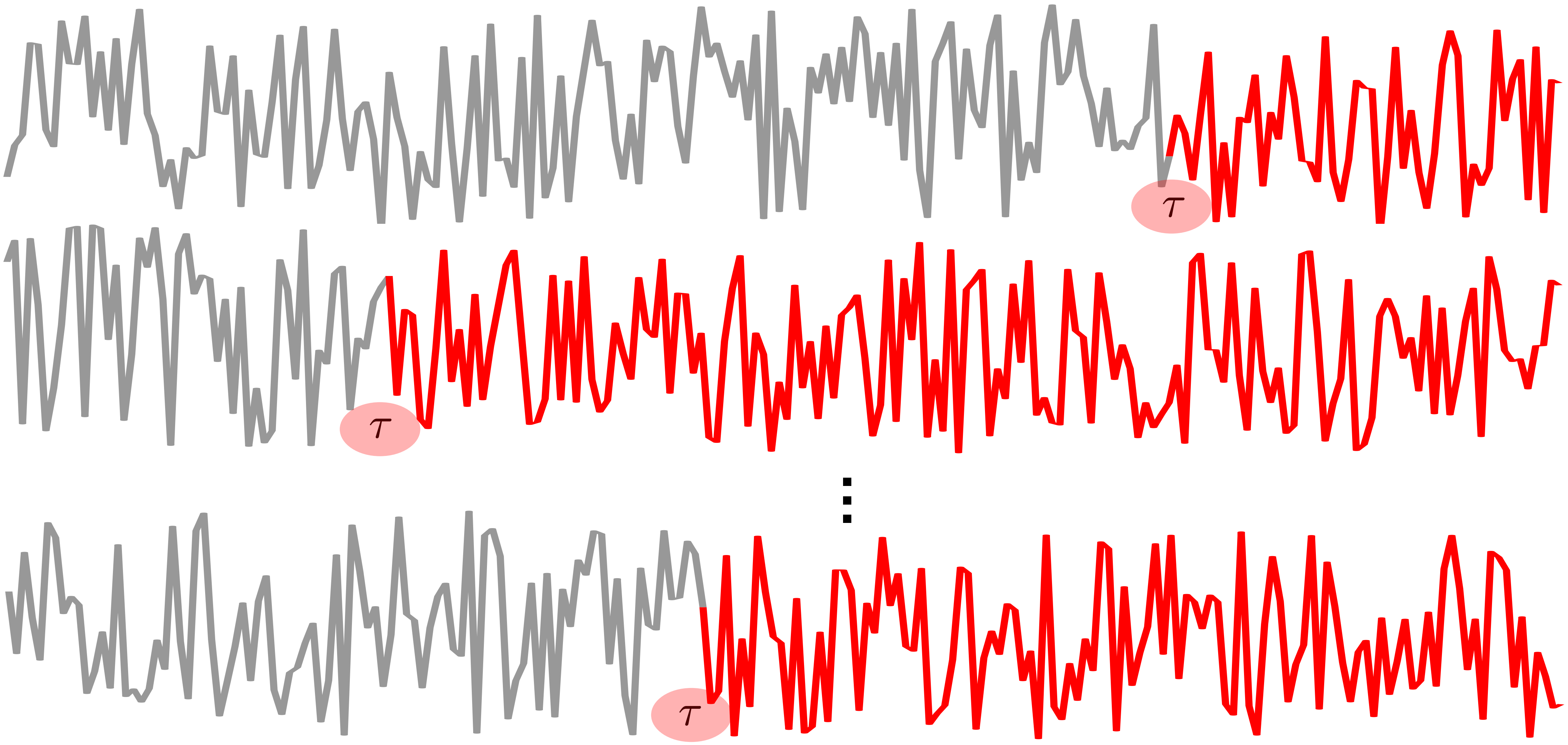
Biased due to false alarms   Extreme case: $T=0$

Conditional Average Detection Delay (small): $\mathbb{E}_1\big[T-\tau \,|\, T>\tau\big]$

Computable ?

What is change-time τ ?

Understanding Changes



Change-time τ **random !** Can this randomness be described by $\mathbb{P}_0$, $\mathbb{P}_1$?

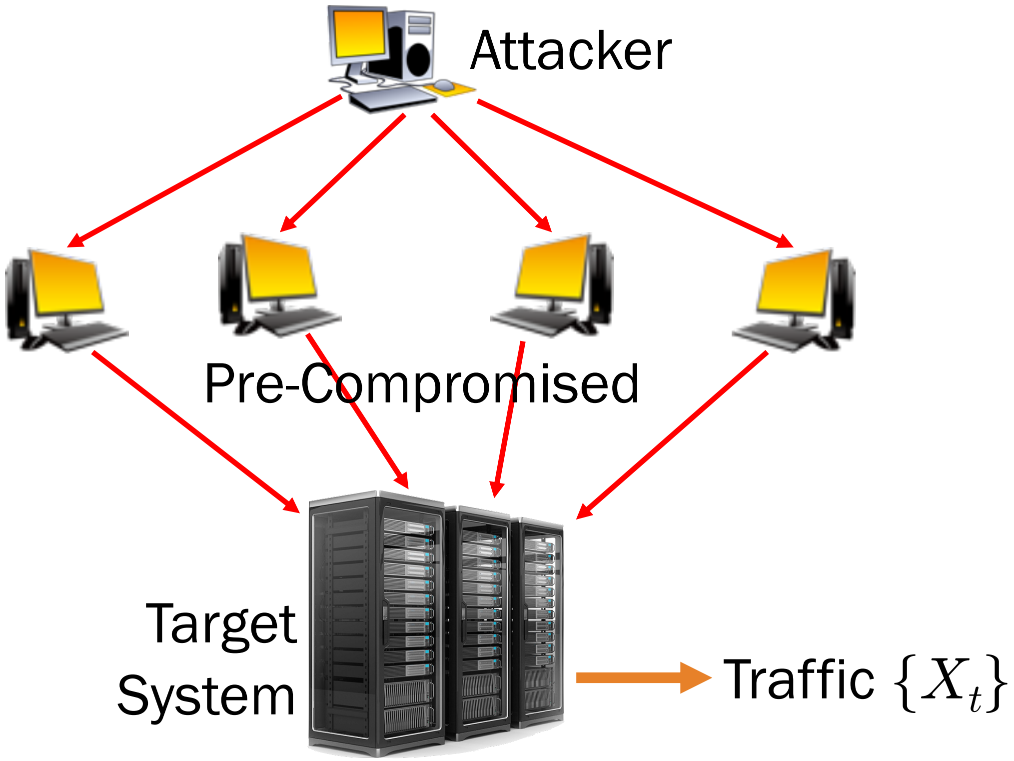
Attack

Vibration measurements $\{X_t\}$ at sensors to detect structural changes

Change imposing mechanism does not use observations $\{X_t\}$

Relies on **independent** data (coordinates of the ball in football game)!

Attack



The system uses traffic measurements $\{X_t\}$ to detect attacks

The attacker has no access to $\{X_t\}$, therefore time of attack is **independent** from observations

Earthquake

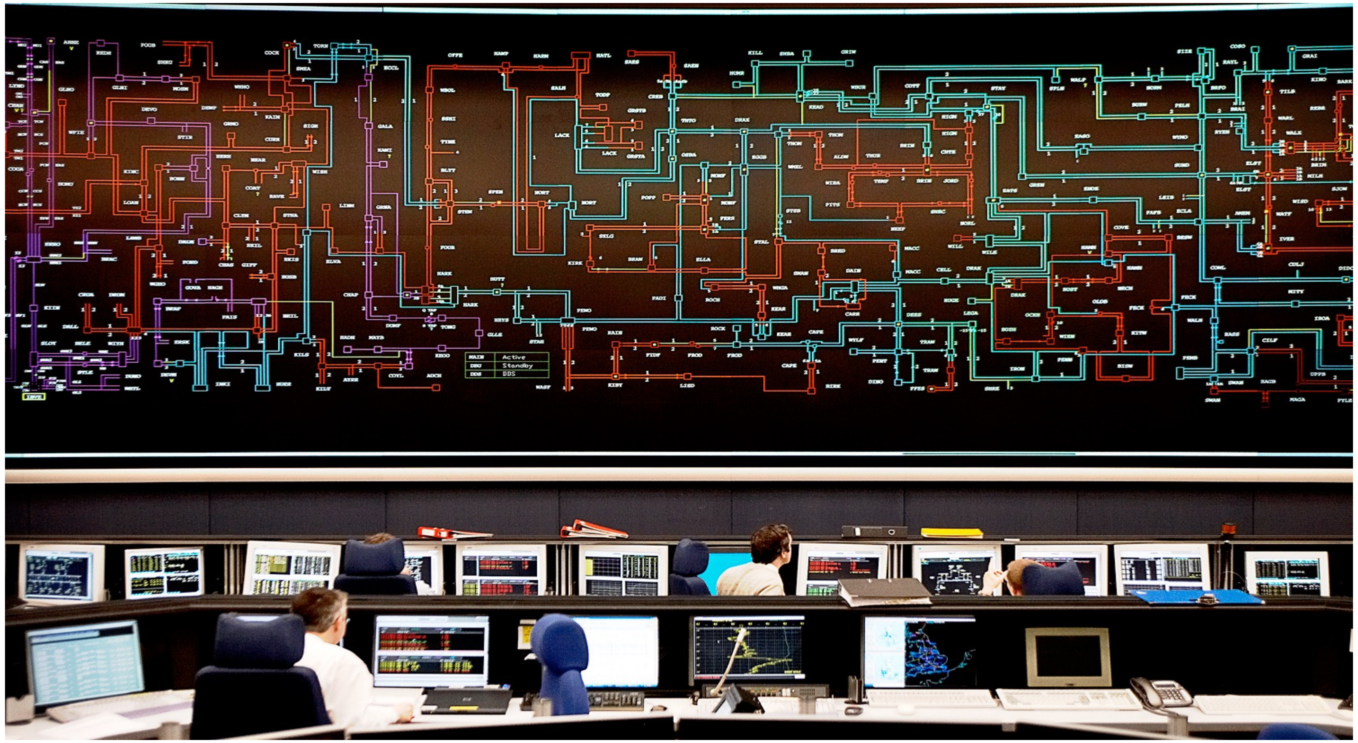
Change imposing mechanism consults $\{Z_t\}$ which **depend** on observations $\{X_t\}$

Vibration measurements $\{X_t\}$ at sensors to detect structural changes

$X_t = AZ_t + W_t, \quad Z_t$ state of the whole structure

Change (crack) when $\|Z_t\|^2 \geq \lambda$

In previous examples detection delay can be arbitrarily large



Power Grid

$\{X_t\}$ measurements at major points

Change in statistical behavior must be detected between τ and $\tau+W$

After $\tau+W$ change produces blackout!

Detection delay is **hard limited**: $W \geq T - \tau > 0$

$\{X_t\}$ Observations collected sequentially to detect the change

Change Imposing Mechanism

- Mechanism applies changes based on information independent from $\{X_t\}$
- Mechanism applies changes based on information dependent on $\{X_t\}$

Delay Constraint

- No hard limit on detection delay
- Detection delay is hard limited

Performance Criteria and Optimum Tests

Change Imposing Mechanisms Independent from Observations (no hard limit)

$$\mathbb{P}(\tau = t) = \pi_t$$ sequence of numbers

If prior $\{\pi_t\}$ known we can compute $\mathbb{E}_1\big[T-\tau \,|\, T>\tau\big]$

When $\{\pi_t\}$ unknown follow a **worst-case analysis**

Pollak Criterion

$$J(T) = \sup_{\{\pi_t\}} \mathbb{E}_1\big[T-\tau \,|\, T>\tau\big] \;=\; \sup_{t\geq 0} \mathbb{E}_1\big[T-t \,|\, T>t\big]$$

Optimize T by solving constrained optimization (1983)

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

subject to: $\mathbb{E}_0\big[T\big] \geq \gamma$

For i.i.d. observations before and after the change
Pollak proposed the Shiryaev-Roberts-Pollak (SRP) Test

Specially design initial value S_0

At each time t with new sample X_t update statistic S_t

$$S_t = (1 + S_{t-1}) \frac{\mathsf{f}_1(X_t)}{\mathsf{f}_0(X_t)}$$

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

Select ν to satisfy false alarm constraint: $\mathbb{E}_0[T_{\mathsf{SRP}}] = \gamma$

(1983) Asymptotically optimum for γ large (tending to infinity)

Strongest sense of asymptotic optimality

NOT exactly optimum (2010 counterexample)

Change Imposing Mechanisms Dependent on Observations
(no hard limit)

$$\mathbb{P}(\tau = t) = \pi_t(X_1, X_2, \ldots, X_t)$$

The probabilities depend on the realization

When $\{\pi_t(X_1, X_2, \ldots, X_t)\}$ unknown follow a **worst-case analysis**

$$J(T) = \sup_{\{\pi_t\}} \mathbb{E}_1\left[T - \tau \,|\, T > \tau\right] = \sup_{t \geq 0} \sup_{X_1, \ldots, X_t} \mathbb{E}_1\left[T - t \,|\, T > t, X_1, \ldots, X_t\right]$$

Lorden Criterion

Optimize T by solving constrained optimization (1971)

$$\inf_T J(T) = \inf_T \sup_{t \geq 0} \sup_{X_1, \ldots, X_t} \mathbb{E}_1\left[T - t | T > t, X_1, \ldots, X_t\right]$$

subject to: $\mathbb{E}_0\left[T\right] \geq \gamma$

For i.i.d. data before and after the change apply **CUSUM Test**

At each time t with new sample X_t update CUSUM statistic S_t

$$S_t = \max\{S_{t-1}, 0\} + \log\left(\frac{f_1(X_t)}{f_0(X_t)}\right), \quad S_0 = 0$$

$$T_{\text{CUSUM}} = \inf\{t > 0 : S_t \geq \nu\}$$

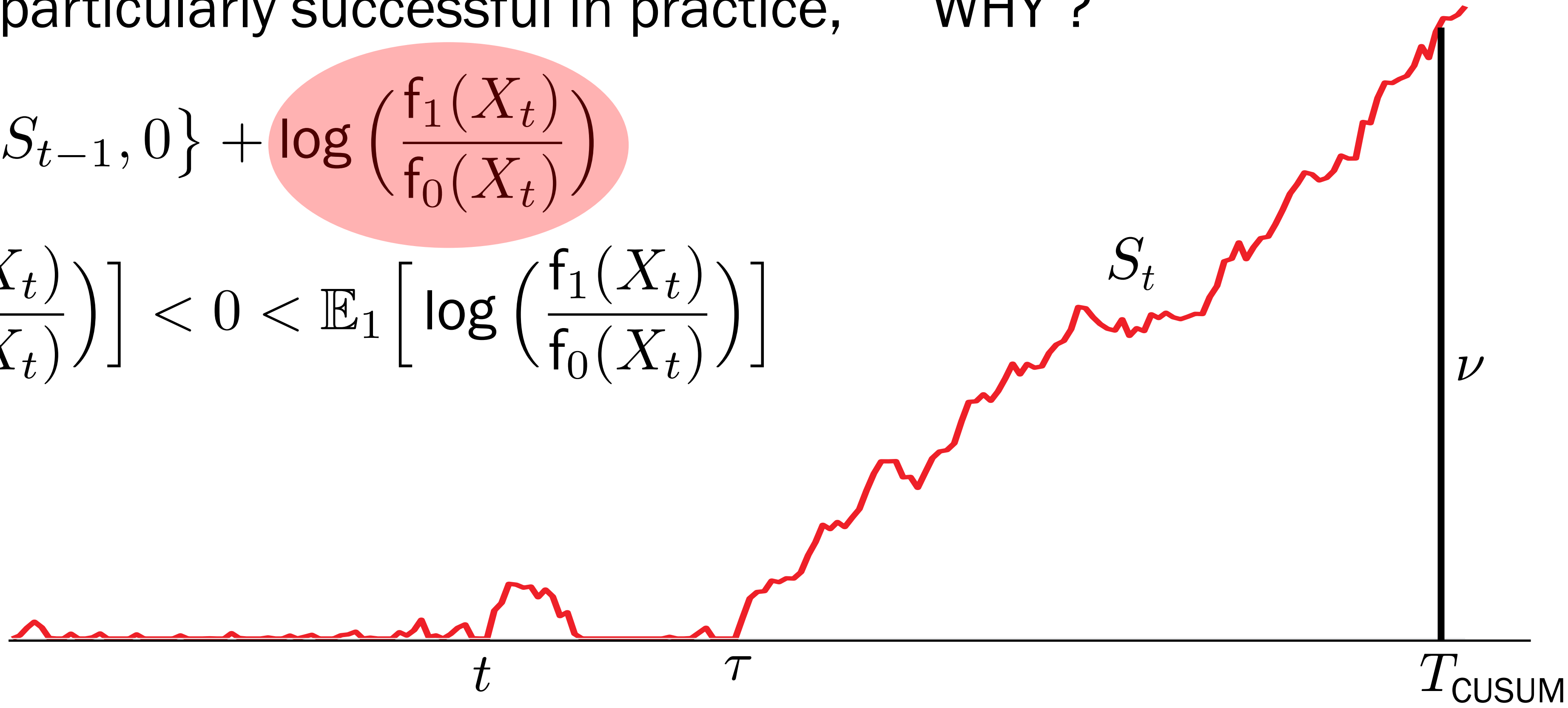
Select ν to safisfy false alarm constraint: $\mathbb{E}_0\left[T_{\text{CUSUM}}\right] = \gamma$

Exact optimality (1986)

CUSUM Test particularly successful in practice,   WHY ?

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

$$\mathbb{E}_0\left[\log\left(\frac{\mathsf{f}_1(X_t)}{\mathsf{f}_0(X_t)}\right)\right] < 0 < \mathbb{E}_1\left[\log\left(\frac{\mathsf{f}_1(X_t)}{\mathsf{f}_0(X_t)}\right)\right]$$



Prototype for other data models

$$S_t = \max\{S_{t-1},0\} + \log\left(\frac{\mathsf{f}_1(X_t|X_{t-1},\ldots)}{\mathsf{f}_0(X_t|X_{t-1},\ldots)}\right), \quad S_0 = 0$$

No exact optimality

Asymptotic optimality for false alarm values γ large (tending to infinity)

Hard limited detection delay

In certain applications necessary to detected between τ and $\tau+W$
Stopping after $\tau+W$ is **no detection** (too late)

Interested in T such that $\tau<T\leq\tau+W$

$$\mathbb{P}_1\big(\tau<T\leq\tau+W\,|\,T>\tau\big) = \mathbb{P}_1\big(T\leq\tau+W\,|\,T>\tau\big)$$

Change mechanism independent from observations

Leads to $$J(T) = \inf_{t\geq 0} \mathbb{P}_1\big(T \leq t+W \mid T>t\big)$$

Pollak-like Criterion

$$\sup_T J(T) = \sup_T \inf_{t\geq 0} \mathbb{P}_1\big(T \leq t+W|T>t\big)$$

subject to: $\mathbb{E}_0\big[T\big] \geq \gamma$

Change mechanism dependent on observations

Lorden-like Criterion

Leads to $$J(T) = \inf_{t\geq 0} \inf_{X_1,\ldots,X_t} \mathbb{P}_1\big(T \leq t+W \mid T>t, X_1,\ldots,X_t\big)$$

$$\sup_T J(T) = \sup_T \inf_{t\geq 0} \inf_{X_1,\ldots,X_t} \mathbb{P}_1\big(T \leq t+W|T>t, X_1,\ldots,X_t\big)$$

subject to: $\mathbb{E}_0\big[T\big] \geq \gamma$

Solution for arbitrary W ? **NO**

Only for $W = 1$, Immediate detection with the first post-change sample

$$\sup_T \inf_{t \geq 0} \mathbb{P}_1\big(T = t+1 \,|\, T > t\big)$$

or

$$\sup_T \inf_{t \geq 0} \inf_{X_1,\ldots,X_t} \mathbb{P}_1\big(T = t+1 \,|\, T > t, X_1, \ldots, X_t\big)$$

subject to: $\mathbb{E}_0\big[T\big] \geq \gamma$

For i.i.d. data before and after the change optimum is the **Shewhart Test**

$$T_{\mathsf{sh}} = \inf\left\{t > 0 : \frac{\mathsf{f}_1(X_t)}{\mathsf{f}_0(X_t)} \geq \nu\right\}$$

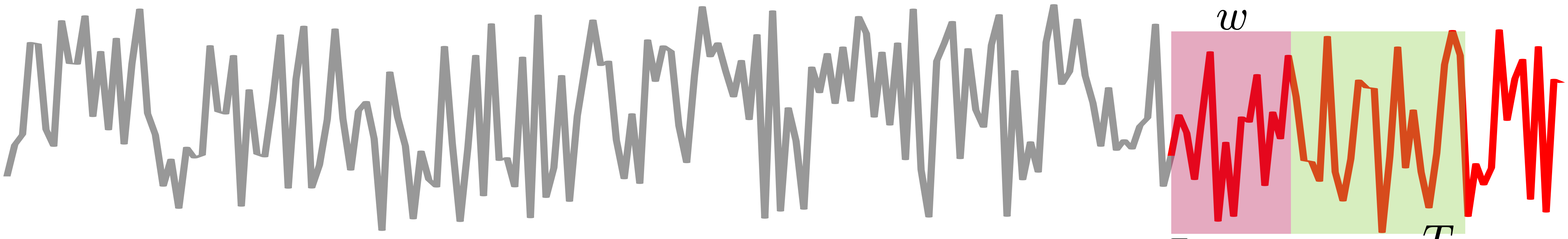
Advanced Versions

Unknown Parameters

For i.i.d. data before and after the change: Probability densities $\mathsf{f}_0(X)$, $\mathsf{f}_1(X,\theta)$

At every time t use sliding window of previous data $X_{t\text{-}1},\ldots,X_{t\text{-}w}$ to estimate θ

$$S_t = \max\{S_{t-1},0\} + \log\left(\frac{\mathsf{f}_1(X_t,\hat{\theta}_{t-1})}{\mathsf{f}_0(X_t)}\right) \qquad T_{\text{WCUSUM}} = \inf\{t>0:\ S_t \geq \nu\}$$



$$\mathbb{E}_0\left[T_{\text{CUSUM}}\right] = \gamma \to \infty, \quad \mathbb{E}_1\left[T_{\text{CUSUM}}\right] \sim \log\gamma \to \infty$$

$$\text{Optimum} \quad w \sim \sqrt{\log\gamma}, \quad \mathbb{E}_1\left[T_{\text{WCUSUM}}\right] = \mathbb{E}_1\left[T_{\text{CUSUM}}\right] + O\left(\sqrt{\log\gamma}\right)$$

Data-Driven Version

Densities $f_0(X_t | X_{t\text{-}1},\ldots)$, $f_1(X_t | X_{t\text{-}1},\ldots)$ are completely unknown. Instead:

Training Data

$$\{X_1^0,\ldots,X_{n_0}^0\} \quad \text{sampled from} \quad f_0(X_t|X_{t-1},\ldots)$$

$$\{X_1^1,\ldots,X_{n_1}^1\} \quad \text{sampled from} \quad f_1(X_t|X_{t-1},\ldots)$$

DO NOT estimate individual densities $f_0(X_t | X_{t\text{-}1},\ldots)$, $f_1(X_t | X_{t\text{-}1},\ldots)$

Use Machine Learning techniques (neural networks) to estimate directly

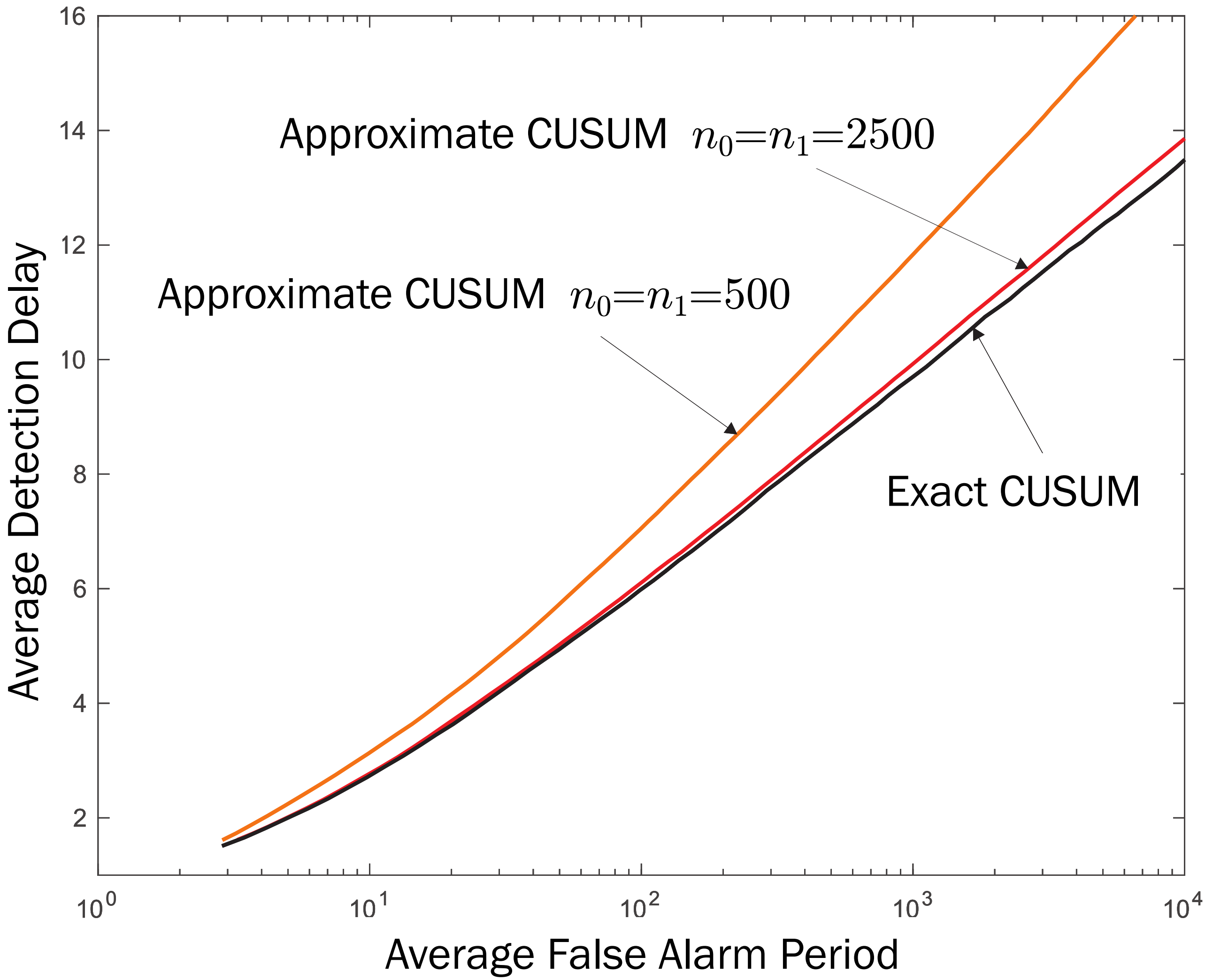
$$u(X_t, X_{t-1},\ldots) \approx \log\left(\frac{f_1(X_t|X_{t-1},\ldots)}{f_0(X_t|X_{t-1},\ldots)}\right)$$

Approximate CUSUM statistic: $S_t = \max\{S_{t-1}, 0\} + u(X_t, X_{t-1},\ldots)$

Before and after change: Markovian process of unit memory

$f_0(X_t | X_{t-1})$

$f_1(X_t | X_{t-1})$



No claim of optimality of any type

Simulations suggest asymptotic optimality if n_0, n_1 suitable functions of γ