Sequential techniques

for

Hypothesis testing & Change detection

George V. Moustakides
University of Patras

Outline

- Sequential hypothesis testing
- The Sequential Probability Ratio Test (SPRT) for optimum hypothesis testing
- Intrusion detection in wireless networks
- Sequential change detection
- Performance criteria and optimum detection rules
- Lorden's criterion and the CUSUM test
- Decentralized detection of changes

Sequential hypothesis testing

Conventional binary hypothesis testing (fixed sample size): Collection of observations $\xi_1,\ldots,\xi_K$

$$\mathbb{H}_0\colon \xi_1,\ldots,\xi_K \sim f_0(\xi_1,\ldots,\xi_K);$$
$$\mathbb{H}_1\colon \xi_1,\ldots,\xi_K \sim f_1(\xi_1,\ldots,\xi_K);$$

Decision rule $D(\xi_1,\ldots,\xi_K)\in\{0,1\}$

$\mathbb{P}(D=1 \mid \mathbb{H}_1)$ (Correct decision)

$\mathbb{P}(D=1 \mid \mathbb{H}_0)$ (Type I error)

$\mathbb{P}(D=0 \mid \mathbb{H}_1)$ (Type II error)

$\mathbb{P}(D=0 \mid \mathbb{H}_0)$ (Correct decision)

Bayes formulation

Pr. Err.(D) = $\mathbb{P}(\mathbb{H}_0)\mathbb{P}(D{=}1|\mathbb{H}_0)+\mathbb{P}(\mathbb{H}_1)\mathbb{P}(D{=}0|\mathbb{H}_1)$

$$\min_D \text{Pr. Err.}(D)$$

Neyman-Pearson formulation

$\max_D \mathbb{P}(D{=}1|\mathbb{H}_1)$ subject to $\mathbb{P}(D{=}1|\mathbb{H}_0) \leqslant \alpha$

Likelihood ratio test:

$$\frac{f_1(\xi_1,\ldots,\xi_K)}{f_0(\xi_1,\ldots,\xi_K)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \gamma$$

For i.i.d.:

$$u_K = \sum_{n=1}^{K} \log\left(\frac{f_1(\xi_n)}{f_0(\xi_n)}\right) \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtrless}} \log(\gamma) = \gamma'$$

Sequential binary hypothesis testing

Observations $\xi_1,\ldots,\xi_t,\ldots$ are supplied **sequentially**.

$$\mathbb{H}_0\colon \xi_1,\ldots,\xi_t,\ldots \sim f_0(\xi_1,\ldots,\xi_t,\ldots)$$
$$\mathbb{H}_1\colon \xi_1,\ldots,\xi_t,\ldots \sim f_1(\xi_1,\ldots,\xi_t,\ldots)$$

Time	Observations	
1	ξ_1	$D(\xi_1)$
2	ξ_1,ξ_2	$D(\xi_1,\xi_2)$
...	...	...
t	$\xi_1,\ldots,\xi_t$	$D(\xi_1,\ldots,\xi_t)$
...	...	...

Decide reliably **as soon as possible**.

We apply a two-rule scheme:

1st Rule

Time	Observations
1	ξ_1
2	ξ_1,ξ_2
...	...
T	$\xi_1,\ldots,\xi_T$

Can ξ_1 make a reliable decision?

$T(\xi_1,\ldots,\xi_t) = \{\text{stop,continue}\}$

Time T is **RANDOM**

We **stop** receiving observations

2nd Rule

Decision Rule

$D(\xi_1,\ldots,\xi_T) \in \{0,1\}$

WHY sequential?

For the same level of confidence with a sequential test we need, in the average, **(significantly) less samples** than a fixed sample size test, to reach a decision.

The Sequential Probability Ratio Test (SPRT) (Wald 1947)

Changes with time

$$u_t = \log\left(\frac{f_1(\xi_1,\ldots,\xi_t)}{f_0(\xi_1,\ldots,\xi_t)}\right)$$

For i.i.d.

$$u_t = u_{t-1} + \log\left(\frac{f_1(\xi_t)}{f_0(\xi_t)}\right)$$

We define **two** thresholds $A < 0 < B$



Stopping rule: $T = \inf_t \{t : u_t \notin (A, B)\}$

Decision rule: $D(\xi_1, \ldots, \xi_T) = \begin{cases} 1 & \text{if } u_T \geq B \\ 0 & \text{if } u_T \leq A \end{cases}$

Remarkable optimality property of SPRT

$$\min_{T,D} \mathbb{E}[T|\mathbb{H}_0] \quad \textbf{AND} \quad \min_{T,D} \mathbb{E}[T|\mathbb{H}_1]$$

subject to

$$\mathbb{P}[D_T = 1|\mathbb{H}_0] \leq \alpha; \quad \mathbb{P}[D_T = 0|\mathbb{H}_1] \leq \beta$$

- Optimum for i.i.d. observations (Wald and Wolfowitz, 1948)
- Brownian Motion with constant drift (Shiryayev, 1967)
- Homogeneous Poisson (Peskir, Shiryayev, 2000)
- **Open problems:** Dependent observations, multiple hypothesis testing

Misbehavior detection in wireless networks

(with Radosavac and Baras)

MAC Layer: When the channel is not in use, nodes wait a random (back-off) time and then reserve the channel.

- The node with the smaller back-off time reserves the channel first.
- Back-off times of legitimate nodes are distributed according to the **known** uniform distribution $f_0 = U[0, W]$.
- Attacker's goal is to reserve the channel more often than the legitimate users. Back-off distribution $f_1 = ?$ is **unknown**.

Use back-off time measurements to detect attacker!

For each node we measure **back-off times** (observations) sequentially and we decide whether it is legitimate $(\mathbb{H}_0)$ or attacker $(\mathbb{H}_1)$.

Candidate test: **SPRT**

Not directly applicable, since we don't know f_1

Quantification of an "attack"

N legitimate nodes have probability $1/N$ of reserving the channel.

A node is characterized as "attacker" if its probability of reserving the channel is **at least** η/N, where $\eta>1$.

Example: $\eta = 1.1$ means that a node "attacks" if it reserves the channel 10% more than a legitimate node.

Probability of reserving the channel $\geqslant \eta/N$

$$\int_0^W x f_1(x) dx \leq \epsilon \frac{W}{2}$$

Defines a **CLASS** $\mathcal{F}$ of possible attack densities

where $\epsilon < 1$ a quantity that depends on η.

Optimization problem

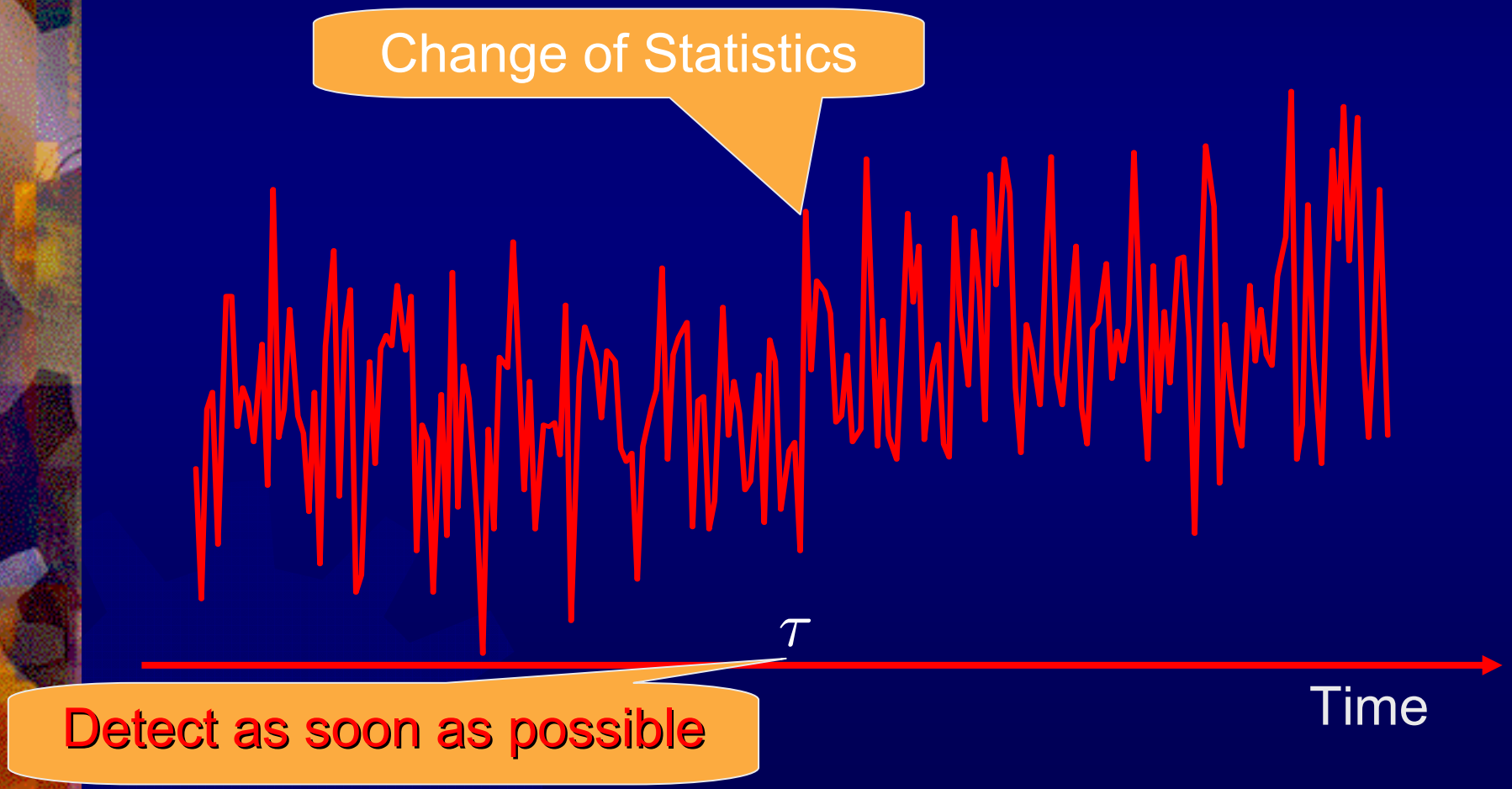
$$\min_{T,D} \max_{f_1 \in \mathcal{F}} \mathbb{E}[T|\mathbb{H}_1]$$ subject to

$$\mathbb{P}[D_T = 1|\mathbb{H}_0] \leq \alpha; \quad \max_{f_1 \in \mathcal{F}} \mathbb{P}[D_T = 0|\mathbb{H}_1] \leq \beta$$

SPRT with $$f_1^\star(x) = \begin{cases} Ce^{-\mu x} & 0 \leq x \leq W \\ 0 & x > W \end{cases}$$

The Sequential change detection problem

Also known as the **Disorder problem** or the **Change-Point problem** or the **Quickest Detection problem.**



Applications

- Monitoring of quality of manufacturing process (1930's)
- Biomedical Engineering
- Electronic Communications
- Econometrics
- Seismology
- Speech & Image Processing
- Vibration monitoring
- Security monitoring (fraud detection)
- Spectrum monitoring
- Scene monitoring
- Network monitoring and diagnostics (router failures, intruder detection)
- Databases .....

Mathematical setup

We are observing sequentially a process $\{\xi_t\}$ with the following statistics:

$$\xi_t \quad \sim f_0 \quad \text{for} \quad 0 < t \leqslant \tau$$
$$\sim f_1 \quad \text{for} \quad \tau < t$$

Goal: Detect the change time τ **"as soon as possible"**

- Change time τ : deterministic (but unknown) or random
- Densities f_0 , f_1 : known

The observation process $\{\xi_t\}$ is available sequentially.

Interested in **sequential detection schemes**.

- At every time instant t we perform a test to decide whether to stop (and issue an alarm) or continue sampling.
- The test at time t must be based on the **available information up to time t** (and not any future information).

Any sequential detection scheme is nothing but a **stopping rule T** that decides when to stop.

Overview of existing results

Optimality criteria

They must take into account two quantities:

- Detection delay T - τ
- Frequency of false alarms

Possible approaches: Bayesian and Min-max

Bayesian approach (Shiryayev 1978)

The change time τ is random with geometric prior.

$$\mathrm{Pro}[\tau = t] = (1\text{-}\varpi)\varpi^t$$

For any stopping rule T define the criterion:

$$J(T) = c\mathbb{E}[\ (T\text{ - }\tau)^+\ ] + \mathbb{P}[\ T \leqslant \tau\ ]$$

Optimization problem: $\inf_T J(T)$

Define the statistics: $\pi_t = \mathbb{P}[\ \tau \leqslant t \mid \xi_1,\ldots,\xi_t]$

Stopping rule: $T_S = \inf_t \{\ t\colon \pi_t \geqslant \nu\ \}$

- Discrete time: when $\{\xi_t\}$ is i.i.d. and there is a change in the pdf from $f_0(\xi)$ to $f_1(\xi)$.

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

- Continuous time: when $\{\xi_t\}$ is a Brownian Motion and there is a change in the constant drift; or a Poisson process and there is a change in the constant rate.

Stochastic differential equation

Min-max approach (Pollak, 1985)

The change time τ is deterministic but unknown.

For any stopping rule T define the criterion:

$$J(T) = \sup_{\tau} \mathbb{E}_1[\ (T - \tau)^+ \mid T > \tau\ ]$$

Optimization problem: $\inf_T J(T)$;

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

Discrete time: when $\{\xi_t\}$ is i.i.d. and there is a change in the pdf from $f_0(\xi)$ to $f_1(\xi)$.

Compute the statistics: $S_t = (S_{t-1} + 1)\ \dfrac{f_1(\xi_t)}{f_0(\xi_t)}$.

Stopping rule: $T_{\mathrm{P}} = \inf_t \{\ t\colon S_t \geqslant \nu\ \}$ Mei (2006)

CUSUM test and Lorden's criterion

Page (1954) introduced the CUmulative SUM (CUSUM) test for i.i.d. observations.

Suppose we are given $\xi_1,...,\xi_t$. Form a likelihood ratio test for the following two hypotheses:

$\mathbb{H}_0$: All observations are under the nominal regime

$\mathbb{H}_1$: There is a change at $\tau < t$

Assume τ **unknown**

$$\max_{0\le\tau\le t} \sum_{n=\tau+1}^{t} \log\left(\frac{f_1(\xi_n)}{f_0(\xi_n)}\right) \gtrless \nu$$

$$\sum_{n=1}^{t} \log\left(\frac{f_1(\xi_n)}{f_0(\xi_n)}\right) - \min_{0\le\tau\le t} \sum_{n=1}^{\tau} \log\left(\frac{f_1(\xi_n)}{f_0(\xi_n)}\right) \gtrless \nu$$

Define the CUSUM process y_t as follows:

$$y_t = u_t - m_t$$

where

$$u_t = \log\left(\frac{f_1(\xi_1,\ldots,\xi_t)}{f_0(\xi_1,\ldots,\xi_t)}\right)$$

$$m_t = \inf_{0\leqslant s\leqslant t} u_s\ .$$

The CUSUM stopping rule:

$$\mathcal{S} = \inf_t \{\ t\colon y_t \geqslant \nu\ \}$$

For the i.i.d. case we have a convenient recursion:

$$y_t = \left(y_{t-1} + \log\left(\frac{f_1(\xi_t)}{f_0(\xi_t)}\right)\right)^+$$


u_t
m_t
S
ML estimate of τ

Min-max criterion (Lorden, 1971):

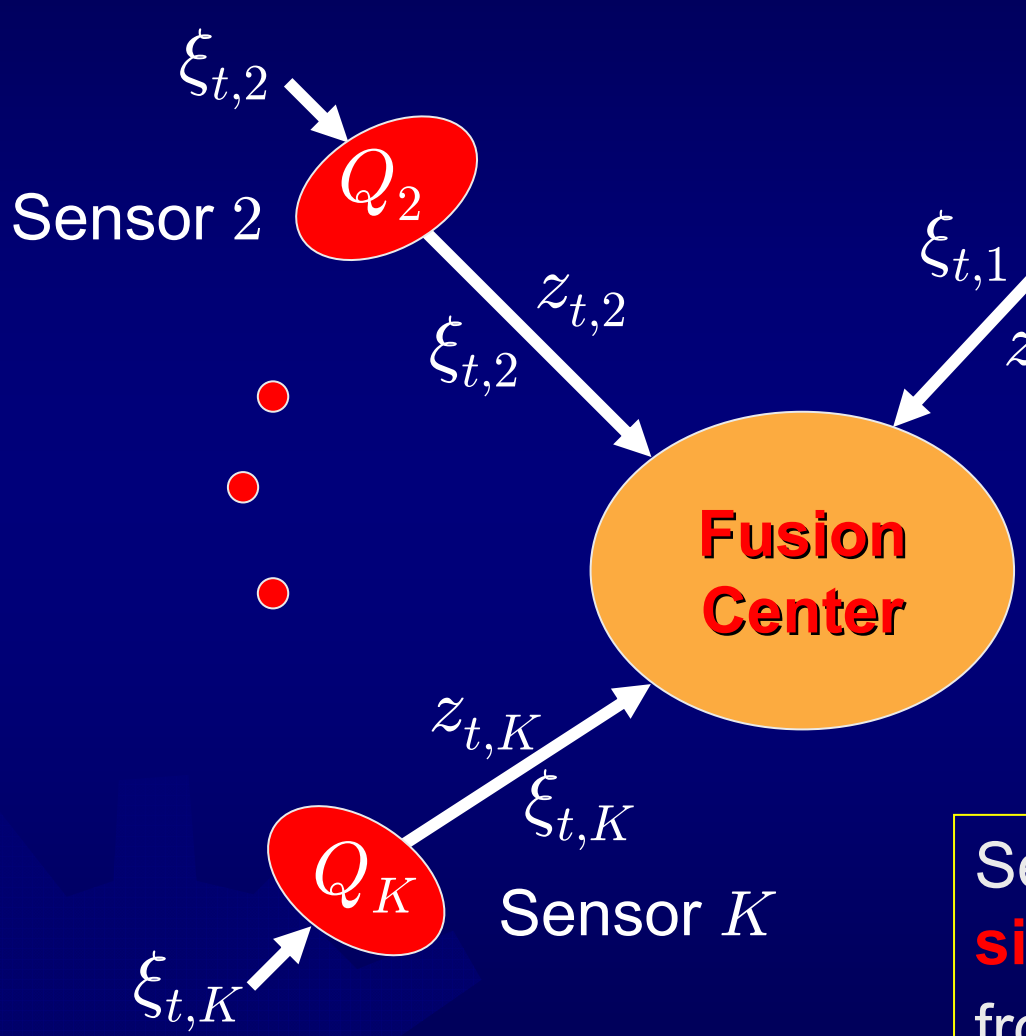
$$J(T) = \sup_{\tau} \sup_{\xi_1,\ldots,\xi_\tau} \mathbb{E}_1[(T-\tau)^+ | \xi_1, \ldots, \xi_\tau]$$

Optimization problem: $\inf_T J(T)$;

subject to: $\mathbb{E}_0[\, T \,] \geqslant \gamma$.

- For i.i.d. observations (Lorden, 1971), asymptotic optimality.
- For i.i.d. observations (Moustakides 1986 and Ritov 1990).
- Brownian Motion with constant drift (Shiryayev 1996, Beibel 1996). Ito processes (Moustakides 2004)
- Homogeneous Poisson (Moustakides, >2009)
- Change-time models and performance criteria (Moustakides 2008)
- **Open problems:** Dependency, multiple change possibilities

Decentralized change detection



Centralized Test

High communication load

Decentralized Test

Quantization scheme

Sequential detection of **simultaneous** change from $f_{0,i}$ to $f_{1,i}$.

$$J(Q,T) = \sup_{\tau} \sup_{\xi_1,\ldots,\xi_\tau} \mathbb{E}_1[(T-\tau)^+ | \xi_1,\ldots,\xi_\tau]$$

Optimization problem: $\inf_{Q,T} J(Q,T)$;

subject to: $\mathbb{E}_0[\,T\,] \geqslant \gamma\,.$

For **given** $Q=\{Q_1,\ldots,Q_K\}$ the optimum T is the CUSUM rule $\mathcal{S}_Q$ that uses sequentially the sequence of quantized observation vectors $Z_t=[z_{t,1},\ldots,z_{t,K}]$. This leads to

$$J(Q,T) \geqslant J(Q,\mathcal{S}_Q)=J(Q)$$

Minimization over quantization: Based on methodology developed by Tsitsiklis (1993), Mei (2006) proved that

$$z_i = Q_i(\xi) = \begin{cases} 1 & \text{if } f_{1,i}(\xi)/f_{0,i}(\xi) \geq \lambda_i \\ 0 & \text{otherwise} \end{cases}$$

Using these quantizers, performance measure becomes a function of the quantization thresholds

$$J(\lambda_1,\ldots,\lambda_K)$$

which must be minimized over the λ_i.

For any **given** combination of thresholds we have an **integral equation** satisfied by $J(\lambda_1,\ldots,\lambda_K)$. Therefore, this function is well defined.

The optimum thresholds are obtained by performing the last minimization



School of Electrical Engineering



Electrical Engineering

People & Organization »
Research »
Education »
News
Centers
Publications
Seminars
Diversity
Open positions
Intranet

Seminar on Sequential Techniques by George Moustakides

» 2007-10-08
Sequential techniques for hypothesis testing and change detection - seminar by Professor George M. Moustakides. kl. 13:15, Thursday 11th Oktober at Seminar room, plan 8

We are going to make an overview of the techniques developed over the last 60 years for solving two of the most important problems in sequential analysis, namely sequential hypothesis testing and sequential detection of changes. We are going to focus on the Sequential Probability Ratio Test (SPRT) which solves the first problem and the Cumulative Sum (CUSUM) test used for the second. Specifically we are going to discuss existing optimality results for the two tests, mathematical formulas for estimating the corresponding performance and finally several interesting questions that still remain open. In the last part of our talk we are going to introduce the results developed for distributed hypothesis testing and change detection. Both distributed versions have recently attracted significant attention because they find application in sensor and ad-hoc networks.

Back

Published by: EE School of Electrical Engineering
Webmaster, webmaster@ee.kth.se
Last updated: Tuesday, 07-Nov-2006 10:23:31 CET