Annual DHS University Network Summit 2008

Asymptotic Theory of Sequential Change Detection and Identification

Kazutoshi Yamazaki

Joint work with Savas Dayanik and Warren Powell
Department of Operations Research and Financial Engineering, Princeton University

,

1 Motivation

We study the joint problem of detection and identification of a sudden and unobservable change in the statistical pattern of an information sequence to one of several alternative patterns. This problem has applications in various fields including sensor management, bio-surveillance, threat detection and identification. We propose a practical online change detection and identification strategy that is asymptotically optimal under the fixed-error and the Bayesian formulations.

Suppose we observe an information sequence, and it is crucial to detect a change in the threat level as early as possible and identify its cause as accurately as possible. The inferred threat is initially low, but at some unknown and unobservable time, the threat level jumps suddenly to one of pre-described categories. To maintain the public safety, it is important to detect this change as early as possible, and at the same time, assess the nature of the change as accurately as possible in order to take the most appropriate countermeasures. This boils down to solving optimally the tradeoff between the detection delay cost and the false alarm and misdiagnosis costs to the society.

2 Problem description

A system is subject to be failed at some unobservable disorder time θ by some unknown cause μ.

Cause μ can be any one of $\mathcal{M} \triangleq \{1, 2, ..., M\}$ with $\mathbb{P}\{\mu = i\} = \nu_i$.

Disorder time θ is zero-modified geometric distributed;

$$\mathbb{P}\{\theta = 0\} = p_0 \quad \mathbb{P}\{\theta = t | \theta > 0\} = (1-p)^{t-1}p.$$

We want to

– detect the disorder time θ as early as possible, and

– identify its cause μ as accurately as possible,

based on a sequence of observations $X_1, X_2, ...$, such that

$$\underbrace{X_1,\ X_2,\ \cdots,\ X_{\theta-1}}_{\text{w/ density } f_0(\cdot)},\ \underbrace{X_\theta,\ X_{\theta+1},\ \cdots}_{\text{w/ density } f_\mu(\cdot)}$$

conditionally independent given θ and μ.

We want a decision rule (τ, d) that declares at time τ that the disorder has been triggered by cause d so as to optimize the tradeoff among

– the detection delay loss, $\mathbb{E}[(\tau-\theta)^+]$ or $\mathbb{E}[\tau - \theta | \tau \geq \theta]$,

– the terminal decision loss:

- the false alarm cost, $\mathbb{P}\{\text{false alarm under } (\tau, d)\}$,
- the misdiagnosis cost, $\mathbb{P}\{\text{misidentification under } (\tau, d)\}$.

We consider two formulations:

– Fixed-error formulation pursues a decision rule (τ, d) that minimizes the expected delay, $\mathbb{E}[\tau - \theta | \tau \geq \theta]$, subject to bounds on false alarm and misdiagnosis probabilities,

$$R_{ji}(\tau, d) \leq r_{ji},\ i \in \mathcal{M}, j \in \mathcal{M} \cup \{0\} \backslash \{i\}$$

where

$$R_{ji}(\tau, d) \triangleq \begin{cases} \mathbb{P}\{d = i, \tau < \theta\}, & j = 0, \\ \mathbb{P}\{d = i, \mu = j, \theta \leq \tau < \infty\}, & j \in \mathcal{M} \backslash \{i\}. \end{cases}$$

– Bayes formulation pursues a decision rule (τ, d) that minimizes a Bayes risk, a linear combination of the losses described above,

$$R^{(c,a)}(\tau, d) \triangleq c\mathbb{E}\left[(\tau-\theta)^+\right] + \kappa^{(a)}(\tau, d),$$

where

$$\kappa^{(a)}(\tau, d) \triangleq \sum_{i \in \mathcal{M}} a_{0i}\mathbb{P}\{d = i, \tau < \theta\} + \sum_{i \in \mathcal{M}} \sum_{j \in \mathcal{M} \backslash \{i\}} a_{ji}\mathbb{P}\{\mu = j, d = i, \theta \leq \tau < \infty\}$$

is the terminal decision loss of the Bayes risk associated with every decision rule (τ, d), for given constants c and a.

3 Need for simple decision rules

- Obtaining an optimal decision rule is computationally infeasible in general.
- Dayanik et al. [2] showed that the Bayesian formulation can be reduced to an optimal stopping problem. However, unless the number of causes M is small, the problem cannot be solved due to the so-called "curse of dimensionality".
- We therefore need a simple decision rule.

3.1 Simple decision rule

Let

$\Pi_n^{(0)}$: posterior probability that the change has not happened, and

$\Pi_n^{(i)}$: posterior probability that the change has been triggered by cause i,

after observing $X_1 \ldots X_n$; namely

$$\Pi_n^{(0)} \triangleq \mathbb{P}\{\theta > n | \mathcal{F}_n\} \quad \text{and} \quad \Pi_n^{(i)} \triangleq \mathbb{P}\{\theta \leq n, \mu = i | \mathcal{F}_n\}, \quad i \in \mathcal{M} \text{ and } n \geq 0,$$

where $(\mathcal{F}_n)_{n \geq 0}$ is the filtration generated by $(X_n)_{n \geq 0}$.

We define decision rule (τ_A, d_A) as the following

$$\tau_A \triangleq \inf\left\{n \geq 1 : \Pi_n^{(i)} > \frac{1}{1+A_i} \text{ for some } i \in \mathcal{M}\right\} \quad \text{and} \quad d_A \in \arg\max_{i \in \mathcal{M}} \Pi_{\tau_A}^{(i)}.$$

- Namely, the rule stops when one of the posterior probability processes exceeds a certain threshold and declares that the disorder has been triggered by the corresponding cause.
- The implementation of the decision rule is easy: we only need to choose the values of A.
- We show that by choosing the values of A appropriately, the decision rule (τ_A, d_A) is asymptotically optimal
  – as $\|r\| \triangleq \max_{i \in \mathcal{M},\ j \in \mathcal{M} \cup \{0\} \backslash \{i\}} r_{ji} \downarrow 0$ in the fixed-error formulation, and
  – as $c \downarrow 0$ in the Bayesian formulation.

4 Results

4.1 Asymptotic of each loss as $A \downarrow 0$

In order to choose the values of A for the asymptotic optimalities, it is crucial to know how the expected delay loss and the terminal decision loss behave as $A \downarrow 0$. We can in fact obtain their approximations when A is small. Let $\mathbb{P}_i$ and $\mathbb{P}_i^{(t)}$ denote, respectively, the conditional probabilities given $\mu = i$ and given $\mu = i$ and $\theta = t$. Define

$$l_i(j) \triangleq \begin{Bmatrix} q(i,0) - \log(1-p), & j = 0 \\ \min\{q(i,j), q(i,0) - \log(1-p)\}, & j \in \mathcal{M} \backslash \{i\} \end{Bmatrix}$$

in terms of the Kullback-Leibler distance

$$q(i,j) \triangleq \int_E \log\frac{f_i(x)}{f_j(x)} f_i(x) m(dx), \quad i \neq j \in \mathcal{M} \cup \{0\}$$

between densities f_i and f_j.

The approximation of the decision delay loss:

$$\mathbb{E}_i[\tau_A] = \frac{-\log A_i}{\min_{j \in \mathcal{M} \cup \{0\} \backslash \{i\}} l(i,j)} + o(\|A\|) \quad \text{as } \|A\| \downarrow 0.$$

The asymptotics of $\kappa^{(a)}(\tau, d)$ can be obtained by nonlinear renewal theory as in Baum and Veeravalli [1].

The approximation of the terminal decision loss:

$$\kappa^{(a)}(\tau_A, d_A) = \sum_{i \in \mathcal{M}} \nu_i \kappa_i(\tau_A, d_A) = \sum_{i \in \mathcal{M}} \left(\nu_i a_{j_i^* i} \mathbb{E}\left[e^{-W_i}\right]\right) A_i + o(\|A\|).$$

given that $j_i^* \in \arg\min_{j \in \mathcal{M}} l_i(j)$ is unique for every $i \in \mathcal{M}$.

Here, W_i is a random variable such that $\mathbb{E}_i[g(W_i)] = \mathbb{E}_i^{(0)}\left[g(W_i^{(0)})\right]$ for every continuous and bounded function $g : \mathbb{R} \mapsto \mathbb{R}$ where $W_i^{(t)}$ is a random variable with the distribution

$$\mathbb{P}_i^{(t)}(W_i^{(t)} \leq w) = \frac{\int_0^w \mathbb{P}_i^{(t)}\left\{\sum_{l=t}^{T_i^{(t)}} \log\frac{f_i(X_l)}{f_{j_i^*}(X_l)} > s\right\} ds}{\mathbb{E}_i^{(t)}\left[\sum_{l=t}^{T_i^{(t)}} \log\frac{f_i(X_l)}{f_{j_i^*}(X_l)}\right]}, \quad w \in \mathbb{R},\ t \geq 0,$$

where $T_i^{(t)} \triangleq \inf\left\{n \geq t : \sum_{l=t}^n \log\frac{f_i(X_l)}{f_{j_i^*}(X_l)} > 0\right\}$.

4.2 Asymptotic Optimality

The asymptotic optimality in the fixed-error formulation can be obtained by setting $A_i(r) \triangleq \frac{r_{j_i^* i}}{\nu_i}$ if the family $(\tau_A/(-\log A_i))_{A>0}$ is uniformly integrable.

Asymptotic Optimality in the fixed-error formulation:

$$\lim_{\|r\| \downarrow 0} \frac{\inf_{(\tau,d) \in \Delta(r)} \mathbb{E}_i[\tau - \theta | \tau \geq \theta]}{\mathbb{E}_i\left[\tau_{A(r)} - \theta \middle| \tau \geq \theta\right]} = 1.$$

The asymptotic optimality in the Bayesian formulation can be obtained by setting $A_i(c) \triangleq \frac{c}{\mathbb{E}_i e^{-W_i} \min_{j \in \mathcal{M} \cup \{0\} \backslash \{i\}} l_i(j)}$.

Asymptotic Optimality in the Bayesian formulation:

$$\lim_{c \downarrow 0} \frac{\inf_{(\tau,d) \in \Delta} R_i^{(c,a)}(\tau, d)}{R_i^{(c,a)}(\tau_{A(c)}, d_{A(c)})} = 1.$$

4.3 Numerical Results (in the Bayesian formulation)

Consider the case where $\mathcal{M} = \{1, 2\}$ with $\nu_1 = \nu_2 = 0.5$, f_i is the density of the normal distribution with mean λ_i and variance 1 for every $i \in \mathcal{M} \cup \{0\}$ and θ is geometric with success probability 0.05. Figure 1 illustrates the decision rule $(\tau_{A(c)}, d_{A(c)})$ and the optimal decision rule (τ^*, d^*) with a sample trajectory of the posterior probability process $\Pi(\omega)$.

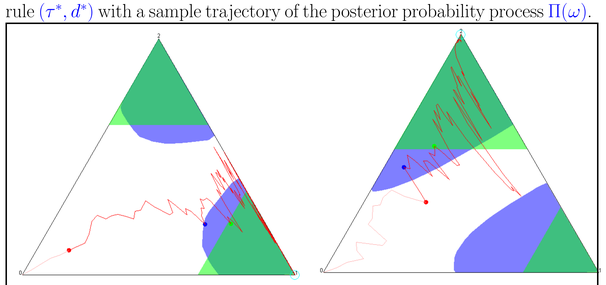

Figure 1. The blue region and the green region correspond to the optimal stopping region and the stopping region under $(\tau_{A(c)}, d_{A(c)})$, respectively, when (a) $\lambda_0 = 0.0, \lambda_1 = -0.1, \lambda_2 = 0.1$, $c = 0.01$ (b) $\lambda_0 = 0.0, \lambda_1 = -0.5, \lambda_2 = 0.1$, $c = 0.05$ along with a sample realization ω. The red point shows the disorder time $\theta(\omega)$. The blue and green points show $\tau^*(\omega)$ and $\tau_{A(c)}(\omega)$, respectively.

Table 2 shows the strong performance of $(\tau_{A(c)}, d_{A(c)})$ in comparison with the optimal decision rule (τ^*, d^*) for various values of c when $a_{ji} = 1$ when $i \neq j$.

c	$R(\tau_{A(c)}, d_{A(c)})$	$R(\tau^*, d^*)$	$R(\tau_{A(c)}, d_{A(c)})/R(\tau^*, d^*)$
0.05	0.865517	0.860915	1.005344
0.01	0.623135	0.604794	1.030325
0.005	0.490472	0.470261	1.042977
0.001	0.196816	0.194001	1.014512
0.0005	0.118715	0.118339	1.003172

Table 2. The comparison between decision rule $(\tau_{A(c)}, d_{A(c)})$ and the optimal decision rule in the Bayesian formulation.

Acknowledgment

The authors gratefully acknowledge support from the U.S. Department of Homeland Security through the Center for Dynamic Data Analysis for Homeland Security administered through ONR grant number N00014-07-1-0150 to Rutgers University.

References

[1] C. W. Baum and V. V. Veeravalli. A sequential procedure for multihypothesis testing. *IEEE Trans. Inform. Theory*, 40(6):1994–2007, 1994.

[2] S. Dayanik, C. Goulding, and V. Poor. Bayesian sequential change diagnosis. *Mathematics of Operations Research*, to appear.