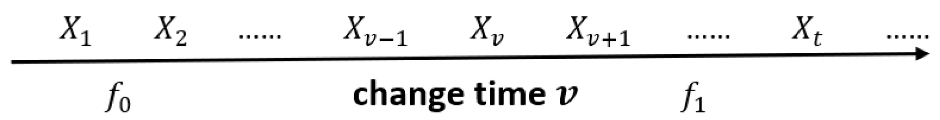
A Computationally Efficient Algorithm for Quickest Change Detection in Anonymous Heterogeneous Sensor Networks

Zhongchang Sun † Qunwei Li* Ruizhi Zhang‡ Shaofeng Zou †

†University at Buffalo *Ant Financial, China ‡ University of Nebraska-Lincoln

1. Problem Formulation

Quickest Change Detection:



- a stochastic process under observation.
- a change point ν at which the statistical property of the process undergoes a change.
- a decision maker that detects the change.
- $X_1, ..., X_t$: Observations from time 1 to t.

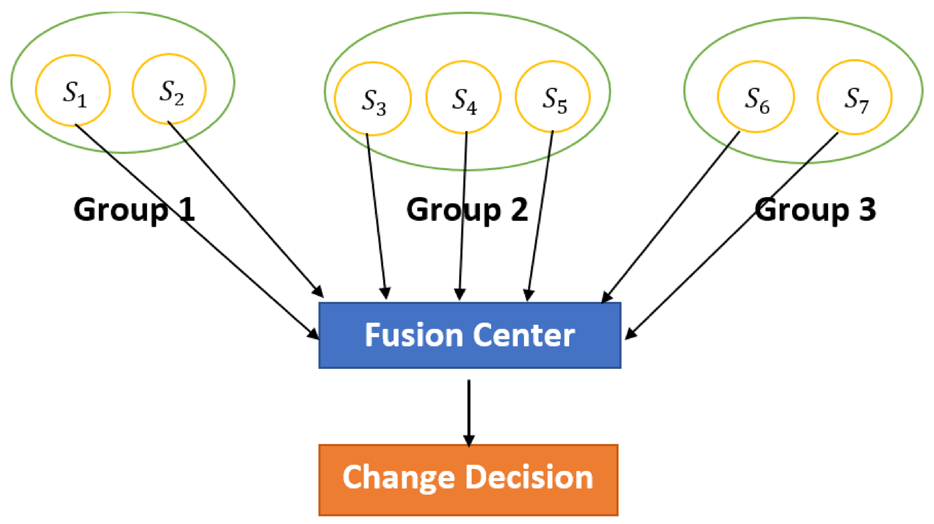
Minimax Setting: No prior knowledge of change time ν. Use worst-case average detection delay (WADD) and average running length (ARL) to evaluate a stoping rule τ.

$$\mathrm{WADD}(\tau) \triangleq \sup_{\nu\geq 1} \operatorname{ess\,sup} \mathbb{E}^{\nu}\left[(\tau-\nu)^+ | \mathbf{X}[1,\nu-1]\right],$$

$$\mathrm{ARL}(\tau) \triangleq \mathbb{E}^{\infty}[\tau],$$

Goal: Minimize WADD subject to constraint on ARL.

Anonymous Heterogeneous Networks:
A network consists of n sensors and a fusion center. Fusion center collects samples for each sensor and make decision.



- The distributions of the observations in group k are $p_{\theta,k}$, $\theta \in \{0, 1\}$.
- $\sigma(i) \in \{1, ..., K\}$: label of group that X_i comes from, are unknown to fusion center.
- $S_{n,\lambda}$: collection of all labelings.

Goal: Detect the change in anonymous heterogeneous sensor networks as quickly as possible subject to flase alarm constraint.

2. Related Work

[1] Z. Sun, S. Zou, and Q. Li, "Quickest change detection in anonymous heterogeneous sensor networks," in Proc. IEEE Int. Conf. Acoust. Speech Signal Process.(ICASSP), 2020.
[2] W. N. Chen and I. H. Wang, "Anonymous heterogeneous distributed detection: Optimal decision rules, error exponents, and the price of anonymity," IEEE Trans. Inform. Theory, 2019.

3. Mixture CuSum

Problem: Unordered samples $X^n[t]$ are observed sequentially from an anonymous heterogeneous sensor network. $X^n[t]$ follows the distribution $\mathbb{P}_{\theta,\sigma_t} \triangleq \prod_{i=1}^n p_{\theta,\sigma_t(i)}$. σ is *unknown*.
Before the change, $\theta = 0$, after the change, $\theta = 1$. Consider an unknown change point ν. Define

$$\mathrm{WADD}(\tau) \triangleq \sup_{\nu\geq 1}\sup_{\Omega} \operatorname{ess\,sup} \mathbb{E}_{\Omega}^{\nu}\left[(\tau-\nu)^+ | \mathbf{X}^n[1,\nu-1]\right], \mathrm{WARL}(\tau) \triangleq \inf_{\Omega} \mathbb{E}_{\Omega}^{\infty}[\tau],$$

where $\Omega = \{\sigma_1, \sigma_2, ..., \sigma_\infty\}$, $\mathbb{E}_{\Omega}^{\nu}$ denotes expectation when change point is ν, and group assignment is Ω.

Mixture CuSum algorithm: $T^* = \inf\left\{t : \max_{1\leq k\leq t} \sum_{i=k}^t \log \frac{\sum_{\sigma\in S_{n,\lambda}} \mathbb{P}_{1,\sigma}(X^n[i])}{\sum_{\sigma\in S_{n,\lambda}} \mathbb{P}_{0,\sigma}(X^n[i])} \geq b\right\}.$

Lemma: [1] Mixture CuSum is exactly optimal.
The computationally complexity of T^* increases exponentially with n which limits its applications in large networks. This motivates the need for computationally efficient algorithm.

4. A Computationally Efficient Algorithm

Lemma 1: [2] $\frac{\sum_{\sigma\in S_{n,\lambda}} \mathbb{P}_{1,\sigma}(X^n)}{\sum_{\sigma\in S_{n,\lambda}} \mathbb{P}_{0,\sigma}(X^n)} = \frac{\mathbb{P}_{1,\sigma}\left(T(\Pi_{X^n})\right)}{\mathbb{P}_{0,\sigma}\left(T(\Pi_{X^n})\right)}$.

- Π_{X^n} is the empirical distribution of X^n, $T(\Pi_{X^n})$ is the type class of Π_{X^n}.

Lemma 2: [2] $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{\theta,\sigma}\left(T(Q_n)\right) = -\inf_{\substack{(U_1,...,U_K)\in(\mathcal{P}_{\mathcal{X}})^K \\ \alpha^T\mathbf{U}=Q}} \sum_{k=1}^K \alpha_k D(U_k||p_{\theta,k})$.

- $\mathcal{P}_n$ is the set of types with denominator n, $Q_n \in \mathcal{P}_n$ is a sequence of distributions with $\lim_{n\to\infty} Q_n = Q$.

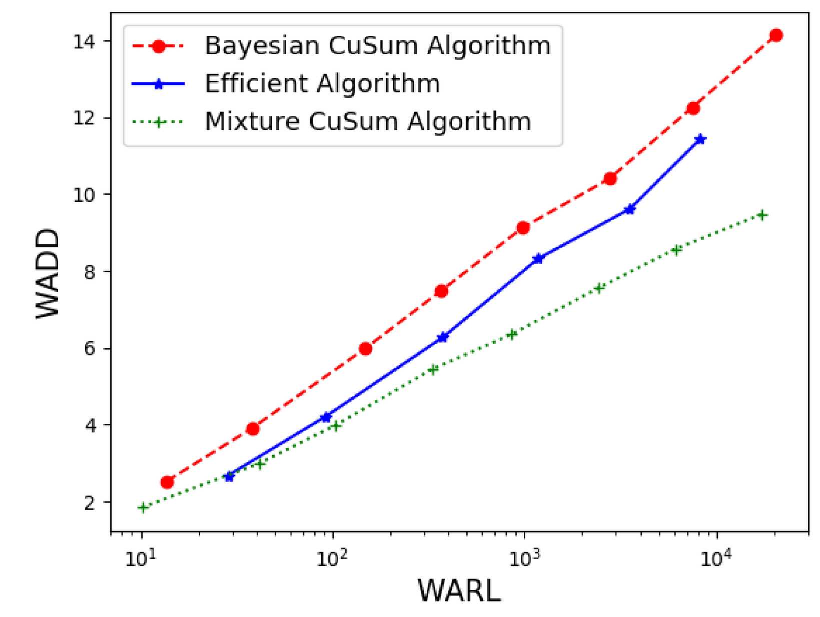
Efficient Test: $W[t] = (t - \hat{\nu}_t + 1)n\left[f_{\mathbf{P}_0}(\alpha, \Pi_{\mathbf{X}^n[\hat{\nu}_t,t]}) - f_{\mathbf{P}_1}(\alpha, \Pi_{\mathbf{X}^n[\hat{\nu}_t,t]})\right]$, $\tau_e = \inf\left\{t \geq 1 : W[t] \geq b\right\}$.

- $\hat{\nu}_t$ is the estimation of change point at t, $\Pi_{\mathbf{X}^n[\hat{\nu}_t,t]}$ is the empirical distribution of samples from $\hat{\nu}_t$ to t.
- $f_{\mathbf{P}_\theta}(\alpha, \Pi_{\mathbf{X}^n[\hat{\nu}_t,t]}) = \inf_{(U_1,...,U_K)\in(\mathcal{P}_{\mathcal{X}})^K, \alpha^T\mathbf{U}=\Pi_{\mathbf{X}^n[\hat{\nu}_t,t]}} \sum_{k=1}^K \alpha_k D(U_k||p_{\theta,k})$, $\mathbf{P}_\theta = [p_{\theta,1}\cdots p_{\theta,K}]^T$.
- $\hat{\nu}_0 = 0$; if $W[t] \leq 0$, $\hat{\nu}_{t+1} = t + 1$, if $W[t] > 0$, $\hat{\nu}_{t+1} = \hat{\nu}_t$.
- if $W[t] \leq 0$, $\Pi_{\mathbf{X}^n[\hat{\nu}_{t+1},t+1]} = \Pi_{\mathbf{X}^n[t+1]}$, if $W[t] > 0$, $\Pi_{\mathbf{X}^n[\hat{\nu}_{t+1},t+1]} = \frac{(t-\hat{\nu}_t+1)\Pi_{\mathbf{X}^n[\hat{\nu}_t,t]}+\Pi_{\mathbf{X}^n[t+1]}}{t-\hat{\nu}_{t+1}+2}$.
- Computation of mixture CuSum is converted into optimization problems thus is more efficient.

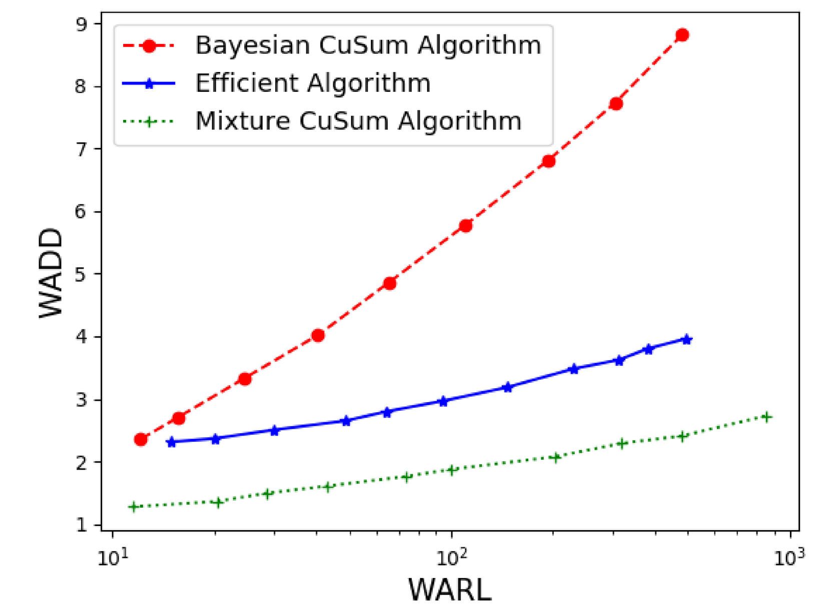
Theorem: For any Ω, $\mathbb{E}_{\Omega}^{\infty}[\tau_e(b)] \geq \frac{e^b}{\left(\frac{b}{h}+1\right)\left(\prod_k \left|\mathcal{P}_{\frac{b}{h}n_k}\right|\right)}$, where $\Gamma \triangleq \{\mu \in \mathcal{P}_{\mathcal{X}} | f_{\mathbf{P}_0}(\alpha,\mu) > f_{\mathbf{P}_1}(\alpha,\mu)\}$, $h = \inf_{\substack{(U_1,...,U_K)\in(\mathcal{P}_{\mathcal{X}})^K \\ \alpha^T\mathbf{U}\in\Gamma}} \sum_{k=1}^K n_k D(U_k||P_{0,k})$. A threshold b can be chosen for false alarm control.

5. Numerical Results

Comparison of three algorithms, WADD v.s. WARL.



- $K = 2, n_1 = 1, n_2 = 1$
- $f_0 = \mathcal{B}(10, 0.5), g_0 = \mathcal{B}(10, 0.5)$
- $f_1 = \mathcal{B}(10, 0.3), g_1 = \mathcal{B}(10, 0.7)$
- Bayesian approach: group assignments in Bayesian setting.



- $K = 2, n_1 = 4, n_2 = 4$

Comparison of the computational complexity, running time v.s. number of sensors.