SEQUENTIAL WASSERSTEIN UNCERTAINTY SETS FOR MINIMAX ROBUST ONLINE CHANGE DETECTION

Yiran Yang, Liyan Xie

School of Data Science, The Chinese University of Hong Kong, Shenzhen

SCHOOL OF DATA SCIENCE 數據科學學院

香港中文大學(深圳) The Chinese University of Hong Kong, Shenzhen

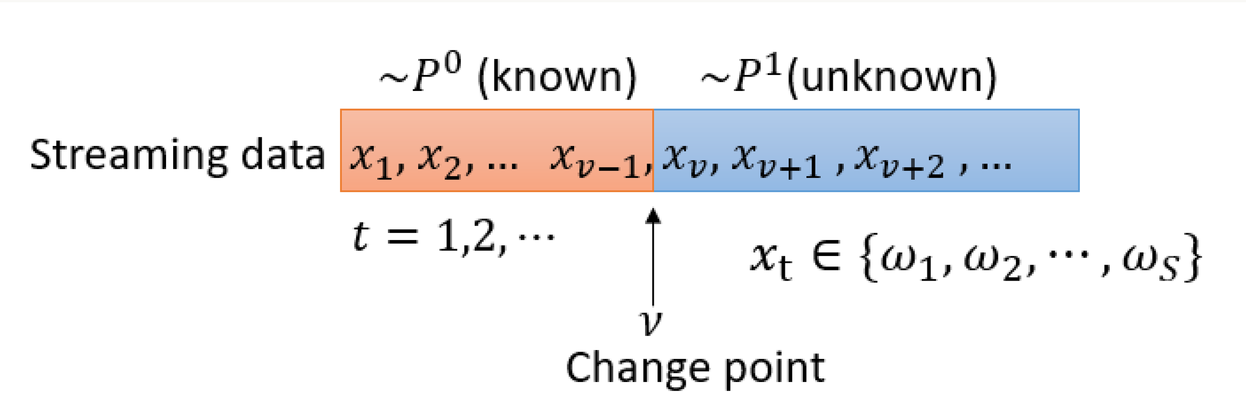
Overview

- **Robust Change Detection** for unknown post-change distributions via **sequential Wasserstein uncertainty sets**;
- The robust CUSUM statistics are constructed by solving the **least favorable distribution** (LFD) via **online convex optimization**;
- Experiments on **i.i.d.** and **non-stationary** post-change distributions.

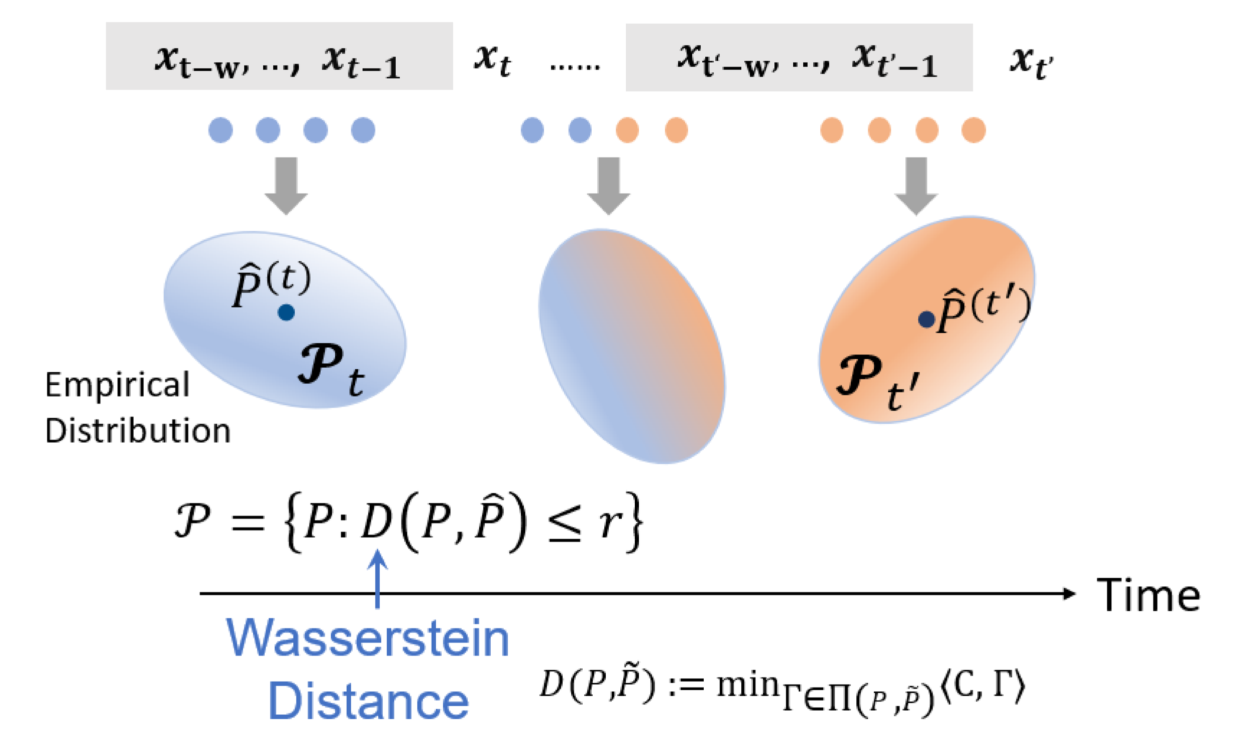
Problem Setup

Goal: Detect **distribution changes** in streaming data (assume finite support of size S)

1. Minimize the detection delay;
2. Control the false alarm rate;
3. Robust Detection under unknown post-change regime.



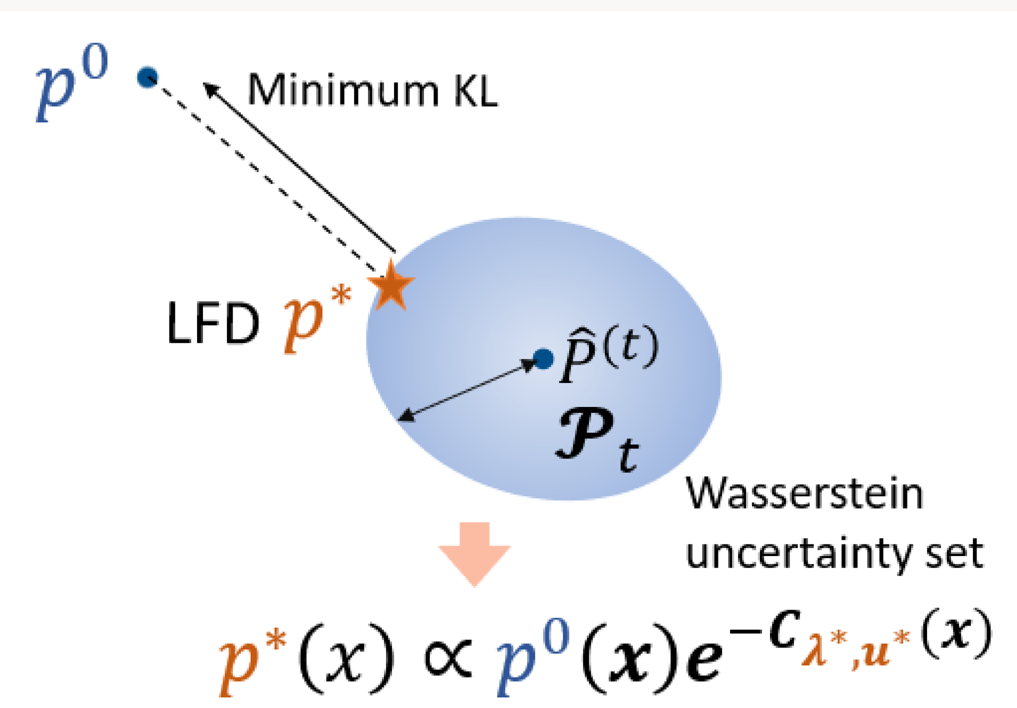
Post-change distributional uncertainties: Sequential sets $\{\mathcal{P}_t\}_{t\in\mathbb{N}}$:



Proposed Method: Online Optimization for LFD and the Robust CUSUM Test

LFD is given by

$$P_t^* = \arg\min_{P\in\mathcal{P}_t} \mathrm{KL}(P\|P^0),$$

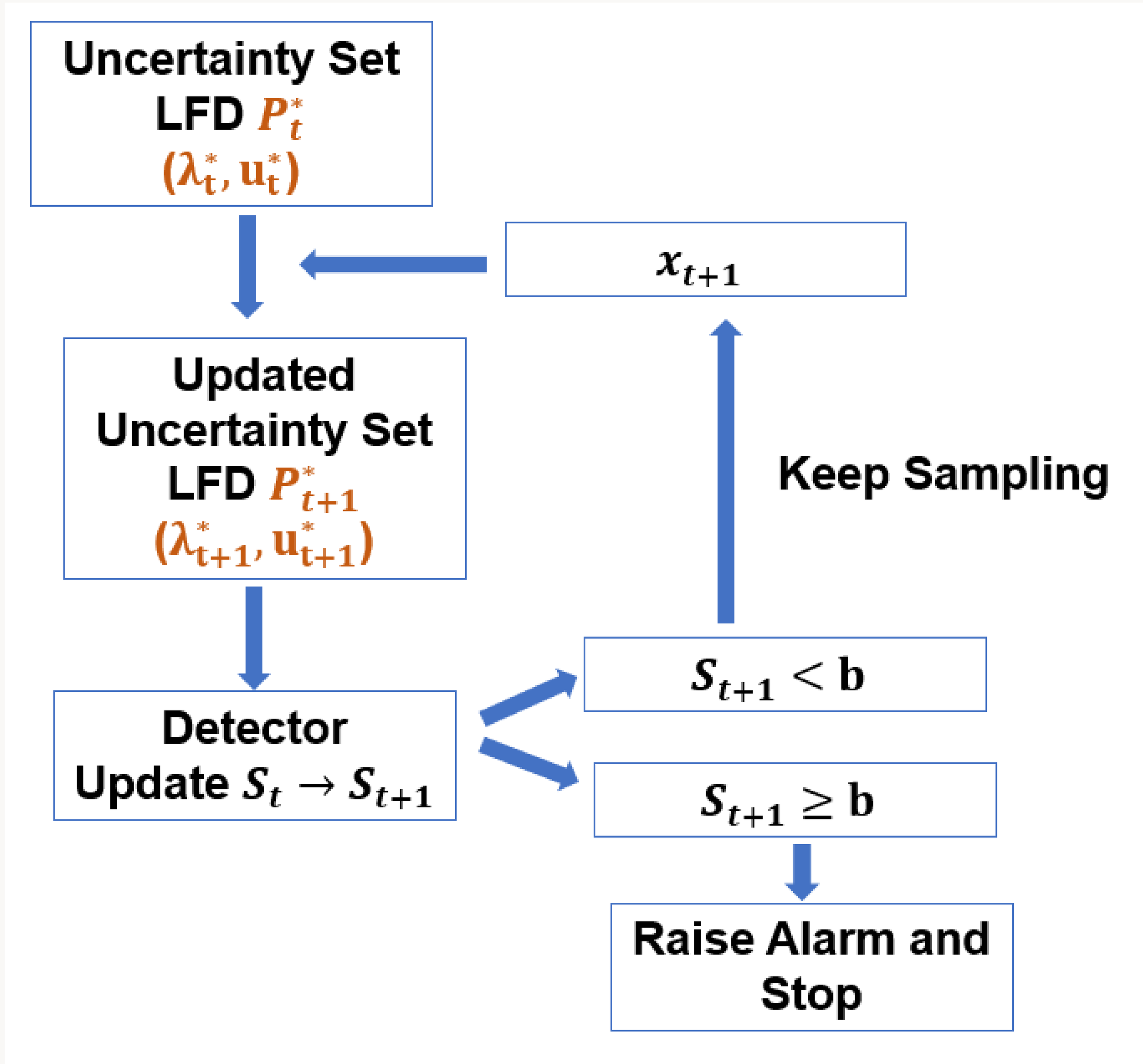


λ_t^*, u_t^*: optimizers of the **dual problem**:

$$\max_{\lambda\ge 0, u\in\mathbb{R}^S}\Big\{-\lambda r_t + \sum_{i=1}^{S}\hat{P}_i^{(t)}u_i - \log\rho(\lambda,u)\Big\},$$

$$\rho(\lambda,u) := \sum_{i=1}^{S} P_i^0 e^{-C_i(\lambda,u)},\ \ C_i(\lambda,u) := \min_{1\le j\le S}\{\lambda c(\omega_i,\omega_j) - u_j\}.$$

- Dual variables λ_t^*, u_t^* updated by **gradient descent**.



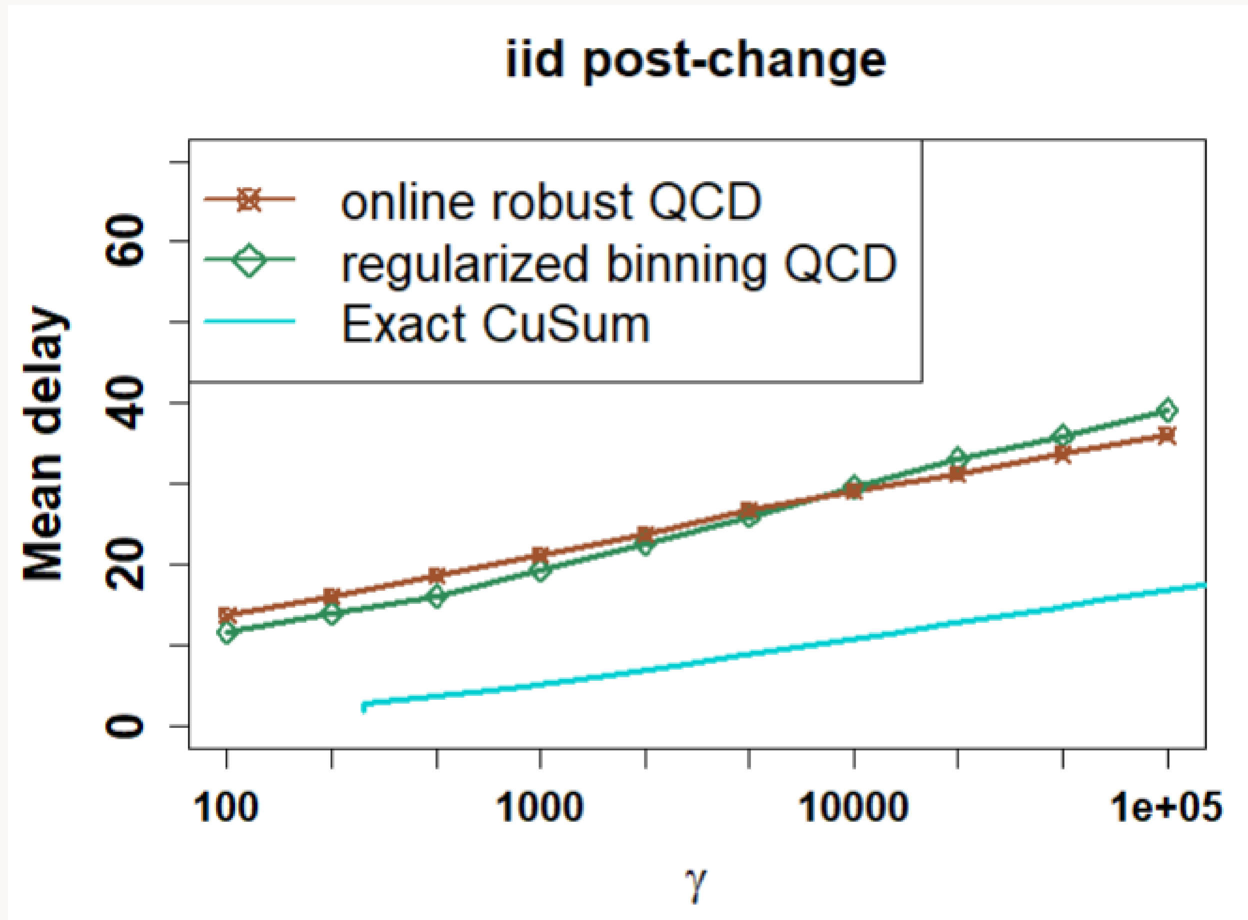
- CUSUM Statistic updated by

$$\mathcal{S}_t = (\mathcal{S}_{t-1})^+ + \log\frac{P_t^*}{P^0}(x_t)$$
$$= (\mathcal{S}_{t-1})^+ - C_{i(x_t)}(\lambda_t^*, u_t^*) - \log\rho(\lambda_t^*, u_t^*).$$

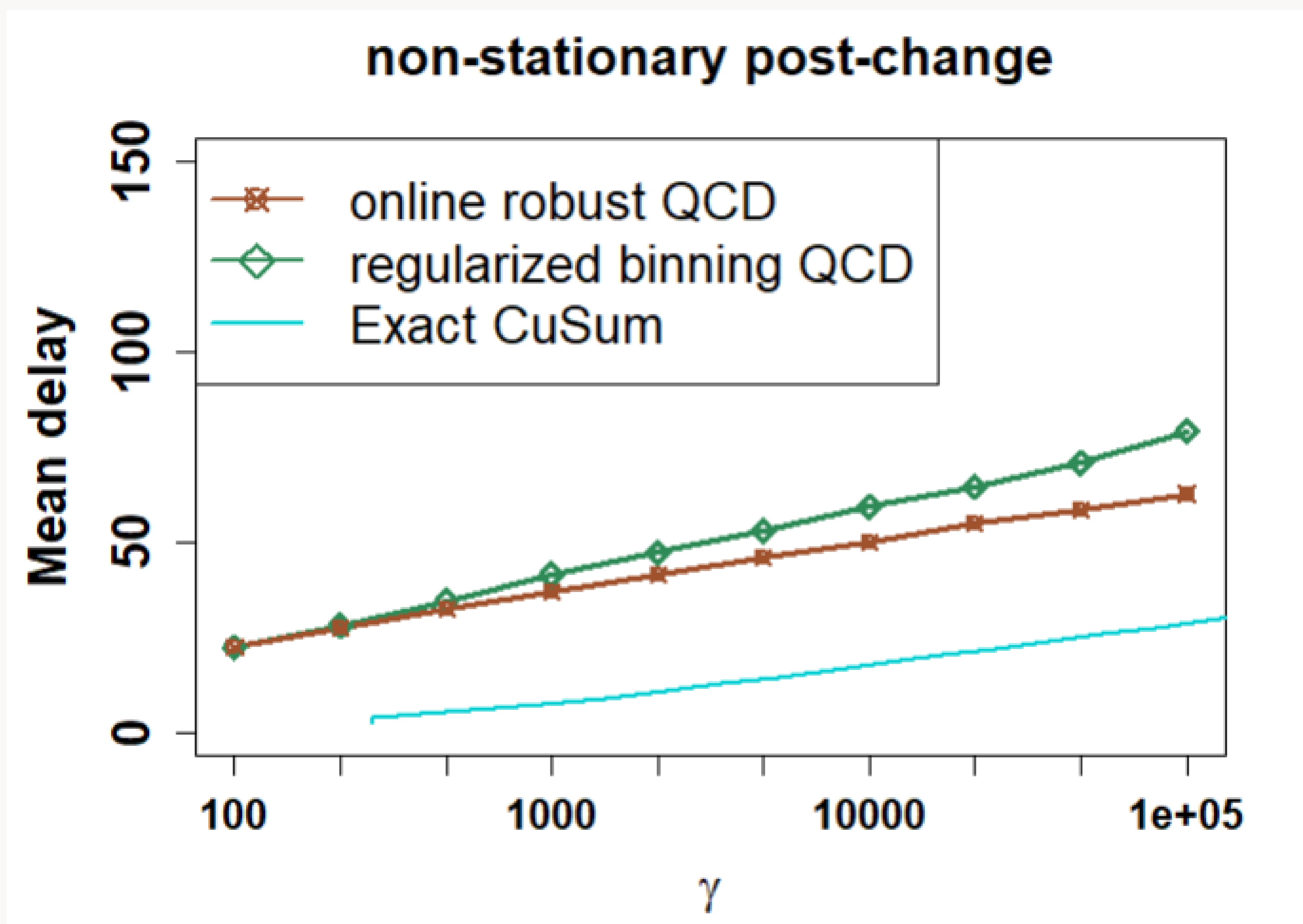
- Threshold b determined by controlling the false alarm rate.

Numerical Results

Comparison of detection delay of the proposed robust CUSUM and the baseline method (regularized binning QCD , Lau, T. S. et al., 2018)



Average Delay vs. Average Run Length (γ) for i.i.d. post-change distribution. The window size is $w = 10$ in both methods.



Average Delay vs. Average Run Length (γ) for non-stationary post-change distributions. The window size is $w = 10$ in both methods.