

# Robust Hypothesis Testing Using Wasserstein Uncertainty Sets

Rui Gao\*, Liyan Xie†, Yao Xie†, Huan Xu†

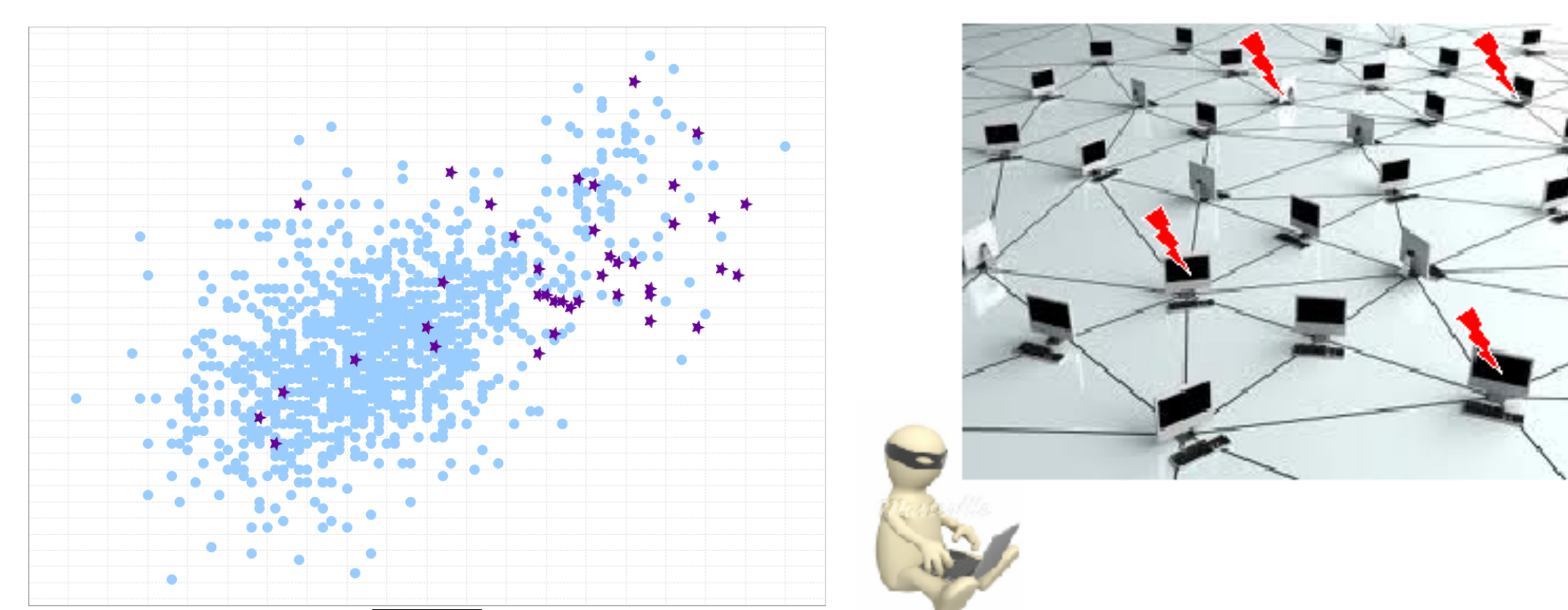
\*Department of Information, Risk, and Operations Management, McCombs School of Business, University of Texas at Austin  
 †H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology

## Overview

- Propose a data-driven, computationally efficient framework for non-parametric robust hypothesis testing
- Develop a nearly-optimal detector and provide an elegant interpretation of the least favorable distributions

## Introduction

- Hypothesis testing with limited samples



Imbalanced classification

Anomaly detection

- Existing robust approaches require density estimation
- Huber's  $\epsilon$ -contamination sets [1]

$$\mathcal{P}_i = \{(1 - \epsilon)q_i + \epsilon f_i : f_i \text{ is any probability density}\}$$

- May not work well in high dimensions
- Parametric hypothesis test via convex optimization by Juditsky and Nemirovski [2]
$$\mathcal{P}_i = \{P_\theta : \theta \in \Theta_i\}$$
- Restrictive convexity assumptions on the parameters
- Goal: find a data-driven non-parametric test that is simple, optimal, robust, and scalable!

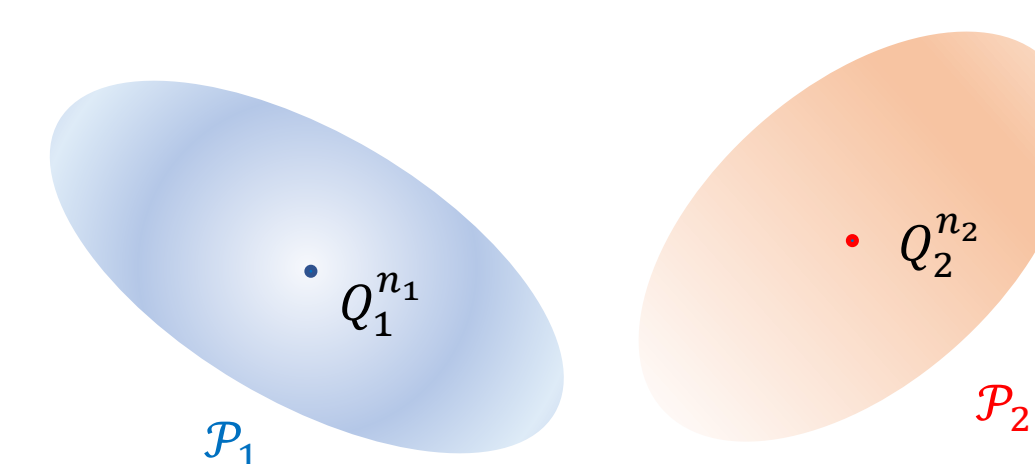
## Problem Setup

- Given an observation  $\omega \in \Omega$ , we would like to decide

$$\begin{aligned} H_1 &: \omega \sim P_1, & P_1 \in \mathcal{P}_1 \\ H_2 &: \omega \sim P_2, & P_2 \in \mathcal{P}_2 \end{aligned}$$

- Wasserstein uncertainty sets:

$$\begin{aligned} \mathcal{P}_1 &= \{P : \mathcal{W}(P, Q_1^{n_1}) \leq \theta_1\} \\ \mathcal{P}_2 &= \{P : \mathcal{W}(P, Q_2^{n_2}) \leq \theta_2\} \end{aligned}$$



where  $Q_1^{n_1}$  and  $Q_2^{n_2}$  are empirical distributions, and  $\mathcal{W}$  denotes the 1-Wasserstein distance

- Problem: find a test  $T : \Omega \rightarrow \{1, 2\}$  to minimize the worst-case type I and type II errors

$$\inf_{T: \Omega \rightarrow \{1, 2\}} \max \left\{ \sup_{P_1 \in \mathcal{P}_1} P_1\{\omega : T(\omega) = 2\}, \sup_{P_2 \in \mathcal{P}_2} P_2\{\omega : T(\omega) = 1\} \right\}$$

- Challenges:

- Non-convex in  $T$
- Infinite-dimensional optimization over  $T$
- Infinite-dimensional optimization over  $P_1, P_2$

## Detector-based Test

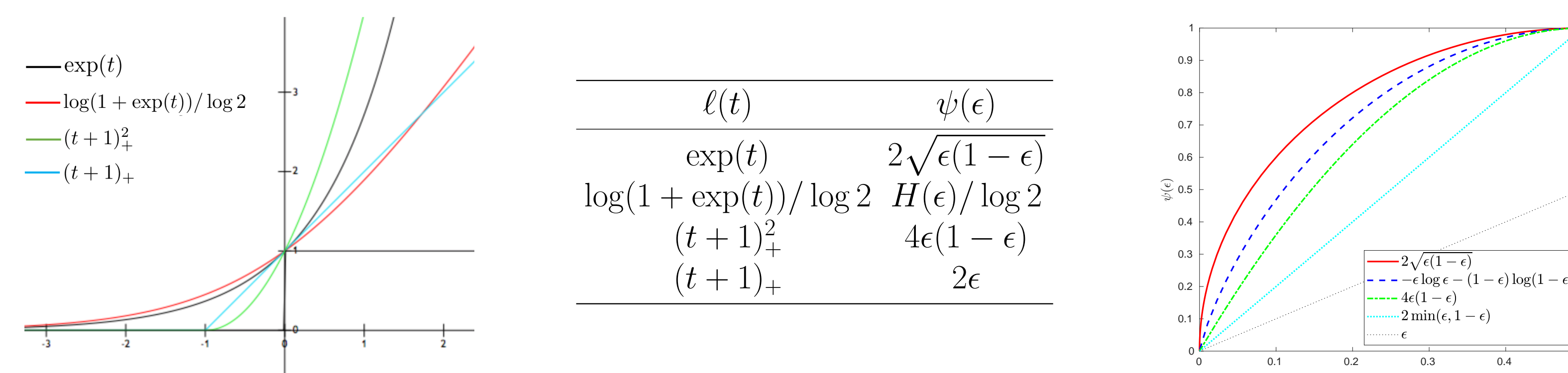
- Consider a detector  $\phi$

$$\begin{aligned} \phi(\omega) \leq 0 &\Rightarrow T_\phi(\omega) = 1 \Rightarrow \text{Claim } H_1 \\ \phi(\omega) > 0 &\Rightarrow T_\phi(\omega) = 2 \Rightarrow \text{Claim } H_2 \end{aligned}$$

- Convex relaxation: convex surrogate  $\ell$  for the 0-1 loss:

$$\inf_{\phi: \Omega \rightarrow \mathbb{R}} \sup_{P_1 \in \mathcal{P}_1, P_2 \in \mathcal{P}_2} \mathbb{E}_{P_1}[\ell \circ (-\phi)(\omega)] + \mathbb{E}_{P_2}[\ell \circ \phi(\omega)]$$

- Result I [Near-optimality]:** Whenever there exists a feasible solution  $T$  of the original problem with objective value  $\leq \epsilon \in (0, 1/2)$ , there exist a feasible solution  $\phi$  of the relaxed problem with objective value  $\leq \psi(\epsilon)$



## Convex Exact Reformulation

- Result II [Strong duality]**  $\inf_{\phi: \Omega \rightarrow \mathbb{R}} \sup_{P_1 \in \mathcal{P}_1, P_2 \in \mathcal{P}_2} \mathbb{E}_{P_1}[\ell \circ (-\phi)(\omega)] + \mathbb{E}_{P_2}[\ell \circ \phi(\omega)]$  is equivalent to

$$\begin{aligned} \max_{\substack{p_1, p_2 \in \mathbb{R}_+^{n_1+n_2} \\ \gamma_1, \gamma_2 \in \mathbb{R}_+^{(n_1+n_2) \times \mathbb{R}_+^{(n_1+n_2)}}}} & \sum_{l=1}^{n_1+n_2} (p_1^l + p_2^l) \psi\left(\frac{p_1^l}{p_1^l + p_2^l}\right) \\ \text{subject to} & \sum_{l=1}^{n_1+n_2} \sum_{m=1}^{n_1+n_2} \gamma_k^{lm} \|\omega^l - \omega^m\| \leq \theta_k, \quad k = 1, 2, \\ & \sum_{m=1}^{n_1+n_2} \gamma_k^{lm} = Q_k^{n_k}(\omega^l), \quad 1 \leq l \leq n_1 + n_2, \quad k = 1, 2, \\ & \sum_{l=1}^{n_1+n_2} \gamma_k^{lm} = p_k^m, \quad 1 \leq m \leq n_1 + n_2, \quad k = 1, 2 \end{aligned}$$

- Statistical interpretation

- Minimizing the divergence  $D(P_1, P_2)$  over  $P_1 \in \mathcal{P}_1, P_2 \in \mathcal{P}_2$  that are supported on the  $n_1 + n_2$  empirical points

- Complexity

- Independent of the dimension of  $\Omega$ , and nearly independent of sample size  $O(\ln n_1 + \ln n_2)$  for most cases

- Proof Sketch

$$\textcircled{1} \inf \sup = \sup \inf$$

- Derived using results on Wasserstein distributionally robust optimization [3]

$$\textcircled{2}$$

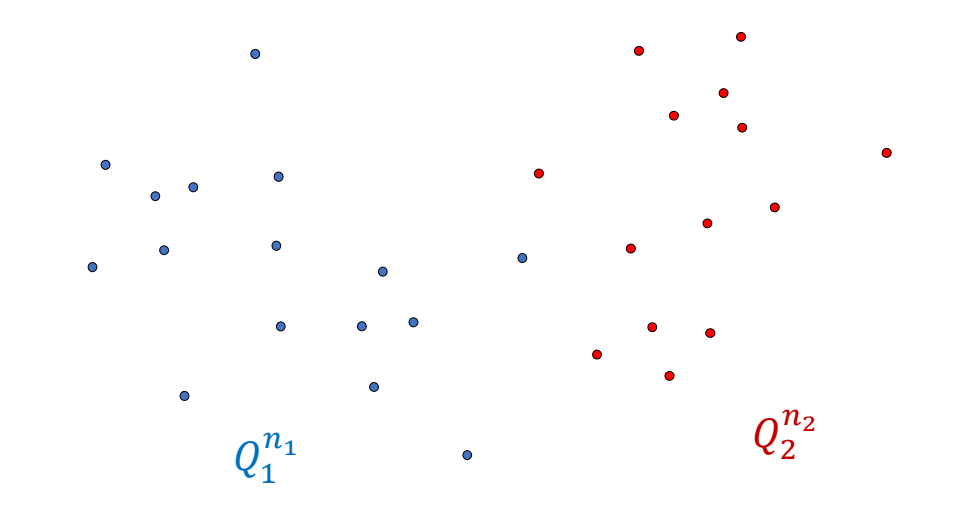
$$\inf_{\phi: \Omega \rightarrow \mathbb{R}} \mathbb{E}_{P_1}[\ell \circ (-\phi)(\omega)] + \mathbb{E}_{P_2}[\ell \circ \phi(\omega)] = \int_{\Omega} \psi\left(\frac{dP_1}{d(P_1+P_2)}\right) d(P_1 + P_2)$$

- Structural characteristics of the least favorable distributions for

$$\sup_{P_1 \in \mathcal{P}_1, P_2 \in \mathcal{P}_2} \int_{\Omega} \psi\left(\frac{dP_1}{d(P_1+P_2)}\right) d(P_1 + P_2)$$

## Least Favorable Distributions and Optimal Detector

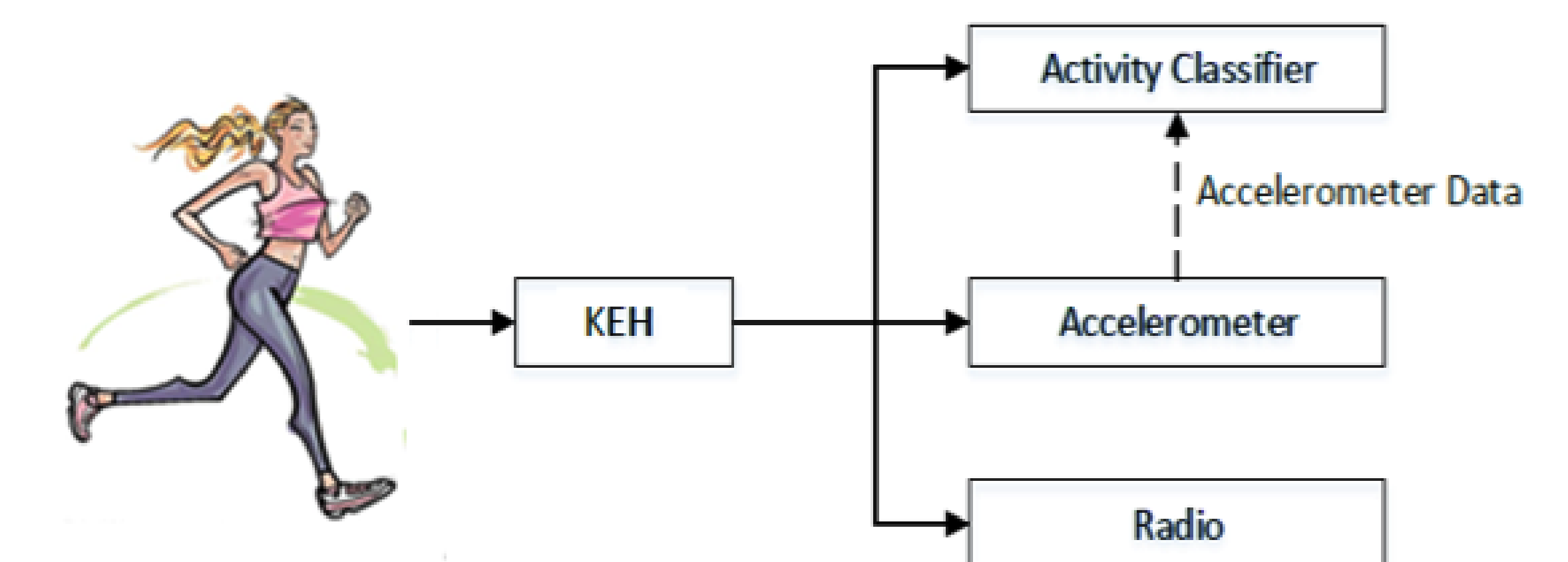
- The least favorable distributions  $P_1^*, P_2^*$  have an identical support on  $\text{supp } Q_1^{n_1} \cup \text{supp } Q_2^{n_2}$
- The worst-case risk is  $2(1 - D(P_1^*, P_2^*))$ , where  $P_1^*, P_2^*$  depend on  $\ell, Q_1^{n_1}, Q_2^{n_2}$ , and  $\theta_1, \theta_2$
- Optimal detector



$\ell(t)$	$\phi(\omega)$	$1 - 1/2 \text{ Risk}$
$\exp(t)$	$\log \sqrt{p_1^*(\omega)/p_2^*(\omega)}$	$H^2(P_1^*, P_2^*)$
$\log(1 + \exp(t))/\log 2$	$\log(p_1^*(\omega)/p_2^*(\omega))$	$JS(P_1^*, P_2^*)/\log 2$
$(t+1)_+^2$	$1 - 2\frac{p_1^*(\omega)}{p_1^*(\omega) + p_2^*(\omega)}$	$\chi^2(P_1^*, P_2^*)$
$(t+1)_+$	$\text{sgn}(p_1^*(\omega) - p_2^*(\omega))$	$TV(P_1^*, P_2^*)$

## Numerical Example

- Human activity detection



Credit: CSIRO Research

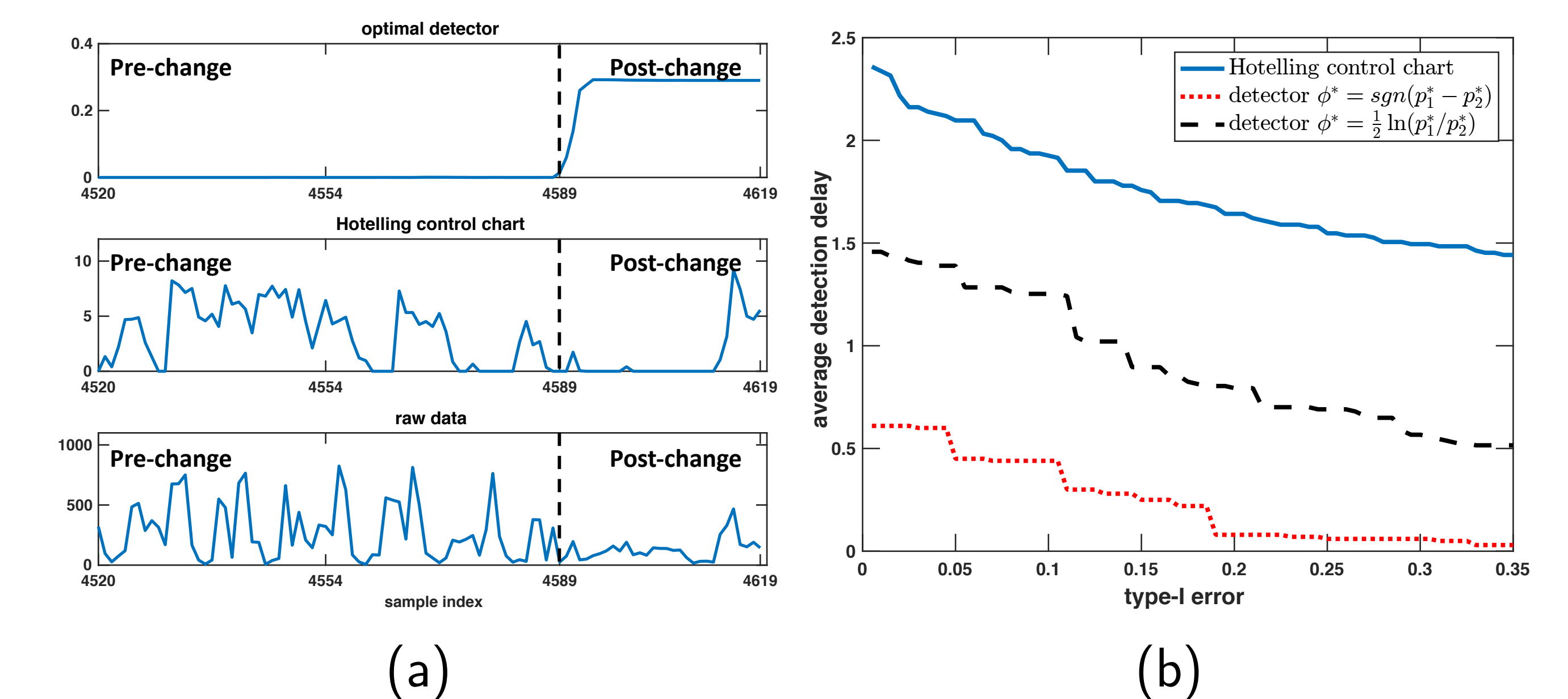


Figure: Jogging vs. Walking, the average is taken over 100 sequences of data.

## References

- Peter J Huber. *A robust version of the probability ratio test*. The Annals of Mathematical Statistics, 36(6):1753-1758, 1965.
- Anatoli Juditsky Alexander Goldenshluger and Arkadi Nemirovski. *Hypothesis testing by convex optimization*. Electron. J. Statist., 9(2):1645-1712, 2015.
- Rui Gao and Anton Kleywegt, *Distributionally robust stochastic optimization with Wasserstein distance*. arXiv preprint arXiv:1604.02199, 2016.

arXiv

