Data-driven Robust Change Detection under Wasserstein Uncertainty Sets

Liyan Xie*, Yuchen Liang[†], and Venugopal V. Veeravalli[⋆]

*Department of Industrial and Systems Engineering, University of Minnesota,
[†]Department of Electrical and Computer Engineering, The Ohio State University
[⋆]Department of Electrical and Computer Engineering, UIUC

ACMS, University of Notre Dame, Sep 2024

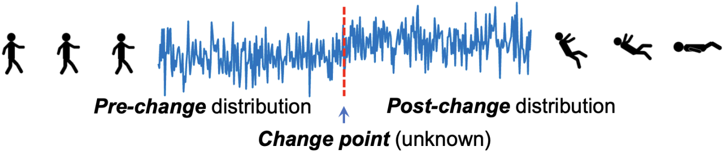
Outline

- Motivation: Data-Driven Robust Change Detection
- Proposed: Distributionally Robust CuSum Test (DR-CuSum)
- Theoretical Results: Asymptotic Optimality
- Extension to Online Distributional Uncertainties
- Simulation Examples

Outline

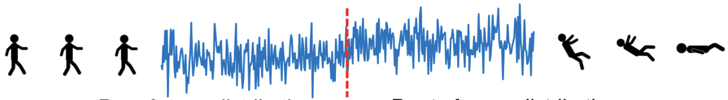
- Motivation: Data-Driven Robust Change Detection
- Proposed: Distributionally Robust CuSum Test (DR-CuSum)
- Theoretical Results: Asymptotic Optimality
- Extension to Online Distributional Uncertainties
- Simulation Examples

Distributional Shifts in Online Data



$$x_1, \ldots, x_{\tau-1} \overset{iid}{\sim} f_0(\text{known}), \quad x_\tau, x_{\tau+1}, \ldots \overset{iid}{\sim} f_1(\text{unknown}).$$

Distributional Shifts in Online Data

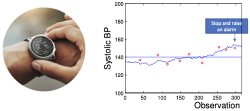


Pre-change distribution ***Post-change*** distribution

Change point (unknown)

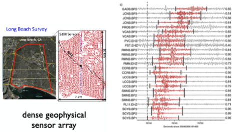
$$x_1, \ldots, x_{\tau-1} \overset{iid}{\sim} f_0(\text{known}), \quad x_\tau, x_{\tau+1}, \ldots \overset{iid}{\sim} f_1(\text{unknown}).$$

- Application examples



Bloop pressure monitoring

Anomaly detection



Seismic event monitoring

Typical Detection Approaches

- CUSUM (for known f_1):

$$S_t = \max\{S_{t-1}, 0\} + \log \frac{f_1(x_t)}{f_0(x_t)}.$$

Page (1954). Continuous inspection schemes. Biometrika.
Lai (1998).Information bounds and quick detection of parameter changes in stochastic systems. IEEE TIT.
X., Moustakides, Xie (2023). Window-Limited CUSUM for Sequential Change Detection. IEEE TIT.

Typical Detection Approaches

- CUSUM (for known f_1):

$$S_t = \max\{S_{t-1}, 0\} + \log \frac{f_1(x_t)}{f_0(x_t)}.$$

- Generalized likelihood ratio test (assuming $f_1(\cdot;\theta), \theta \in \Theta$):

$$S_t^{GLR} = \max_{0 \le k < t} \sup_{\theta \in \Theta} \sum_{i=k+1}^{t} \log \frac{f_1(x_i;\theta)}{f_0(x_i)}.$$

(also non-parametric versions based on DRE)

Page (1954). Continuous inspection schemes. Biometrika.
Lai (1998).Information bounds and quick detection of parameter changes in stochastic systems. IEEE TIT.
X., Moustakides, Xie (2023). Window-Limited CUSUM for Sequential Change Detection. IEEE TIT.

Typical Detection Approaches

- CUSUM (for known f_1):

$$S_t = \max\{S_{t-1}, 0\} + \log \frac{f_1(x_t)}{f_0(x_t)}.$$

- Generalized likelihood ratio test (assuming $f_1(\cdot;\theta), \theta \in \Theta$):

$$S_t^{GLR} = \max_{0 \le k < t} \sup_{\theta \in \Theta} \sum_{i=k+1}^{t} \log \frac{f_1(x_i;\theta)}{f_0(x_i)}.$$

(also non-parametric versions based on DRE)

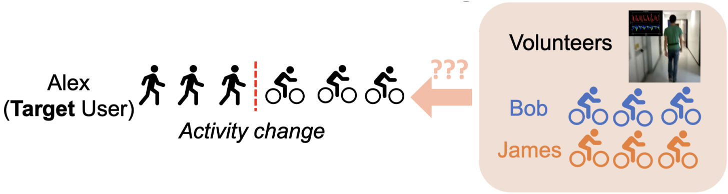
- Window-limited CUSUM

$$\mathcal{S}_t = \max\{\mathcal{S}_{t-1}, 0\} + \log \frac{f_1(x_t;\hat{\theta}_{t-1})}{f_0(x_t)},\ \hat{\theta}_{t-1} = \hat{\theta}_{t-1}(x_{t-w}, \ldots, x_{t-1}), t > w.$$

Page (1954). Continuous inspection schemes. Biometrika.
Lai (1998).Information bounds and quick detection of parameter changes in stochastic systems. IEEE TIT.
X., Moustakides, Xie (2023). Window-Limited CUSUM for Sequential Change Detection. IEEE TIT.

Motivation

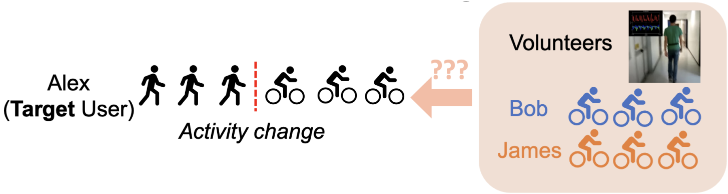
When there are pre-collected training data,
can we utilize them effectively?



- Yes! Data distributions among different people are similar.

Motivation

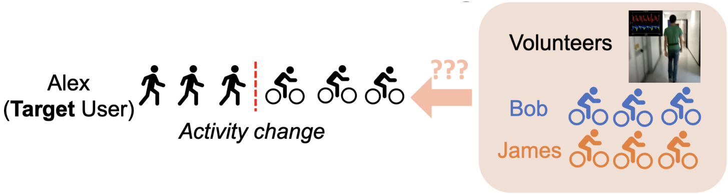
When there are pre-collected training data,
can we utilize them effectively?



- Yes! Data distributions among different people are similar, *but different*
- But, better in a smart way. Training data can be *limited*.

Motivation

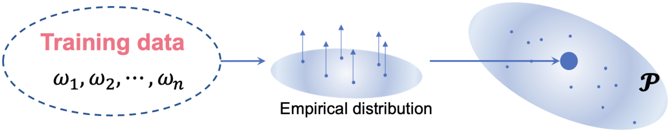
When there are pre-collected training data,
can we utilize them effectively?



- Yes! Data distributions among different people are similar, *but different*
- But, better in a smart way. Training data can be *limited*.
- Use training data in a **robust** way.

Data-Driven Uncertainty Sets

A fully *data-driven* and *non-parametric* construction,
assuming training data $\{\omega_1, \ldots, \omega_n\} \overset{\text{iid}}{\sim} P$.

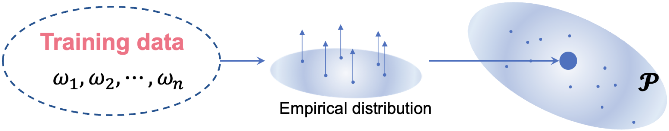


Data-Driven Uncertainty Sets

A fully *data-driven* and *non-parametric* construction,
assuming training data $\{\omega_1, \ldots, \omega_n\} \overset{\text{iid}}{\sim} P$.

- The uncertainty model:

$$\mathcal{P} = \{P : \mathsf{W}_s(P, \hat{P}) \leq r_s\},$$



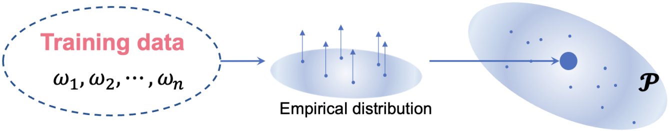
Data-Driven Uncertainty Sets

A fully *data-driven* and *non-parametric* construction, assuming training data $\{\omega_1, \ldots, \omega_n\} \overset{\text{iid}}{\sim} P$.

- The uncertainty model:

$$\mathcal{P} = \{P : \mathsf{W}_s(P, \hat{P}) \leq r_s\},$$

  - $\hat{P}$: **empirical distribution** of $\{\omega_1, \ldots, \omega_n\}$



Data-Driven Uncertainty Sets

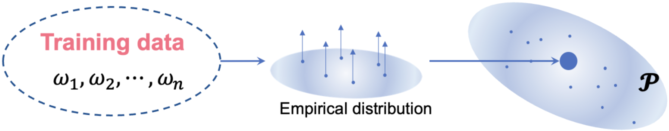
A fully *data-driven* and *non-parametric* construction,
assuming training data $\{\omega_1, \ldots, \omega_n\} \overset{\text{iid}}{\sim} P$.

- The uncertainty model:

$$\mathcal{P} = \{P : \mathsf{W}_s(P, \hat{P}) \leq r_s\},$$

  - $\hat{P}$: **empirical distribution** of $\{\omega_1, \ldots, \omega_n\}$
  - Wasserstein distance $\mathsf{W}_s(\cdot, \cdot)$:

$$\mathsf{W}_s(P, Q) := \left\{ \min_{\gamma \in \Gamma(P,Q)} \mathbb{E}_{(\omega, \omega') \sim \gamma} \left[c(\omega, \omega')^s \right] \right\}^{1/s}$$



Problem Setup of Robust Change Detection

- Data Model:

$$X_k \overset{\text{iid}}{\sim} Q, \quad k = 1, 2, \ldots, \nu - 1,$$
$$X_k \overset{\text{iid}}{\sim} P, \quad k = \nu, \nu + 1, \ldots$$

Problem Setup of Robust Change Detection

- Data Model:

$$X_k \overset{\text{iid}}{\sim} Q, \quad k = 1, 2, \ldots, \nu - 1,$$
$$X_k \overset{\text{iid}}{\sim} P, \quad k = \nu, \nu + 1, \ldots$$

  - The change-point ν is *unknown* but *deterministic*

Problem Setup of Robust Change Detection

- Data Model:

$$X_k \overset{\text{iid}}{\sim} Q, \quad k = 1, 2, \ldots, \nu - 1,$$
$$X_k \overset{\text{iid}}{\sim} P, \quad k = \nu, \nu + 1, \ldots$$

  - The change-point ν is *unknown* but *deterministic*
  - Q is known; P is unknown, and $P \in \mathcal{P}$ (**uncertainty set**)

Problem Setup of Robust Change Detection

- Data Model:

$$X_k \overset{\text{iid}}{\sim} Q, \quad k = 1, 2, \ldots, \nu - 1,$$
$$X_k \overset{\text{iid}}{\sim} P, \quad k = \nu, \nu + 1, \ldots$$

  - The change-point ν is *unknown* but *deterministic*
  - Q is known; P is unknown, and $P \in \mathcal{P}$ (**uncertainty set**)

- Goal: Detect change-point ν as quickly as possible, while controlling the false alarm rate.

Problem Setup of Robust Change Detection

- Data Model:

$$X_k \overset{\text{iid}}{\sim} Q, \quad k = 1, 2, \ldots, \nu - 1,$$
$$X_k \overset{\text{iid}}{\sim} P, \quad k = \nu, \nu + 1, \ldots$$

  - The change-point ν is *unknown* but *deterministic*
  - Q is known; P is unknown, and $P \in \mathcal{P}$ (**uncertainty set**)

- Goal: Detect change-point ν as quickly as possible, while controlling the false alarm rate.
- Method: The detection is performed through a *stopping time* τ on the observation sequence.

Problem Setup (continued): Asymptotically Minimax Robust QCD

- Minimax problem:

$$\inf_{\tau \in C(\gamma)} \sup_{P \in \mathcal{P}} \mathsf{WADD}^{P}(\tau)$$

Problem Setup (continued): Asymptotically Minimax Robust QCD

- Minimax problem:

$$\inf_{\tau \in C(\gamma)} \sup_{P \in \mathcal{P}} \mathsf{WADD}^P(\tau)$$

  - $\mathcal{C}(\gamma) = \{\tau : \mathbb{E}_\infty[\tau] \geq \gamma\}$

Problem Setup (continued): Asymptotically Minimax Robust QCD

- Minimax problem:

$$\inf_{\tau \in C(\gamma)} \sup_{P \in \mathcal{P}} \mathsf{WADD}^P(\tau)$$

  - $\mathcal{C}(\gamma) = \{\tau : \mathbb{E}_\infty[\tau] \geq \gamma\}$
  - $\mathsf{WADD}^P(\tau) = \sup_{\nu \geq 1} \operatorname{ess\,sup} \mathbb{E}_\nu^P\left[(\tau - \nu + 1)^+ | \mathcal{F}_{\nu-1}\right].$

Problem Setup (continued): Asymptotically Minimax Robust QCD

- Minimax problem:

$$\inf_{\tau \in C(\gamma)} \sup_{P \in \mathcal{P}} \mathsf{WADD}^P(\tau)$$

  - $\mathcal{C}(\gamma) = \{\tau : \mathbb{E}_\infty[\tau] \geq \gamma\}$
  - $\mathsf{WADD}^P(\tau) = \sup_{\nu \geq 1} \operatorname{ess\,sup} \mathbb{E}_\nu^P\left[(\tau - \nu + 1)^+ | \mathcal{F}_{\nu-1}\right].$

- A solution $\tau^* \in C(\gamma)$ is called *first-order asymptotic minimax robust* if

$$\sup_{P \in \mathcal{P}} \mathsf{WADD}^P(\tau^*) = \inf_{\tau \in C(\gamma)} \sup_{P \in \mathcal{P}} \mathsf{WADD}^P(\tau) \cdot (1 + o(1)),$$

here $o(1) \to 0$ as $\gamma \to \infty$.

Related Work

- Classial QCD
  - Seminal works [Page, 1954] [Lorden, 1971] [Pollak, 1985] [Moustakides, 1986] [Lai, 1998], etc.

Related Work

- Classial QCD
  - Seminal works [Page, 1954] [Lorden, 1971] [Pollak, 1985] [Moustakides, 1986] [Lai, 1998], etc.
  - Recent books [Tartakovsky et al, 2015] [Tartakovsky, 2020]

Related Work

- Classial QCD
  - Seminal works [Page, 1954] [Lorden, 1971] [Pollak, 1985] [Moustakides, 1986] [Lai, 1998], etc.
  - Recent books [Tartakovsky et al, 2015] [Tartakovsky, 2020]
  - Data-driven and non-parametric methods: [Moustakides and Basioti, 2019] [Liang and Veeravalli, 2023] [Li et al, 2015], etc.

Related Work

- Classial QCD
    - Seminal works [Page, 1954] [Lorden, 1971] [Pollak, 1985] [Moustakides, 1986] [Lai, 1998], etc.
    - Recent books [Tartakovsky et al, 2015] [Tartakovsky, 2020]
    - Data-driven and non-parametric methods: [Moustakides and Basioti, 2019] [Liang and Veeravalli, 2023] [Li et al, 2015], etc.
- Minimax Robust Hypothesis Testing or QCD
    - Joint stochastic boundedness [Unnikrishnan, Veeravalli, and Meyn, 2011], Weak stochastic boundedness [Molloy and Ford, 2017]

Related Work

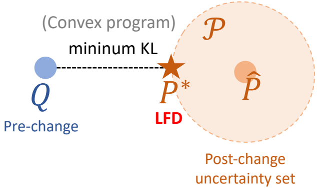
- Classial QCD
  - Seminal works [Page, 1954] [Lorden, 1971] [Pollak, 1985] [Moustakides, 1986] [Lai, 1998], etc.
  - Recent books [Tartakovsky et al, 2015] [Tartakovsky, 2020]
  - Data-driven and non-parametric methods: [Moustakides and Basioti, 2019] [Liang and Veeravalli, 2023] [Li et al, 2015], etc.
- Minimax Robust Hypothesis Testing or QCD
  - Joint stochastic boundedness [Unnikrishnan, Veeravalli, and Meyn, 2011], Weak stochastic boundedness [Molloy and Ford, 2017]
  - Different Uncertainty Sets: ϵ-contamination [Crow and Schwartz, 1994], KL divergence sets [Levy, 2008], Wasserstein sets [Gao et al, 2018], kernel MMD sets [Sun and Zou, 2021], Sinkhorn uncertainty sets [Wang and Xie, 2022], moment-based sets [Liang and Veeravalli, 2022], etc.

Outline

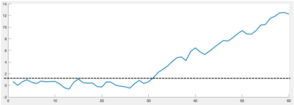
- Motivation: Data-Driven Robust Change Detection
- Proposed: Distributionally Robust CuSum Test (DR-CuSum)
- Theoretical Results: Asymptotic Optimality
- Extension to Online Distributional Uncertainties
- Simulation Examples

Method Overview

Solving LFD → **DR-CuSum Test**



CuSum based on P^* and Q



Finding LFDs — Weak Stochastic Boundedness

LFD — a representative distribution within uncertainty set on which the stopping time reaches the worst-case performance.

Molloy and Ford, "Misspecified and asymptotically minimax robust quickest change detection," IEEE Transactions on Information Theory, 2017.

Finding LFDs — Weak Stochastic Boundedness

LFD — a representative distribution within uncertainty set on which the stopping time reaches the worst-case performance.

- *Weak stochastic boundedness:* $(\mathcal{P}_0, \mathcal{P}_1)$ is weakly stochastically bounded by (P_0^*, P_1^*) if
  1. $\mathsf{KL}(P_1^*||P_0^*) \leq \mathsf{KL}(P_1||P_0^*) - \mathsf{KL}(P_1||P_1^*),\ \forall P_1 \in \mathcal{P}_1$;
  2. $\mathbb{E}^{P_0}\left[\frac{dP_1^*}{dP_0^*}(X)\right] \leq \mathbb{E}^{P_0^*}\left[\frac{dP_1^*}{dP_0^*}(X)\right] = 1,\ \forall P_0 \in \mathcal{P}_0$.
- (P_0^*, P_1^*) are called least favorable distributions (LFDs).

Molloy and Ford, "Misspecified and asymptotically minimax robust quickest change detection," IEEE Transactions on Information Theory, 2017.

Finding LFDs — Weak Stochastic Boundedness

LFD — a representative distribution within uncertainty set on which the stopping time reaches the worst-case performance.

- *Weak stochastic boundedness:* $(\mathcal{P}_0, \mathcal{P}_1)$ is weakly stochastically bounded by (P_0^*, P_1^*) if
  1. $\mathsf{KL}(P_1^* || P_0^*) \leq \mathsf{KL}(P_1 || P_0^*) - \mathsf{KL}(P_1 || P_1^*),\ \forall P_1 \in \mathcal{P}_1$;
  2. $\mathbb{E}^{P_0}\left[\frac{dP_1^*}{dP_0^*}(X)\right] \leq \mathbb{E}^{P_0^*}\left[\frac{dP_1^*}{dP_0^*}(X)\right] = 1,\ \forall P_0 \in \mathcal{P}_0$.
- (P_0^*, P_1^*) are called least favorable distributions (LFDs).
- Corollary: For the singleton $\mathcal{Q} = \{Q\}$ and a *convex* set of distributions $\mathcal{P}$, $(\mathcal{Q}, \mathcal{P})$ is weakly stochastically bounded by the pair of distributions (Q, P^*), where

$$P^* = \arg\min_{P \in \mathcal{P}} \mathsf{KL}(P || Q).$$

Molloy and Ford, "Misspecified and asymptotically minimax robust quickest change detection," IEEE Transactions on Information Theory, 2017.

Finding LFDs — Weak Stochastic Boundedness

LFD — a representative distribution within uncertainty set on which the stopping time reaches the worst-case performance.

- *Weak stochastic boundedness:* $(\mathcal{P}_0, \mathcal{P}_1)$ is weakly stochastically bounded by (P_0^*, P_1^*) if
  1. $\mathsf{KL}(P_1^*||P_0^*) \le \mathsf{KL}(P_1||P_0^*) - \mathsf{KL}(P_1||P_1^*),\ \forall P_1 \in \mathcal{P}_1$;
  2. $\mathbb{E}^{P_0}\left[\frac{dP_1^*}{dP_0^*}(X)\right] \le \mathbb{E}^{P_0^*}\left[\frac{dP_1^*}{dP_0^*}(X)\right] = 1,\ \forall P_0 \in \mathcal{P}_0.$
- (P_0^*, P_1^*) are called least favorable distributions (LFDs).
- Corollary: For the singleton $\mathcal{Q} = \{Q\}$ and a *convex* set of distributions $\mathcal{P}$, $(\mathcal{Q}, \mathcal{P})$ is weakly stochastically bounded by the pair of distributions (Q, P^*), where

$$P^* = \arg\min_{P \in \mathcal{P}} \mathsf{KL}(P||Q).$$

- In our case:

$$\min_P \mathsf{KL}(P||Q),\ \text{such that } \mathsf{W}_s(P, \widehat{P}) \le r_s.$$

Molloy and Ford, "Misspecified and asymptotically minimax robust quickest change detection," IEEE Transactions on Information Theory, 2017.

Solving LFDs

Theorem (X., Liang, Veeravalli, 2024)

The pdf of the LFD satisfies $p^*(x) \propto q(x)e^{-C_{\lambda^*,u^*}(x)}$, *where* $q(x)$ *is the pre-change pdf and* $\lambda^* \geq 0, u^* \in \mathbb{R}^n$ *are the optimal solution to*

$$\max_{\lambda \geq 0, u \in \mathbb{R}^n} \left\{ -\lambda r_s + \frac{1}{n}\sum_{i=1}^{n} u_i - \log \eta(\lambda, u) \right\},$$

where $\eta(\lambda, u) = \mathbb{E}_Q[e^{-C_{\lambda,u}(x)}]$, $C_{\lambda,u}(x) = \min_{1 \leq i \leq n} \{\lambda c^s(x, \omega_i) - u_i\}$.

X., Liang, and Veeravalli, International Conference on Artificial Intelligence and Statistics (AISTATS), 2024.

Solving LFDs

Theorem (X., Liang, Veeravalli, 2024)

The pdf of the LFD satisfies $p^*(x) \propto q(x)e^{-C_{\lambda^*,u^*}(x)}$, *where* $q(x)$ *is the pre-change pdf and* $\lambda^* \geq 0, u^* \in \mathbb{R}^n$ *are the optimal solution to*

$$\max_{\lambda \geq 0, u \in \mathbb{R}^n} \Big\{ -\lambda r_s + \frac{1}{n}\sum_{i=1}^{n} u_i - \log \eta(\lambda, u) \Big\},$$

where $\eta(\lambda, u) = \mathbb{E}_Q[e^{-C_{\lambda,u}(x)}]$, $C_{\lambda,u}(x) = \min_{1 \leq i \leq n} \{\lambda c^s(x, \omega_i) - u_i\}$.

- The optimization problem above is *convex* and can be solved efficiently by off-the-shelf solvers.

X., Liang, and Veeravalli, International Conference on Artificial Intelligence and Statistics (AISTATS), 2024.

Solving LFDs

Theorem (X., Liang, Veeravalli, 2024)

The pdf of the LFD satisfies $p^*(x) \propto q(x)e^{-C_{\lambda^*,u^*}(x)}$, *where* $q(x)$ *is the pre-change pdf and* $\lambda^* \geq 0, u^* \in \mathbb{R}^n$ *are the optimal solution to*

$$\max_{\lambda \geq 0, u \in \mathbb{R}^n} \left\{ -\lambda r_s + \frac{1}{n}\sum_{i=1}^{n} u_i - \log \eta(\lambda, u) \right\},$$

where $\eta(\lambda, u) = \mathbb{E}_Q[e^{-C_{\lambda,u}(x)}]$, $C_{\lambda,u}(x) = \min_{1 \leq i \leq n} \{\lambda c^s(x, \omega_i) - u_i\}$.

- The optimization problem above is *convex* and can be solved efficiently by off-the-shelf solvers.
- When Q is also unknown but samples available, we may use sample average approximation to estimate $\eta(\lambda, u) = \mathbb{E}_Q[e^{-C_{\lambda,u}(x)}]$ in the objective.

X., Liang, and Veeravalli, International Conference on Artificial Intelligence and Statistics (AISTATS), 2024.

A Gaussian Example

- Setting:
  - $c(x, x') = \|x - x'\|_2$, $s = 2$
  - $Q : \mathcal{N}(0, 1)$

A Gaussian Example

- Setting:
  - $c(x, x') = \|x - x'\|_2$, $s = 2$
  - $Q : \mathcal{N}(0, 1)$

- Closed-form solution of LFD when $n = 1$:
  - $C_{\lambda,u}(x) = \lambda(x - \omega_1)^2 - u$
  - $\eta(\lambda, u) = \frac{1}{\sqrt{1+2\lambda}} e^{-\frac{\lambda}{1+2\lambda}\omega_1^2 + u}$

$$\Rightarrow \lambda^* = \begin{cases} \frac{\omega_1^2}{\sqrt{1+4r\omega_1^2} - 1} - \frac{1}{2}, & \text{if } r \leq 1 + \omega_1^2 \\ 0 & \text{o.w.} \end{cases}$$

A Gaussian Example

- Setting:
  - $c(x, x') = \|x - x'\|_2$, $s = 2$
  - $Q : \mathcal{N}(0, 1)$

- Closed-form solution of LFD when $n = 1$:
  - $C_{\lambda,u}(x) = \lambda(x - \omega_1)^2 - u$
  - $\eta(\lambda, u) = \frac{1}{\sqrt{1+2\lambda}} e^{-\frac{\lambda}{1+2\lambda}\omega_1^2 + u}$

$$\Rightarrow \lambda^* = \begin{cases} \frac{\omega_1^2}{\sqrt{1+4r\omega_1^2 - 1}} - \frac{1}{2}, & \text{if } r \leq 1 + \omega_1^2 \\ 0 & \text{o.w.} \end{cases}$$

- For general $n > 1$, an efficient LFD-solving algorithm is also available based on the simplification of $\eta(\lambda, u)$.

Detection Procedure: DR-CuSum based on LFDs

[Molloy and Ford, 2017] Suppose $(\{Q\}, \mathcal{P})$ is weakly stochastically bounded by (Q, P^*). Then the CuSum test based on (Q, P^*) with threshold $b = |\log \gamma|$ is first-order asymptotically minimax optimal.

DR-CuSum is the **CuSum** test based on LFD P^*

Detection Procedure: DR-CuSum based on LFDs

[Molloy and Ford, 2017] Suppose $(\{Q\}, \mathcal{P})$ is weakly stochastically bounded by (Q, P^*). Then the CuSum test based on (Q, P^*) with threshold $b = |\log \gamma|$ is first-order asymptotically minimax optimal.

DR-CuSum is the **CuSum** test based on LFD P^*

- Recursive implementation of DR-CuSum statistic:

$$S_k = (S_{k-1})^+ + \log \frac{p^*(X_k)}{q(X_k)}, \quad S_0 = 0$$

LFD is exponential tilting:

$$\log \frac{p^*(X_k)}{q(X_k)} = -C_{\lambda^*, u^*}(X_k) - \log \eta(\lambda^*, u^*)$$

Detection Procedure: DR-CuSum based on LFDs

[Molloy and Ford, 2017] Suppose $(\{Q\}, \mathcal{P})$ is weakly stochastically bounded by (Q, P^*). Then the CuSum test based on (Q, P^*) with threshold $b = |\log \gamma|$ is first-order asymptotically minimax optimal.

DR-CuSum is the **CuSum** test based on LFD P^*

- Recursive implementation of DR-CuSum statistic:

$$S_k = (S_{k-1})^+ + \log \frac{p^*(X_k)}{q(X_k)}, \quad S_0 = 0$$

LFD is exponential tilting:

$$\log \frac{p^*(X_k)}{q(X_k)} = -C_{\lambda^*, u^*}(X_k) - \log \eta(\lambda^*, u^*)$$

- Stopping rule:

$$\tau_{\mathsf{DR}} = \inf \{k \in \mathbb{N} : S_k \geq b\},$$

Extension: Multiple Uncertainty Sets

- M post-change scenarios. For each $m = 1, 2, \ldots, M$:

Extension: Multiple Uncertainty Sets

- M post-change scenarios. For each $m = 1, 2, \ldots, M$:
  - Training samples $\{\omega_1^{(m)}, \ldots, \omega_{n_m}^{(m)}\}$ and empirical distribution $\widehat{P}_{n_m}^{(m)}$

Extension: Multiple Uncertainty Sets

- M post-change scenarios. For each $m = 1, 2, \ldots, M$:
  - Training samples $\{\omega_1^{(m)}, \ldots, \omega_{n_m}^{(m)}\}$ and empirical distribution $\widehat{P}_{n_m}^{(m)}$
  - Uncertainty set $\mathcal{P}_m := \{P \in \mathscr{P} : \mathsf{W}_s(P, \widehat{P}_n^{(m)}) \leq r_s^{(m)}\}$.

Extension: Multiple Uncertainty Sets

- M post-change scenarios. For each $m = 1, 2, \ldots, M$:
  - Training samples $\{\omega_1^{(m)}, \ldots, \omega_{n_m}^{(m)}\}$ and empirical distribution $\widehat{P}_{n_m}^{(m)}$
  - Uncertainty set $\mathcal{P}_m := \{P \in \mathscr{P} : \mathsf{W}_s(P, \widehat{P}_n^{(m)}) \leq r_s^{(m)}\}$.
  - Find LFD similarly:
    $p_m^*(x) = q(x) \exp\big\{ - C_{\lambda_m^*, u_m^*}^{(m)}(x) - \eta^{(m)}(\lambda_m^*, u_m^*)\big\}$

Extension: Multiple Uncertainty Sets

- M post-change scenarios. For each $m = 1, 2, \ldots, M$:
  - Training samples $\{\omega_1^{(m)}, \ldots, \omega_{n_m}^{(m)}\}$ and empirical distribution $\widehat{P}_{n_m}^{(m)}$
  - Uncertainty set $\mathcal{P}_m := \{P \in \mathscr{P} : \mathsf{W}_s(P, \widehat{P}_n^{(m)}) \leq r_s^{(m)}\}$.
  - Find LFD similarly:
    $p_m^*(x) = q(x) \exp\big\{ - C_{\lambda_m^*, u_m^*}^{(m)}(x) - \eta^{(m)}(\lambda_m^*, u_m^*)\big\}$
  - Individual CuSum Statistics:

$$S_k^{(m)} = (S_{k-1}^{(m)})^+ + \log \frac{p_m^*(X_k)}{q(X_k)}, \quad S_0^{(m)} = 0.$$

Extension: Multiple Uncertainty Sets

- M post-change scenarios. For each $m = 1, 2, \ldots, M$:
  - Training samples $\{\omega_1^{(m)}, \ldots, \omega_{n_m}^{(m)}\}$ and empirical distribution $\widehat{P}_{n_m}^{(m)}$
  - Uncertainty set $\mathcal{P}_m := \{P \in \mathscr{P} : \mathsf{W}_s(P, \widehat{P}_n^{(m)}) \leq r_s^{(m)}\}$.
  - Find LFD similarly:
    $p_m^*(x) = q(x) \exp\big\{ - C_{\lambda_m^*, u_m^*}^{(m)}(x) - \eta^{(m)}(\lambda_m^*, u_m^*)\big\}$
  - Individual CuSum Statistics:

$$S_k^{(m)} = (S_{k-1}^{(m)})^+ + \log \frac{p_m^*(X_k)}{q(X_k)}, \quad S_0^{(m)} = 0.$$

- Denote $\mathcal{P} := \cup_{m=1}^M \mathcal{P}_{n_m}^{(m)}$ as the union of all the uncertainty sets.

Extension: Multiple Uncertainty Sets

- M post-change scenarios. For each $m = 1, 2, \ldots, M$:
  - Training samples $\{\omega_1^{(m)}, \ldots, \omega_{n_m}^{(m)}\}$ and empirical distribution $\widehat{P}_{n_m}^{(m)}$
  - Uncertainty set $\mathcal{P}_m := \{P \in \mathscr{P} : \mathsf{W}_s(P, \widehat{P}_n^{(m)}) \leq r_s^{(m)}\}$.
  - Find LFD similarly:
    $p_m^*(x) = q(x) \exp\left\{ -C_{\lambda_m^*, u_m^*}^{(m)}(x) - \eta^{(m)}(\lambda_m^*, u_m^*)\right\}$
  - Individual CuSum Statistics:

$$S_k^{(m)} = (S_{k-1}^{(m)})^+ + \log \frac{p_m^*(X_k)}{q(X_k)}, \quad S_0^{(m)} = 0.$$

- Denote $\mathcal{P} := \cup_{m=1}^{M} \mathcal{P}_{n_m}^{(m)}$ as the union of all the uncertainty sets.

- Overall Stopping Rule:

$$\tau_b^* := \inf\left\{k \in \mathbb{N} : \max_{m=1,\ldots,M} S_k^{(m)} \geq b\right\},$$

Asymptotic Optimality

- Denote the smallest KL divergence

$$I^* = \min_{i=1,2,\ldots,M} \mathsf{KL}(P_i^* || Q).$$

- The test τ_b^* with threshold $b = \log(M\gamma)$ solves the problem $\inf_{\tau \in C(\gamma)} \sup_{P \in \mathcal{P}} \mathsf{WADD}^P(\tau)$ asymptotically as $\gamma \to \infty$, with worst-case delay

$$\begin{aligned} \sup_{P \in \mathcal{P}} \mathsf{WADD}^P(\tau_b^*) &= \inf_{\tau' \in C(\gamma)} \sup_{P \in \mathcal{P}} \mathsf{WADD}^P(\tau') \cdot (1 + o(1)) \\ &= \frac{\log \gamma}{I^*} \cdot (1 + o(1)). \end{aligned}$$

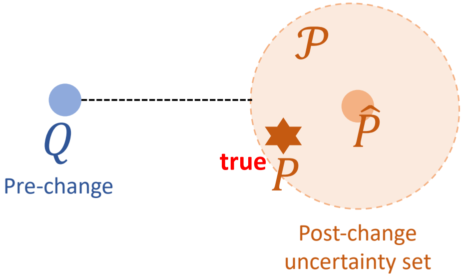
Choice of Radius

Radius: tradeoff between robustness and detection performance.

Choice of Radius

Radius: tradeoff between robustness and detection performance.

Guidline:

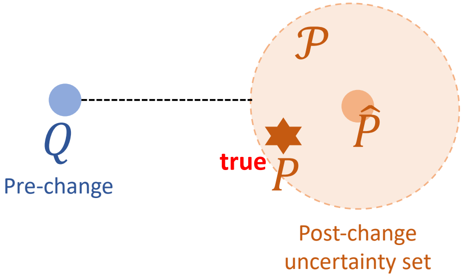


Choice of Radius

Radius: tradeoff between robustness and detection performance.

Guidline:

C1 The pre-change distribution Q is excluded from $\mathcal{P}$ (w.h.p)



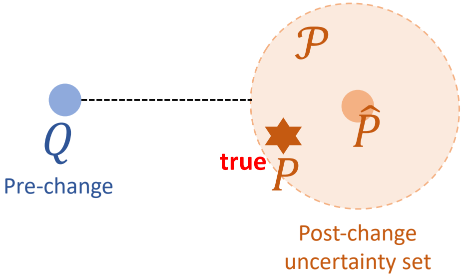
Choice of Radius

Radius: tradeoff between robustness and detection performance.

Guidline:

C1 The pre-change distribution Q is excluded from $\mathcal{P}$ (w.h.p)

C2 The true post-change distribution is included in $\mathcal{P}$ (w.h.p)



Choice of Radius (continued)

Assumptions:

- Empirical samples $\{\omega_1, \ldots, \omega_n\} \overset{iid}{\sim} P$ (true post-change)
- P satisfies $T_s(c)$ transportation-cost inequality:
  $\mathsf{W}_s(P, Q) \leq \sqrt{2c\mathsf{KL}(Q||P)}, \forall Q \ll P$.

Choice of Radius (continued)

Assumptions:

- Empirical samples $\{\omega_1, \ldots, \omega_n\} \overset{iid}{\sim} P$ (true post-change)
- P satisfies $T_s(c)$ transportation-cost inequality: $\mathsf{W}_s(P, Q) \leq \sqrt{2c\mathsf{KL}(Q||P)}, \forall Q \ll P$.
  - Then we have the concentration [Bolley, Guillin, and Villani, 2007]:

$$\mathbb{P}^{\Omega}\{\mathsf{W}_s(P, \widehat{P}_n) > \epsilon\} \leq e^{-\gamma_s n\epsilon^2/2c'}.$$

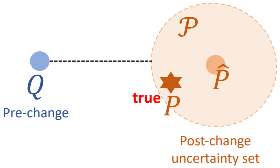
Choice of Radius (continued)

Assumptions:

- Empirical samples $\{\omega_1, \dots, \omega_n\} \overset{iid}{\sim} P$ (true post-change)
- P satisfies $T_s(c)$ transportation-cost inequality:
  $\mathsf{W}_s(P, Q) \leq \sqrt{2c\mathsf{KL}(Q||P)}, \forall Q \ll P$.
    - Then we have the concentration [Bolley, Guillin, and Villani, 2007]:

$$\mathbb{P}^{\Omega}\{\mathsf{W}_s(P, \widehat{P}_n) > \epsilon\} \leq e^{-\gamma_s n\epsilon^2/2c'}.$$

Fix $\delta \in (0, 1)$ and $s \in [1, 2]$.



Choice of Radius (continued)

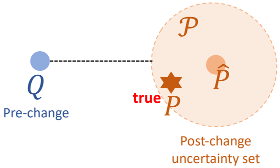
Assumptions:

- Empirical samples $\{\omega_1, \ldots, \omega_n\} \overset{iid}{\sim} P$ (true post-change)
- P satisfies $T_s(c)$ transportation-cost inequality:
  $\mathsf{W}_s(P, Q) \le \sqrt{2c\mathsf{KL}(Q||P)}, \forall Q \ll P$.
  - Then we have the concentration [Bolley, Guillin, and Villani, 2007]:

$$\mathbb{P}^{\Omega}\{\mathsf{W}_s(P, \widehat{P}_n) > \epsilon\} \le e^{-\gamma_s n \epsilon^2 / 2c'}.$$

Fix $\delta \in (0, 1)$ and $s \in [1, 2]$.
C1: W. p. at least $1 - \delta$ we have $Q \notin \mathcal{P}$, if

$$r_s \le \overline{r}_{\delta,n} := \mathsf{W}_s(P, Q) - \sqrt{\frac{2|\log \delta| c}{\gamma_s n}}$$



Choice of Radius (continued)

Assumptions:

- Empirical samples $\{\omega_1, \ldots, \omega_n\} \overset{iid}{\sim} P$ (true post-change)
- P satisfies $T_s(c)$ transportation-cost inequality:
  $\mathsf{W}_s(P, Q) \leq \sqrt{2c\mathsf{KL}(Q||P)}, \forall Q \ll P$.
  - Then we have the concentration [Bolley, Guillin, and Villani, 2007]:

$$\mathbb{P}^{\Omega}\{\mathsf{W}_s(P, \widehat{P}_n) > \epsilon\} \leq e^{-\gamma_s n \epsilon^2 / 2c'}.$$

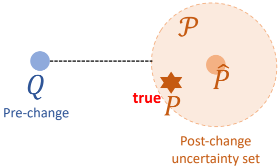
Fix $\delta \in (0,1)$ and $s \in [1,2]$.

C1: W. p. at least $1-\delta$ we have $Q \notin \mathcal{P}$, if

$$r_s \leq \overline{r}_{\delta,n} := \mathsf{W}_s(P, Q) - \sqrt{\frac{2|\log\delta|c}{\gamma_s n}}$$

C2: W. p. at least $1-\delta$ we have $P \in \mathcal{P}$, if

$$r_s \geq \underline{r}_{\delta,n} := \sqrt{\frac{2|\log\delta|c}{\gamma_s n}}.$$

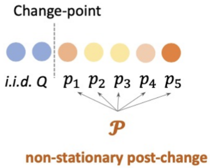


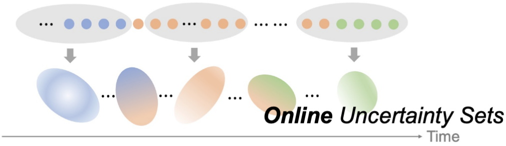
$$\underline{r}_{\delta,n} \leq r_s \leq \overline{r}_{\delta,n}$$

Outline

- Motivation: Data-Driven Robust Change Detection
- Proposed: Distributionally Robust CuSum Test (DR-CuSum)
- Theoretical Results: Asymptotic Optimality
- Extension to Online Distributional Uncertainties
- Simulation Examples

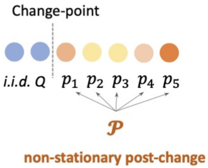
Sequential Uncertainty Sets

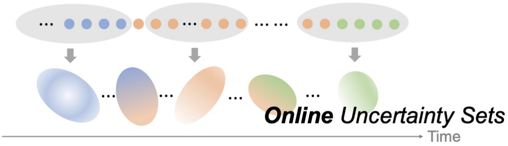


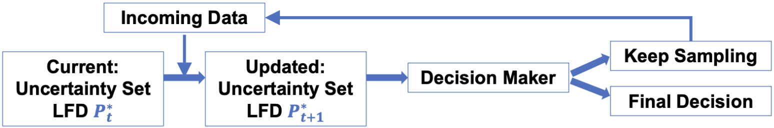


Yang, X, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2024.

Sequential Uncertainty Sets



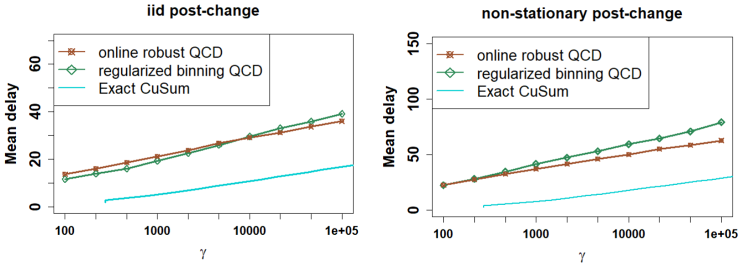




Yang, X, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2024.

Sequential Uncertainty Sets

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



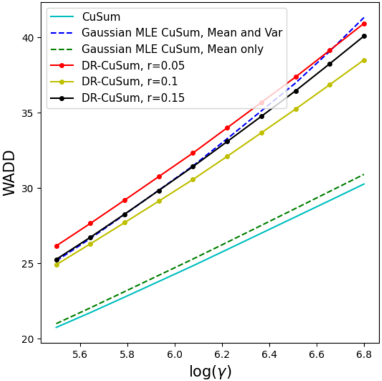
Average Delay vs. Average Run Length (γ) for i.i.d. (Left) and non-stationary (Right) post-change distributions. The window size is $w = 10$ in both methods.

Outline

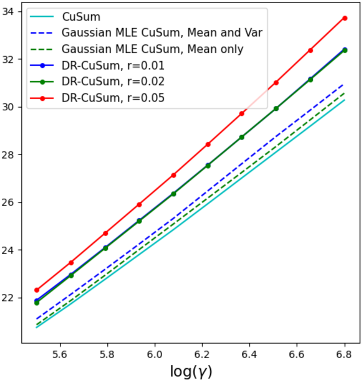
- Motivation: Data-Driven Robust Change Detection
- Proposed: Distributionally Robust CuSum Test (DR-CuSum)
- Theoretical Results: Asymptotic Optimality
- Extension to Online Distributional Uncertainties
- Simulation Examples

Simulation Results

- Exact-CuSum
- CuSum with Gaussian MLE mean
- CuSum with Gaussian MLE mean and variance;
- DR-CuSum with various radii.



$n = 25;$



$n = 150$

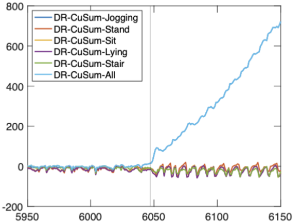
Averaged over 30 sets of different training samples

A Real Data Example

WISDM Dataset

Much Smaller Delay

	DR-CuSum	NGLR-CuSum
$n = 10,\ r = 2$	25.61 (4.08)	81.26 (0.51)
$n = 20,\ r = 1.5$	16.88 (2.74)	74.77 (13.62)



True post-change: Jogging

Conclusion

Summary:

- Data-Driven Minimax Robust QCD test under Wasserstein uncertainty sets
- LFD enjoys exponential tilting form and can be efficiently calculated via Convex Optimization
- Both offline and online uncertainty sets are studied

X., Liang, and Veeravalli (2024) "Distributionally Robust Quickest Change Detection using Wasserstein Uncertainty Sets," International Conference on Artificial Intelligence and Statistics (AISTATS).

Conclusion

Summary:

- Data-Driven Minimax Robust QCD test under Wasserstein uncertainty sets
- LFD enjoys exponential tilting form and can be efficiently calculated via Convex Optimization
- Both offline and online uncertainty sets are studied

Future Directions:

X., Liang, and Veeravalli (2024) "Distributionally Robust Quickest Change Detection using Wasserstein Uncertainty Sets," International Conference on Artificial Intelligence and Statistics (AISTATS).

Conclusion

Summary:

- Data-Driven Minimax Robust QCD test under Wasserstein uncertainty sets
- LFD enjoys exponential tilting form and can be efficiently calculated via Convex Optimization
- Both offline and online uncertainty sets are studied

Future Directions:

- Theoretical guarantee under non-stationary post-change distributions

X., Liang, and Veeravalli (2024) "Distributionally Robust Quickest Change Detection using Wasserstein Uncertainty Sets," International Conference on Artificial Intelligence and Statistics (AISTATS).

Conclusion

Summary:

- Data-Driven Minimax Robust QCD test under Wasserstein uncertainty sets
- LFD enjoys exponential tilting form and can be efficiently calculated via Convex Optimization
- Both offline and online uncertainty sets are studied

Future Directions:

- Theoretical guarantee under non-stationary post-change distributions
- Change Diagnosis under multiple post-change scenarios

X., Liang, and Veeravalli (2024) "Distributionally Robust Quickest Change Detection using Wasserstein Uncertainty Sets," International Conference on Artificial Intelligence and Statistics (AISTATS).

Conclusion

Summary:

- Data-Driven Minimax Robust QCD test under Wasserstein uncertainty sets
- LFD enjoys exponential tilting form and can be efficiently calculated via Convex Optimization
- Both offline and online uncertainty sets are studied

Future Directions:

- Theoretical guarantee under non-stationary post-change distributions
- Change Diagnosis under multiple post-change scenarios
- More general uncertainty sets

X., Liang, and Veeravalli (2024) "Distributionally Robust Quickest Change Detection using Wasserstein Uncertainty Sets," International Conference on Artificial Intelligence and Statistics (AISTATS).

Backup: Derivation of LFD

- Consider $s = 1$.
- $\Pi(P, \widehat{P})$: set of joint distributions with marginals P and $\widehat{P}$; each one can be characterized by the mixed joint density:

$$\pi(x, \omega_i) = \frac{1}{n} f_i(x), \text{ where } f_i(x) \geq 0, \int f_i d\mu(x) = 1, \ \forall i.$$

  Let $p(x) := \sum_{i=1}^n \pi(x, \omega_i) = \frac{1}{n} \sum_{i=1}^n f_i(x)$.
- By Lagrange duality:

$$\begin{aligned} &\inf \left\{ \mathsf{KL}(P||Q) : P \in \mathcal{P}_\mu, \mathsf{W}_1(P, \widehat{P}) \leq r \right\} \\ &= \max_{\lambda \geq 0} \inf_{P \in \mathcal{P}_\mu} \left(\mathsf{KL}(P||Q) + \lambda \mathsf{W}_1(P, \widehat{P}) - \lambda r \right), \end{aligned}$$

- Replace Wasserstein distance as $\mathsf{W}_1(P, \widehat{P}) = \inf_{\pi \in \Pi(P, \widehat{P})} \sum_{i=1}^n \int c(x, \omega_i) \pi(x, \omega_i) d\mu(x)$, yields

$$\max_{\lambda \geq 0} -\lambda r + \inf_{\pi} \int \left(\Big(\sum_{i=1}^n \pi(x, \omega_i) \Big) \log \frac{\sum_{i=1}^n \pi(x, \omega_i)}{q(x)} + \sum_{i=1}^n \lambda c(x, \omega_i) \pi(x, \omega_i) \right) d\mu(x).$$

Derivation of LFD (continued)

- The Lagrangian dual of the inner inf problem can be solved with

$$\frac{1}{n}\sum_{i=1}^{n} f_i^*(x) = q(x)e^{-\min_i(\lambda c(x,\omega_i)-u_i)-\check{\eta}-1},$$

  with $\check{\eta}+1 = \log\left(\int_{\mathscr{X}} q(x)e^{-\min_i(\lambda c(\omega,\omega_i)-u_i)}d\mu(x)\right) = \log\eta(\lambda,u)$.
  This gives the optimization problem that yields λ^*, u^*.
- After solving λ^*, u^*, we arrive at LFD solution

$$p^*(x) = q(x)e^{-C_{\lambda^*,u^*}(x)-\eta(\lambda^*,u^*)}.$$

- For general $s \geq 1$, repeat above process using $c^s(x,\omega_i)$.