Collaborative likelihood-ratio estimation over graphs

Alejandro de la Concha, Nicolas Vayatis, Argyris Kalogeratos

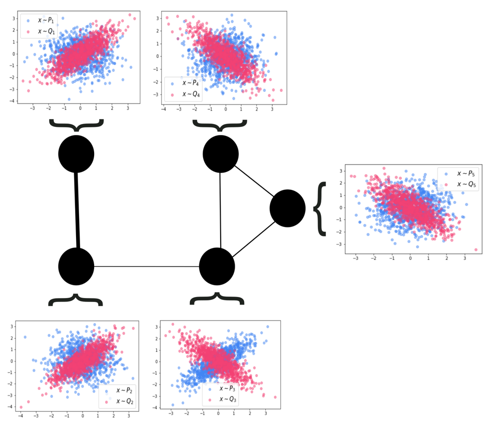
Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, France

Department of Mathematics, University of Luxembourg, 4364 Esch-sur-Alzette, Luxembourg

1. Motivation

Consider a feature space $\mathcal{X} \subset \mathbb{R}^d$ and N data sources, each associated with the nodes of a graph $G = (V, E, W)$. Each node v has access to two i.i.d datasets each of them from two unknown pdfs, p_v and q_v:

$$\begin{cases} \mathbf{X} = \{\mathbf{X}_v\}_{v\in V} = \{\{x_{v,i} : x_{v,i} \sim p_v\}_{i=1}^{n_v}\}_{v\in V} \\ \mathbf{X}' = \{\mathbf{X}'_v\}_{v\in V} = \{\{x'_{v,i} : x'_{v,i} \sim q_v\}_{i=1}^{n'_v}\}_{v\in V} \end{cases}$$



Goal: For each node $v \in V$, compute the **likelihood-ratio** $r_v(x) = \frac{q_v(x)}{p_v(x)}$ between the two density functions.

Applications of likelihood-ratio estimation (LRE): *Hypothesis Testing* (Neyman-Pearson Lemma), *Sequential Change-point detection*, *Transfer Learning* (*Importance sampling*), ...

Questions:

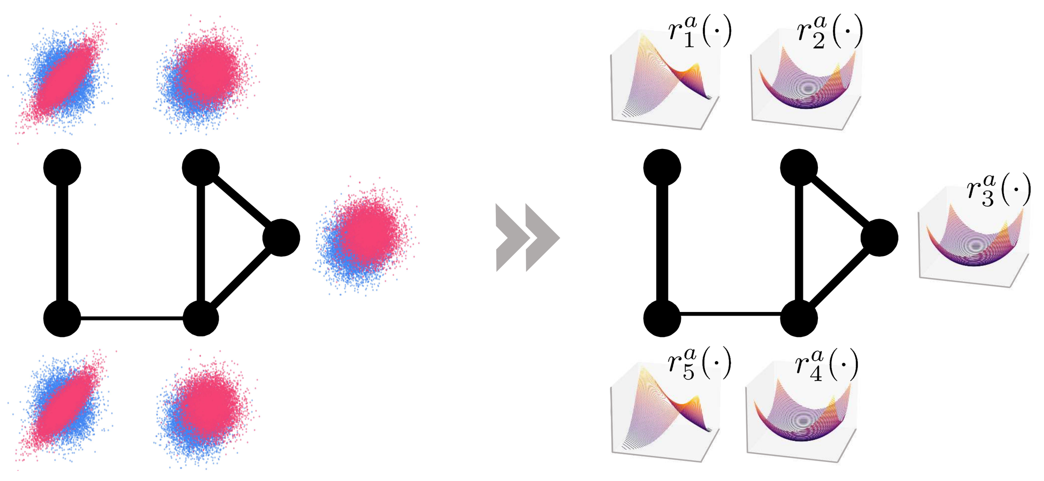
1. LRE techniques are only used on single-source or aggregated data. How can we extend LRE to complex systems such as sensors networks, transport networks, public health surveillance, etc?
2. Can the network topology be exploited to achieve faster convergence rates?

2. Building Blocks

1. **Under what conditions is LRE well defined?**
   The likelihood-ratio r_v exists iff q_v is absolutely continuous with respect to p_v. However, the **relative** likelihood-ratio always exists:
   $$r_v^\alpha(x) = \frac{q_v(x)}{(1-\alpha)p_v(x) + \alpha q_v(x)} = \frac{q_v(x)}{p_v^\alpha(x)}$$
   where $\alpha \in (0,1)$ and $\|r_v^\alpha\|_\infty \leq \frac{1}{\alpha}$.
2. **Non-parametric LRE: How can we infer the likelihood-ratio without restrictive assumptions on p_v or q_v?**
   Infer the node-level relative likelihood-ratios $r_v^\alpha(\cdot)$ via the variational formulation of the Pearson's divergence:
   $$\begin{aligned} PE(p_v^\alpha \| q_) &= \int \frac{(r_v^\alpha(x)-1)^2}{2} p_v^\alpha(x)dx \\ &\geq \sup_{f\in} \int f(x)q_v(x)dx \\ &\quad - \int \frac{f^2(x)}{2} p_v^\alpha(x)dx - \frac{1}{2} \end{aligned}$$
3. **How to choose to enhance estimation and theoretical analysis?**
   - We choose $\mathbb{H}$ to be an RKHS with kernel function $K(\cdot,\cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.
   - $\mathbb{H}$ satisfies the reproducing property, that is $\forall x \in \mathcal{X}$ and $f \in \mathbb{H}$: $f(x) = \langle f(\cdot), K(x,\cdot)\rangle_{\mathbb{H}}$.
   - $\phi(x)$ denotes the associated feature map.
4. **How can we incorporate the network topology?**
   Graph Smoothness.
   $$\|\mathbf{f}\|_{\mathbb{G}}^2 = \frac{1}{2}\sum_{u,v\in V} W_{uv}\|f_u - f_v\|_{\mathbb{H}}^2 + \gamma \sum_{v\in V}\|f_v\|_{\mathbb{H}}^2$$

3. GRULSIF: Graph-based Relative Unconstrained Least-Squares Importance Fitting [1]

From node-level datasets to node-level likelihood-ratio functions



Multitasking formulation of LRE

$$\begin{aligned} &\min_{\Theta\in\mathbb{R}^{NL}} \frac{1}{N}\sum_{v\in V} \overbrace{\left((1-\alpha)\frac{\theta_v^\top H_v\theta_v}{2} + \alpha\frac{\theta_v^\top H'_v\theta_v}{2} - h'_v\theta_v\right)}^{\text{PE-divergence node-level}} \\ &+ \overbrace{\frac{\lambda\gamma}{2}\sum_{v\in V}\|\theta_v\|^2}^{\text{node-level regularization term}} + \overbrace{\frac{\lambda}{4}\sum_{u,v\in V} W_{uv}\|\theta_u - \theta_v\|^2}^{\text{graph-level regularization term}} \end{aligned} \tag{1}$$

where:

$$H_v = \frac{1}{n_v}\sum_{x\in\mathbf{X}_v}\phi(x)\phi(x)^\top,\ H'_v = \frac{1}{n'_v}\sum_{x\in\mathbf{X}'_v}\phi(x)\phi(x)^\top,\ h'_v = \frac{1}{n'_v}\sum_{x\in\mathbf{X}'_v}\phi(x).$$

By the reproducing property of $\mathbb{H}$,$\forall v \in V$:

$$r_v^\alpha(x) \approx f_v(x) = \sum_{l=1}^{L}\theta_{v,l}K(x, x_l)$$

Numerical implementation:

$$\begin{aligned} \hat{\theta}_v^{(i)} = (\lambda\gamma + \eta_{vL})^{-1}\Bigg[&\overbrace{\eta_v\hat{\theta}_v^{(i-1)} - \left[\left(\frac{1-\alpha}{N}H_v + \frac{\alpha}{N}H'_v\right)\hat{\theta}_v^{(i-1)} - \frac{h'_v}{N}\right]}^{\text{component depending on node } v} \\ &- \overbrace{\lambda\left(d_v\hat{\theta}_v^{(i-1)} - \sum_{(u,v)\in E} W_{uv}\left(1\{u<v\}\hat{\theta}_u^{(i)} + 1\{u \geq v\}\hat{\theta}_u^{(i-1)}\right)\right)}^{\text{component depending on the graph}}\Bigg] \end{aligned}$$

Cyclic Block Coordinate Gradient Descent and Nyström approximation: $O(N\hat{L}^3 + L\hat{L}^2 + N\hat{L}^2\log^2(N\hat{L}))$, where $\hat{L} \ll L$

4. Theoretical Guarantees

Theorem: Convergence of Collaborative LRE. Consider the case where $n_v = n'_v = n$, the model is well specified in the sense that there exists $\Lambda^2 > 0$ such that $\|\mathbf{f}\|_{\mathbb{G}}^2 \leq \Lambda^2$, and all observations are independent. Then, for any $\delta \in (0,1)$, with probability at least $1-\delta$, the solution to Problem 1, $\hat{\mathbf{f}} = (\hat{f}_1, ..., \hat{f}_N)$, satisfies:

$$\begin{aligned} \frac{1}{N}\sum_{v\in V}\left[PE(p_v^\alpha\|q_v) - PE_v^\alpha(\hat{f}_v)\right] &\leq C_1\rho^* + \frac{C_2}{nN}\log\left(\frac{1}{\delta}\right) \\ \frac{1}{N}\sum_{v\in V}\mathbb{E}_{p^\alpha(x)}[(\hat{f}_v(x) - r_v^\alpha(x))^2] &\leq C'_1\rho^* + \frac{C'_2}{nN}\log\left(\frac{1}{\delta}\right), \end{aligned}$$

where:

$$\zeta^* = \min_{v\in V}\zeta_v\ (\text{recall } \zeta_v > 1, \forall v \in V),\quad s_{\max} = \max_{v\in V} s_v,$$

$$\rho^* \leq 8B_0\sqrt{\frac{\zeta^*+1}{\zeta^*-1}}\left(\mathcal{T}_{\mathsf{G}}\Lambda^2(b + C_a)^{2\zeta^*}\right)^{\frac{1}{1+\zeta^*}} n^{-\frac{\zeta^*}{1+\zeta^*}} N^{-\frac{1}{1+\zeta^*}} s_{\max}^{\frac{1}{1+\zeta^*}},$$

where:

- N:= number of nodes.
- n:= number of observations per node.
- $\mathcal{T}_{\mathsf{G}}$:= encodes graph topology.
- Λ:= encodes the graph smoothness of the vector-valued function $(r_1^\alpha, \ldots, r_N^\alpha)$.
- $\zeta^*, s_{\max}$:= encodes the capacity of the RKHS to approximate each r_1^α.

Take-home messages:

- **Slow eigenvalue decay:** $\zeta^* \approx 1$**:** Collaboration is relevant.
- **Fast eigenvalue decay:** $\zeta^* \to \infty$**:** Simpler approximation problem. Collaboration may not be relevant.
- **Intermediate values:** $\rho^* = O(n^{\frac{-\zeta^*}{1+\zeta^*}}\Lambda^{\frac{2}{1+\zeta^*}}N^{-\frac{1}{1+\zeta^*}})$.

If the space is complex enough so that collaboration matters:

- Λ **small:** implies faster convergence.
- **More nodes:** beneficial if nodes are similar $(O(\Lambda^{\frac{2}{1+\zeta^*}}N^{-\frac{1}{1+\zeta^*}}))$.

5. Examples on Seismic Data [2]

Application: Collaborative Two-Sample Test. Consider a two-sample test at each node $v \in V$:

$$\{\mathrm{H}_{\text{null},v}:\ p_v = q_v \quad \text{vs} \quad \mathrm{H}_{\text{alt},v}:\ p_v \neq q_v\}_{v\in V}$$

which are jointly tested to determine the set:

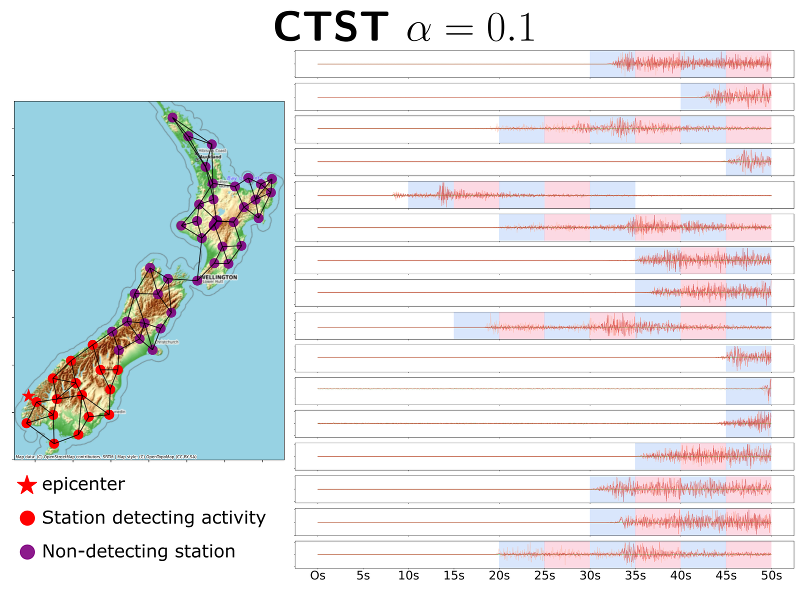
$$R_{\text{MT}} = \{v \in V \,|\, \mathrm{H}_{\text{null},v} \text{ is rejected}\}$$

Goal: Detect the stations and periods during which the stations sense the event.

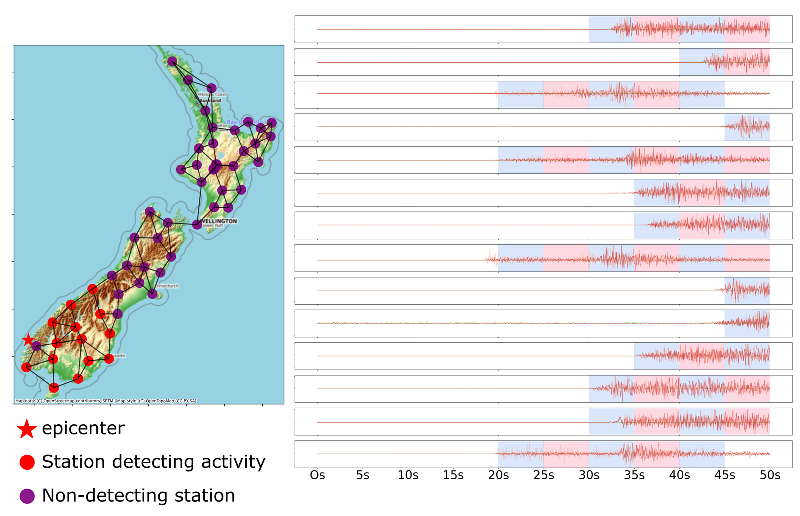
The data. The stations are equipped with accelerometers that provide 3D signals corresponding to three perpendicular directions ($\mathcal{X} \in \mathbb{R}^3$). We compare the waveforms from 50 seconds before to 50 seconds after the event, at a sampling frequency of 100 Hz.

Graph structure. A multiplex graph representing spatial and temporal similarity: $G_{S,T} = (V, E, W)$. The spatial component is built via 3-nearest-neighbors graph based on geographical location. Time is encoded as connections between copies of the same node observed at different time windows.

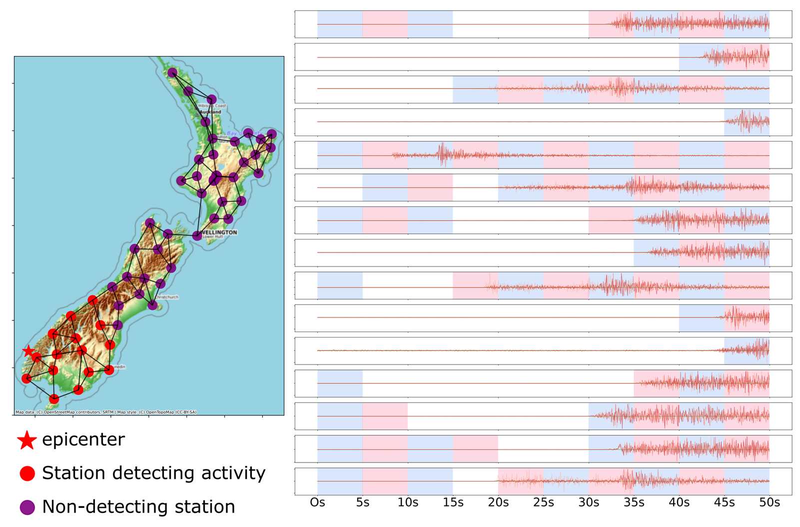
CTST $\alpha = 0.1$



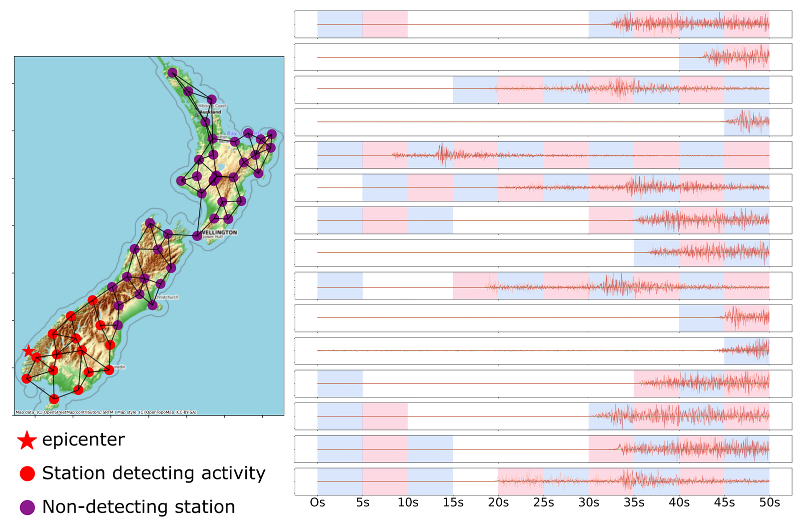
POOL $\alpha = 0.1$



RULSIF $\alpha = 0.1$



MMD



6. Conclusions

1. A novel non-parametric framework for multiple likelihood-ratio estimation, with a detailed and efficient implementation that scales to large graphs.
2. An approach that enhances multiple hypothesis testing when node-level data are multivariate, the differences between the compared distributions are unknown, and some heterogeneity among node-level tests is allowed.

References

[1] A. de la Concha, N. Vayatis, and A. Kalogeratos. Collaborative likelihood-ratio estimation over graphs. *Journal of Machine Learning Research*, 26(259):1–66, 2025.

[2] A. de la Concha, N. Vayatis, and A. Kalogeratos. Collaborative non-parametric two-sample testing. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2025.