Sequential Nonparametric Detection of Anomalous Data Streams

Srikrishna Bhashyam

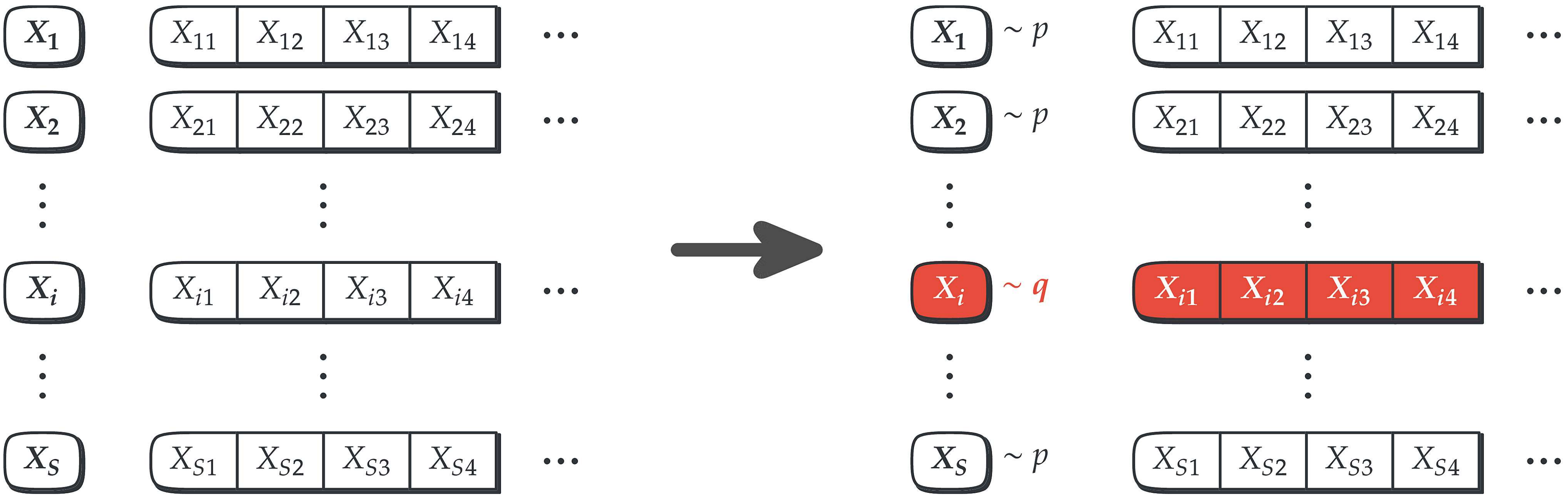
IIT Madras

Joint work with Sreeram C. Sreenivasan

December 2, 2023

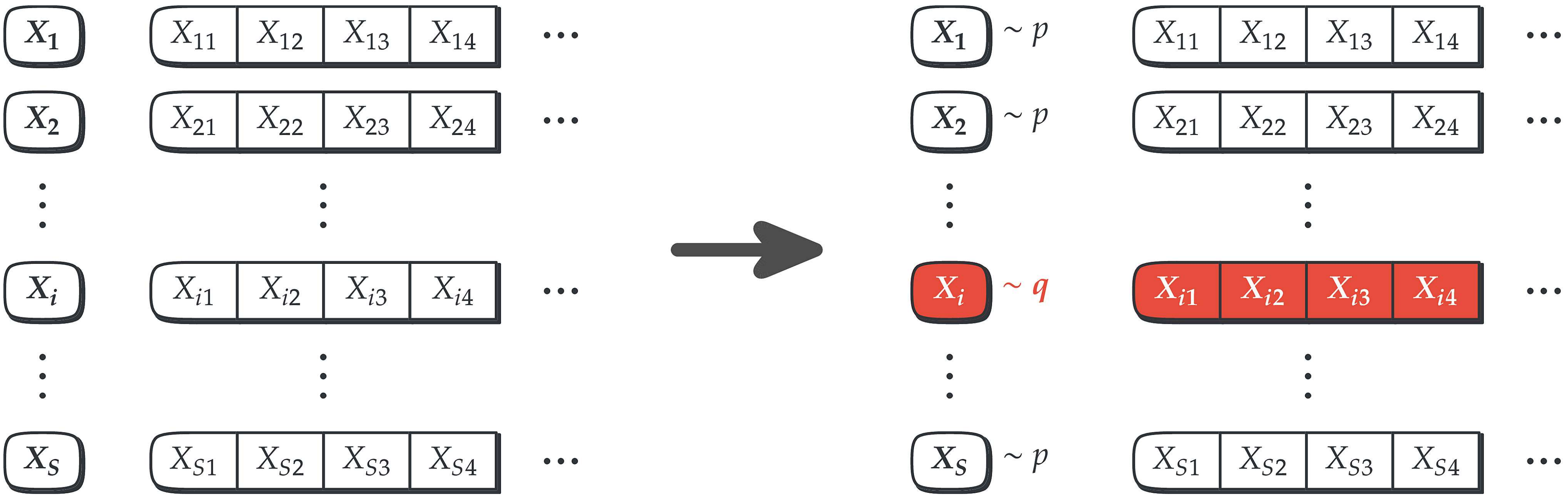
IISc Bangalore

Detection of Anomalous Data Streams


$\boldsymbol{X}_1$
X_{11} X_{12} X_{13} X_{14} ⋯
$\boldsymbol{X}_2$
X_{21} X_{22} X_{23} X_{24} ⋯
$\boldsymbol{X}_i$
X_{i1} X_{i2} X_{i3} X_{i4} ⋯
$\boldsymbol{X}_S$
X_{S1} X_{S2} X_{S3} X_{S4} ⋯
$\boldsymbol{X}_1 \sim p$
$\boldsymbol{X}_2 \sim p$
$\boldsymbol{X}_i \sim q$
$\boldsymbol{X}_S \sim p$

- Each data stream independent and identically distributed (i.i.d.) samples from an unknown distribution
- Typical vs Anomalous: Unknown number of anomalous streams
- Applications: Sensor networks, network monitoring

Hypothesis Testing Setting


$\boldsymbol{X}_1$
X_{11} X_{12} X_{13} X_{14} ⋯
$\boldsymbol{X}_2$
X_{21} X_{22} X_{23} X_{24} ⋯
$\boldsymbol{X}_i$
X_{i1} X_{i2} X_{i3} X_{i4} ⋯
$\boldsymbol{X}_S$
X_{S1} X_{S2} X_{S3} X_{S4} ⋯
$\boldsymbol{X}_1 \sim p$
$\boldsymbol{X}_2 \sim p$
$\boldsymbol{X}_i \sim q$
$\boldsymbol{X}_S \sim p$

Example

- One anomalous stream only: *S* hypotheses
- Hypothesis *i*: The *i* th stream is anomalous

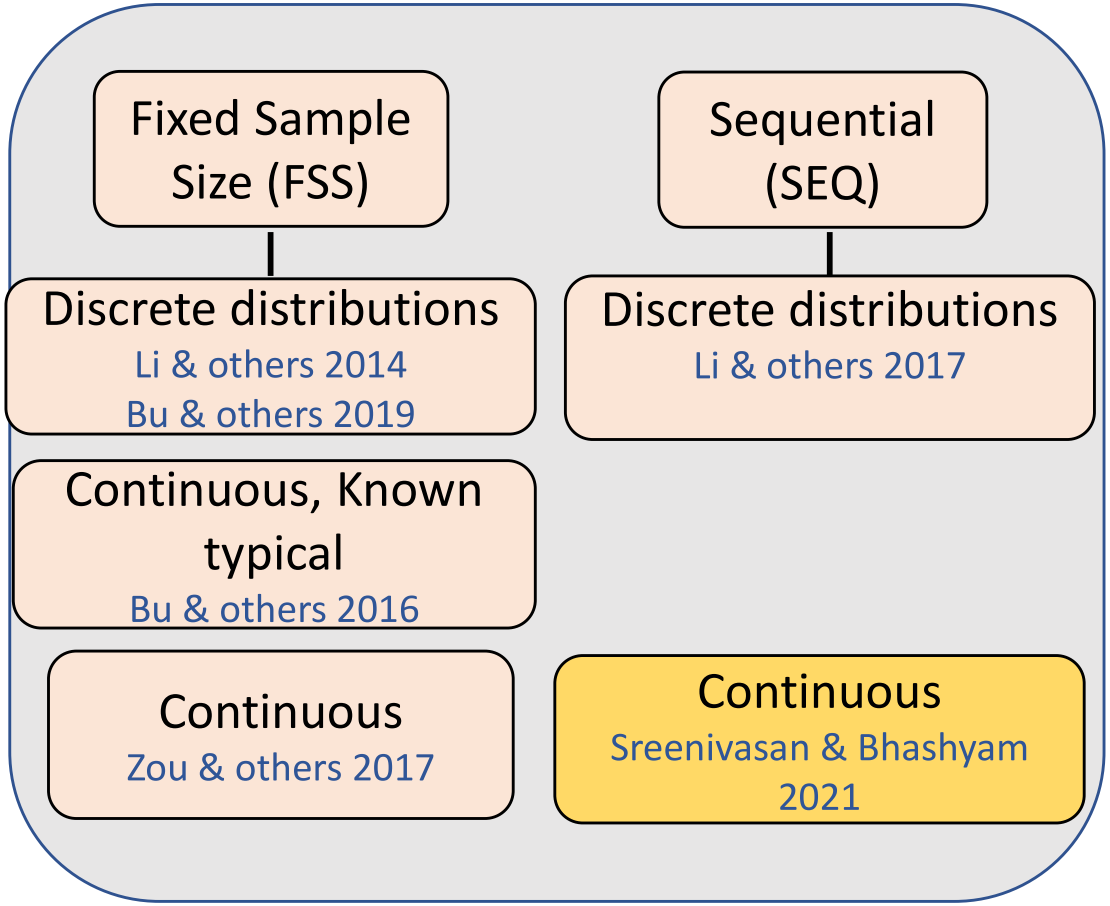
Settings: Fixed sample size (FSS)/Sequential (SEQ), Unknown distributions

Sequential Nonparametric Hypothesis Testing

- Observations arrive sequentially
- One new sample observed in each stream at each time
- Sequential decision rule consists of:
  - A stopping rule (whether or stop or continue sampling)
  - A decision (if stopping, what is the decision)

- Nonparametric: Unknown distributions *p* and *q*
  - $p \neq q$
  - Also called Universal or Distribution-free tests

Closely Related Work: Outlying Sequence Detection



Fixed Sample Size (FSS)

Discrete distributions
Li & others 2014
Bu & others 2019

Continuous, Known typical
Bu & others 2016

Continuous
Zou & others 2017

Sequential (SEQ)

Discrete distributions
Li & others 2017

Continuous
Sreenivasan & Bhashyam 2021

Y. Li, S. Nitinawarat and V. V. Veeravalli, "Universal Outlier Hypothesis Testing," in IEEE Trans. on Information Theory, vol. 60, no. 7, pp. 4066-4082, July 2014.

Y. Bu, S. Zou, Y. Liang and V. V. Veeravalli, "Universal outlying sequence detection for continuous observations," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 2016, pp. 4254-4258.

Y. Bu, S. Zou and V. V. Veeravalli, "Linear-Complexity Exponentially-Consistent Tests for Universal Outlying Sequence Detection," in IEEE Transactions on Signal Processing, vol. 67, no. 8, pp. 2115-2128, 15 April15, 2019.

Y. Li, S. Nitinawarat & V. V. Veeravalli (2017) Universal sequential outlier hypothesis testing, Sequential Analysis, 36:3, 309-344.

S. Zou, Y. Liang, H. V. Poor and X. Shi, "Nonparametric Detection of Anomalous Data Streams," in IEEE Transactions on Signal Processing, vol. 65, no. 21, pp. 5785-5797, 1 Nov.1, 2017.

Performance metrics used

- Performance metrics
  - Universal consistency
  - Universal exponential consistency
  - Error Exponent

- FSS: As number of samples grows
- SEQ: As the expected stopping time/stopping threshold grows
- Sequential tests can stop fast for good realizations
  - Expected number of samples required reduces

Known Results: Discrete, FSS

- Discrete distributions
  - *L* = 1 (single anomaly), γ_i Empirical pmf of stream i, GL test

$$\hat{\imath} = \arg\max_i D(\gamma_i || p)$$

p known

$$\hat{\imath} = \arg\min_i \sum_{j \neq i} D\left(\gamma_j || \frac{\sum_{k \neq i} \gamma_k}{S-1}\right)$$

  - Exponential consistency
  - Optimal error exponent if p known

  - Similar results for *L* >= 1 and known
  - *L* <=1
    - Exponential consistency only under non-null hypothesis

Y. Li, S. Nitinawarat and V. V. Veeravalli, "Universal Outlier Hypothesis Testing," in IEEE Trans. on Information Theory, vol. 60, no. 7, pp. 4066-4082, July 2014.

Known Results: Continuous, FSS, Known typical distribution

- Continuous distributions, *L* = 1 (single anomaly)
- Based on estimated KL divergence

$$\hat{\imath} = \arg\max_{i} \widehat{D}\ (\boldsymbol{Y}_i || p)$$

- Based on estimated Maximum Mean Discrepancy (MMD)

$$\hat{\imath} = \arg\max_{i} \widehat{MMD}\ (\boldsymbol{Y}_i || p)$$

- Exponential consistency in both cases
- One test not always better than the other

Y. Bu, S. Zou, Y. Liang and V. V. Veeravalli, "Universal outlying sequence detection for continuous observations," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 2016, pp. 4254-4258.

Our Work

- Sequential test for
  - Number of anomalous streams L = 1
  - Known or Unknown L: $1 \leq L \leq A$
  - Unknown L: $0 \leq L \leq A$

- Expected number of samples lower than that of FSS test for the same error probability
- Universal consistency (or) Universal exponential consistency

S. C. Sreenivasan and S. Bhashyam, "Sequential Nonparametric Detection of Anomalous Data Streams," in IEEE Signal Processing Letters, vol. 28, pp. 932-936, 2021.

Comparing distributions

- Known distributions
  - Compute likelihood under each distribution

- Unknown distributions + Parametric model for distributions
  - Generalized likelihood instead of likelihood
  - Parameters estimated under each hypothesis and plugged into likelihood

- Unknown distributions, Nonparametric
  - Estimated KL divergence
  - Maximum Mean Discrepancy (MMD)
  - Kolmogorov-Smirnov Distance (KSD)

Maximum Mean Discrepancy (MMD)

$$MMD(p,q) = \sup_{f \in F} \quad E_p[f(X)] - E_q[f(Y)]$$

- $X \sim p$ and $Y \sim q$,
- f a real − valued function from class F
- F: Unit ball in a Reproducing Kernel Hilbert Space (RKHS) with kernel $k(.,.)$
- Estimate with finite number of samples
- Convergence as number of samples grows

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. (2012). A kernel two-sample test. The Journal of Machine Learning Research, 13(1), 723-773.

MMD Estimate and Convergence

$$X_i^n = \{x_{i1}, x_{i2}, \dots, x_{in}\}$$
$$X_j^n = \{x_{j1}, x_{j2}, \dots, x_{jn}\}$$

Gaussian Kernel

$$k(x, y) = e^{-\frac{(x-y)^2}{2\sigma^2}}$$

$$M_u(i, j, n) = \frac{1}{n(n-1)} \sum_{l \neq m} \left(k(x_{il}, x_{im}) + k(x_{jl}, x_{jm}) - k(x_{il}, x_{jm}) - k(x_{jl}, x_{im}) \right)$$

$M_u(i, j, n)$ converges a.s. to *MMD(p, q)* as *n* → ∞

Sequential update with O(*n*) computations

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. (2012). A kernel two-sample test. The Journal of Machine Learning Research, 13(1), 723-773.

Sequential Test: Single Anomaly Case

- Find
  - Stream with maximum minimum distance from other streams
  - Corresponding max-min distance

$$\hat{\imath}(n) = \arg\max_{i} \min_{j \neq i} M_u(i, j, n)$$

$$\Gamma(n) = \max_{i} \min_{j \neq i} M_u(i, j, n)$$

- Compare max-min distance with a threshold

$$\Gamma(n) > \frac{c}{n^{\alpha}}$$

- Choice of alpha

Sequential Test: Multiple Anomaly Case

- Find
  - Subset ***A*** with maximum minimum distance from other subsets
  - Corresponding max-min distance
  - Search over all subsets of size *L* (known *L* or $1 \leq L \leq A$)

$$\hat{\imath}(n) = \arg\max_{\boldsymbol{A}} \min_{i \in \boldsymbol{A}} \min_{j \in \boldsymbol{S} \,|\, \boldsymbol{A}} M_u(i,j,n)$$

$$\Gamma(n) = \max_{\boldsymbol{A}} \min_{i \in \boldsymbol{A}} \min_{j \in \boldsymbol{S} \,|\, \boldsymbol{A}} M_u(i,j,n)$$

- Compare max-min distance with a threshold

$$\Gamma(n) > \frac{c}{n^{\alpha}}$$

Possibility of No Anomalies $0 \le L \le A$

- Additional time-out parameter T_0
  - controls error probability when there are no anomalies
- Use previous test up to T_0
- Stop if number of samples exceeds T_0

$$\Gamma(n) > \frac{c}{n^{0.5}}$$

Properties of the Proposed Test

- Stopping time *N*, Maximal error prob $P_{\max}$

- Finite stopping time $P_i[N < \infty] = 1$ for each i
- Universal consistency $\lim_{C \to \infty} P_{\max} = 0$

- When $L > 0$, we also have universal exponential consistency

$$E_i[N] \leq -\frac{32\ \log P_{\max}}{MMD^4(p, q)}$$

Proof outline: Single Anomaly Case

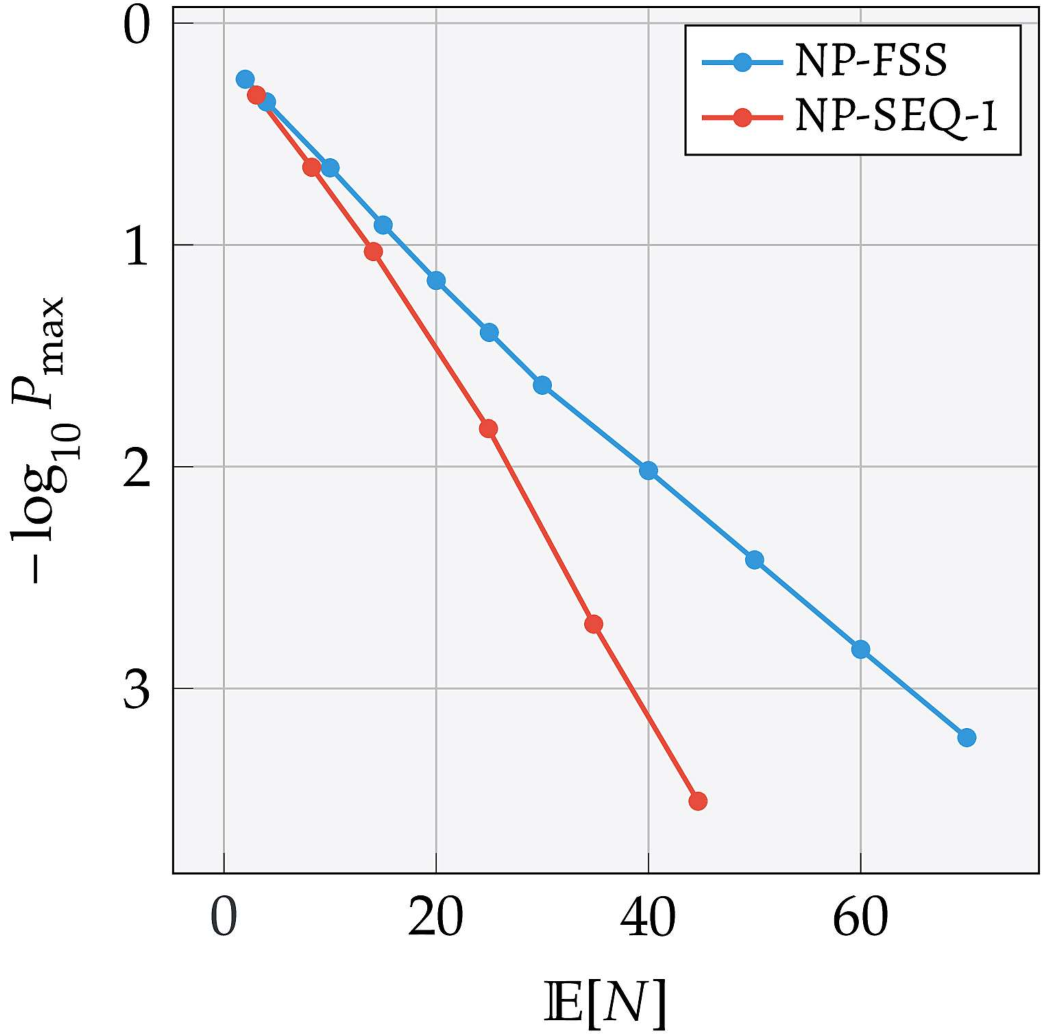
- Finite stopping proof
  - Exponential bound $P_i[N \geq n]$ for $n > n_0$

- Error bound
  - Split into two terms
  - Error when $N > n_0$, Error when $N \leq n_0$
  - Goes to 0 as $C \to \infty$

- $E\left[\left|\frac{N}{C} - \frac{1}{MMD^2(p,q)}\right|\right] \to 0$ as $C \to \infty$

- Combine above results to get exponential consistency

Simulation Results: Single Anomaly



- 5 streams
- N(0,1) and N(1.2,1)

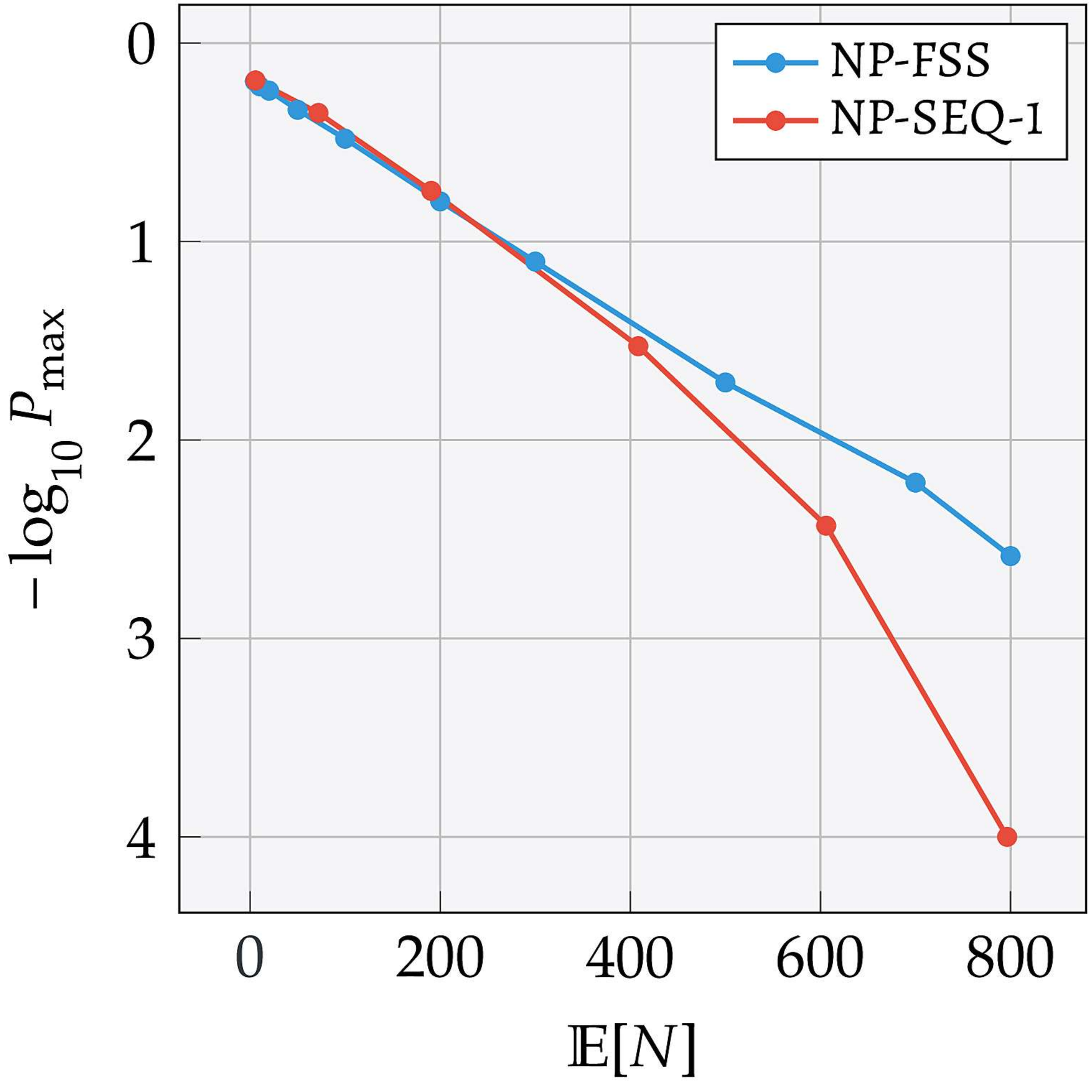
Threshold $\frac{c}{n}$

- NP-FSS: Zou 2017
- NP-SEQ-1: Proposed

- Universal exponential consistency

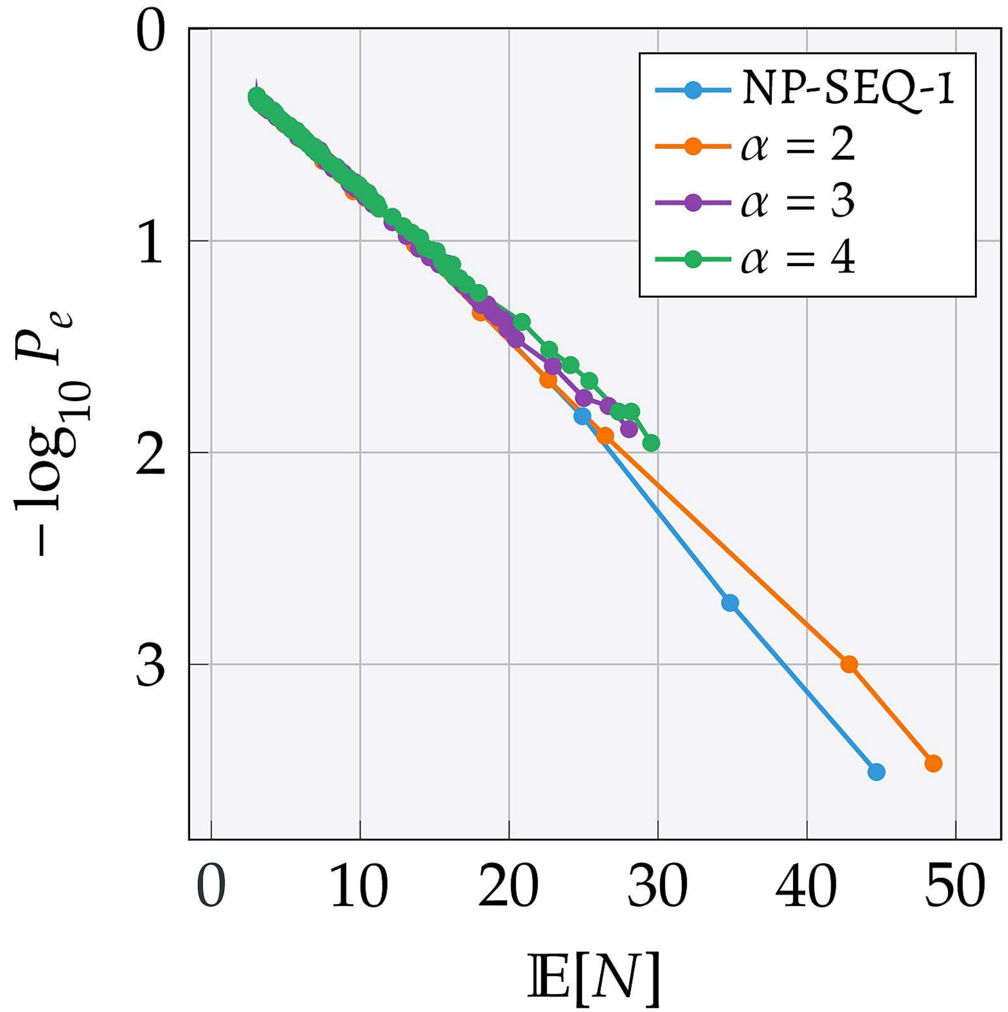
S. Zou, Y. Liang, H. V. Poor and X. Shi, "Nonparametric Detection of Anomalous Data Streams," in IEEE Transactions on Signal Processing, vol. 65, no. 21, pp. 5785-5797, 1 Nov.1, 2017.

Simulation Results: Single Anomaly



- 5 streams
- N(0,1) and L(0,$\frac{1}{\sqrt{2}}$)
- Distributions are closer in this case

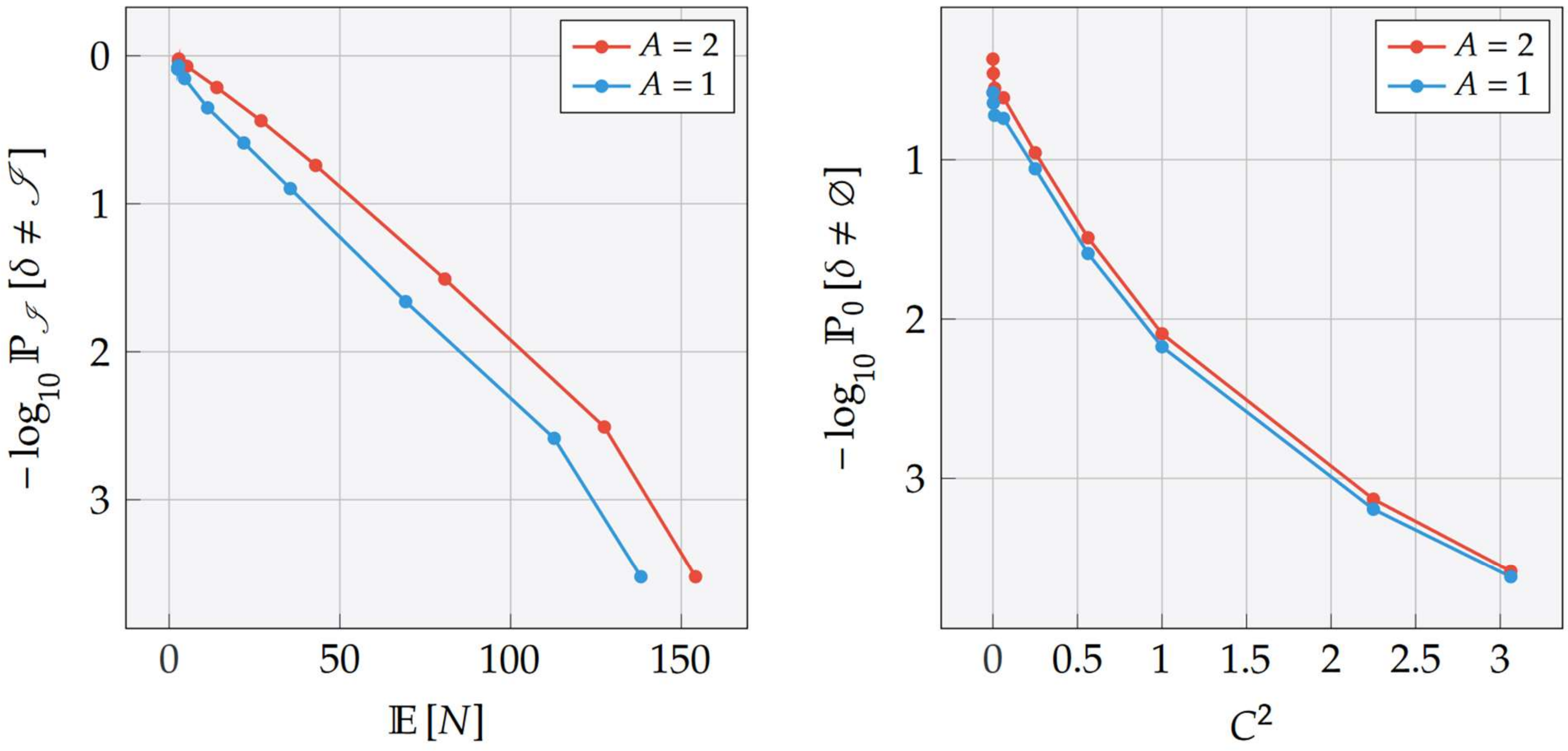
Simulation Results: Single Anomaly



- 5 streams
- N(0,1) and N(1.2,1)
- Higher alpha reduces the threshold faster

Threshold $\frac{c}{n^{\alpha}}$

Simulation Results: $0 \le L \le A$

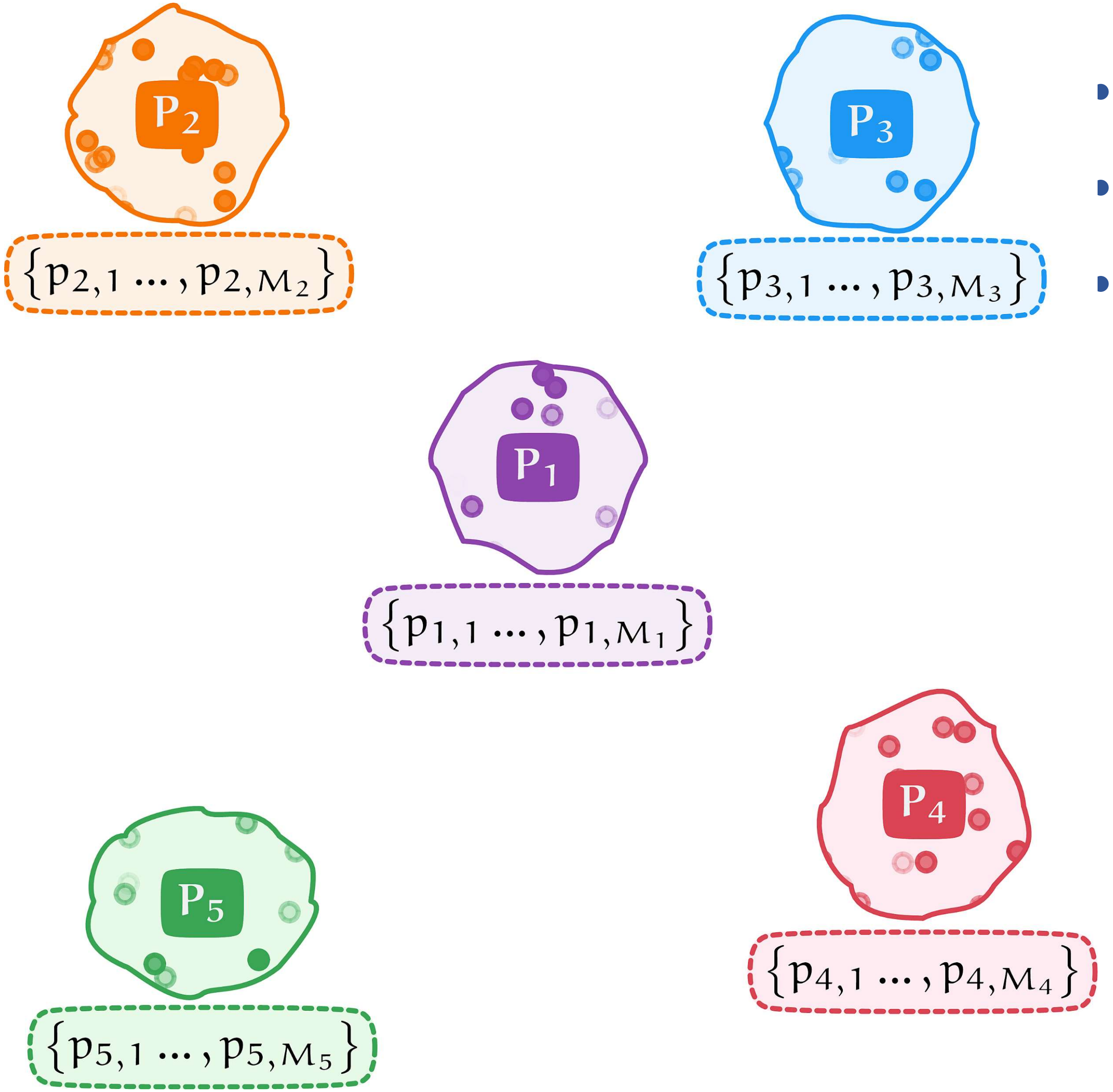


a) Missed-detection error

b) False-alarm error

- 5 streams
- N(0,1) and N(1.2,1)

Clustering



P_2
$\{p_{2,1} \dots, p_{2,M_2}\}$
P_3
$\{p_{3,1} \dots, p_{3,M_3}\}$
P_1
$\{p_{1,1} \dots, p_{1,M_1}\}$
P_4
$\{p_{4,1} \dots, p_{4,M_4}\}$
P_5
$\{p_{5,1} \dots, p_{5,M_5}\}$

- *S* data streams
- *K* clusters
- M_k distributions in cluster *k*

FSS Non-parametric Clustering

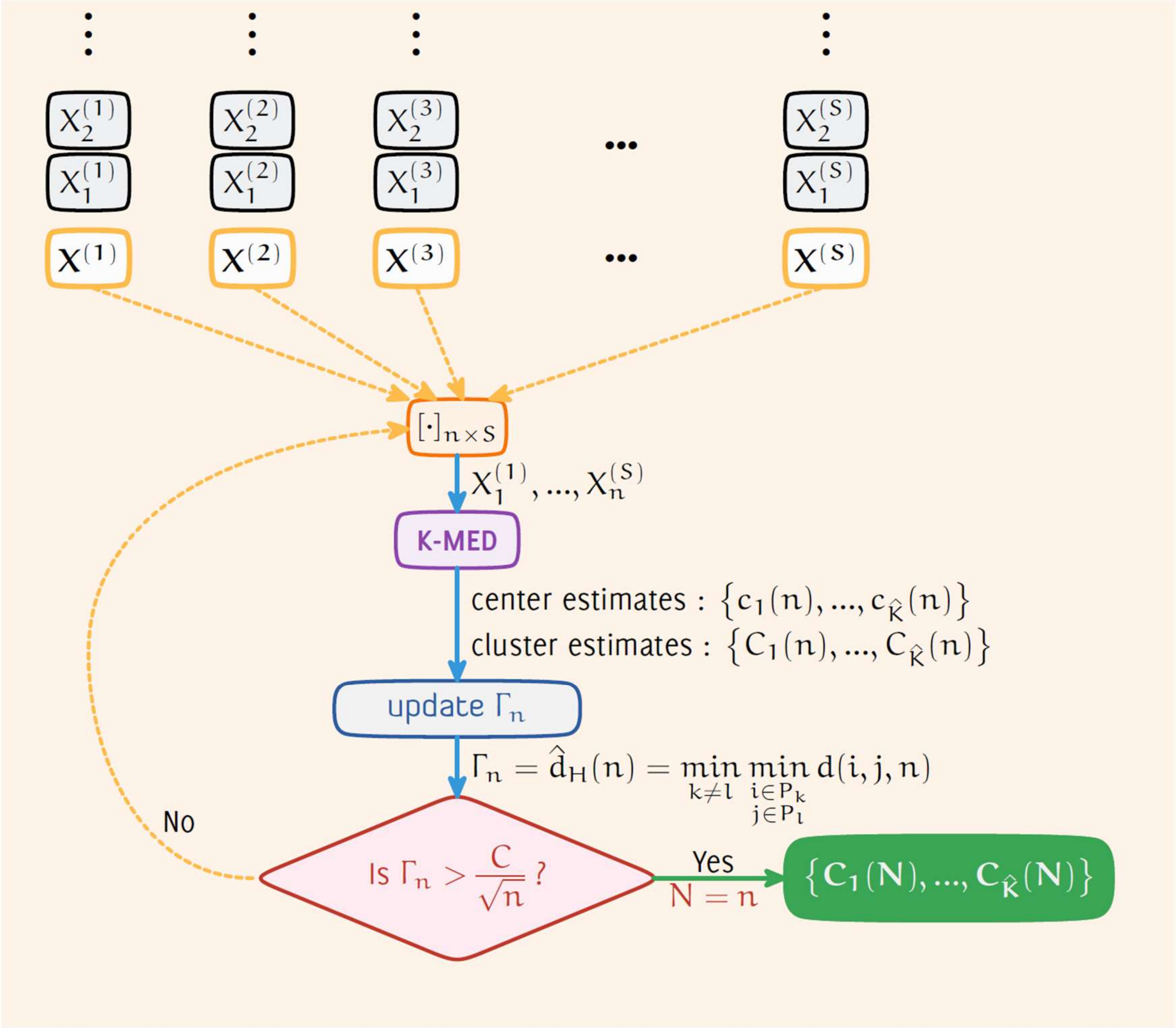
- Use pairwise distances (MMD/KSD)
- Cluster based on k-medoid clustering
  - Number of clusters *K* known (K-MED)
  - Number of clusters *K* unknown

- Steps
  - Center and Cluster initialization
  - Update till convergence

- Universal exponential consistency $(n \to \infty)$

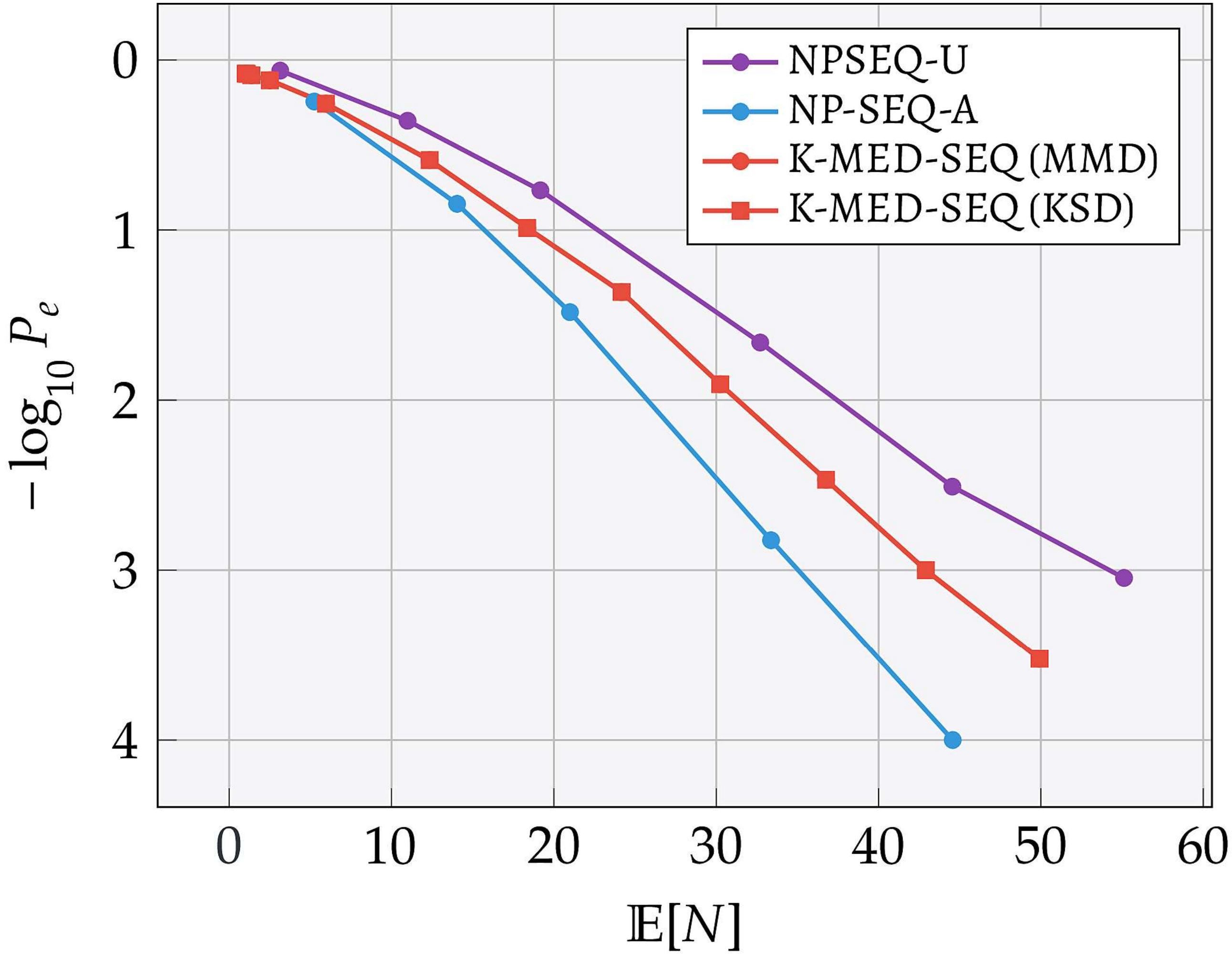
T. Wang, Q. Li, D. J. Bucci, Y. Liang, B. Chen and P. K. Varshney, "K-Medoids Clustering of Data Sequences With Composite Distributions," in IEEE Transactions on Signal Processing, vol. 67, no. 8, pp. 2093-2106, 15 April15, 2019.

Our Work: Sequential Clustering


$X_2^{(1)}$ $X_2^{(2)}$ $X_2^{(3)}$ ... $X_2^{(S)}$
$X_1^{(1)}$ $X_1^{(2)}$ $X_1^{(3)}$ ... $X_1^{(S)}$
$\mathbf{X}^{(1)}$ $\mathbf{X}^{(2)}$ $\mathbf{X}^{(3)}$ ... $\mathbf{X}^{(S)}$
$[\cdot]_{n\times S}$
$X_1^{(1)}, \ldots, X_n^{(S)}$
K-MED
center estimates : $\{c_1(n), \ldots, c_{\hat{K}}(n)\}$
cluster estimates : $\{C_1(n), \ldots, C_{\hat{K}}(n)\}$
update Γ_n
$\Gamma_n = \hat{d}_H(n) = \min_{k\neq l} \min_{\substack{i\in P_k \\ j\in P_l}} d(i,j,n)$
No
Is $\Gamma_n > \frac{C}{\sqrt{n}}$?
Yes
$N = n$
$\{C_1(N), \ldots, C_{\hat{K}}(N)\}$

- Threshold on empirical minimum inter-cluster distance

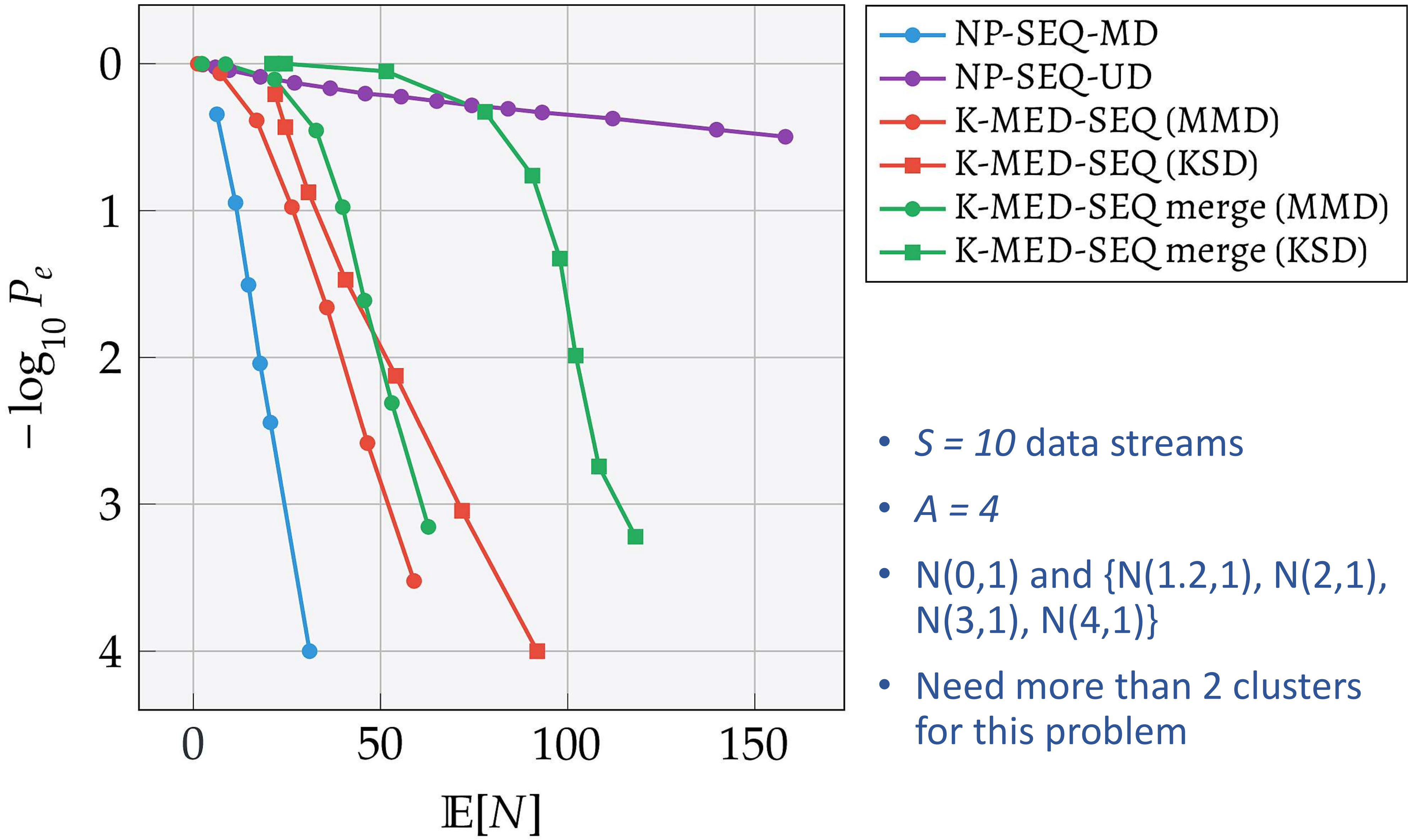
Multiple Anomalies



- *S = 5* data streams
- *A = 2*
- N(0,1) and N(1.2,1)

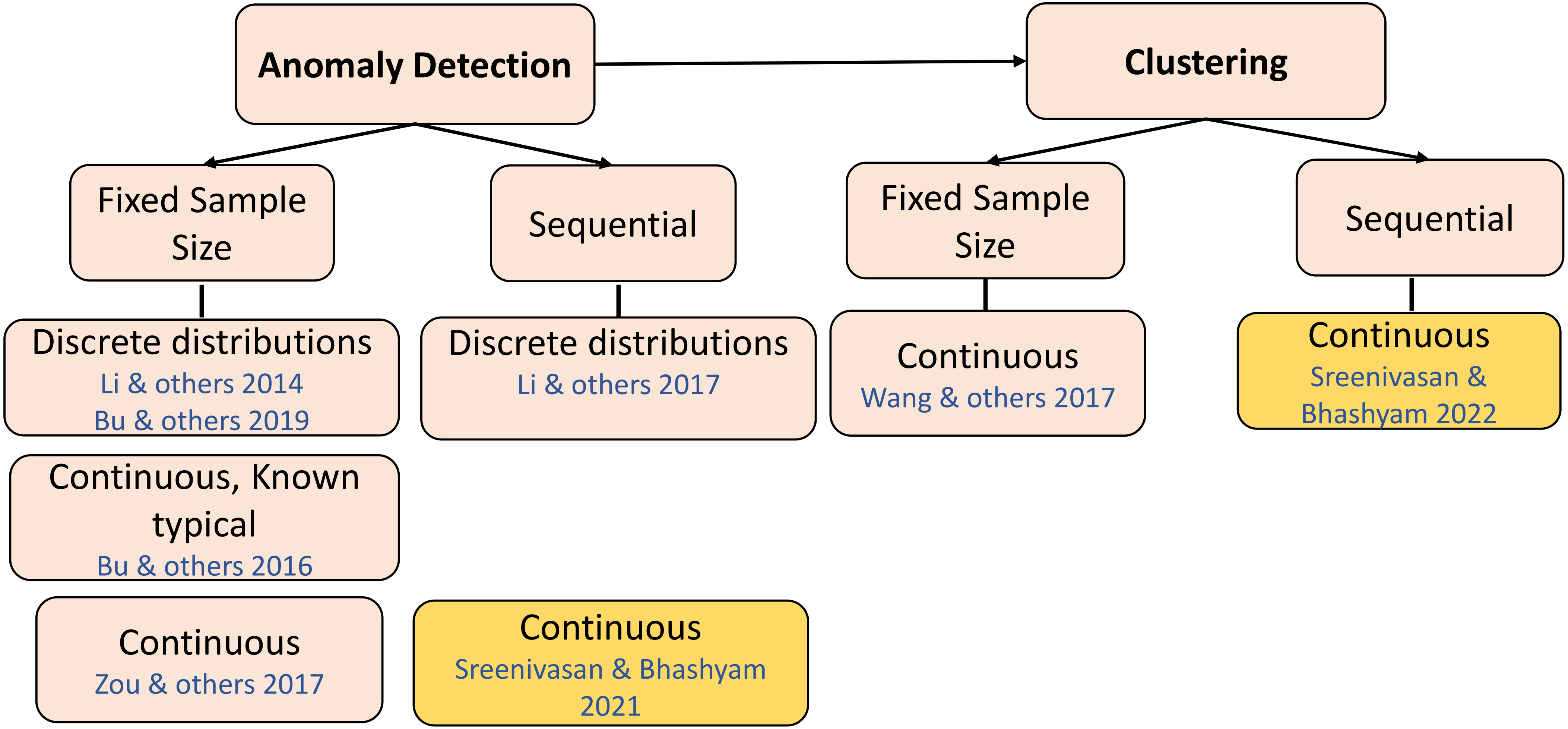
- NP-SEQ-A: Known A
- NP-SEQ-U: Unknown A

Multiple Distinct Anomalies



- *S = 10* data streams
- *A = 4*
- N(0,1) and {N(1.2,1), N(2,1), N(3,1), N(4,1)}
- Need more than 2 clusters for this problem

Summary



Anomaly Detection

Clustering

Fixed Sample Size

Sequential

Fixed Sample Size

Sequential

Discrete distributions
Li & others 2014
Bu & others 2019

Discrete distributions
Li & others 2017

Continuous
Wang & others 2017

Continuous
Sreenivasan & Bhashyam 2022

Continuous, Known typical
Bu & others 2016

Continuous
Zou & others 2017

Continuous
Sreenivasan & Bhashyam 2021

- Universally consistent sequential tests for anomaly detection and clustering

https://www.ee.iitm.ac.in/~skrishna/

Extensions/Current Work

- More general cases
  - $d_L > d_H$, for both FSS and Sequential settings
  - Higher dimensional observations
- Clustering with bandit feedback/controlled sampling
- Lower complexity methods
- More than consistency
  - Error exponent and optimality

https://www.ee.iitm.ac.in/~skrishna/