Sequential Methods for Anomaly Detection and Clustering

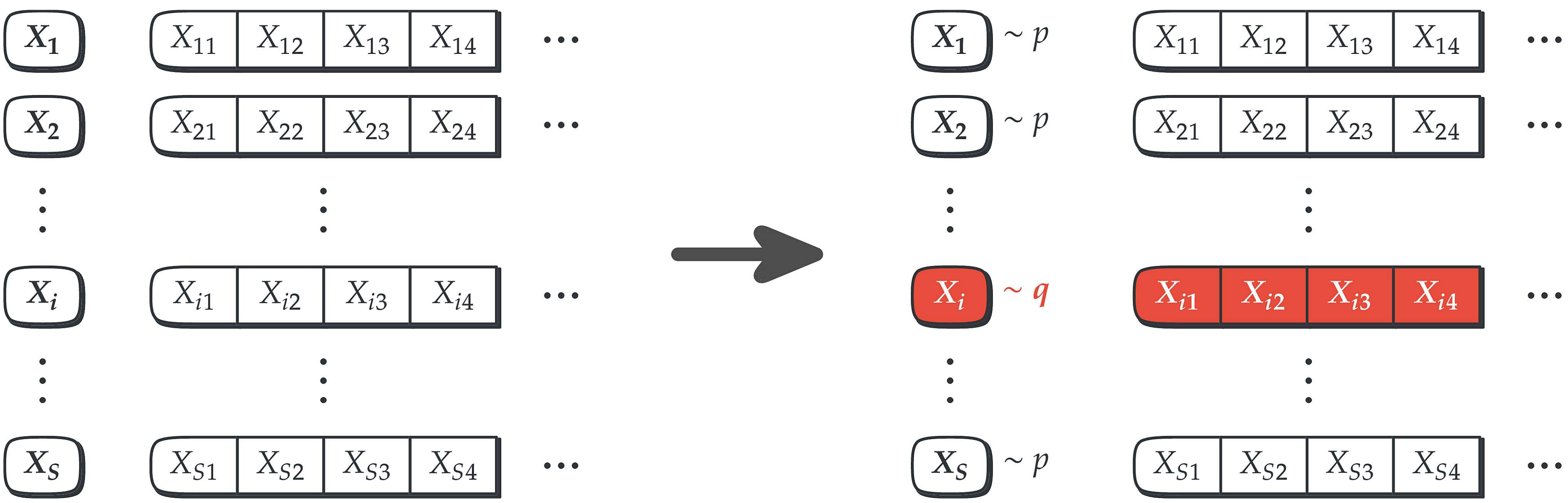
Srikrishna Bhashyam
IIT Madras

Joint work with Sreeram C. Sreenivasan

July 26, 2022
IIIT Sri City, WASDAM 2022

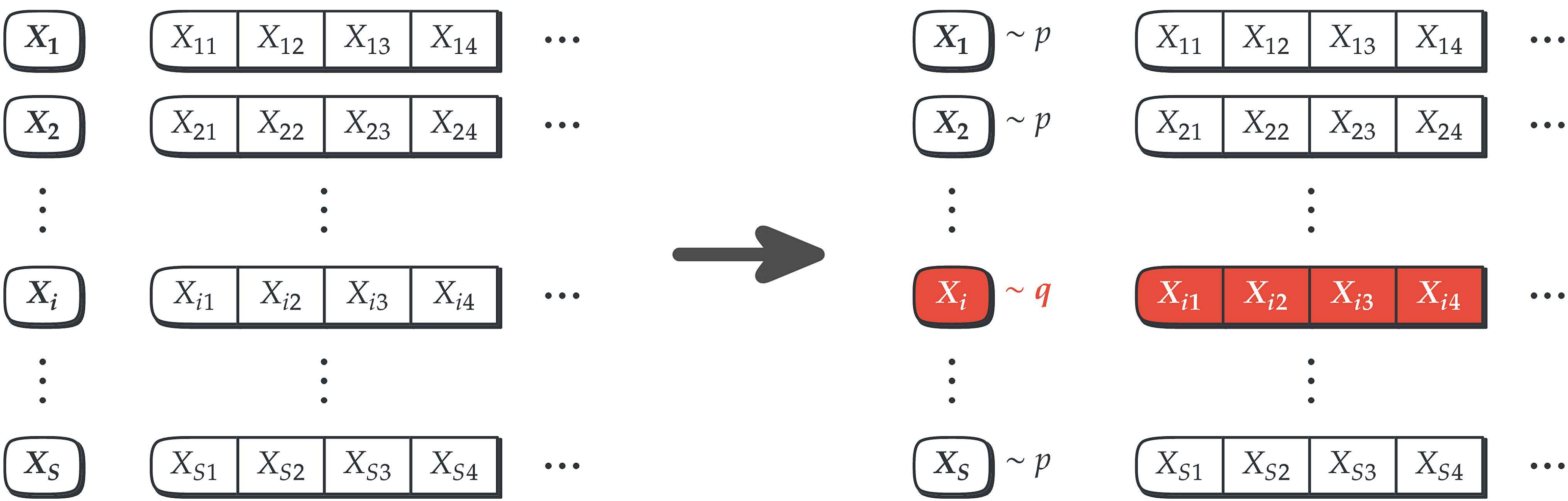
Acknowledgements: Rajesh Sundaresan, IISc

Detection of Anomalous Data Streams



- Each data stream independent and identically distributed (i.i.d.) samples from an unknown distribution
- Applications: Sensor networks, network monitoring

Hypothesis Testing



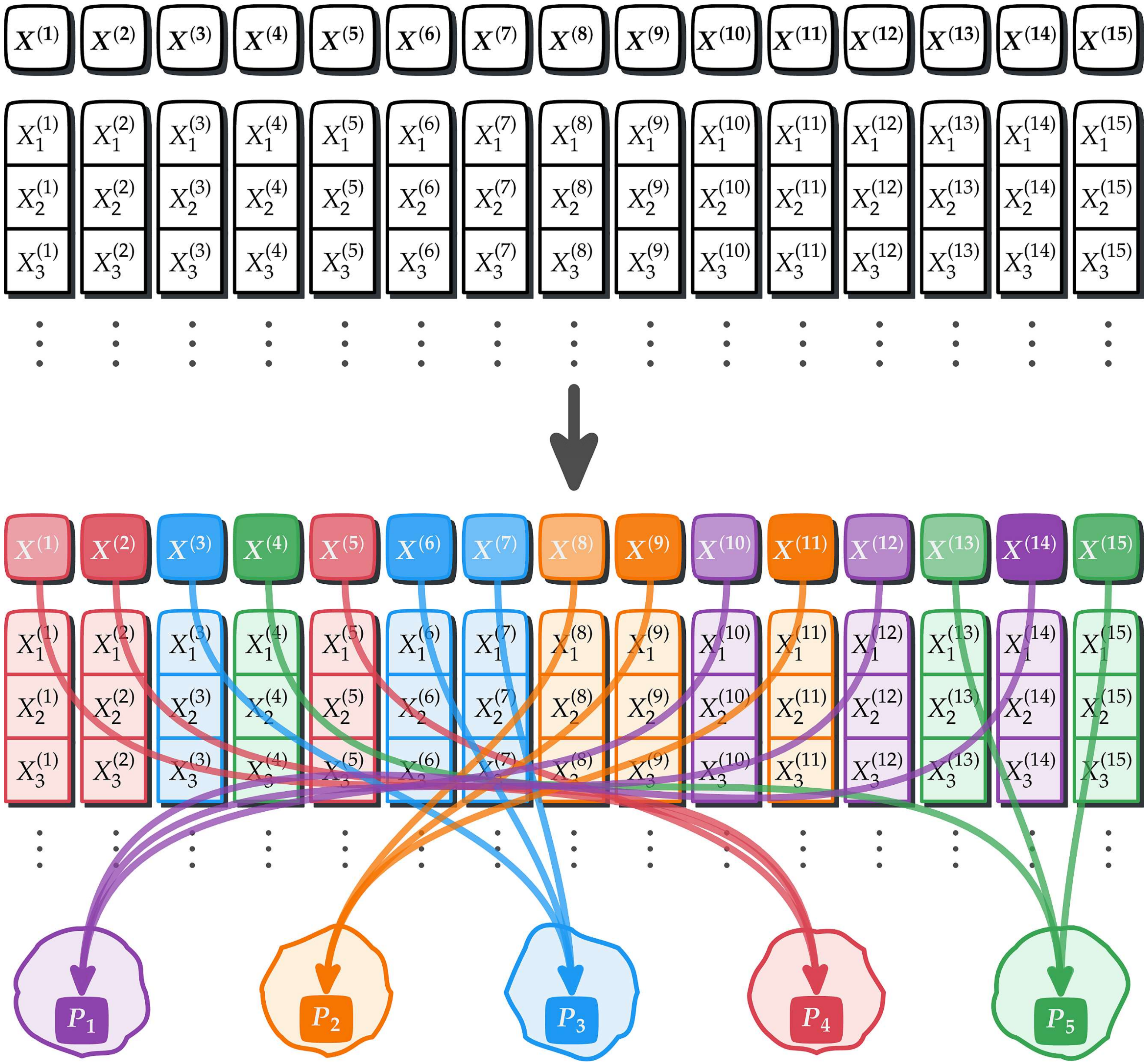
- S hypotheses
- Hypothesis i: The i th stream is anomalous

Nonparametric Sequential 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)

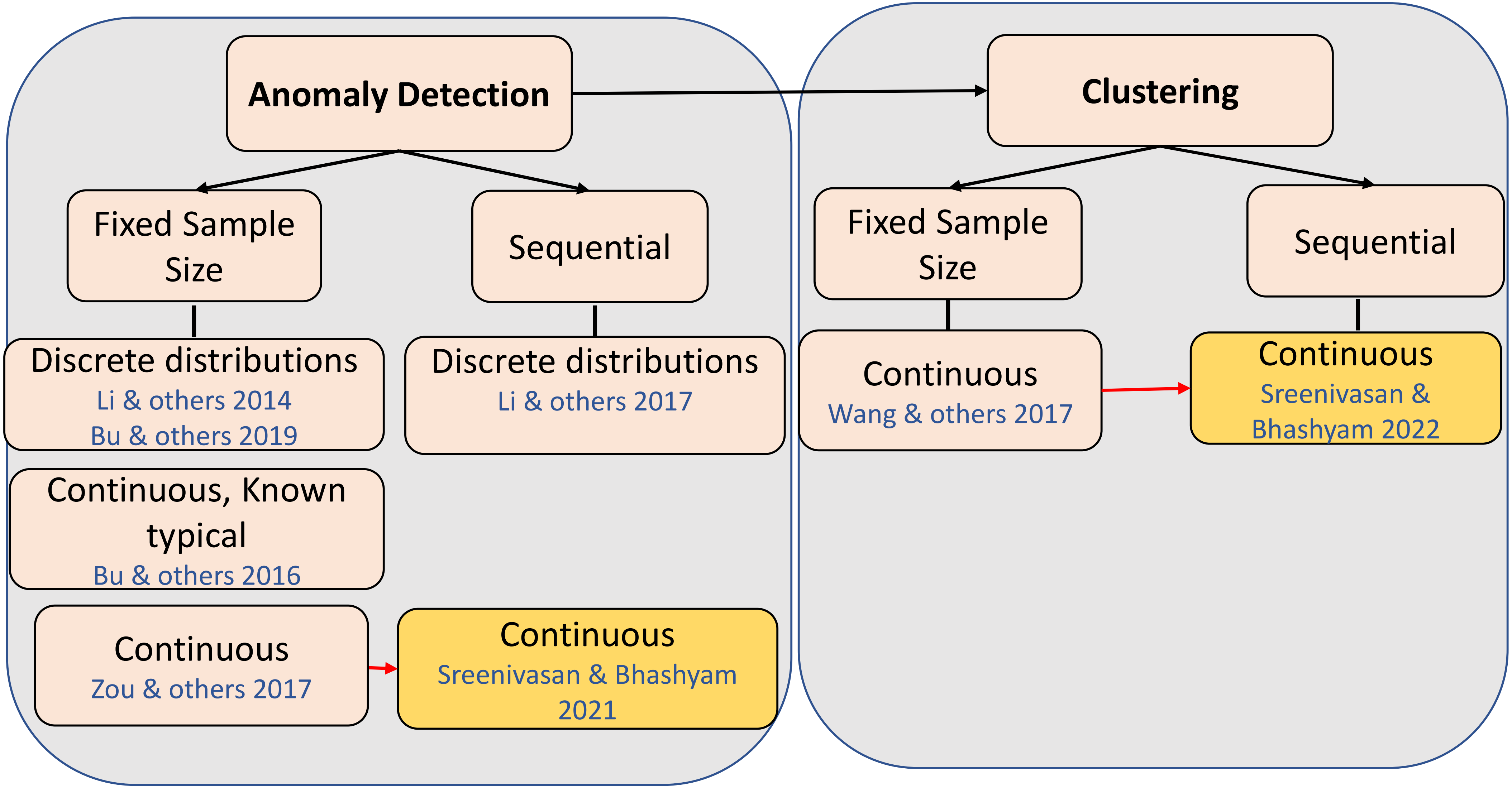
- Nonparametric: Unknown distributions p and q
  - $p \neq q$
  - Also called Universal or Distribution-free tests

Clustering



- Each stream can be from a different distribution
- Distributions for clusters
  - Distributions in the same cluster are closer
- Need to cluster the streams
- Unknown distributions

Closely Related Work



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

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.

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.

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
  - 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 $n \to \infty$

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.

Fixed Sample Size (FSS) vs. Sequential (SEQ)

- Performance metrics
  - Universal consistency
  - Universal exponential consistency
  - Error Exponent
- FSS: As number of samples grows
- SEQ: As the stopping threshold grows

- Sequential tests can stop fast for good realizations
- Expected number of sample required reduces

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.

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

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

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

- Compare max-min distance with a threshold

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

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

- 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

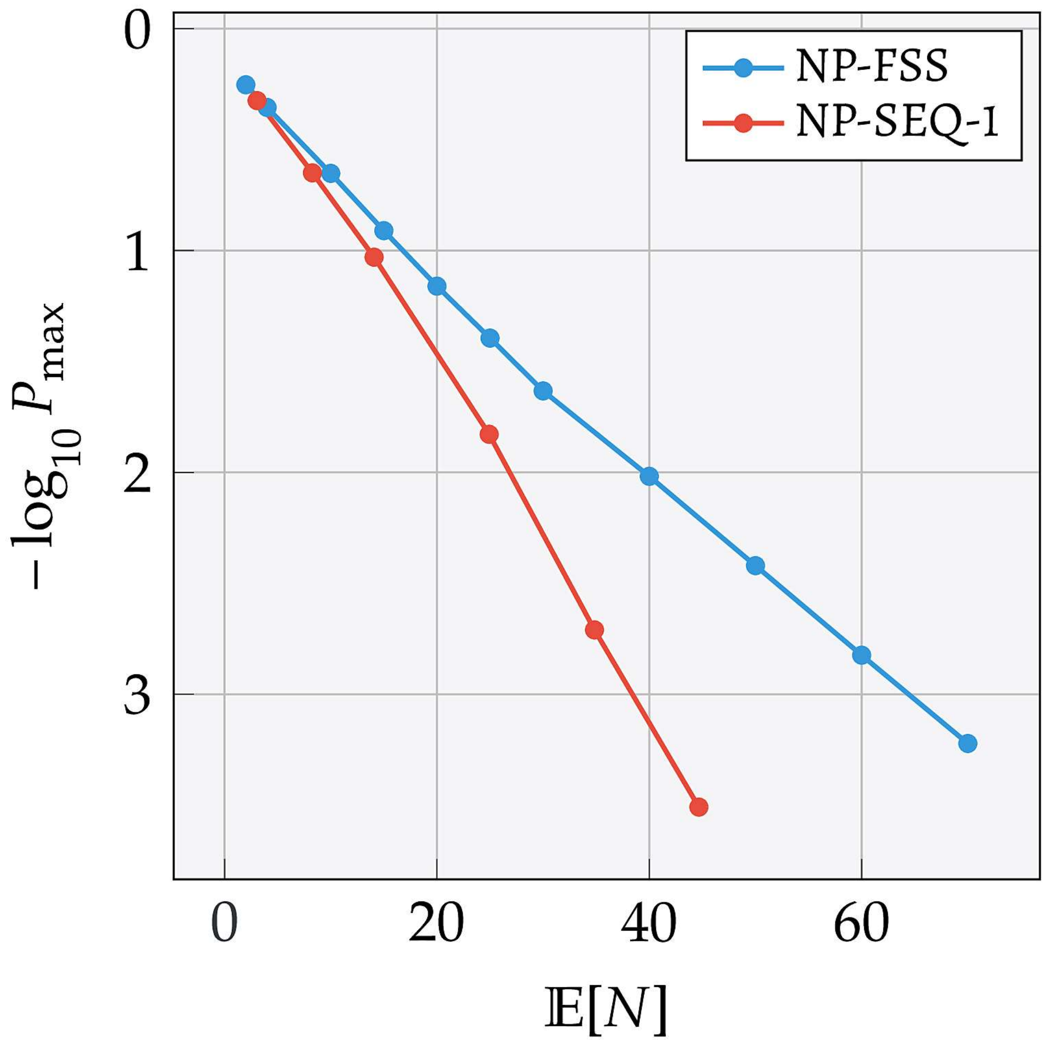
- Finite stopping proof
  - Exponential bound $P_i[N \geq n]$ for $n > n_0$

- 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



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

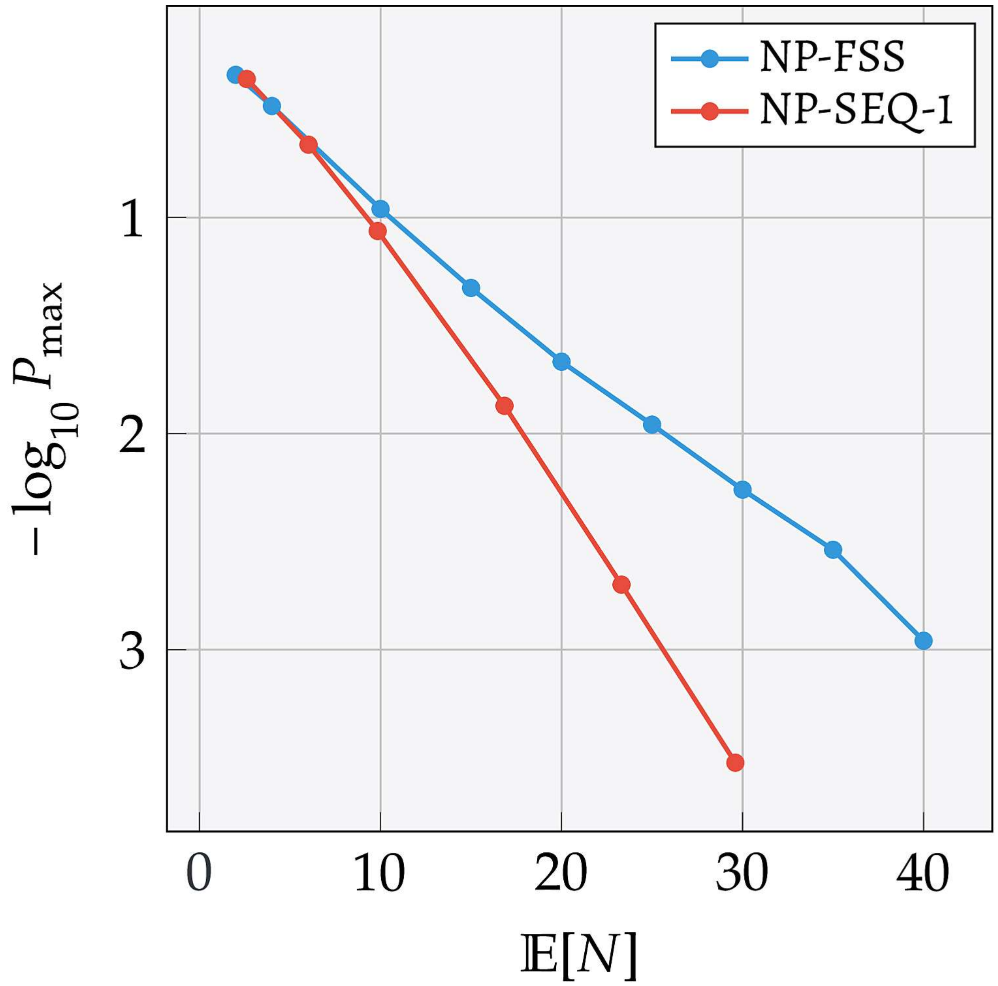
Threshold $\frac{c}{n}$

- 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.

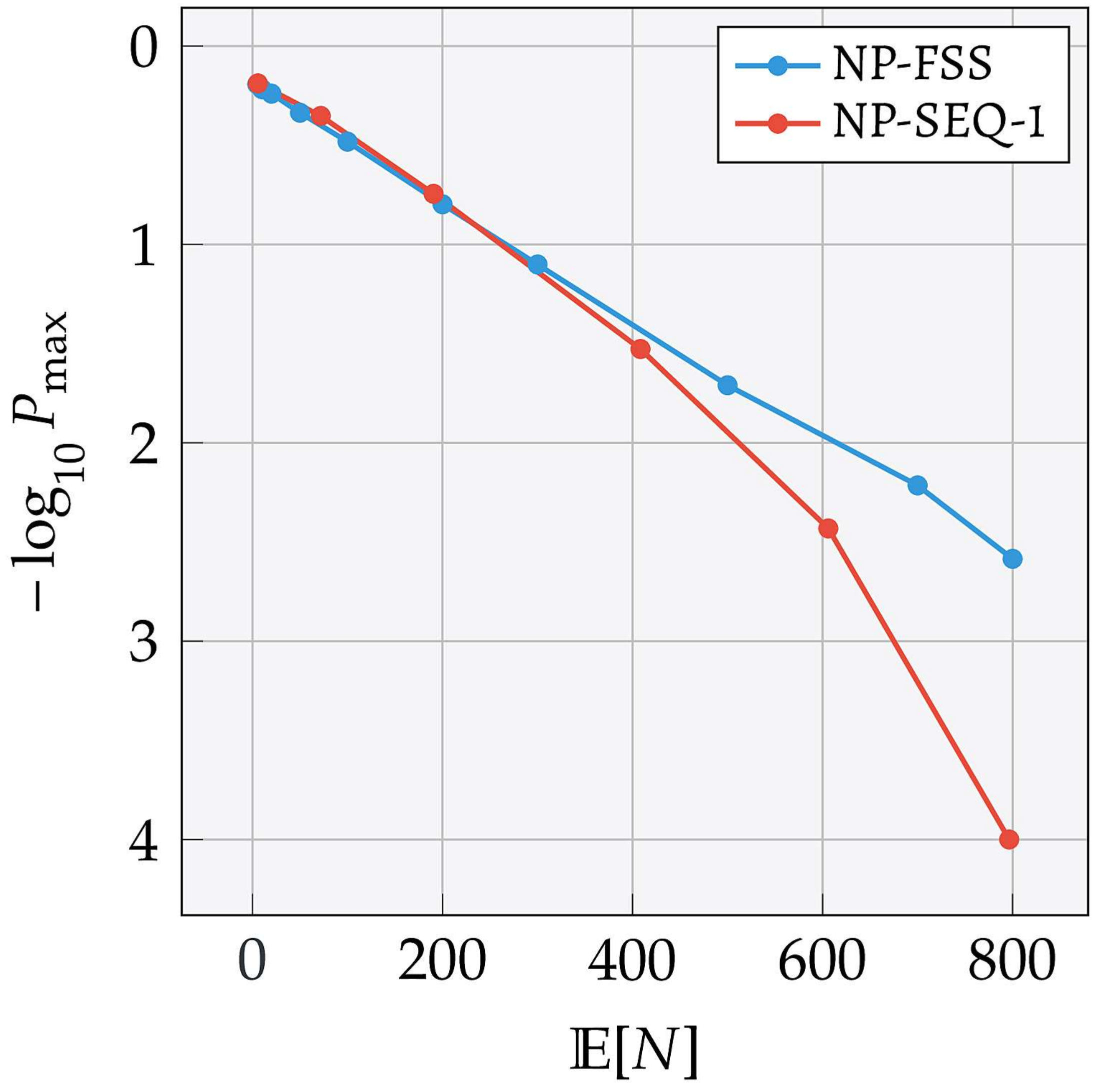
Simulation Results: Single Anomaly



- N(0,1) and N(1,0.5)

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.

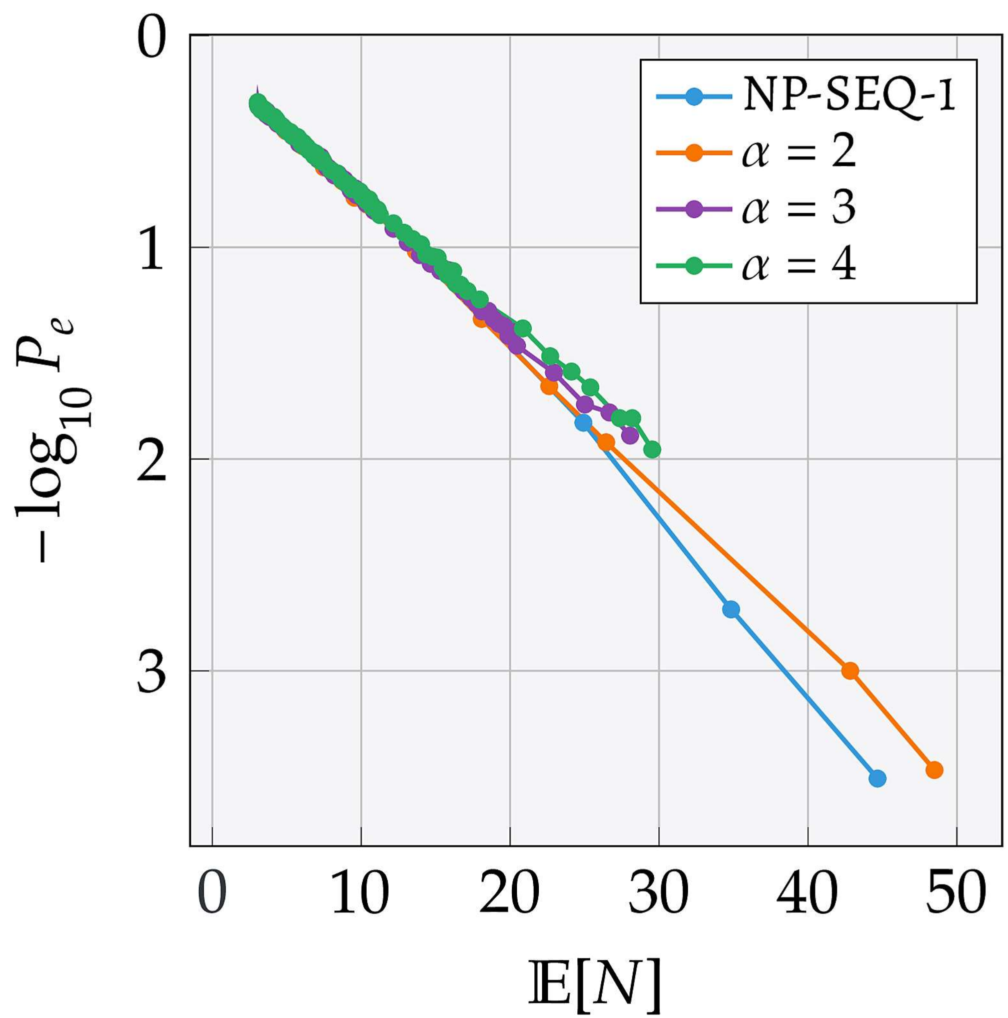
Simulation Results: Single Anomaly



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

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.

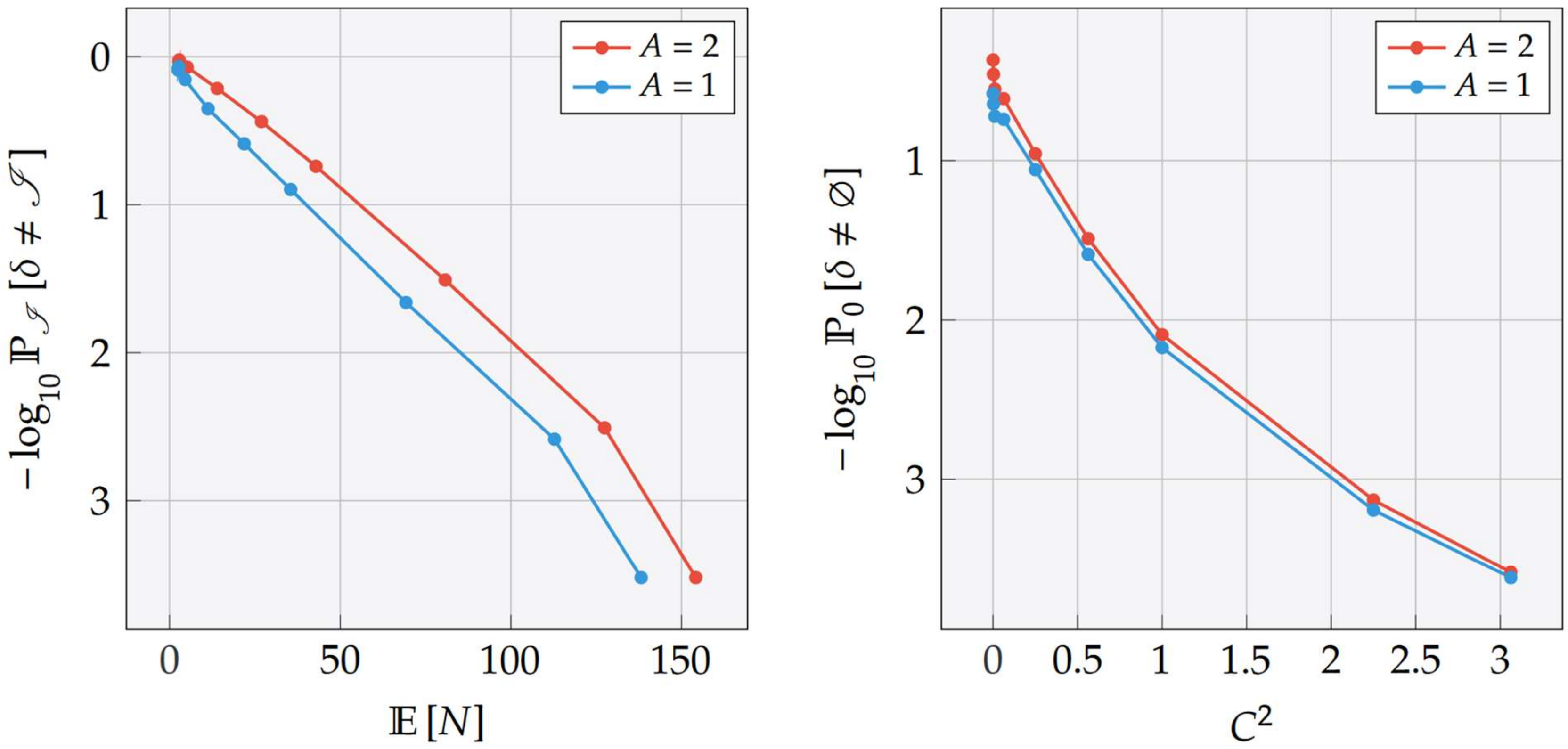
Simulation Results: Single Anomaly



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

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

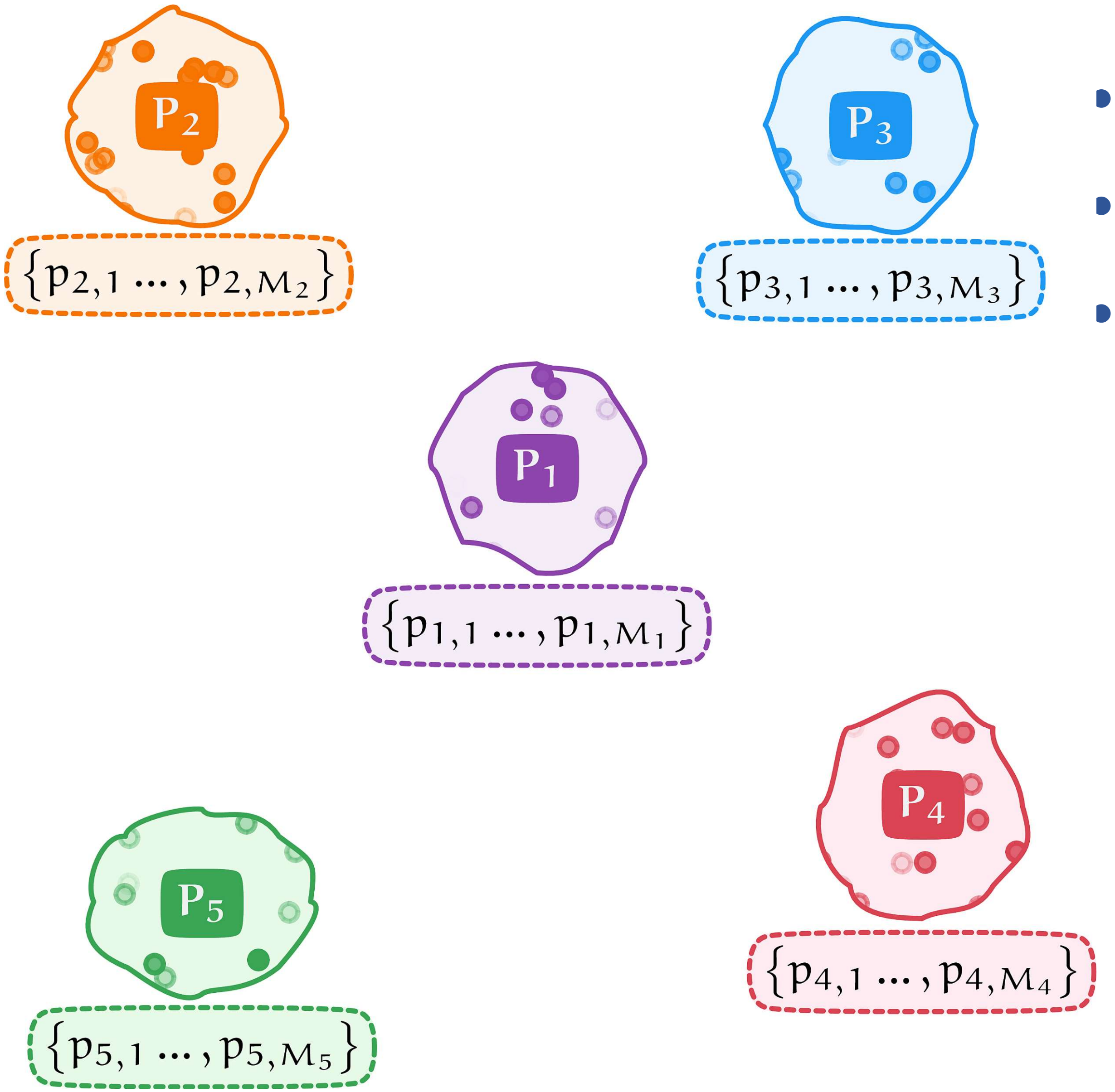
Simulation Results: $0 \leq L \leq A$



a) Missed-detection error

b) False-alarm error

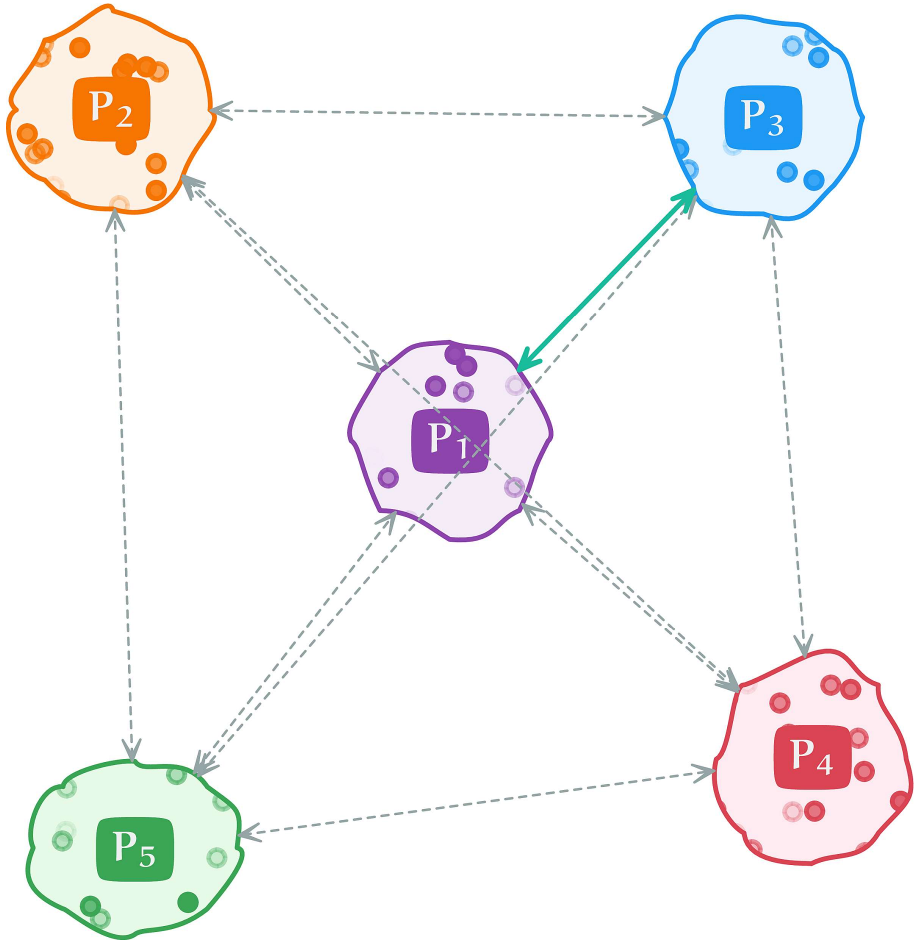
Clustering



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

Assumptions (for the analysis)

- Minimum inter-cluster distance d_H
- Maximum intra-cluster distance $d_L < d_H$



FSS Non-parametric Clustering

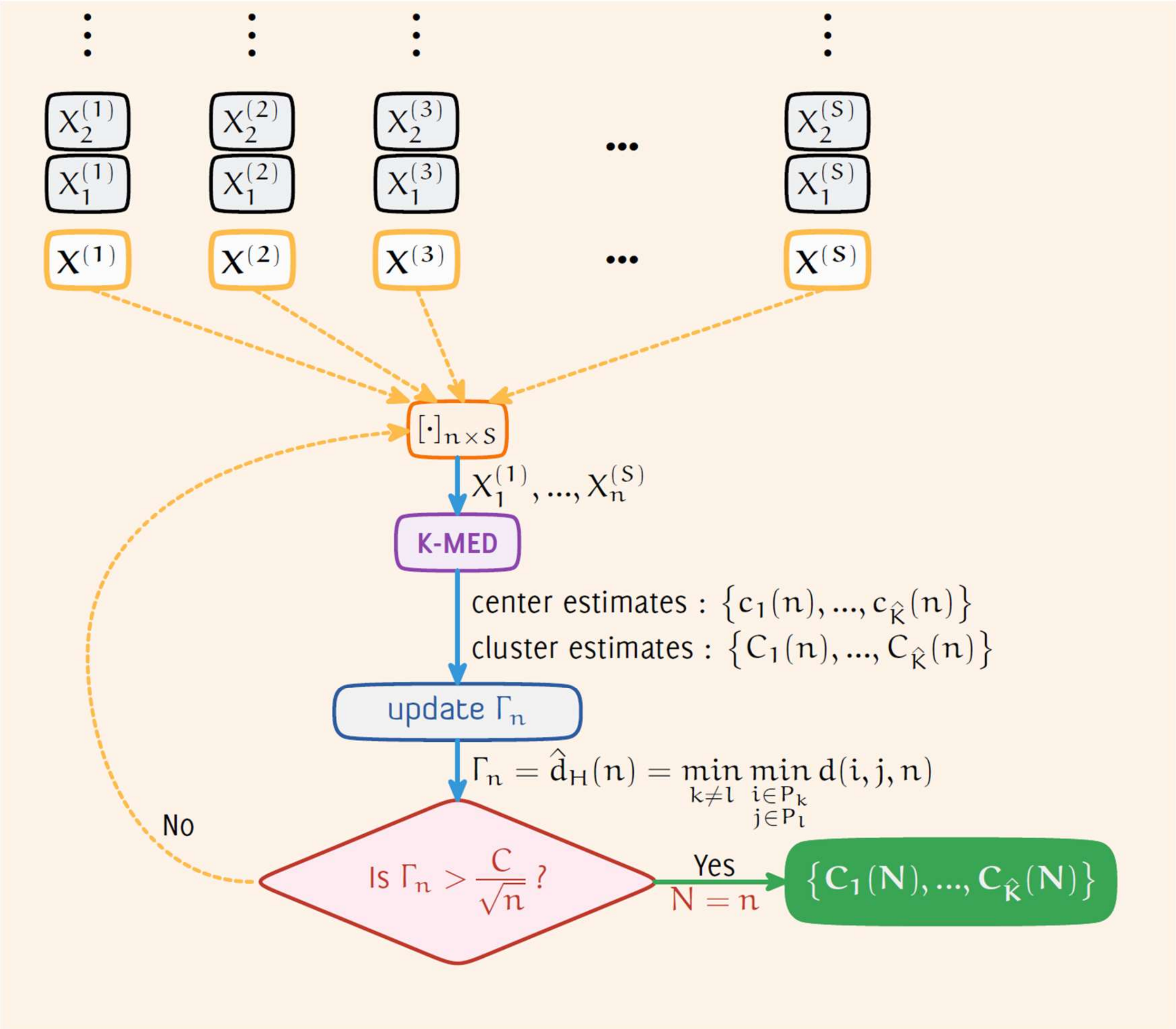
- Use pairwise distances
- Cluster based on k-medoid clustering
  - Number of clusters K known (K-MED)
  - Number of clusters K unknown

- 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



- Threshold on empirical minimum inter-cluster distance

Properties of the Proposed Test

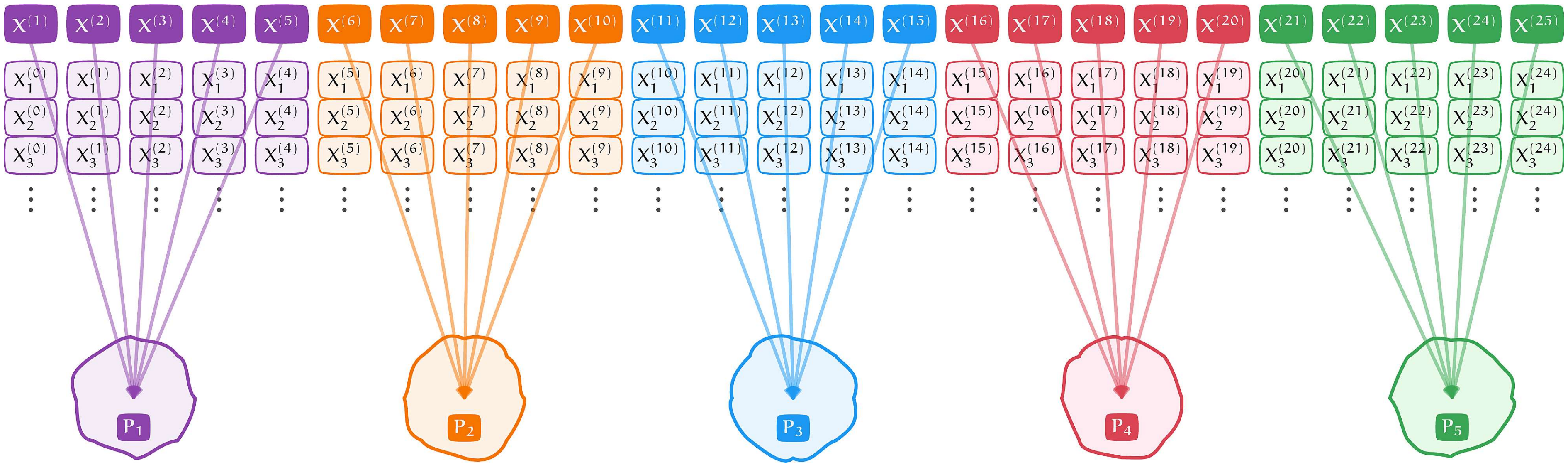
- Stopping time N, Maximal error prob $P_{\max}$

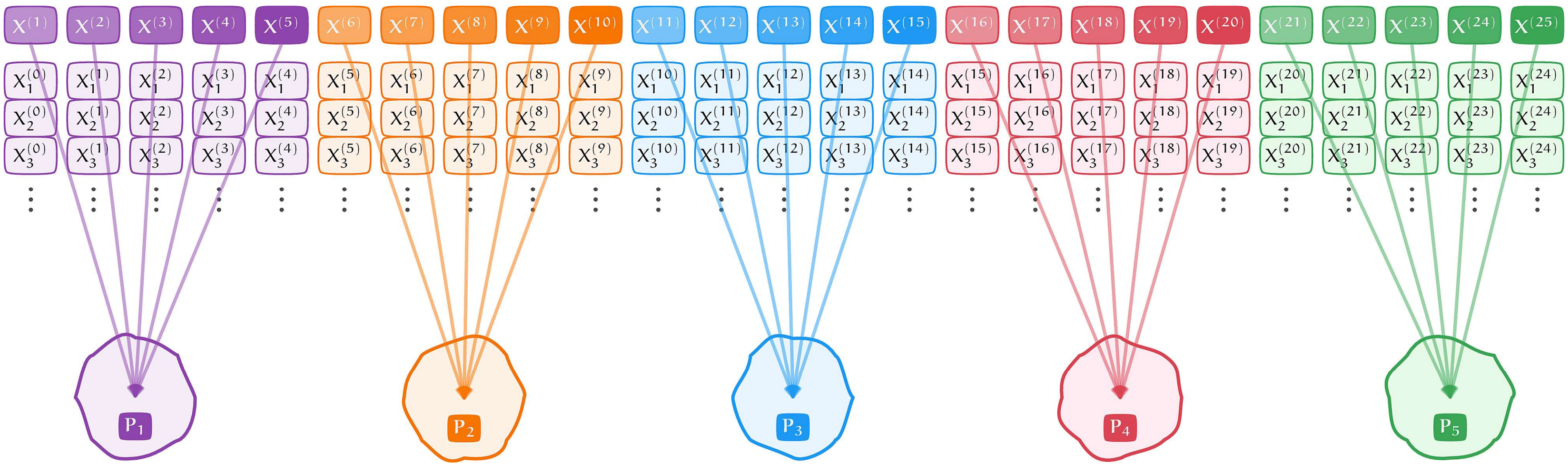
- Finite stopping time
- Universal exponential consistency

- At least 2 clusters assumed

Threshold $\frac{C}{n^{0.5}}$

Simulation Setting

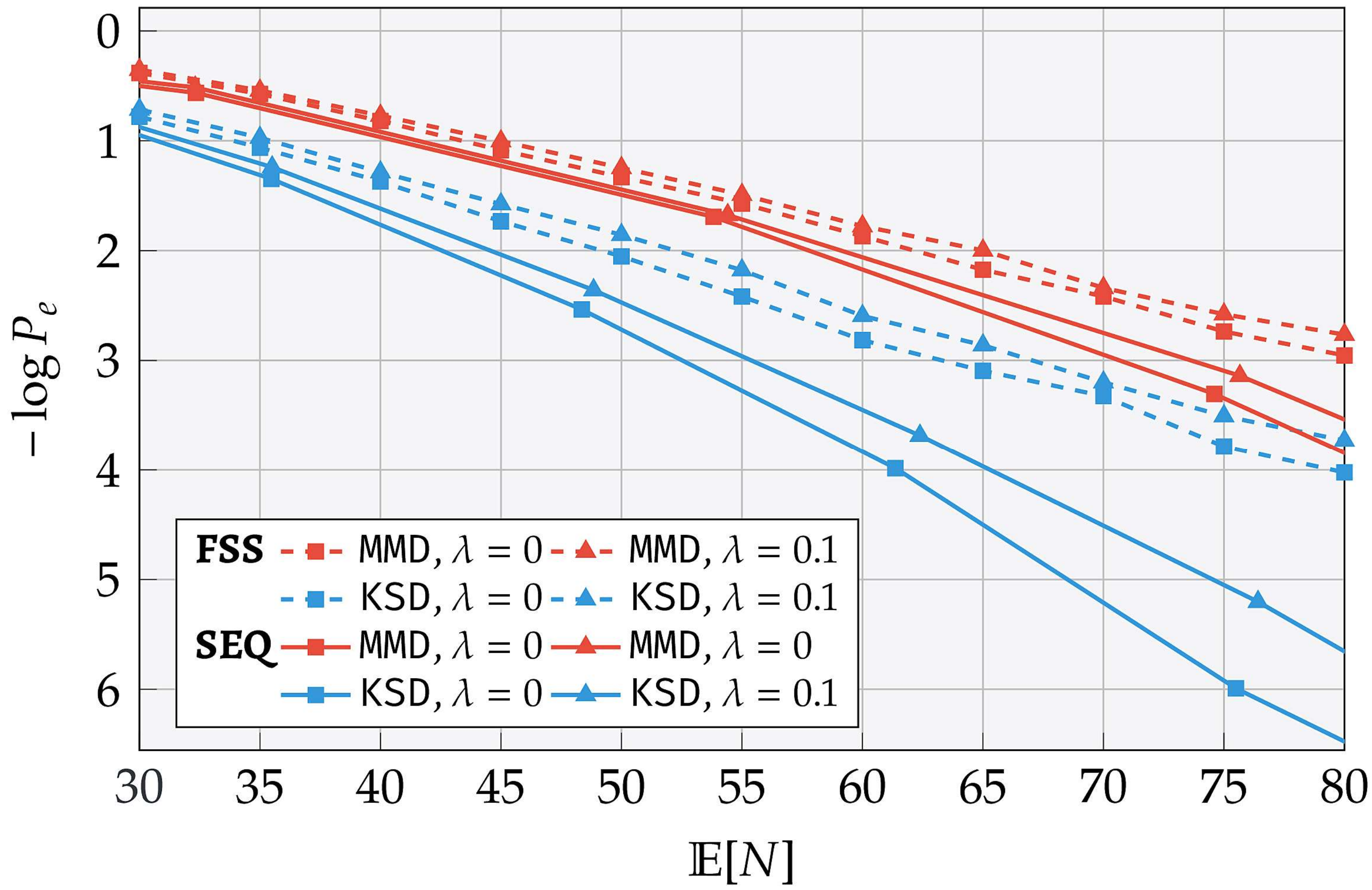




Simulation Setting

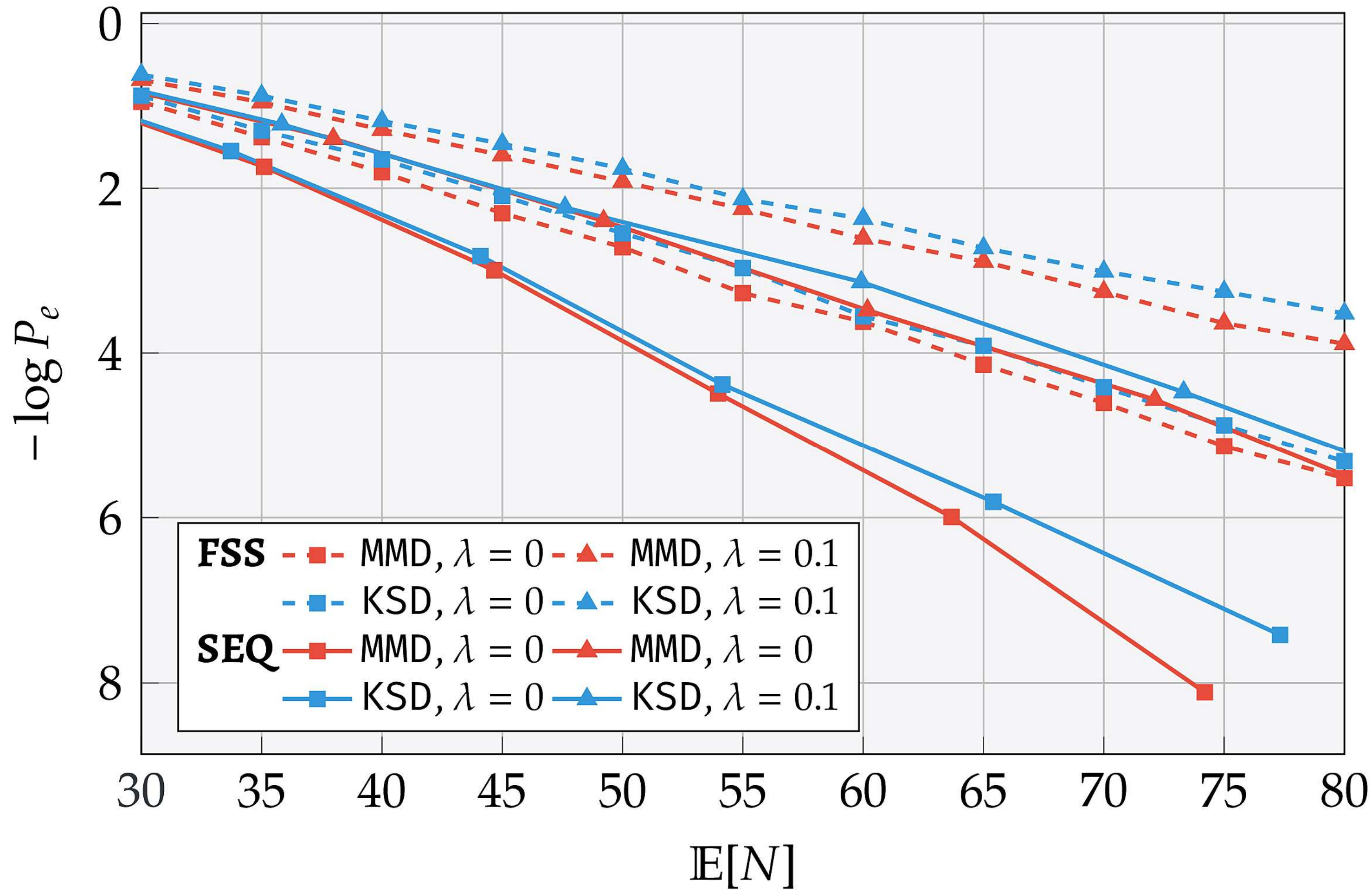
	Gaussian $\mathcal{N}(\mu, 1)$		Gamma $\Gamma(\mu, 1)$	
	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 0$	$\lambda = 0.1$
P_1	$\{0\}$	$\{-0.1, -0.05, 0.0, 0.05, 0.1\}$	$\{1.0\}$	$\{0.9, 0.95, 1.0, 1.05, 1.1\}$
P_2	$\{1\}$	$\{0.9, 0.95, 1.0, 1.05, 1.1\}$	$\{3.5\}$	$\{3.4, 3.45, 3.5, 3.55, 3.6\}$
P_3	$\{2\}$	$\{1.9, 1.95, 2.0, 2.05, 2.1\}$	$\{6.0\}$	$\{5.9, 5.95, 6.0, 6.05, 6.1\}$
P_4	$\{3\}$	$\{2.9, 2.95, 3.0, 3.05, 3.1\}$	$\{8.5\}$	$\{8.4, 8.45, 8.5, 8.55, 8.6\}$
P_5	$\{4\}$	$\{3.9, 3.95, 4.0, 4.05, 4.1\}$	$\{11.0\}$	$\{10.9, 10.95, 11.0, 11.05, 11.1\}$

Results: Known number of clusters



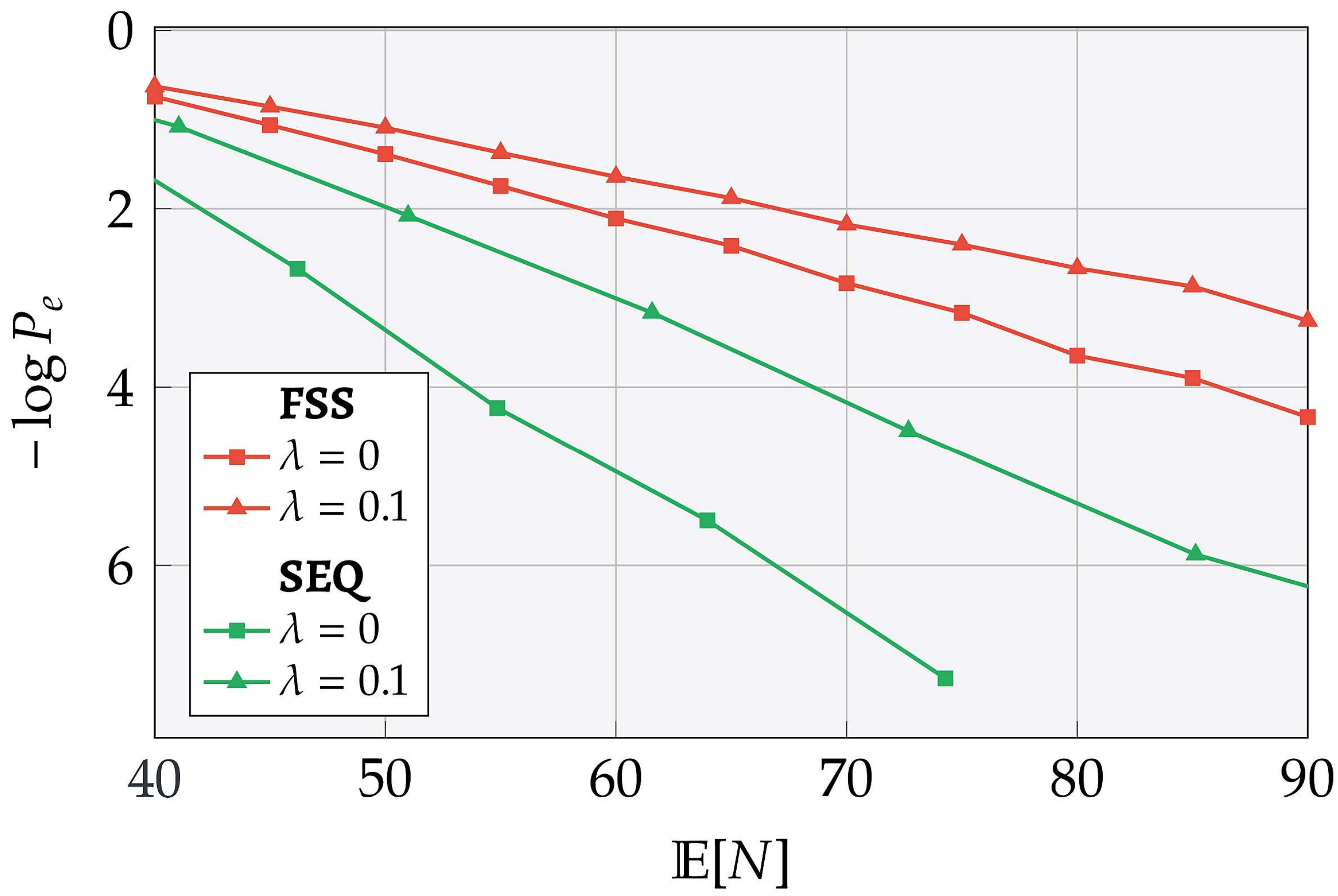
- Gamma distributions case: KSD better than MMD
- Fewer samples required than FSS clustering on average

Results: Known number of clusters



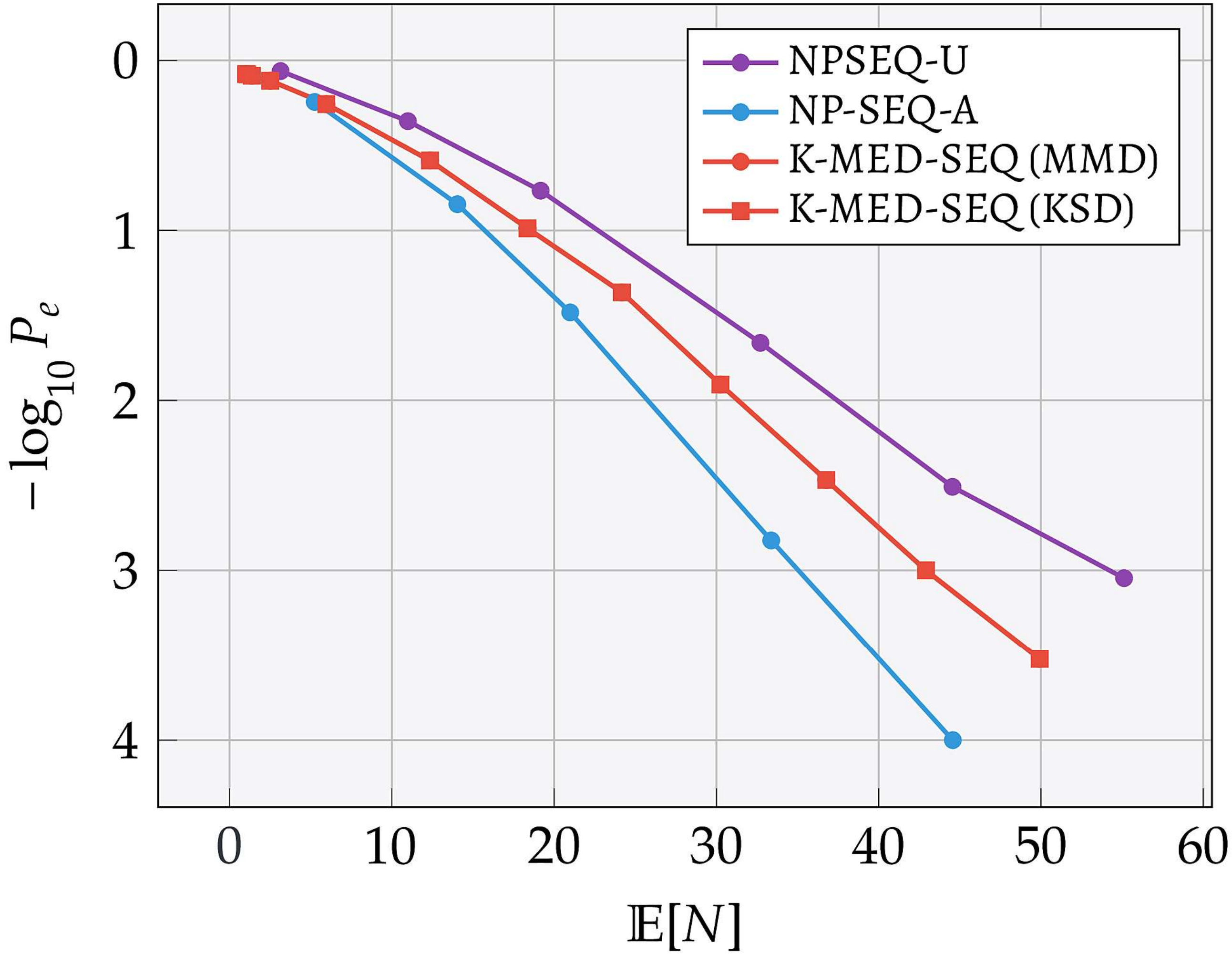
- Gaussian distributions case: MMD better than KSD
- Fewer samples required than FSS clustering on average

Results: Unknown number of clusters



- Gaussian distributions case, K-MED + Merge
- Fewer samples required than FSS clustering on average

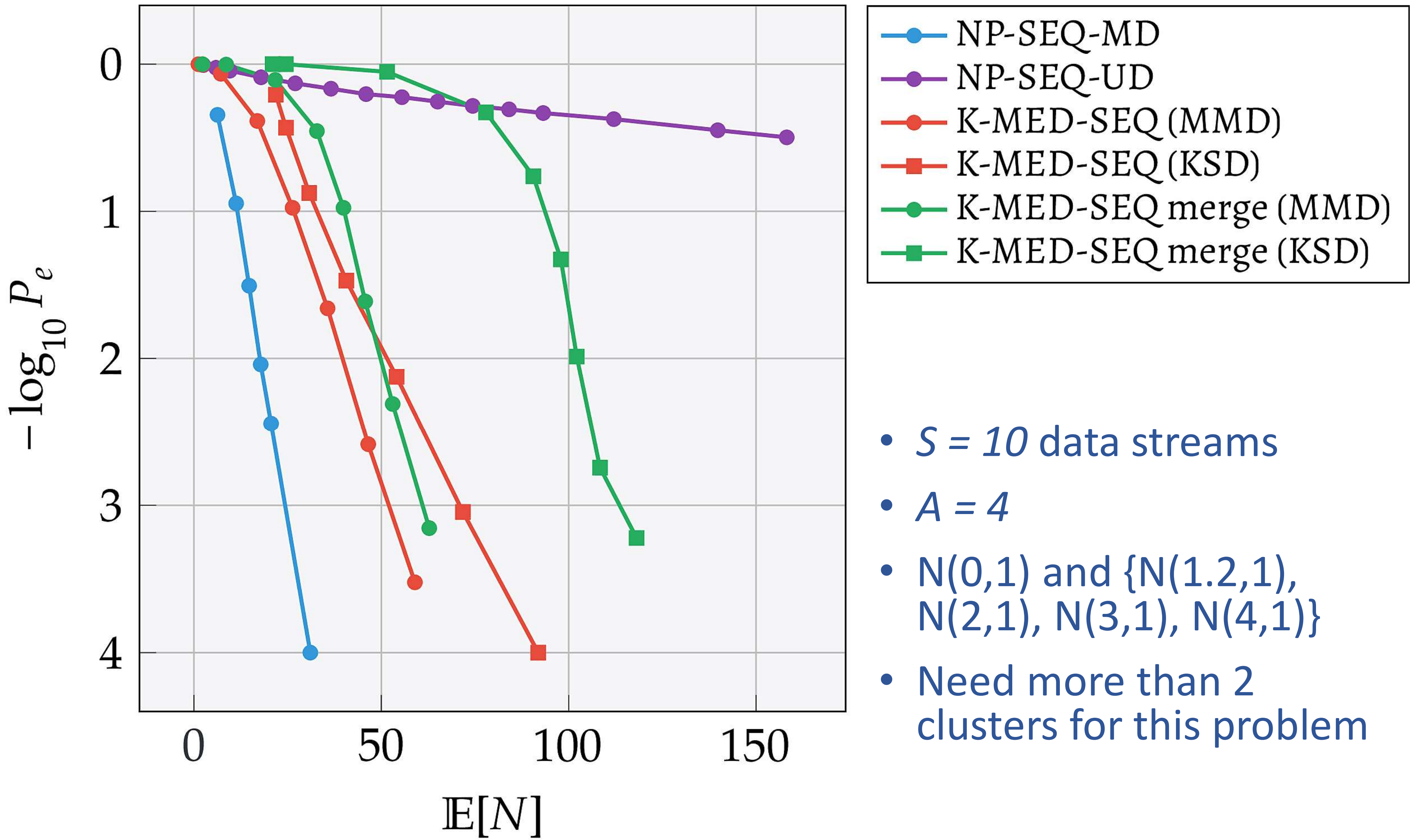
Multiple Anomalies



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

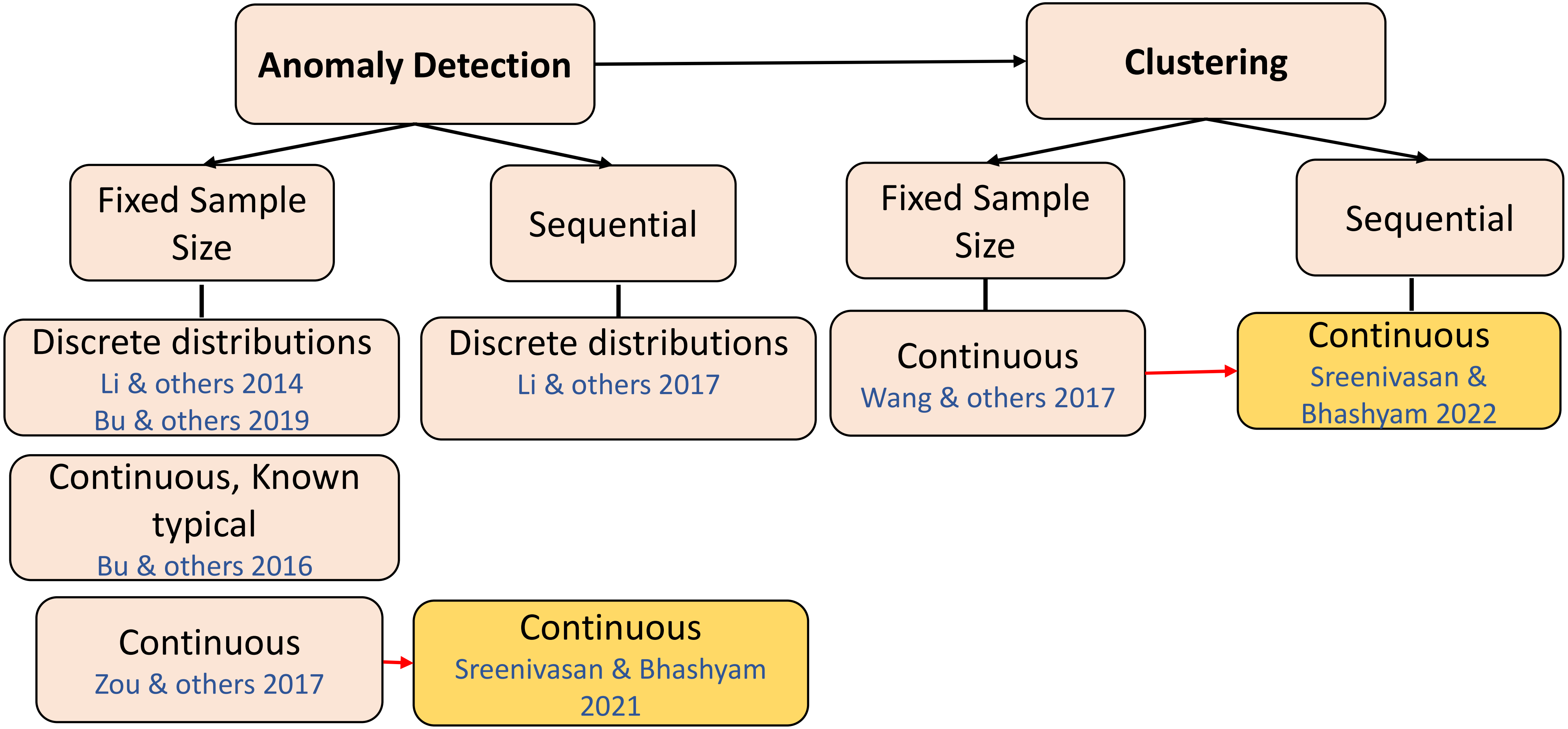
- NP-SEQ-A: Known A
- NP-SEQ-U: Unknown A

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



- Universally consistent sequential tests for anomaly detection and clustering

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

Possible Extensions

- More general cases
  - $d_L > d_H$, for both FSS and Sequential settings
  - Higher dimensional observations

- Clustering with bandit feedback/controlled sampling

- More than consistency
  - Bound on error exponent and optimality
  - Second-order asymptotic analysis

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