Change Detection and Localization With Degree-*k* Nearest Neighbors

Symposium on Data Science & Statistics, 31 May 2019

Matthew A. Hawks

Mathematics Department, United States Naval Academy

mhawks@usna.edu

Outline

- Background
  - Hypothetical Scenario
  - Graphs
    - Minimum Spanning Tree
    - Nearest Neighbor
    - Minimum Non-Bipartite Matching
- Innovations
  - Higher Degree Nearest Neighbors
  - Change Localization

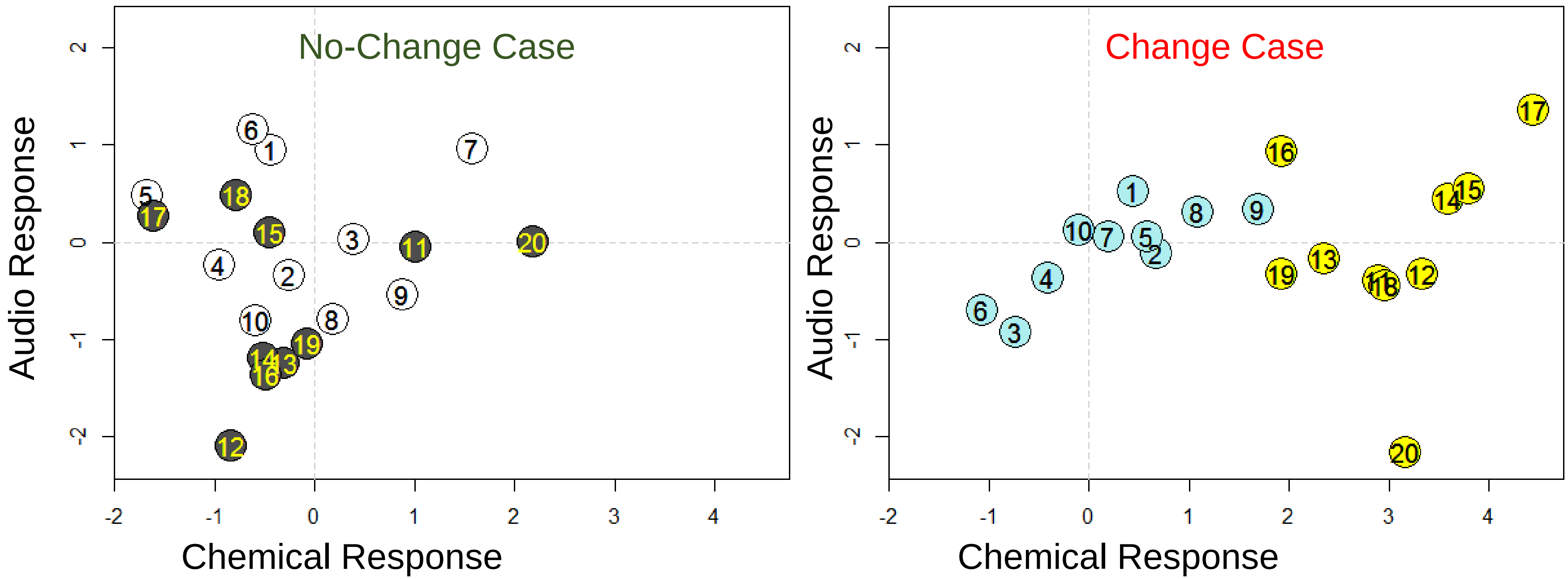
Hypothetical Scenario

Subsea sensors

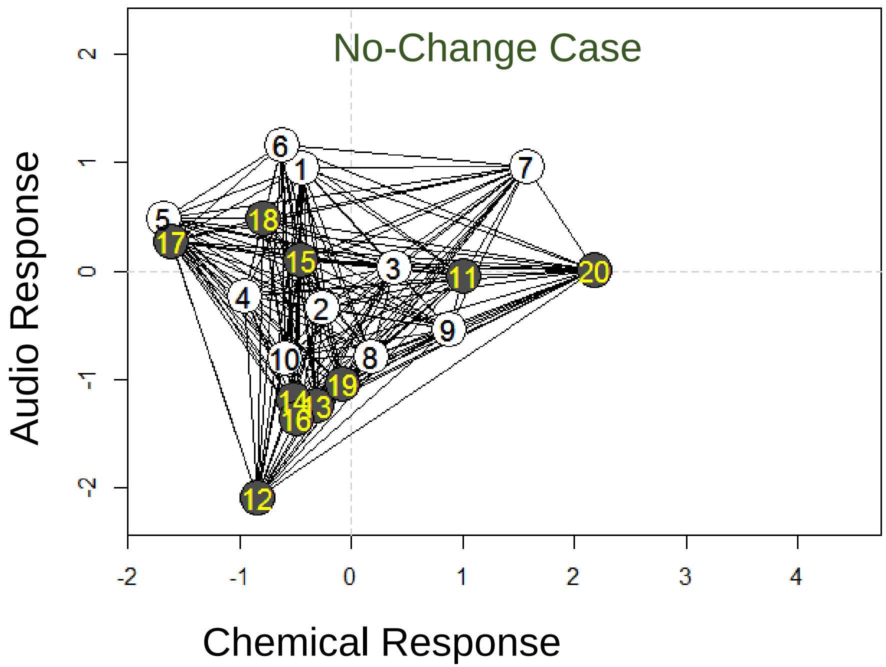
- Two types of measurements are taken over time, as an (x, y) couple
  - Chemical
  - Audio
- Plot measurements on one set of axes
- Label the measurements chronologically
- Treat each data point as vertex

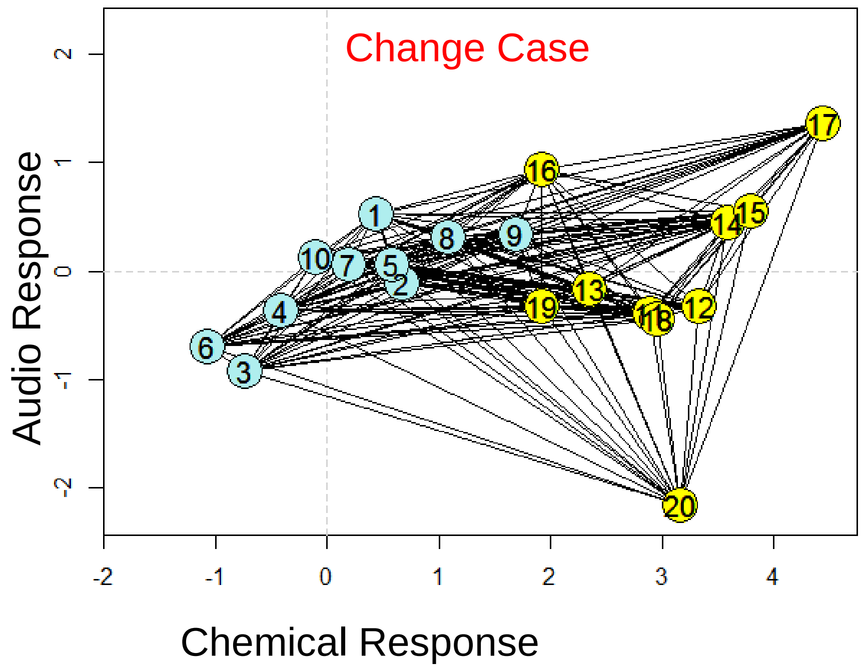
Example – 2 cases, 20 observations each

10 before a potential change, 10 after a potential change



Complete Graph – All possible edges



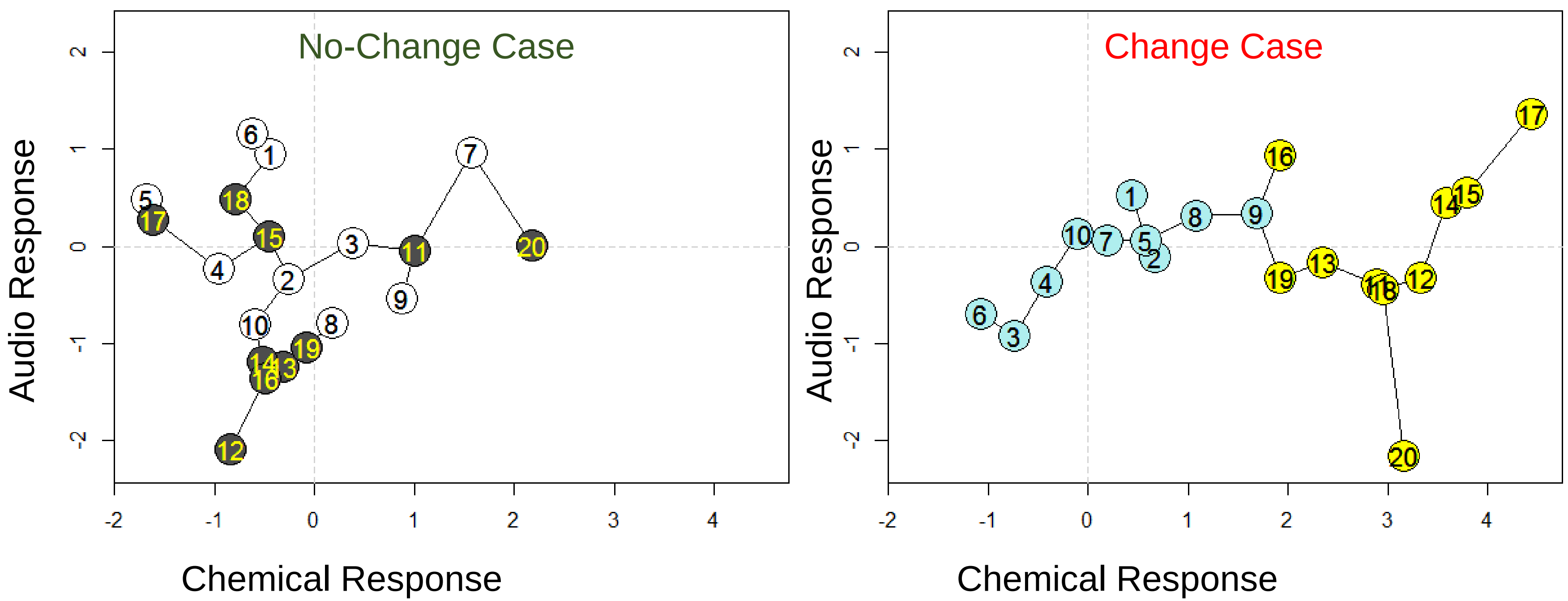


The complete graph is the starting point for extracting graphs

Minimum Spanning Tree (MST)

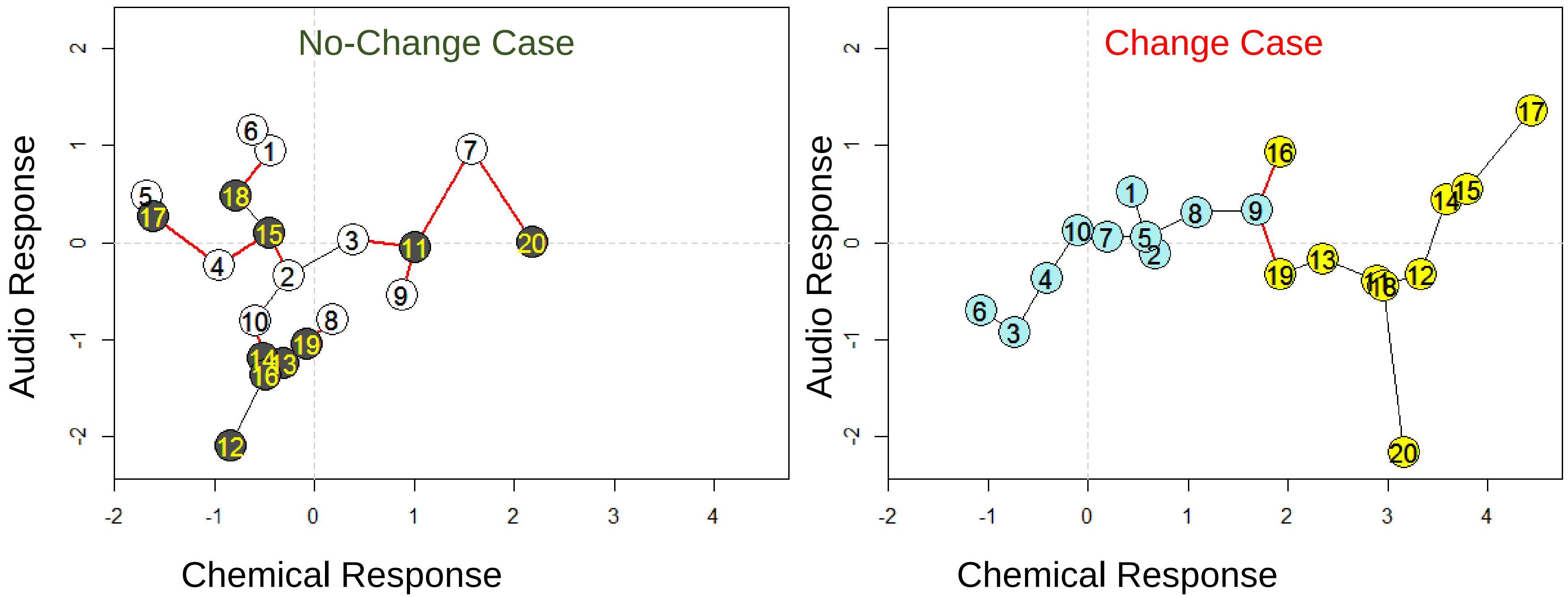
- A ***spanning*** graph connects all vertices – you can get to each vertex without lifting your pencil.
- The particular edges forming a ***Minimum*** Spanning Tree minimize some attribute of the graph
  - Combined edge weight
  - Total cost
  - Distance
- For |Vertices| = N, |MST Edges| = $N - 1$
- Advantage: the greedy algorithm produces the exact solution

Minimum Spanning Tree (MST)



Minimum Spanning Tree (MST)

Color the edges linking earlier data to later data red,
then count the number of trees if the red edges are removed



R: number of remaining trees. *R* ϵ {2, 3, …, *N*}

Generating a Statistic based on the Minimum Spanning Tree

Nomenclature

N: number of observations.

e: MST edge number, from 1 to $N - 1$.

x_e: indicator for the presence of a MST edge connecting different groups.

R: statistic counting the number of trees remaining after removing the linking edges.

$$x_e = \begin{cases} 1 & \text{if edge } e \text{ links vertices from different groups} \\ 0 & \text{otherwise} \end{cases}, \quad \forall e \in [1, N-1]$$

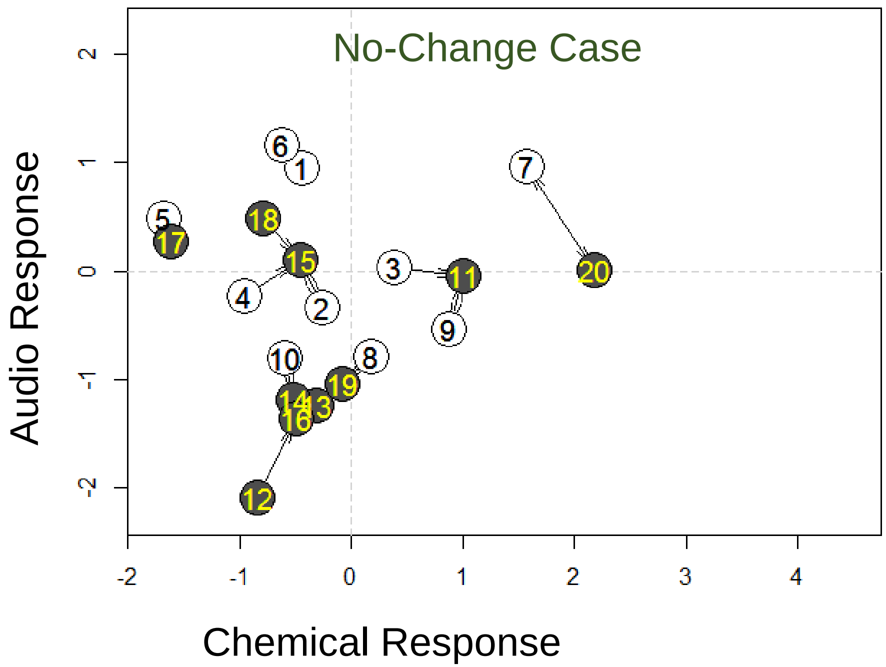
$$R = \sum_{e=1}^{N-1} x_e + 1.$$

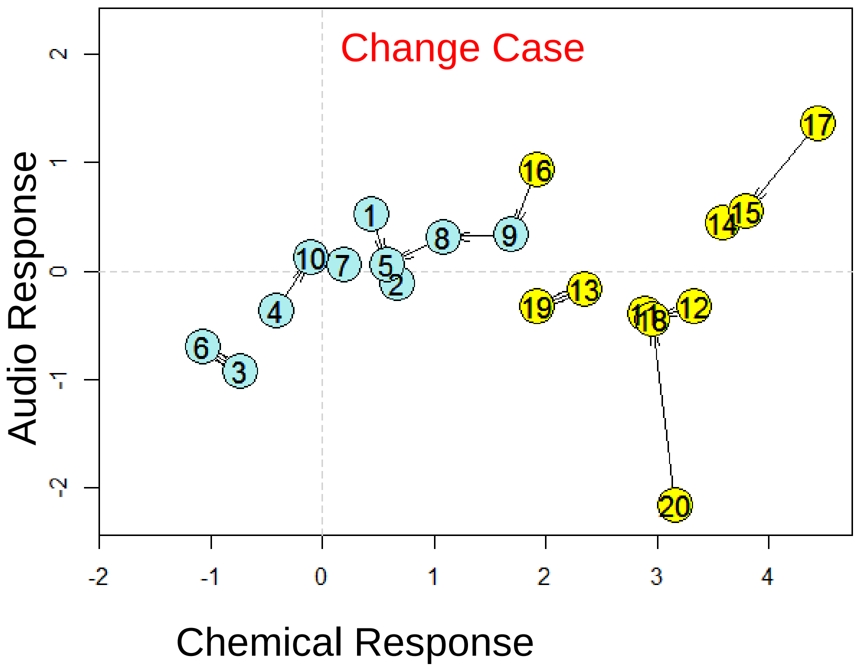
- It is possible to find the probability ($R \leq$ any value) if there is change present.
  - We counted 12 trees (no change) and 3 trees (change). Likelihoods are 0.679 and 0.000093.
- We can reasonably conclude (because R = 3) that a change took place.

Friedman & Rafsky (1979)

Nearest Neighbors

Count edges connecting two vertices from the <u>same</u> group





Generating a Statistic based on the Nearest Neighbors

NN_i : the NN of observation i, $i \in [1, N]$

$$I_i = \begin{cases} 1 \text{ if } NN_i \text{ belongs to the same sample as observation } i \\ 0 \text{ otherwise} \end{cases}$$

$$T_N = \sum_{i=1}^{N} I_i.$$

- It is possible to find the probability (T_N ≥ any value) if there is change present.
  - We counted 8 edges (no change) and 19 edges (change). Likelihoods are 0.752 and 0.0005.
- We can reasonably conclude (because T_N = 19) that a change took place.

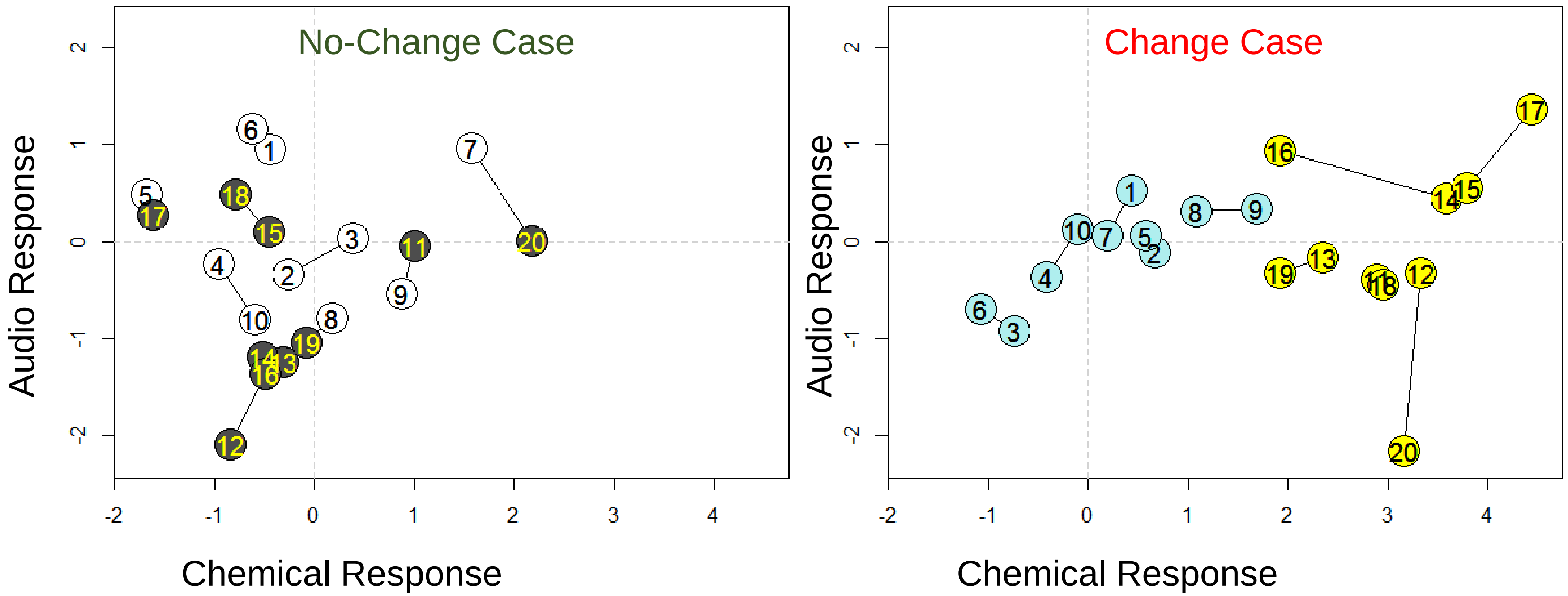
Schilling (1986)
Henze (1988)

Minimum Non-Bipartite Matching

- In Graph Theory, Bipartite Matching is obtained by matching two “partite,” or separate, groups.
  - Workers to jobs
  - Guests to hosts
  - Students to advisors
- Non-bipartite matching is when you form links without regard for original group membership.
- Minimum non-bipartite matching forms links that minimize an attribute – in our case, the total distance of the edges.
- The edge cardinality is $\lfloor N/2 \rfloor$.

Minimum Non-Bipartite Matching

Count edges connecting two observations from different groups



Generating Statistics based on Minimum Non-Bipartite Matching

i, j: observation labels

$\{i, j\}$: edge from observation *i* to observation *j*

x_{ij}: edge indicator

$$x_{ij} = \begin{cases} 1, & \text{if } \{i, j\} \in \text{MNBM} \\ 0, & \text{otherwise} \end{cases}, \quad \forall \{i, j\} \in E.$$

d_{ij}: distance from observiation *i* to observation *j*

$$\boxed{\begin{aligned} &\min_{\mathbf{x}} \sum_{\{i,j\} \in E} d_{ij} x_{ij} \\ &\text{s.t.} \quad \sum_{j \in V} x_{ij} = 1, \ \forall i \in V \end{aligned}}$$

cross-match statistic: $A = \sum_{i=11}^{N} \sum_{j=1}^{10} x_{ij}$

sum of pair-maxima: $T_N = \sum_{i=2}^{N} \sum_{j=1}^{i-1} i x_{ij}.$

- It is possible to find the probability ($A \leq$ any value) if there is change present.
    - We counted 4 edges (no change) and 0 edges (change). Likelihoods are 0.433 and 0.0014.
- We can reasonably conclude (because A = 0) that a change took place.
- We also can use T_N and show the change case is below a critical value.

Rosenbaum (2005)
Ruth (2009)

Research aims

- What happens to the ability to detect change when you use more complex graphs?
  - Introduce more nearest neighbors (degree $k > 3$) to increase detection power

- Can you estimate at what point in the data sequence a change has occurred using these graphs?
  - These techniques can accurately estimate a change point.

Higher Degree Nearest Neighbors

k : degree; number of edges to nearest neighbors

i, j : observation labels $\quad \{i, j\}$: edge from observation i to observation j

x_{ij} : edge indicator $\quad x_{ij} = \begin{cases} 1, & \text{if } \{i,j\} \in k\text{-NN} \\ 0, & \text{otherwise} \end{cases}, \quad \forall \{i,j\} \in E.$

d_{ij} : distance from observiation i to observation j

- For our k-NN subgraph, we derive a statistic T, with an edge score function $h(i, j)$.

$$\boxed{T = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} h(i,j)\, x_{ij}.}$$

- Now $E[T] = kN\bar{h}/2.$

Selection of Edge Score Function

$$\boxed{T = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} h(i,j)\, x_{ij}.}$$

- Investigate Box-Cox transformation family of score functions based on the pair-difference:

$$h(i,j) = \begin{cases} \dfrac{|i-j|^{\lambda} - 1}{\lambda}, & (\lambda \neq 0) \\ \log|i-j|, & (\lambda = 0) \end{cases}, \quad \forall (i,j) \in E,$$

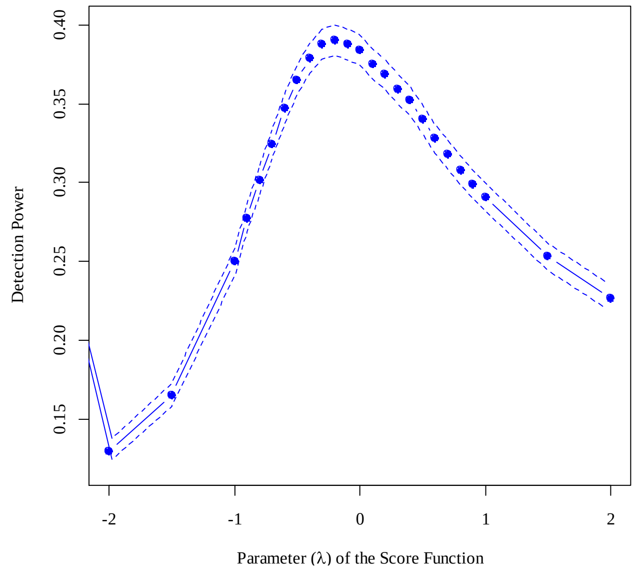
- Is it better to augment, suppress, or let stand the pair-difference?
- Evaluate performance over a range of parameter λ
- Two competing objectives:
  - Maximum detection power
  - Interpretable λ

Simulations

- Simulations vary the following factors:
  - Sampling distribution
    - Normal: $\boldsymbol{\mu} = \mathbf{0}$, $\Sigma = I_p$
    - Outlier-prone normal mixture: $N_p(\mathbf{0}, I_p)$ and $N_p(\mathbf{0}, \psi^2 I_p)$, $\psi > 1$
    - Skew (Weibull): $\eta = 1.5$, $\boldsymbol{\beta} = \mathbf{1}$
  - Dimension: 5 or 20 variables
  - Change parameter
    - mean change in one component: $\mu_1 = \Delta$
    - scale change in all components: $\Sigma = (1 + \Delta)^2 I_p$ or $\psi^2(1 + \Delta)^2 I_p$
    - Weibull scale parameter β change in one component: $\beta_1 = 1 + \Delta$
  - Change magnitude Δ : 0.25, 0.5, 0.75, 1.0
- Simulations use sample size $N = 200$, change point $\tau = 101$, level of test $\alpha = 0.05$, and 10,000 simulations per vignette.
- Critical values based on 100,000 null hypothesis simulations.

Selection of Edge Score Function

Example Plot: Multivariate Normal, Mean Change 0.5



$$h(i,j)=\begin{cases}\dfrac{|i-j|^{\lambda}-1}{\lambda}, & (\lambda\neq 0)\\ \log|i-j|, & (\lambda=0)\end{cases},\quad \forall(i,j)\in E,$$

95% CI for maximum change detection power: λ = [-0.4, 0.1]

Finding: Suppress the pair-difference

Test Approach for Nearest Neighbors

- We have $\boxed{T = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} |i - j|^{-0.3} x_{ij}.}$
- We can also use the asymmetry of **x**.
- Count the number of *k*-nearest neighbors whose index is less than the current index, for all indices.

$$\boxed{K = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} I\{i > j\} x_{ij} = \frac{1}{2}\sum_{i=2}^{N}\sum_{j=1}^{i-1} x_{ij}.}$$

$$E[K] = \frac{kN}{4}.$$

- Combine *T* and *K* for a composite test:

$$\text{Test Decision} = \begin{cases} \text{reject } H_0, \text{ if either } T \text{ or } K \text{ reject at a } \alpha = 0.025 \text{ level of test;} \\ \text{fail to reject } H_0, \text{ otherwise.} \end{cases}$$

Selection of k: kNN

Detection Power as k Varies, N = 200

Simulation Scenario	10	20	30	40	50	60	70	80	90	100
F_{MVN} mean jump 0.75	0.41	0.54	0.61	0.65	0.68	0.69	0.71	0.71	0.73	0.73
F_{MVN} mean drift 0.75	0.16	0.21	0.24	0.27	0.28	0.29	0.30	0.31	0.30	0.31
F_{MVN} mean jump 0.75 p = 20	0.19	0.26	0.29	0.32	0.33	0.36	0.37	0.37	0.37	0.37
F_{MVN} mean drift 0.75 p = 20	0.10	0.12	0.12	0.13	0.13	0.14	0.14	0.14	0.15	0.15
F_{MVN} scale jump 0.25	0.92	0.96	0.96	0.96	0.96	0.96	0.97	0.97	0.97	0.98
F_{MVN} scale drift 0.25	0.67	0.71	0.73	0.72	0.74	0.75	0.76	0.77	0.79	0.80
F_{mix} mean jump 0.75	0.29	0.39	0.43	0.47	0.49	0.49	0.48	0.47	0.45	0.44
F_{mix} mean drift 0.75	0.12	0.16	0.17	0.19	0.19	0.19	0.19	0.19	0.18	0.18
F_{mix} mean jump 0.75 p = 20	0.12	0.13	0.13	0.14	0.13	0.13	0.11	0.12	0.10	0.10
F_{mix} scale jump 0.25	0.67	0.74	0.75	0.77	0.76	0.77	0.77	0.77	0.77	0.75
F_{mix} scale drift 0.25	0.40	0.46	0.47	0.47	0.48	0.50	0.50	0.51	0.51	0.50
F_{Weib} β jump 0.75	0.58	0.74	0.80	0.83	0.85	0.86	0.87	0.88	0.89	0.89
F_{Weib} β drift 0.75	0.33	0.42	0.47	0.51	0.53	0.56	0.56	0.60	0.61	0.62
F_{Weib} β jump 0.75 p = 20	0.41	0.52	0.59	0.63	0.65	0.66	0.68	0.68	0.68	0.67

Table Value = Fraction of 10,000 simulations rejecting H_0.

Dimension $p = 5$, unless otherwise noted.

Using k = N/2 provides nearly optimal results for all scenarios, increasing change detection power compared to lower values of k

Change Detection Power Comparison

Detection Power, N = 200

Simulation Scenario	JJS	kNN
F_{MVN} mean jump 0.50	0.53	0.31
F_{MVN} mean drift 0.75	0.54	0.30
F_{MVN} mean jump 0.75 p = 20	0.64	0.38
F_{MVN} mean drift 1.00 p = 20	0.49	0.29
F_{MVN} scale jump 0.25	0.09	0.98
F_{MVN} scale drift 0.50	0.26	1.00
F_{mix} mean jump 0.50		0.16
F_{mix} mean drift 0.75		0.17
F_{mix} mean jump 0.75 p = 20		0.10
F_{mix} scale jump 0.25		0.74
F_{mix} scale drift 0.25		0.50
F_{Weib} β jump 0.50		0.55
F_{Weib} β drift 0.75		0.61
F_{Weib} β jump 0.50 p = 20		0.32

JJS Type I error greatly exceeds 0.05 for the non-normal distributions

Table Value = Fraction of 10,000 simulations rejecting H_0.

Dimension p = 5, unless otherwise noted.

*k*NN test is inferior to James, James, Sigmund for MVN mean change, but is superior to for scale changes and is not tied to distribution.

Estimate Change Point

Single Abrupt Change

- Prior edge: an edge of the minimum subgraph whose destination observation label is less than the source observation label.
- Observations just before the change point should have more prior edges than observations just after the change point.

$$\boxed{L_i = \sum_{j=1}^{i-1} x_{ij}, \forall i \in \{2, \ldots, N\}.}$$

- Under the null hypothesis, the prior edge count has a hypergeometric distribution:

$$E[L_i] = k\left(\frac{i-1}{N-1}\right)$$

$$V[L_i] = k\left(\frac{i-1}{N-1}\right)\left(\frac{N-i}{N-1}\right)\left(\frac{N-1-k}{N-2}\right).$$

- We produce the standardized prior edge counts: $\boxed{Z_i = \frac{L_i - E[L_i]}{\sqrt{V[L_i]}}.}$
- Optimize the fit of a two-piece regression line to the standardized prior edge counts.

Location Estimation Process

- Obtain the minimum subgraph
- Test for change detection.
- If change is detected, obtain and standardize prior edge count (Z_i).
- For each candidate break point $\tau \in \{\tau_0, \ldots, N - \tau_0\}$, calculate the mean squared error for the two-piece regression line:

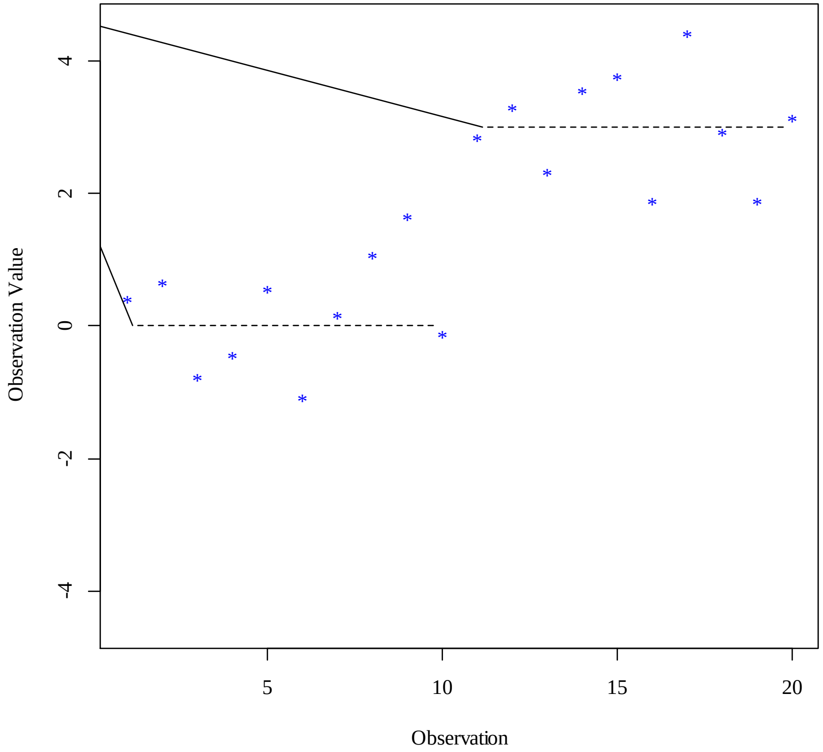
$$MSE(\tau) = \sum_{i=2}^{N} \left(Z_i - \hat{\beta}_{0,\tau} - \hat{\beta}_{1,\tau} i - \hat{\beta}_{2,\tau} I\{i > \tau\} - \hat{\beta}_{3,\tau} i I\{i > \tau\} \right)^2.$$

- Change-point estimator is the break point that minimizes MSE.

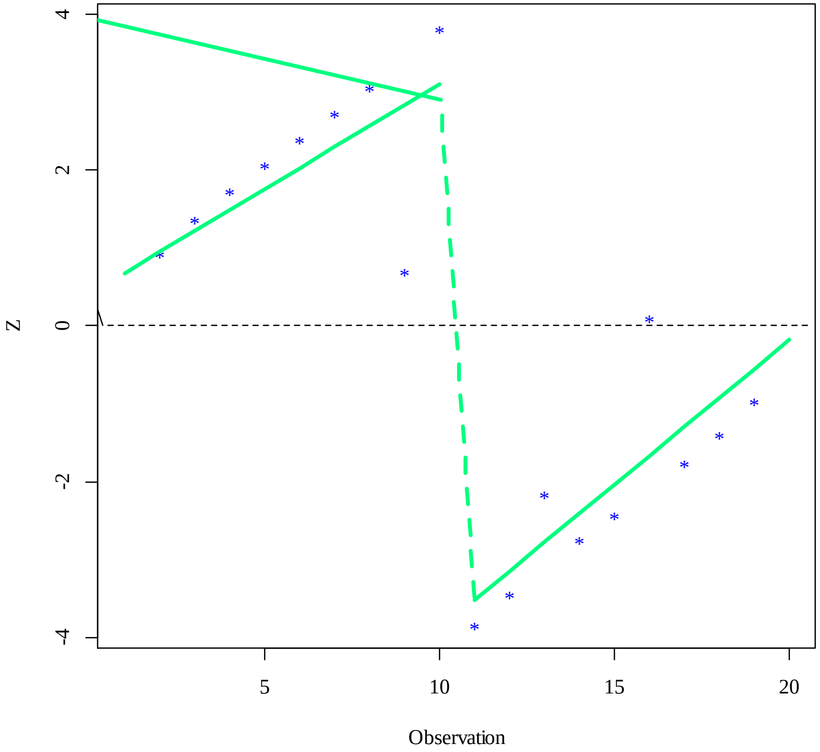
Location Estimation Technique

$N = 20$, $p = 2$, $\Delta = 3$

Value for Changed Variable



Standardized Prior Edge Count



Actual change point

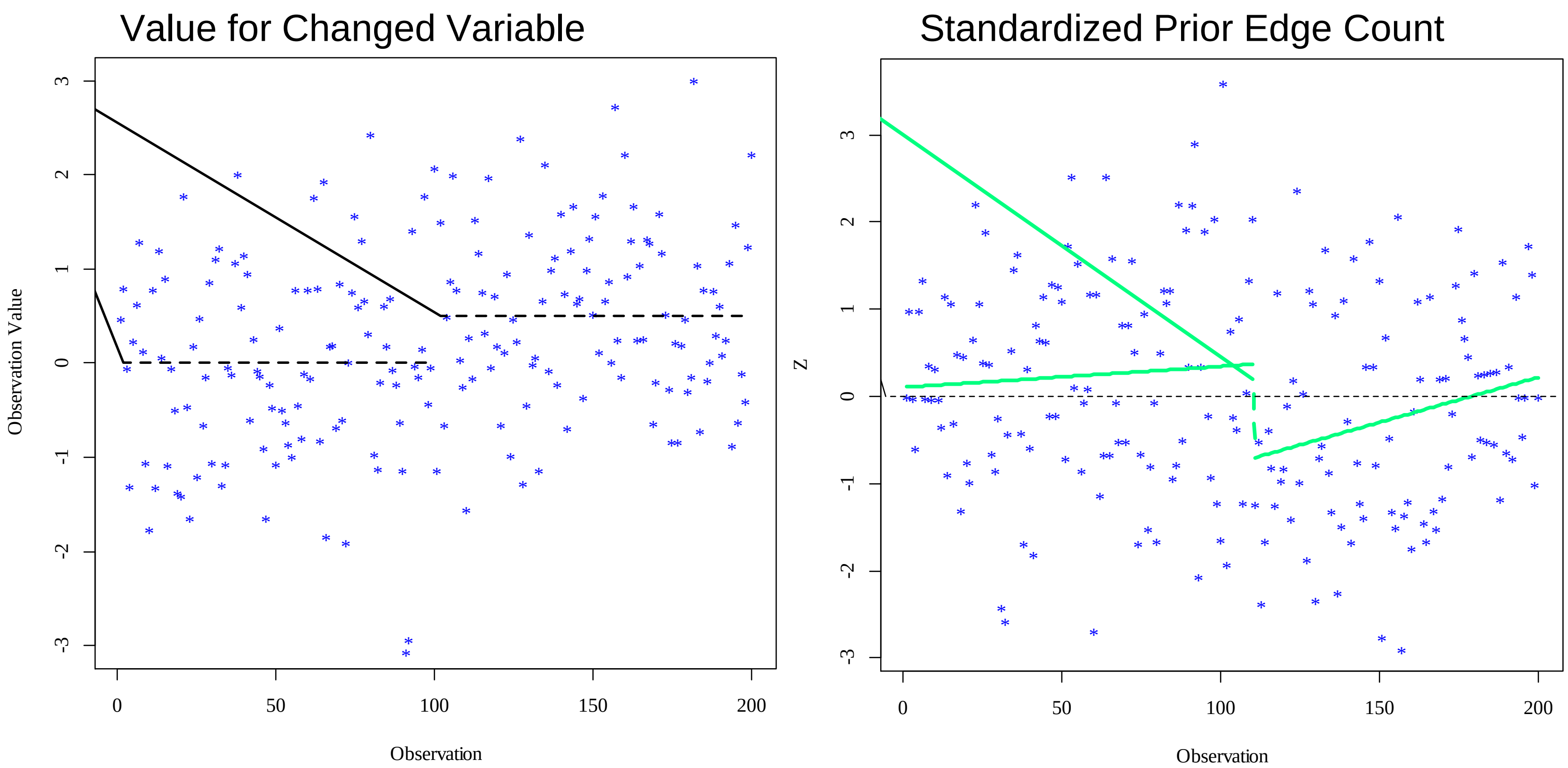
$\tau = 11$

Estimated change point

$\hat{\tau} = 11$

Location Estimation Technique

N = 200, p = 5, Δ = 0.5



Actual change point $\tau = 101$

Estimated change point $\hat{\tau} = 110$

Location Estimation Results

- For success, must both detect a change and estimate accurately: $\hat{\tau} = \tau \pm 0.025N$

F_{MVN} Δ	Mean Jump	Scale Jump
0.25	0.01 (71)	0.23 (28)
0.50	0.08 (40)	0.58 (11)
0.75	0.38 (10)	0.85 (2)
1.00	0.73 (4)	0.96 (1)

F_{mix} Δ	Mean Jump	Scale Jump
0.25	0.01 (69)	0.13 (42)
0.50	0.03 (57)	0.35 (20)
0.75	0.15 (22)	0.58 (12)
1.00	0.42 (9)	0.74 (4)

F_{Weib} Δ	β Jump
0.25	0.02 (56)
0.50	0.21 (22)
0.75	0.51 (9)
1.00	0.69 (5)

Success rate (and interquartile range) from 1000 simulations per vignette

- *k*NN can be used to accurately estimate the change point.

Conclusions & Next Steps

- Change Detection and Localization with higher degree nearest neighbors can work.
  - Using $k = N/2$ provides nearly optimal results for all scenarios, increasing change detection power compared to lower values of k
  - Higher degree nearest neighbors can also be used to accurately estimate the change point.
- Document verification of techniques using UCI Machine Learning Repository data sets.

Selected References

- Friedman, J. H., & Rafsky, L. C. (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *The Annals of Statistics,* 697-717.
- Hawks, M. A. (2015). *Graph-Theoretic Statistical Methods for Detecting and Localizing Distributional Change in Multivariate Data* (Doctoral dissertation, Naval Postgraduate School).
- Henze, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *The Annals of Statistics,* 772–783.
- James, B., James, K. L., & Siegmund, D. (1992). Asymptotic approximations for likelihood ratio tests and confidence regions for a change-point in the mean of a multivariate normal distribution. *Statistica Sinica*, *2*(1), 69-90.
- Rosenbaum, P. R. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(4), 515-530.
- Ruth, D. M. (2009). *Applications of Assignment Algorithms to Nonparametric Tests for Homogeneity* (Doctoral dissertation, Naval Postgraduate School).
- Ruth, D. M., & Koyak, R. A. (2011). Nonparametric tests for homogeneity based on non-bipartite matching. *Journal of the American Statistical Association, 106*(496), 1615-1625.
- Schilling, M. F. (1986). Multivariate two-sample tests based on nearest neighbors. *Journal of the American Statistical Association*, *81*(395), 799–806.

Change Detection and Localization With Degree-*k* Nearest Neighbors

Symposium on Data Science & Statistics, 31 May 2019

Matthew A. Hawks

Mathematics Department, United States Naval Academy

mhawks@usna.edu