Motivation

Design a nonparametric change-point detection method (M-statistic) in both the offline and online setting.

- **Distribution Free:** kernel approaches are distribution free as they provide consistent results over larger classes of data distributions.
- **Efficient to compute:** split data to compute the offline M-statistics in a novel structured way; update M-statistics recursively in the online setting.
- **Analytical way to obtain threshold:** accurately characterize the tail probability of the M-statistics under null hypothesis in both offline and online setting.
- **Powerful:** numerically demonstrate that our algorithm is more powerful and more robust compared to conventional parametric approaches (e.g., Hotelling’s T²).

### MMD

Assume there are two sets with n observations from a domain X, where \( X = \{x_1, x_2, \ldots, x_n\} \) are drawn iid from distribution \( P \), and \( Y = \{y_1, y_2, \ldots, y_n\} \) are drawn iid from distribution \( Q \).

The **maximum mean discrepancy** (MMD) is defined as [1]

\[
\text{MMD}_{X,Y}^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} h(x_i, x_j) - h(y_i, y_j),
\]

where \( h(\cdot) \) is the kernel of the U-statistic defined as

\[
h(x_i, x_j, y_i, y_j) = k(x_i, x_j) + k(y_i, y_j) - k(x_i, y_j) - k(x_j, y_i).
\]

Intuitively, the empirical test statistic \( \text{MMD}^2 \) is expected to be small (close to zero) if \( P = Q \), and large if \( P \) and \( Q \) are far apart.

### Offline M-statistic

**Examples of M-Statistics**

- Search for a location \( B (2 \leq B \leq B_{max}) \) for a change-point.

\[
Z_B := \frac{1}{N} \sum_{i=1}^{N} \text{MMD}^2(x_i(B), y_i(B)) = \frac{1}{NB(B-1)} \sum_{i=2}^{B} \left( \sum_{j=1}^{i-1} h(x_i(B), x_j(B), y_i(B), y_j(B)) \right)
\]

- Detect a change-point whenever the M-statistic exceeds the threshold \( b > 0 \):

\[
M := \max_{B \in [2, \ldots, B_{max}]} Z_B / \sqrt{\text{Var}(Z_B)} > b.
\]

**Online M-statistic**

**Theorem 1. Significance Level Approximation.** When \( b \to \infty \) and \( b/\sqrt{B_{max}} \to c \) for some constant \( c \), the significant level of the offline M-statistic is given by

\[
P \{ \text{MMD} > b \} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(2B-1)/\sqrt{2(2B-1)}} e^{-u^2/2} du + o(1),
\]

where the special function

\[
\nu(u) \approx \frac{1}{2\alpha} \Phi(\alpha/2) - 0.5
\]

is the pdf and \( \Phi(x) \) is the cdf of the standard normal distribution, respectively.

**Theorem 2. Average Run Length (ARL) Approximation.** When \( b \to \infty \) and \( b/\sqrt{B_{max}} \to c' \) for some constant \( c' \), the average run length (ARL) of the stopping time \( T \) is given by

\[
E_T[\nu(u)] = \frac{2}{\sqrt{\nu(0)}} \nu(u/2) \Phi(u/2) + o(1).
\]

**Power Analysis**

<table>
<thead>
<tr>
<th>Case</th>
<th>M-statistic</th>
<th>Hotelling’s T²</th>
<th>GLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.71</td>
<td>0.18</td>
<td>0.03</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>0.88</td>
<td>0.05</td>
</tr>
<tr>
<td>3</td>
<td>0.26</td>
<td>0.07</td>
<td>0.12</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>0.87</td>
<td>0.03</td>
</tr>
<tr>
<td>5</td>
<td>0.44</td>
<td>0.12</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Note: Case 1 to 4, change from \( N(0, I) \) to \( N(0.1, I) \), \( N(0.2, I) \), \( N(0, \Sigma) \), \( N(0.2, \Sigma) \), respectively.

Case 5, change from \( N(0, I) \) to Laplace(0,1).

---