Analysis about Correlation Between MAMCOD p values

Ruoxiao Tang

July 2024

1 Recall: MAMCOD procedure

Generally, MAMCOD contains the following four steps:

- 1. Compute a data-driven score function of the general form.

$$g(X) = g(x, (X_1, X_2, \cdots, X_s), (X_{s+1}, \cdots, X_{n+m})),\ z \in \mathcal{D} \tag{1}$$

which satisfies that g(X) is invariant under any permutation π of $\{s+1, s+2, \cdots, n+m\}$.

- 2. Compute the adaptive scores

$$S_i^k = g(x, (X_1, X_2, \cdots, X_s), (X_{s+1}, \cdots, X_{n+m})),\ i \in \mathcal{D}_k^{cal} \cup \{n+1, \cdots, n+m\} \tag{2}$$

- 3. Consider the p-values $(p_1, p_2, \cdots, p_m)$, obtained by comparing each score of interest $(S_{s+1}^1, \cdots, S_n^K, S_{n+1}^1, S_{n+1}^2, \cdots, S_{n+m}^{K-1}, S_n^K$

$$p_i^k = \frac{1 + |j \in \mathcal{D}_k^{cal} : S_j^k \le S_i^k|}{1 + |\mathcal{D}_k^{cal}|},\ 1 \le k \le K \tag{3}$$

$$p_i = max(p_i^1, \cdots, p_i^K) \tag{4}$$

- 4. Finally apply the BH algorithm to the empirical p-values $(p_1, p_2, \cdots, p_m)$ at the desired level α.

The pseudocode of MAMCOD to get conformal p-values is summarized in Algorithm 1

Algorithm 1 MAMCOD conformal p-value

Input: null samples $\mathcal{D}^{ns}$ containing K different categories of inliers, test point X_{n+1}.
Randomly split $\mathcal{D}^{ns}$ into $\mathcal{D}^{train}$ for training and $\mathcal{D}^{cal}$ for calibration.
Divide $\mathcal{D}^{cal}$ and $\mathcal{D}^{train}$ into K part respectively, making each part containing only one type of inliers: $\mathcal{D}_1^{train}, \cdots, \mathcal{D}_K^{train}$, $\mathcal{D}_1^{cal}, \cdots, \mathcal{D}_K^{cal}$.
Train a data driven score function g with 1.
Compute the adaptive scores for each data in $\mathcal{D}^{cal}$ and X_{n+1} with 2
Compute the standard conformal p-values of each type of inliers with 3
Output: MAMCOD conformal p-value p_i with 4

2 Symbols and Notations

Assume that $(R_1^1, R_2^1, \cdots, R_{n_1}^1, R_{n+1}^1, R_{n+2}^1)$ be the rank of $(S_1^1, S_2^1, \cdots, S_{n_1}^1, S_{n+1}^1, S_{n+2}^1)$ in the ascending order, and let $(R_1^2, R_2^2, \cdots, R_{n_2}^2, R_{n+1}^2, R_{n+2}^2)$ be the rank of $(S_{n_1+1}^2, S_{n_1+2}^2, \cdots, S_n^2, S_{n+1}^2, S_{n+2}^1)$ in the ascending order. Here S_i^k means the scores obtained by applying the k-th OCC to X_i. And S_i follows continuous distribution and different a.s. By definition:

$$\begin{aligned} R_{n+1}^1 &= \left(\sum_{i=1}^{n_1} + \sum_{i=n+1}^{n+2}\right) \mathbb{I}\{S_{n+1}^1 \ge S_i^1\} \qquad R_{n+1}^2 = \left(\sum_{i=n_1+1}^{n+2}\right) \mathbb{I}\{S_{n+1}^2 \ge S_i^2\} \\ R_{n+2}^1 &= \left(\sum_{i=1}^{n_1} + \sum_{i=n+1}^{n+2}\right) \mathbb{I}\{S_{n+2}^1 \ge S_i^1\} \qquad R_{n+2}^2 = \left(\sum_{i=n_1+1}^{n+2}\right) \mathbb{I}\{S_{n+2}^2 \ge S_i^2\} \end{aligned} \tag{5}$$

Therefore, the MAMCOD p-value of test data X_{n+1} can be expressed as:

$$
\begin{aligned}
p_1 &= p_1^1 \vee p_1^2 && (6)\\
&= \frac{R_{n+1}^1 - \mathbb{I}\{S_{n+1}^1 \geq S_{n+2}^1\}}{n_1+1} \vee \frac{R_{n+1}^2 - \mathbb{I}\{S_{n+1}^2 \geq S_{n+2}^2\}}{n_2+1} && (7)\\
&= \frac{(n_2+1)\cdot(R_{n+1}^1 - \mathbb{I}\{S_{n+1}^1 \geq S_{n+2}^1\})}{(n_1+1)\times(n_2+1)} \vee \frac{(n_1+1)\cdot(R_{n+1}^2 - \mathbb{I}\{S_{n+1}^2 \geq S_{n+2}^2\})}{(n_1+1)\times(n_2+1)} && (8)
\end{aligned}
$$

the same expression for p_2 can be defined as:

$$
\begin{aligned}
p_2 &= p_2^1 \vee p_2^2 && (9)\\
&= \frac{R_{n+2}^1 - \mathbb{I}\{S_{n+2}^1 \leq S_{n+2}^1\}}{n_1+1} \vee \frac{R_{n+2}^2 - \mathbb{I}\{S_{n+1}^2 \leq S_{n+2}^2\}}{n_2+1} && (10)\\
&= \frac{(n_2+1)\cdot(R_{n+2}^1 - \mathbb{I}\{S_{n+1}^1 \leq S_{n+2}^1\})}{(n_1+1)\times(n_2+1)} \vee \frac{(n_1+1)\cdot(R_{n+2}^2 - \mathbb{I}\{S_{n+1}^2 \leq S_{n+2}^2\})}{(n_1+1)\times(n_2+1)} && (11)
\end{aligned}
$$

By looking further into the numerator of p_1^1 is a multiple of n_2+1, the numerator of p_1^2 is a multiple of n_1+1.

For convenience, we introduce another notations of ranks. We denote $(r_{1,1}^1, r_{2,1}^1, \cdots, r_{n_1,1}^1, r_{n_1+1,1}^1)$ be the rank of $(S_1^1, S_2^1, \cdots, S_{n_1}^1, S_{n+1}^1)$ and $(r_{1,2}^1, r_{2,2}^1, \cdots, r_{n_1,2}^1, r_{n_1+1,2}^1)$ be the rank of $(S_2^1, S_2^1, \cdots, S_{n_1}^1, S_{n+2}^1)$. Naturally, $(r_{1,1}^2, r_{2,1}^2, \cdots, r_{n_2,1}^2, r_{n_2+1,}^2$ stands for the rank of $(S_{n_1+1}^2, S_{n_1+2}^2, \cdots, S_n^2, S_{n+1}^2)$, and $(r_{1,2}^2, r_{2,2}^2, \cdots, r_{n_2,2}^2, r_{n_2+1,2}^2)$ stands for the rank of $(S_{n_1+1}^2, S_{n_1+2}^2, \cdots, S_n^2, S_{n+}^2$ Therefore:

$$
(R_{n+1}^1 - \mathbb{I}\{S_{n+1}^1 \geq S_{n+2}^1\}) = r_{n_1+1,1}^1 \quad (R_{n+2}^1 - \mathbb{I}\{S_{n+1}^1 \leq S_{n+2}^1\}) = r_{n_1+1,2}^1 \tag{12}
$$

3 Analysis that Correlation between Two p values are Non-Negatively Correlated under i.i.d assumption when existing only two type of inliers

This section is to prove the correlation between two MAMCOD pvalues are non-negative correlated under i.i.d assumption.

Consider the expression of correlation between two MAMCOD p-values, we can divide the correlation into four parts by Total Probability Formula:

$$
\begin{aligned}
Cov[p_1, p_2 | n+1, n+2 \in \mathcal{H}_0] = &\ \mathbb{P}\left[n+1, n+2 \in \mathcal{H}_{0,1}\right] \cdot Cov[p_1, p_2 | n+1, n+2 \in \mathcal{H}_{0,1}]\\
&+ \mathbb{P}\left[n+1, n+2 \in \mathcal{H}_{0,2}\right] \cdot Cov[p_1, p_2 | n+1, n+2 \in \mathcal{H}_{0,2}]\\
&+ \mathbb{P}\left[n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}\right] \cdot Cov[p_1, p_2 | n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}]\\
&+ \mathbb{P}\left[n+1 \in \mathcal{H}_{0,2}, n+2 \in \mathcal{H}_{0,1}\right] \cdot Cov[p_1, p_2 | n+1 \in \mathcal{H}_{0,2}, n+2 \in \mathcal{H}_{0,1}]
\end{aligned} \tag{13}
$$

In order to prove this covariance is non-negative, we only need to prove the above four parts are bigger or equal to zero.

3.1 $n+1, n+2 \in \mathcal{H}_{0,1}$

in this subsection we need to prove $Cov[p_1, p_2 | n+1, n+2 \in \mathcal{H}_{0,1}] \geq 0$, we break it into three possible situations: $p_1^1 > p_1^2$, $p_1^1 < p_1^2$ and $p_1^1 = p_1^2$

3.1.1 $p_1^1 > p_1^2$

$$
\begin{aligned}
Cov[p_1, p_2 | n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 > p_1^2] &= \mathbb{P}[p_2^1 > p_2^2] \cdot Cov[p_1, p_2 | n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 > p_1^2, p_2^1 > p_2^2]\\
&= \mathbb{P}[p_2^1 < p_2^2] \cdot Cov[p_1, p_2 | n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 > p_1^2, p_2^1 < p_2^2]\\
&= \mathbb{P}[p_2^1 = p_2^2] \cdot Cov[p_1, p_2 | n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 > p_1^2, p_2^1 = p_2^2]
\end{aligned} \tag{14}
$$

- $\boldsymbol{p_2^1 > p_2^2}$

the first covariance can be written as:

$$Cov[p_1, p_2|n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 > p_1^2, p_2^1 > p_2^2] = \mathbb{E}[p_1^1, p_2^1] - \mathbb{E}[p_1^1] \cdot \mathbb{E}[p_2^1] \tag{15}$$

By convinience, we omit the "conditional on $n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 > p_1^2, p_2^1 > p_2^2$" of expectation $\mathbb{E}[p_1^1, p_2^1]$ and $\mathbb{E}[p_1^1] \cdot \mathbb{E}[p_2^1]$, Same below.

Therefore, by *Lemma 1* of *Stephen Bates(2023)*,

$$\mathbb{E}[p_1^1, p_2^1] = \frac{1}{(n_1+1)(n_1+2)} \left[\sum_{j=1}^{n_1+1} (\frac{j}{n_1+1})^2 + \left(\sum_{j=1}^{n_1+1} \frac{j}{n_1+1} \right)^2 \right] \tag{16}$$

$$\mathbb{E}[p_1^1] \cdot \mathbb{E}[p_2^1] = \frac{1}{(n_1+1)^2} \left[\sum_{j=1}^{n_1+1} \frac{j}{n_1+1} \right]^2 \tag{17}$$

$$Cor(p_1, p_2) = \frac{1}{n_1+2} > 0 \tag{18}$$

- $\boldsymbol{p_2^1 = p_2^2}$

$$Cov[p_1, p_2|n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 > p_1^2, p_2^1 = p_2^2] = \mathbb{E}[p_1^1, p_2^1] - \mathbb{E}[p_1^1] \cdot \mathbb{E}[p_2^1] \tag{19}$$

in this way, the result is just the same as $p_2^1 > p_2^2$

- $\boldsymbol{p_2^1 < p_2^2}$

$$\begin{aligned} Cov[p_1, p_2|n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 > p_1^2, p_2^1 < p_2^2] &= \mathbb{E}[p_1^1, p_2^2] - \mathbb{E}[p_1^1] \cdot \mathbb{E}[p_2^2] \\ &= \sum_{0 < u_1, u_2 \leq 1} u_1 u_2 \left(\mathbb{P}[p_1^1 = u_1, p_2^2 = u_2] - \mathbb{P}[p_1^1 = u_1] \mathbb{P}[p_2^2 = u_2] \right) \end{aligned} \tag{20}$$

we claim that $p_1^1 \perp\!\!\!\perp p_2^2$ conditional on $n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 > p_1^2, p_2^1 < p_2^2$. Since,

$$p_1^1 = \frac{\sum_{i=1}^{n_1} \mathbb{I}[S_i^1 \leq S_{n+1}^1] + 1}{n_1+1} = \frac{r_{n_1+1,1}^1}{n_1+1} \qquad p_2^2 = \frac{\sum_{i=n_1+1}^{n} \mathbb{I}[S_i^2 \leq S_{n+2}^2] + 1}{n_2+1} \tag{21}$$

Since $S_i^1, \cdots, S_{n_1}^1, S_{n+1}^1$ are exchangeable, that means:

$$r_{n_1+1,1}^1 \sim Unif(1, 2 \cdots, n_1+1) \qquad r_{n_1+1,1}^1 \perp\!\!\!\perp p_2^2 \tag{22}$$

therefore, it comes after $p_1^1 = \frac{r_{n_1+1,1}^1}{n_1+1}$ and

$$\begin{aligned} Cov[p_1, p_2|n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 > p_1^2, p_2^1 < p_2^2] &= \mathbb{E}[p_1^1, p_2^2] - \mathbb{E}[p_1^1] \cdot \mathbb{E}[p_2^2] \\ &= \mathbb{E}[p_1^1] \cdot \mathbb{E}[p_2^2] - \mathbb{E}[p_1^1] \cdot \mathbb{E}[p_2^2] \\ &= 0 \end{aligned} \tag{23}$$

In conclusion:

$$Cov[p_1, p_2|n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 > p_1^2] = \frac{1}{n_1+2} \cdot \mathbb{P}[p_2^1 \geq p_2^2] \geq 0 \tag{24}$$

3.1.2 $p_1^1 = p_1^2$

Normally, we break into three possible situations:

- $\boldsymbol{p_2^1 > p_2^2}$

by symmetry, this is the same case with $p_1^1 > p_2^2, p_2^1 = p_2^2$ and

$$Cor(p_1, p_2 \,|\, n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 = p_1^2, p_2^1 > p_2^2) = \frac{1}{n_1 + 2} \tag{25}$$

- $\boldsymbol{p_2^1 = p_2^2}$

$$Cor(p_1, p_2 \,|\, n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 = p_1^2, p_2^1 = p_2^2) = \mathbb{E}[p_1^1, p_2^1] - \mathbb{E}[p_1^1] \cdot \mathbb{E}[p_2^1] = \frac{1}{n_1 + 2} \tag{26}$$

- $\boldsymbol{p_2^1 < p_2^2}$

$$Cor(p_1, p_2 \,|\, n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 = p_1^2, p_2^1 < p_2^2) = \mathbb{E}[p_1^1, p_2^2] - \mathbb{E}[p_1^1] \cdot \mathbb{E}[p_2^2] = 0 \tag{27}$$

So,

$$Cov[p_1, p_2 | n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 = p_1^2] = \frac{1}{n_1 + 2} \cdot \mathbb{P}[p_2^1 \geq p_2^2] \geq 0 \tag{28}$$

3.1.3 $p_1^1 < p_1^2$

- $\boldsymbol{p_2^1 \geq p_2^2}$

  By symmetry, this is the same case with $p_1^1 \geq p_1^2, p_2^1 < p_2^2$:

$$Cov[p_1, p_2 | n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 < p_1^2, p_2^1 \geq p_2^2] = \mathbb{E}[p_1^2, p_2^1] - \mathbb{E}[p_1^2] \cdot \mathbb{E}[p_2^1] = 0 \tag{29}$$

- $\boldsymbol{p_2^1 < p_2^2}$

$$\begin{aligned} Cov[p_1, p_2 | n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 < p_1^2, p_2^1 < p_2^2] &= \mathbb{E}[p_1^2, p_2^2] - \mathbb{E}[p_1^2] \cdot \mathbb{E}[p_2^2] \\ &= \sum_{0 < u_1, u_2 \leq 1} u_1 u_2 \left(\mathbb{P}[p_1^2 = u_1, p_2^2 = u_2] - \mathbb{P}[p_1^2 = u_1] \mathbb{P}[p_2^2 = u_2] \right) \end{aligned} \tag{30}$$

  According to *Theorem A.1(ii)* of *AdaDetect*, p_2^2 can be written as a version of p_1^2:

$$\begin{aligned} p_2^2 &= C_{p_1^2, p_2^2} + \mathbb{I}\{S_{n+1}^2 > S_{n+2}^2\}/(n_2+1) = C_{p_1^2, p_2^2} + \mathbb{I}\{S_{((n_2+1)p_1^2)}^2 > S_{n+2}^2\}/(n_2+1) \\ C_{p_1^2, p_2^2} &= (n_2+1)^{-1} \sum_{s \in \{S_{n_2+1}^2, S_{n_2+2}^2, \cdots, S_n^2, S_{n+1}^2\}} \mathbb{I}\{s \leq S_{n+2}^2\} \end{aligned} \tag{31}$$

  and the set $\{S_{n_1+1}^2, \cdots, S_n^2, S_{n+1}^2\} = \{S_{(1)}^2, \cdots, S_{(n_2)}^2, S_{(n_2+1)}^2\}$ (with $S_{(1)}^2 < S_{(2)}^2 < \cdots < S_{(n_2+1)}^2$), therefore, p_2^2 is a non-decreasing function of p_1^2, thus non-negatively correlated.

In conclusion,

$$\begin{aligned} Cor[p_1, p_2 \,|\, n+1, n+2 \in \mathcal{H}_{0,1}] = &\frac{1}{n_1+2} \cdot \mathbb{P}[p_1^1 \geq p_1^2 \,|\, n+1, n+2 \in \mathcal{H}_{0,1}] \mathbb{P}[p_2^1 \geq p_2^2 \,|\, n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 \geq p_1^2] \\ &+ Cor[p_1, p_2 | n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 < p_1^2, p_2^1 < p_2^2] \cdot \mathbb{P}[p_1^1 < p_1^2 \,|\, n+1, n+2 \in \mathcal{H}_{0,1}] \\ &\cdot \mathbb{P}[p_2^1 < p_2^2 \,|\, n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 < p_1^2] \\ = &\frac{1}{n_1+2} \cdot \mathbb{P}[p_2^1 \geq p_2^2, p_1^1 \geq p_1^2 \,|\, n+1, n+2 \in \mathcal{H}_{0,1}] \\ &+ Cor[p_1, p_2 | n+1, n+2 \in \mathcal{H}_{0,1}, p_1^1 < p_1^2, p_2^1 < p_2^2] \cdot \mathbb{P}[p_1^1 < p_1^2, p_2^1 < p_2^2 \,|\, n+1, n+2 \in \mathcal{H}_{0,1}] \\ \geq &\, 0 \end{aligned} \tag{32}$$

3.2 $n+1, n+2 \in \mathcal{H}_{0,2}$

This case is symmetric to $n+1, n+2 \in \mathcal{H}_{0,1}$. Similarly, the correlation value is:

$$
\begin{aligned}
Cor[p_1, p_2 \,|\, n+1, n+2 \in \mathcal{H}_{0,2}] = & \frac{1}{n_2+2} \cdot \mathbb{P}[p_2^1 \leq p_2^2, p_1^1 \leq p_1^2 \,|\, n+1, n+2 \in \mathcal{H}_{0,2}] \\
& + Cor[p_1, p_2 | n+1, n+2 \in \mathcal{H}_{0,2}, p_1^1 > p_1^2, p_2^1 > p_2^2] \cdot \mathbb{P}[p_1^1 > p_1^2, p_2^1 > p_2^2 \,|\, n+1, n+2 \in \mathcal{H}_{0,2}] \\
& \geq 0
\end{aligned} \tag{33}
$$

3.3 $n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}$

3.3.1 $p_1^1 \geq p_1^2$

- $\boldsymbol{p_2^1 \leq p_2^2}$

  In this situation:

$$
Cov[p_1, p_2 \,|\, n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}, p_1^1 \geq p_1^2, p_2^1 \leq p_2^2] = 0 \tag{34}
$$

  Since p_1^1 is independent with p_2^2

- $\boldsymbol{p_2^1 > p_2^2}$

$$
Cov[p_1, p_2 \,|\, n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}, p_1^1 \geq p_1^2, p_2^1 > p_2^2] = \mathbb{E}[p_1^1, p_2^1] - \mathbb{E}[p_1^1]\mathbb{E}[p_2^1] \tag{35}
$$

  Use the same idea of *Theorem A.1(ii)* of *AdaDetect*, $p_2^1$is a non-negative function of p_1^1, thus the correlation is non-negative.

3.3.2 $p_1^1 < p_1^2$

- $\boldsymbol{p_2^1 \leq p_2^2}$

$$
Cov[p_1, p_2 \,|\, n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}, p_1^1 < p_1^2, p_2^1 \leq p_2^2] = \mathbb{E}[p_1^2, p_2^2] - \mathbb{E}[p_1^2]\mathbb{E}[p_2^2] \tag{36}
$$

  The same method of *Theorem A.1(ii)* of *AdaDetect*, the correlation is non-negative.

- $\boldsymbol{p_2^1 > p_2^2}$

$$
Cov[p_1, p_2 \,|\, n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}, p_1^1 < p_1^2, p_2^1 > p_2^2] = \mathbb{E}[p_1^2, p_2^1] - \mathbb{E}[p_1^2]\mathbb{E}[p_2^1] \tag{37}
$$

  Since the randomness of p_1^2 and p_2^1 comes from two different data set with no intersection, p_1^2 and p_2^1 are independent, thus the correlation is zero.

Therefore,

$$
\begin{aligned}
Cor[p_1, p_2 \,|\, n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}] = & \, Cor[p_1, p_2 | n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}, p_1^1 \geq p_1^2, p_2^1 > p_2^2] \\
& \cdot \mathbb{P}[p_1^1 \geq p_1^2, p_2^1 > p_2^2 \,|\, n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}] \\
& + Cor[p_1, p_2 | n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}, p_1^1 < p_1^2, p_2^1 \leq p_2^2] \\
& \cdot \mathbb{P}[p_1^1 < p_1^2, p_2^1 \leq p_2^2 \,|\, n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}] \\
& \geq 0
\end{aligned} \tag{38}
$$

3.4 $n+1 \in \mathcal{H}_{0,2}, n+2 \in \mathcal{H}_{0,1}$

This situation is symmetric to that of $n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}$ and the correlation can be written as:

$$
\begin{aligned}
Cor[p_1, p_2 \,|\, n+1 \in \mathcal{H}_{0,2}, n+2 \in \mathcal{H}_{0,1}] = & \, Cor[p_1, p_2 | n+1 \in \mathcal{H}_{0,2}, n+2 \in \mathcal{H}_{0,1}, p_1^1 \leq p_1^2, p_2^1 < p_2^2] \\
& \cdot \mathbb{P}[p_1^1 \leq p_1^2, p_2^1 < p_2^2 \,|\, n+1 \in \mathcal{H}_{0,2}, n+2 \in \mathcal{H}_{0,1}] \\
& + Cor[p_1, p_2 | n+1 \in \mathcal{H}_{0,2}, n+2 \in \mathcal{H}_{0,1}, p_1^1 > p_1^2, p_2^1 \geq p_2^2] \\
& \cdot \mathbb{P}[p_1^1 > p_1^2, p_2^1 \geq p_2^2 \,|\, n+1 \in \mathcal{H}_{0,2}, n+2 \in \mathcal{H}_{0,1}] \\
& \geq 0
\end{aligned} \tag{39}
$$

Conclusion 1: every part of $Cov[p_1, p_2 | n+1, n+2 \in \mathcal{H}_0]$ is non-negative, thus the correlation between two p-values is non-negative under i.i.d assumption.

Conclusion 2: under i.i.d assumption, when two p values are relied on the same set of calibration data, the correlation is non-negative. When two p values are relied on different calibration data, the correlation is zero. Therefore, it can be easily extended to K classes of inliers.

4 PRDS proof of p values with K types of inliers under i.i.d assumption

Definition 1 We say that a family of p-values$(p_i, 1 \leq i \leq m)$ is *PRDS* on $\mathcal{H}_0$ if, for any $i \in \mathcal{H}_0$ and non-decreasing measurable set $D \subset [0,1]^m$, the function $u \in [0,1] \longrightarrow \mathbb{P}((p_j, 1 \leq j \leq m) \in D | p_i = u)$ is nondecreasing.

Proof We claim that at is suffice to show $u \in [0,1] \longrightarrow \mathbb{P}((p_j, 1 \leq j \leq m) \in D \,|\, p_i^{k_0} = u)$ for every $k_0 \in [K]$, since $p_i = \sum_{k=1}^{K} p_i^k \mathbb{I}\{\bigcup_j (a_{k,j}, b_{k,j})\}$ for any partitions $\bigcup_k \bigcup_j (a_{k,j}, b_{k,j})$ of $[0,1]$.

Note that $p_j = max\{p_j^1, \cdots, p_j^K\}$, so $((p_j^k, 1 \leq j \leq m, k \neq k_0))$ is independent with $p_i^{k_0}$ under i.i.d assumption.

For those p_j^k that $k = k_0$, by *Theorem A.1(ii)* of *AdaDetect*:

$$
\begin{aligned}
p_j^k &= C_{i,j}^k + \mathbb{I}\{S_{n+i}^k > S_{n+j}^k\}/(n_k+1) = C_{i,j}^k + \mathbb{I}\{S_{((n_k+1)p_i^k)}^k > S_{n+j}^k\}/(n_k+1) \\
C_{i,j}^k &= (n_k+1)^{-1} \sum_{s \in \{S_l^k, S_{n+i}^k \,|\, l \in \mathcal{D}_k^{cal}\}} \mathbb{I}\{s \leq S_{n+j}^k\}
\end{aligned} \tag{40}
$$

Therefore, p_j^k is nondecreasing in $p_i^{k_0}$ when $k = k_0$, and $\mathbb{P}((p_j^k, 1 \leq j \leq m) \in D \mid p_i^{k_0} = u)$ is nondecreasing in u for every $k_0 \in [K]$. Thus, $\mathbb{P}((p_j, 1 \leq j \leq m) \in D \mid p_i = u)$ is nondecreasing in u.

5 Analysis that Correlation between Two p values are Non-Negatively Correlated under exchangeable assumption

Proof outline: According to de Finetti's Theorem, every set of exchangeable variables is i.i.d with respect to some conditional probability measure, so we can convert to the i.i.d cases by conditional on this conditional probability measure.

Let's begin with some assumptions first:

Assumption 1 *Given $1 \leq k \leq K$, $(\mathcal{D}_k^{ns}, X_i, i \in \mathcal{H}_{0,k})$ is exchangeable conditionally on $(\mathcal{D}^{ns} \backslash \mathcal{D}_k^{ns}, X_i, i \notin \mathcal{H}_{0,k})$.*

Assumption 2 *Given $1 \leq k, \tilde{k} \leq K$ $(S_i^k, i \in \mathcal{D}_{\tilde{k}}^{cal} \cup \mathcal{D}_{\tilde{k}}^{test})$ is separately exchangeable conditionally on $(S_i^{\tilde{k}}, i \in (\mathcal{D}^{cal} \cup \mathcal{D}^{test}) \backslash (\mathcal{D}_k^{cal} \cup \mathcal{D}_k^{test}))$.*

Assumption 3 $(S_{s+1}^1, \cdots, S_n^K, S_{n+1}^1, S_{n+1}^2, \cdots, S_{n+m}^{K-1}, S_{n+m}^K)$ *has no ties almost surely.*

According to de Finetti's theorem, if $(\mathcal{D}_k^{ns}, X_i, i \in \mathcal{H}_{0,k})$ is exchangeable conditionally on $(\mathcal{D}^{ns} \backslash \mathcal{D}_k^{ns}, X_i, i \notin \mathcal{H}_{0,k})$, then there exists a random η such that $(\mathcal{D}_k^{ns}, X_i, i \in \mathcal{H}_{0,k})$ are i.i.d conditional on $\mathcal{C}_k := (\eta, \mathcal{D}^{ns} \backslash \mathcal{D}_k^{ns}, X_i, i \notin \mathcal{H}_{0,k})$