Analysis about Correlation Between MAMCOD p values

Ruoxiao Tang

July 2024

1 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

$$Cov[p_1, p_2] \tag{1}$$

WLOG, let suppose $n+1 \in \mathcal{H}_{0,1}, n+2 \in \mathcal{H}_{0,2}$

$$\begin{aligned} Cov[p_1, p_2] =& \mathbb{P}[p_1^1 \geq p_1^2, p_2^1 > p_2^2] Cov[p_1, p_2 | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \\ &+ \mathbb{P}[p_1^1 \geq p_1^2, p_2^1 \leq p_2^2] Cov[p_1, p_2 | p_1^1 \geq p_1^2, p_2^1 \leq p_2^2] \\ &+ \mathbb{P}[p_1^1 < p_1^2, p_2^1 > p_2^2] Cov[p_1, p_2 | p_1^1 < p_1^2, p_2^1 > p_2^2] \\ &+ \mathbb{P}[p_1^1 < p_1^2, p_2^1 < p_2^2] Cov[p_1, p_2 | p_1^1 < p_1^2, p_2^1 < p_2^2] \end{aligned} \tag{2}$$

1.1 If p_1 and p_2 are conditional on the same set of calibration data

If p_1 and p_2 are conditional on the same set of calibration data, like in the first cases that $p_1^1 \geq p_1^2, p_2^1 > p_2^2$

$$\begin{aligned} \mathbb{P}[p_1^1 \geq p_1^2, p_2^1 > p_2^2] Cov[p_1, p_2 | p_1^1 \geq p_1^2, p_2^1 > p_2^2] =& \mathbb{P}[p_1^1 \geq p_1^2, p_2^1 > p_2^2] \mathbb{E}[p_1^1 p_2^1 | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \\ &- \mathbb{P}[p_1^1 \geq p_1^2, p_2^1 > p_2^2] \mathbb{E}[p_1^1 | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \mathbb{E}[p_2^1 | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \end{aligned} \tag{3}$$

and

$$\begin{aligned} &\mathbb{P}[p_1^1 \geq p_1^2, p_2^1 > p_2^2] \mathbb{E}[p_1^1 p_2^1 | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \\ &= \mathbb{P}[p_1^1 \geq p_1^2, p_2^1 > p_2^2] \cdot \sum_{i=1}^{n_1+1} \sum_{j=1}^{n_1+1} \frac{i}{n_1+1} \frac{j}{n_1+1} \mathbb{P}[p_1^1 = \frac{i}{n_1+1}, p_2^1 = \frac{j}{n_1+1} | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \\ &= \sum_{i=1}^{n_1+1} \sum_{j=1}^{n_1+1} \frac{i}{n_1+1} \frac{j}{n_1+1} \mathbb{P}[p_1^1 = \frac{i}{n_1+1}, p_2^1 = \frac{j}{n_1+1}, p_1^1 \geq p_1^2, p_2^1 > p_2^2] \\ &= \sum_{i=1}^{n_1+1} \sum_{j=1}^{n_1+1} \frac{i}{n_1+1} \frac{j}{n_1+1} \mathbb{P}[p_1^1 = \frac{i}{n_1+1} | p_2^1 = \frac{j}{n_1+1}, p_1^1 \geq p_1^2, p_2^1 > p_2^2] \cdot \mathbb{P}[p_2^1 = \frac{j}{n_1+1}, p_1^1 \geq p_1^2, p_2^1 > p_2^2] \end{aligned} \tag{4}$$

$$\begin{aligned} &\mathbb{P}[p_1^1 \geq p_1^2, p_2^1 > p_2^2] \mathbb{E}[p_1^1 | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \mathbb{E}[p_2^1 | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \\ &= \mathbb{P}[p_1^1 \geq p_1^2, p_2^1 > p_2^2] \left\{ \sum_{i=1}^{n_1+1} \frac{i}{n_1+1} \mathbb{P}[p_1^1 = \frac{i}{n_1+1} \mathbb{P}[p_1^1 = \frac{i}{n_1+1} | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \cdot \right\} \left\{ \sum_{j=1}^{n_1+1} \frac{j}{n_1+1} \mathbb{P}[p_2^1 = \frac{j}{n_1+1} | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \right\} \\ &= \sum_{i=1}^{n_1+1} \sum_{j=1}^{n_1+1} \frac{i}{n_1+1} \frac{j}{n_1+1} \mathbb{P}[p_1^1 = \frac{i}{n_1+1} | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \cdot \mathbb{P}[p_2^1 = \frac{j}{n_1+1}, p_1^1 \geq p_1^2, p_2^1 > p_2^2] \end{aligned} \tag{5}$$

therefore,

$$
\begin{aligned}
&\mathbb{P}[p_1^1 \geq p_1^2, p_2^1 > p_2^2] Cov[p_1, p_2 | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \\
&= \mathbb{P}[p_1^1 \geq p_1^2, p_2^1 > p_2^2] \left\{ \mathbb{E}[p_1^1 | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \mathbb{E}[p_2^1 | p_1^1 \geq p_1^2, p_2^1 > p_2^2] - \mathbb{E}[p_1^1 | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \mathbb{E}[p_2^1 | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \right\} \\
&= \sum_{i=1}^{n_1+1} \sum_{j=1}^{n_1+1} \frac{i}{n_1+1} \frac{j}{n_1+1} \mathbb{P}[p_2^1 = \frac{j}{n_1+1}, p_1^1 \geq p_1^2, p_2^1 > p_2^2] \cdot \\
&\left(\mathbb{P}[p_1^1 = \frac{i}{n_1+1} | p_2^1 = \frac{j}{n_1+1}, p_1^1 \geq p_1^2, p_2^1 > p_2^2] - \mathbb{P}[p_1^1 = \frac{i}{n_1+1} | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \right)
\end{aligned} \tag{6}
$$

Therefore, we need to prove:

$$
\mathbb{P}[p_1^1 = \frac{i}{n_1+1} | p_2^1 = \frac{j}{n_1+1}, p_1^1 \geq p_1^2, p_2^1 > p_2^2] \geq \mathbb{P}[p_1^1 = \frac{i}{n_1+1} | p_1^1 \geq p_1^2, p_2^1 > p_2^2] \tag{7}
$$

1.2 If p_1 and p_2 are conditional on different set of calibration data

Suppose that $p_1^1 \geq p_1^2, p_2^1 \leq p_2^2$ Similarly, we only need to prove that:

$$
\mathbb{P}[p_1^1 = \frac{i}{n_1+1} | p_2^2 = \frac{j}{n_1+1}, p_1^1 \geq p_1^2, p_2^1 \leq p_2^2] \geq \mathbb{P}[p_1^1 = \frac{i}{n_1+1} | p_2^2 \geq p_1^2, p_2^1 \leq p_2^2] \tag{8}
$$

Although p_2^2 is independent of p_1^1, $\{p_2^2 = \frac{j}{n_1+1}, p_2^1 \leq p_2^2\} \neq \{p_2^1 \leq p_2^2\}$ and p_1^1 is not independent of p_2^1, therefore, the two probabilities are not equal.