C  Details of the non-parametric inference

Non-parametric analysis conducted in ICML 2020 consists of two steps that we now discuss.

Step 1. Matching  First, we conduct matching of (submission, reviewer) pairs by executing the following procedure separately for each submission. Working with a given submission $S$, we consider two groups of reviewers assigned to $S$: cited and uncited. Next, we attempt to find cited reviewer $R_{ctd}$ and uncited reviewer $R_{unctd}$ that are similar in terms of expertise, preference, and seniority characteristics. More formally, in terms of variables we introduced in Table 3, reviewers $R_{ctd}$ and $R_{unctd}$ should satisfy all of the following criteria with respect to $S$:

- Self-reported expertise of reviewers in reviewing submission $S$ is the same:
  \[
  \text{expertiseSRExp}_{ctd} = \text{expertiseSRExp}_{unctd}
  \]

- Self-reported confidence of reviewers in their evaluation of submission $S$ is the same:
  \[
  \text{expertiseSRConf}_{ctd} = \text{expertiseSRConf}_{unctd}
  \]

- Textual overlap between submission $S$ and papers of each of the reviewers differ by at most 0.1:
  \[
  |\text{expertiseText}_{ctd} - \text{expertiseText}_{unctd}| \leq 0.1
  \]

- Reviewers’ bids on submission $S$ satisfy one of the two conditions:
  1. Both bids have value 3 ("In a pinch"):  
     \[
     \text{prefBid}_{ctd} = \text{prefBid}_{unctd} = 3
     \]
  2. Both bids have values greater than 3 (4-"Willing" or 5-"Eager"): 
     \[
     \text{prefBid}_{ctd} \in \{4, 5\} \quad \text{and} \quad \text{prefBid}_{unctd} \in \{4, 5\}
     \]

- Reviewers belong to the same seniority group:
  \[
  \text{seniority}_{ctd} = \text{seniority}_{unctd}
  \]

We run this procedure for all submissions in the pool. If for submission $S$ there are no reviewers $R_{ctd}$ and $R_{unctd}$ that satisfy these criteria, we remove submission $S$ from the non-parametric analysis. Overall, we let $K$ denote the number of such 1-1 matched pairs obtained and introduce the set of triples that the remaining analysis operates with:

\[
\left\{ \left( S^{(i)}, R_{ctd}^{(i)}, R_{unctd}^{(i)} \right) \right\}_{i=1}^{K}.
\]

Each triple in this set consists of submission $S$ and two reviewers $R_{ctd}$ and $R_{unctd}$ that (i) are assigned to $S$ and (ii) satisfy the aforementioned conditions with respect to $S$. Within each submission, each reviewer can be a part of only one triple.

Let us now consider two (submission, reviewer) pairs associated with a given triple. Observe that these pairs share the submission, thereby sharing the value of unobserved characteristic quality. Additionally, the criteria used to select reviewers $R_{ctd}$ and $R_{unctd}$ ensures that characteristics expertise, preference, and seniority are also similar across these pairs. Crucially, while being equal on all four characteristics, these pairs have different values of the citation indicator.
Step 2. Permutation test  Having constructed the set of triples \( \mathcal{T} \), we now compare scores given by cited and uncited reviewers within these triples. Specifically, consider triple \( i \in \{1, \ldots, K\} \) and let \( Y_{ctd}^{(i)} \) (respectively, \( Y_{unctd}^{(i)} \)) be the score given by cited reviewer \( R_{ctd}^{(i)} \) (respectively, uncited reviewer \( R_{unctd}^{(i)} \)) to submission \( S^{(i)} \). Then the test statistic \( \tau \) of our analysis is defined as follows:

\[
\tau = \frac{1}{K} \sum_{i=1}^{K} \left( Y_{ctd}^{(i)} - Y_{unctd}^{(i)} \right).
\]

To quantify the significance of the difference between scores given by cited and uncited reviewers, we execute the permutation test. Specifically, at each of the 10,000 iterations, we independently permute the citation indicator within each triple \( i \in \{1, \ldots, K\} \). For each permuted sample, we recompute the value of the test statistic (6) and finally check whether the actual value of the test statistic \( \tau \) appears to be “too extreme” for the significance level 0.05.