Review of: "Distributional Matrix Completion via Nearest Neighbors in the Wasserstein Space"

Zeyu Wu[1]

1 School of Reliability and Systems Engineering, Beihang University, Beijing, China

Potential competing interests: No potential competing interests to declare.

The authors proposed a novel framework for distributional matrix completion, leveraging optimal transport and the Wasserstein distance to impute missing distributions in a matrix. The method combines theoretical guarantees and practical applicability, demonstrating advantages in distributional accuracy and computational efficiency. While the paper offers valuable contributions, several aspects require clarification or further improvement. Here are some suggestions:

1. In Section 2, the paper states that "The Wasserstein barycenter also has a simple closed-form solution," but it does not specify the conditions under which this closed-form solution is applicable. Providing a clear explanation of these conditions would ensure the accurate application of the method.
2. In Section 4, the paper directly applies the location-scale model to the case study of school performance statistics. While this simplifies the computation, the paper does not provide sufficient justification for this assumption. Offering a more thorough rationale or evidence to support this assumption would enhance its credibility.
3. In Figure 3(a) of Section 5, the blue line exhibits an upward trend, which contradicts the expected error decay as the sample size increases. However, the paper does not provide any explanation. Please give more interpretation.
4. In Figure 3(b) of Section 5, the last two data points are higher than the earlier ones, yet the authors have fitted a downward-sloping line to the data. This choice appears inconsistent with the observed trend, and the paper does not explain it.
5. In Figure 4 of Section 5, the DIST-NN method provides both prediction curves and confidence bands, while the bootstrap method only provides confidence bands without a corresponding prediction curve. This inconsistency makes it difficult to directly compare the results of the two methods.
6. In Figure 5 of Section 5, Scalar NN performs better than DIST-NN in estimating the standard deviation, which is unexpected given the paper's emphasis on the advantages of DIST-NN. However, the paper does not provide an explanation for this observation.
7. The comparison experiments focus on traditional methods such as scalar matrix completion and do not include comparisons with more recent distribution-based techniques, such as those using Wasserstein barycenters for similar tasks.
8. In the final part of Section 6, the paper mentions the limitations of the method's assumptions and suggests that the approach could be applied in non-i.i.d. sampling settings. However, the main text does not provide detailed explanations or demonstrations of this claim. Including more specific examples or discussions in the main body of the paper would clarify this point and enhance the practical applicability of the method.