A Proofs and Remarks

Proposition 1. *The Wasserstein-1 distance between two empirical distributions* $\mathbf{P}$ *and* $\mathbf{Q}$ *with samples* $\{X_i\}$ *and* $\{Y_i\}$ *of size* n *in* $\mathbb{R}^1$ *can be written as:*

$$W_1(\mathbf{P}, \mathbf{Q}) \;=\; \frac{1}{n}\sum_{i=1}^{n}\big|X_{(i)} - Y_{(i)}\big|, \tag{3}$$

where $X_{(i)}$ *and* $Y_{(i)}$ *denote the* i*-th order statistics (sorted samples).*

Proof. The proof chiefly relies on the work of [4]. Given that $\mathbf{P}$ and $\mathbf{Q}$ are discrete, real-valued random variables, we are able to reduce the Optimal Transport Problem to the following structure:

$$W_1(\mathbf{P}, \mathbf{Q}) = \inf\left\{\frac{1}{n}\sum_i |X_i - Y_{\sigma(i)}| : \; \sigma \in S_n\right\}. \tag{4}$$

Here, S_n describes the space of permutations for $n-$tuples. In the following, we want to show that Equation 4 is solved by Equation 3. Without loss of generality, let $X_1^{\downarrow}, X_2^{\downarrow}, \ldots, X_n^{\downarrow}$ describe the *descending* ordering of $(X_i)_{i=1}^n$, i.e., $X_i^{\downarrow} \geq X_{i+1}^{\downarrow}$, $i = 1, \ldots, n-1$. Next, let

$$f_i : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto |X_i^{\downarrow} - x|. \tag{5}$$

One may quickly verify that $f_{i+1} - f_i$ is a non-decreasing function for any $i \in \{1, \ldots, n-1\}$. Now, utilizing the result of Problem 5.4. in [3], we obtain that for any permutation $\sigma \in S_n$,

$$\sum_{i=1}^{n} f_i(Y_i^{\downarrow}) = \sum_{i=1}^{n}|X_i^{\downarrow} - Y_i^{\downarrow}| \leq \sum_{i=1}^{n} f_i(Y_{\sigma(i)}^{\downarrow}). \tag{6}$$

Due to commutativity of the sum, the actual ordering along i becomes irrelevant as long as the pairings remain intact. This concludes our assertion. □

□

Remark 1. On a side note, in [1], two conceptual variants to measuring the distance between two time series sequences are proposed. The first one requires that the number of observations (i.e. the total mass) of both time series is equal. They are then transformed into histograms of relative probabilities via normalization and the W-1 distance is calculated. The second one does not assume equal total masses and is applied "'directly"' to the frequency-histograms (EMD). We will, however, only focus on the Wasserstein distance when referring to this step. Note that in the first case, the EMD would yield an identical result (cf. [2]). Moreover, as the authors of [1] conclude, in case of much total mass and/or approx. similar sequence-lengths, the computation of W1 is more efficient and leads to similar results compared to EMD.

References

1. Chapela-Campa, D., Benchekroun, I., Baron, O., Dumas, M., Krass, D., Senderovich, A.: A framework for measuring the quality of business process simulation models. Information Systems **127**, 102447 (2025)
2. Levina, E., Bickel, P.: The earth mover's distance is the mallows distance: some insights from statistics. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. vol. 2, pp. 251–256 vol.2 (2001)
3. Steele, J.M.: The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities. Cambridge University Press (2004)
4. Villani, C., Society, A.M.: Topics in Optimal Transportation. Graduate studies in mathematics, American Mathematical Society (2003)