

Total Variation Diameter of Rank-Preserving Probability Distributions

A short note on how far probabilities can move while rankings remain unchanged

J. R. Landers

A classifier, ranking model, or scoring system may assign a probability to each of n examples. In practice, we often care about the induced ordering: which examples rank first, second, and so on. Calibration and other post-processing methods can change the probabilities while leaving much of that ordering intact. This raises a basic geometric question:

How different can two probability distributions be if they induce exactly the same ranking?

Let $p = (p_1, \dots, p_n)$ be a probability vector, $p_i \geq 0$ and $\sum_i p_i = 1$. After relabeling the outcomes, suppose its ranking is

$$p_1 > p_2 > \dots > p_n.$$

Define the corresponding strict *rank chamber*

$$\mathcal{R}_n = \left\{ p : p_1 > p_2 > \dots > p_n \geq 0, \quad \sum_{i=1}^n p_i = 1 \right\}.$$

Every pair $p, q \in \mathcal{R}_n$ agrees on the complete ordering of the outcomes. Their total variation distance is

$$d_{\text{TV}}(p, q) = \frac{1}{2} \sum_{i=1}^n |p_i - q_i|,$$

with $0 \leq d_{\text{TV}} \leq 1$.

Key result (rank-chamber TV diameter). For $n \geq 2$,

$$\sup_{p, q \in \mathcal{R}_n} d_{\text{TV}}(p, q) = 1 - \frac{1}{n}.$$

If weak rankings are allowed, $p_1 \geq p_2 \geq \dots \geq p_n$, the same value is attained as a maximum.

Why this is true. Consider the closure of the chamber,

$$\overline{\mathcal{R}_n} = \left\{ p : p_1 \geq p_2 \geq \dots \geq p_n \geq 0, \quad \sum_i p_i = 1 \right\}.$$

It is a convex polytope whose vertices are

$$v^{(k)} = \left(\underbrace{\frac{1}{k}, \dots, \frac{1}{k}}_{k \text{ entries}}, 0, \dots, 0 \right), \quad k = 1, \dots, n.$$

These vertices range from complete concentration,

$$v^{(1)} = (1, 0, \dots, 0),$$

to complete uniformity,

$$v^{(n)} = \left(\frac{1}{n}, \dots, \frac{1}{n} \right).$$

Because total variation is convex, a maximum over the compact product polytope $\overline{\mathcal{R}_n} \times \overline{\mathcal{R}_n}$ is attained at a pair of vertices. For $k \leq \ell$,

$$d_{\text{TV}}(v^{(k)}, v^{(\ell)}) = 1 - \frac{k}{\ell}.$$

The largest value occurs at $k = 1$ and $\ell = n$, giving $1 - 1/n$. In the strict chamber, the two limiting distributions contain ties, so the value is approached arbitrarily closely rather than attained. Hence the result is stated using a supremum.

The intuition is simpler than the geometry. One distribution can be made almost a point mass,

$$p \approx (1, 0, \dots, 0),$$

while another can be made almost uniform,

$$q \approx \left(\frac{1}{n}, \dots, \frac{1}{n} \right).$$

Tiny perturbations make both vectors strictly decreasing, so both still rank the examples in exactly the same order. Yet

$$d_{\text{TV}}(p, q) \longrightarrow \frac{1}{2} \left[\left(1 - \frac{1}{n} \right) + (n-1) \frac{1}{n} \right] = 1 - \frac{1}{n}.$$

The first distribution effectively says “the top example dominates.” The second says “the examples are almost equally likely.” Ranking alone cannot distinguish these very different probabilistic statements.

Large- n consequence. Since

$$\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n} \right) = 1,$$

perfect rank agreement becomes compatible with nearly maximal total variation distance as the number of ranked examples grows. For $n = 10$ the supremum is 0.9; for $n = 100$ it is 0.99.

Connection to probability calibration. This distinction is directly relevant to post-hoc calibration. A fitted classifier may produce scores or probabilities that are transformed to improve their probabilistic interpretation. Common calibration maps are monotone or order-aware; for example, scikit-learn’s `CalibratedClassifierCV` supports sigmoid, isotonic, and temperature calibration. Such transformations motivate a useful separation between two notions of agreement:

$$\text{rank agreement:} \quad p_i > p_j \iff q_i > q_j,$$

versus

$$\text{distributional agreement:} \quad d_{\text{TV}}(p, q) \text{ is small.}$$

The result shows that the first gives essentially no worst-case guarantee on the second when n is large. A calibration or post-processing step can preserve every ordering decision while, in principle, producing a probability vector very far from the original one in total variation.

Takeaway. Preserving rank preserves decisions about *order*; it does not preserve the *geometry of probability mass*. For large ranked sets, two distributions can agree perfectly on ordering and still be almost maximally separated in total variation.