On the weighted total variation distance between probability measures

December 28, 2022

Abstract

This is a short note on the weighted total variation distance between probability measures on \mathbb{R}^d . Everything is well-known but I was not able to find this material in the literature as it is stated here. Please write me for any error/typo/comment at quentin.cormier@inria.fr.

Given $\phi : \mathbb{R}^d \to \mathbb{R}_+$ a non-negative measurable function, we define the weighted total variation distance between two probability measures ν and μ by

$$d_{TV}^{\phi}(\nu,\mu) := \int \phi(x) |\nu - \mu|(dx).$$

Theorem 1. The weighted total variation distance satisfies

$$d_{TV}^{\phi}(\nu,\mu) = \inf_{\mathcal{L}(X)=\nu,\mathcal{L}(Y)=\mu} \mathbb{E}(\phi(X) + \phi(Y))\mathbb{1}_{X \neq Y}.$$

In addition, the Kantorovich duality theorem holds

$$d_{TV}^{\phi}(\nu,\mu) = \sup_{g} \int g(x)(\nu-\mu)(dx),$$

where the supremum is taken over all the measurable functions $g: \mathbb{R}^d \to \mathbb{R}$ s.t.

$$\forall x, y, \quad |g(x) - g(y)| \le (\phi(x) + \phi(y)) \mathbb{1}_{x \neq y}. \tag{1}$$

Remark 2. This result states that the weighted total variation distance is equal to the Wasserstein distance W_1 for the following peculiar semi-metric

$$d(x,y) := (\phi(x) + \phi(y)) \mathbb{1}_{x \neq y}.$$

See also [CD18, Ch. 5].

Remark 3. The standard (non-weighted) total variation distance is obtained by choosing $\phi \equiv 1/2$. In that case, the result simplifies to

$$d_{TV}^{1/2}(\nu,\mu) = \inf_{\mathcal{L}(X)=\nu,\mathcal{L}(Y)=\mu} \mathbb{P}(X \neq Y) = \sup_{A \in \mathcal{B}(\mathbb{R}^d)} \nu(A) - \mu(A),$$

where $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel sets of \mathbb{R}^d .

Proof. Step 1. We prove that $d_{TV}^{\phi}(\nu,\mu) \leq \sup_g \int g(x)(\nu-\mu)(dx)$.

Let *m* be a non-negative measure such that $\nu \ll m$ and $\mu \ll m$ (one can choose $m = \nu + \mu$). The Radon–Nikodym theorem provides non-negative measurable functions h_{ν} and h_{μ} such that

$$\nu(dx) = h_{\nu}(x)m(dx) \quad \text{and} \quad \mu(dx) = h_{\mu}(x)m(dx).$$

Let $B := \{h_{\nu} \geq h_{\mu}\}$ and $\overline{B} := \mathbb{R}^d \setminus B$. We choose $g(x) := \phi(x) \mathbb{1}_B(x) - \phi(x) \mathbb{1}_{\overline{B}}(x)$. It satisfies (1) and

$$\int \phi(x)|\nu - \mu|(dx) = \int g(x)(\nu - \mu)(dx)$$

Step 2. We prove that $\sup_g \int g(x)(\nu-\mu)(dx) \leq \inf_{\mathcal{L}(X)=\nu,\mathcal{L}(Y)=\mu} \mathbb{E}(\phi(X) + \phi(Y))\mathbb{1}_{X \neq Y}$.

Let g such that (1) holds. Consider π any coupling between ν and μ , we have

$$\begin{split} \int g(x)\nu(dx) &- \int g(y)\mu(dy) = \int \int (g(x) - g(y))\pi(dx, dy) \\ &\leq \int \int |g(x) - g(y)|\pi(dx, dy) \\ &\leq \int \int (\phi(x) + \phi(y)) \mathbb{1}_{x \neq y}\pi(dx, dy) \end{split}$$

Taking the infimum of the right-hand side over all the couplings gives the result. Step 3. We prove that $\inf_{\mathcal{L}(X)=\nu,\mathcal{L}(Y)=\mu} \mathbb{E}(\phi(X) + \phi(Y))\mathbb{1}_{X\neq Y} \leq d_{TV}^{\phi}(\nu,\mu)$. We use the following explicit coupling taken from [Vil09]

$$\pi(dx, dy) := (\nu \wedge \mu)(dx)\delta_x(dy) + \frac{1}{a}(\nu - \mu)_+(dx)(\nu - \mu)_-(dy), \qquad (2)$$

where $a := \int (\nu - \mu)_+ = \int (\nu - \mu)_-$ and $\nu \wedge \mu := \nu - (\nu - \mu)_+$. Let $(X, Y) \sim \pi$. Using the definition of π ,

$$\mathbb{E}\phi(X)\mathbb{1}_{X\neq Y} + \mathbb{E}\phi(Y)\mathbb{1}_{X\neq Y} = \frac{1}{a}\int\int (\phi(x) + \phi(y))(\nu - \mu)_{+}(dx)(\nu - \mu)_{-}(dy)$$
$$= \int \phi(x)(\nu - \mu)_{+}(dx) + \int \phi(y)(\nu - \mu)_{-}(dy)$$
$$= d_{TV}^{\phi}(\nu, \mu).$$

Combining the three steps ends the proof.

As an application, using the coupling (2), we can bound the usual Wasserstein distance W_p $(p \ge 1)$ by the total variation with weight $\phi(x) = |x|^p$, see [Vil09, Th. 6.15].

The Kantorovich duality theorem can be rephrased as a supremum over measurable functions g bounded by ϕ , as shown in the next lemma.

Lemma 4. It holds that:

$$d_{TV}^{\phi}(\mu,\nu) = \sup_{g} \int g(x)(\nu-\mu)(dx),$$
(3)

where the supremum is taken over measurable functions g with $|g(x)| \leq \phi(x)$.

Proof. The following argument is taken from [HM11]. First if $|g(x)| \leq \phi(x)$ everywhere, it holds that $|g(x) - g(y)| \leq (\phi(x) + \phi(y)) \mathbb{1}_{x \neq y}$. Conversely, assume that g is such that for all $x \neq y$, $|g(x) - g(y)| \leq \phi(x) + \phi(y)$. Let $c = \inf_x \phi(x) - g(x)$. We have $g(x) + c \leq g(x) + \phi(x) - g(x) = \phi(x)$. In addition,

$$g(x) + c = \inf_{y} \phi(y) - g(y) + g(x) \ge \inf_{y} \phi(y) - [\phi(x) + \phi(y)] = -\phi(x).$$

Altogether, it holds that $|g(x) + c| \le \phi(x)$ everywhere. Because adding a constant to g does not change the value of the right-hand side part of (3), the result is proved.

References

- [CD18] R. Carmona and F. Delarue. Probabilistic theory of mean field games with applications. I. Vol. 83. Probability Theory and Stochastic Modelling. Mean field FBSDEs, control, and games. Springer, Cham, 2018, pp. xxv+713.
- [HM11] M. Hairer and J. C. Mattingly. "Yet another look at Harris' ergodic theorem for Markov chains". English. In: Seminar on stochastic analysis, random fields and applications VI. Centro Stefano Franscini, Ascona (Ticino), Switzerland, May 19–23, 2008. Basel: Birkhäuser, 2011, pp. 109–117. DOI: 10.1007/978-3-0348-0021-1_7.
- [Vil09] C. Villani. Optimal transport. Vol. 338. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Old and new. Springer-Verlag, Berlin, 2009, pp. xxii+973. DOI: 10.1007/978-3-540-71050-9.