Lower Bounds for the Total Variation Distance Between Arbitrary Distributions with Given Means and Variances

Tomohiro Nishiyama
Email: htam0ybboh@gmail.com

Abstract

For arbitrary two probability measures on real d-space with given means and variances (covariance matrices), we provide lower bounds for their total variation distance.

I. INTRODUCTION

The f-divergence \([12]\) is an important class of divergence measures, defined by means of convex functions \(f\), includes many important divergences such as the total variation (TV) distance and the Kullback-Leibler (KL) divergence \([7]\). Given means and variances of two probability measures, closed-form lower bounds for the f-divergence are useful in practice because they can be directly evaluated based on only moments of distributions. These bounds are now beginning to be applied to nonequilibrium physics \([5]\), \([13]\), \([4]\).

In our previous works, we provided tight lower bounds for the KL-divergence and the squared Hellinger distance \([6]\) on real line \([11]\), \([8]\). We generalized these bounds for the asymmetric \(\alpha\)-divergence \([1]\) with \(\alpha \in [-1, 2]\) \([9]\). We also provided a sufficient condition that an arbitrary symmetric f-divergence has a closed-form expression for lower bound \([10]\). Although the TV distance is a symmetric f-divergence, it does not satisfy this sufficient condition. On the other hand, for the TV distance between single Gaussians or Gaussian mixtures on real d-space, the closed-form expressions for lower bounds have recently been derived \([2]\), \([3]\).

In this work, we study closed-form lower bounds for the TV distance between arbitrary probability measures on real d-space with given means and variances (covariance matrices). In particular, we provide a nearly-tight bound for two probability measures on real line with approximately equal variances. For the shared variance case, this bound is tight. If probability measures have different variances, we leave a tight bound as an open problem.

II. MAIN RESULTS

A. Preliminaries

We provide definitions which are used in this paper.

Definition 1. Let \(P\) and \(Q\) be probability measures defined on a common measurable space \((\mathcal{A}, \mathcal{F})\). Let \(\mu\) be a dominating measure of \(P\) and \(Q\) (i.e., \(P, Q \ll \mu\)), and let \(p := \frac{dP}{d\mu}\) and \(q := \frac{dQ}{d\mu}\) be the densities of \(P\) and \(Q\) with respect to \(\mu\). The total variation (TV) distance between \(P\) and \(Q\) is given by

\[
d_{TV}(P, Q) := \frac{1}{2} \int |p - q|d\mu = \sup_{\mathcal{F} \in \mathcal{F}} |P(\mathcal{F}) - Q(\mathcal{F})|.
\]

Definition 2. Let \(P\) and \(Q\) be probability measures on \(\mathbb{R}\). Let \(m_P, m_Q, \sigma_P^2\), and \(\sigma_Q^2\) denote the means and the variances of \(X \sim P\) and \(Y \sim Q\), i.e.,

\[
m_P := \mathbb{E}[X], \quad m_Q := \mathbb{E}[Y],
\]

\[
\sigma_P^2 := \mathbb{E}[(X - m_P)^2], \quad \sigma_Q^2 := \mathbb{E}[(Y - m_Q)^2].
\]

A set of pairs of probability measures \((P, Q)\) with given means \((m_P, m_Q)\) and variances \((\sigma_P^2, \sigma_Q^2)\) is defined as \(\mathcal{P}[m_P, \sigma_P; m_Q, \sigma_Q]\). Similarly, a set of pairs of probability measures \((P, Q)\) on \(\mathbb{R}^d\) with given means \((m_P, m_Q)\) and covariance matrices \((\Sigma_P, \Sigma_Q)\) is defined as \(\mathcal{P}[m_P, \Sigma_P; m_Q, \Sigma_Q]\), where \(\Sigma_P := \mathbb{E}[(X - m_P)(X - m_P)^T]\) and \(\Sigma_Q := \mathbb{E}[(Y - m_Q)(Y - m_Q)^T]\).
B. Lower bounds for the TV distance

Our main result is the following lower bound between a pair of arbitrary probability measures on real line with given means and variances.

**Theorem 1.** Let \((P, Q) \in \mathcal{P}[m_P, \sigma_P; m_Q, \sigma_Q]\).

(a) If \(m_P \neq m_Q\), then

\[
\frac{a^2}{2(\sigma_P^2 + \sigma_Q^2) + a^2} \leq \inf d_{TV}(P, Q) \leq \frac{a^2}{(\sigma_P + \sigma_Q)^2 + a^2},
\]

where \(a := m_P - m_Q\).

Furthermore, if \(\sigma_P = \sigma_Q =: \sigma\), the inequality \((1)\) reduces to

\[
\min d_{TV}(P, Q) = \frac{a^2}{4\sigma^2 + a^2}.
\]

(b) When \(\sigma > 0\), the lower bound on the right side of \((2)\) is attained by \(P\) and \(Q\) which are defined on a common three-element set \(\{x_1, x_2, x_3\}\), and

\[
p := \frac{a^2}{4\sigma^2 + a^2} \in (0, 1),
\]

\[
P = (1 - p, p, 0), \quad Q = (1 - p, 0, p),
\]

\[
x_1 = m_P - \frac{a}{2}, \quad x_2 = m_P + \frac{2\sigma^2}{a}, \quad x_3 = m_Q - \frac{2\sigma^2}{a}.
\]

(c) If \(m_P = m_Q\), then

\[
\inf d_{TV}(P, Q) = 0.
\]

The minimum and infimum in \((1), (2), \text{and } (6)\) are taken over all \((P, Q) \in \mathcal{P}[m_P, \sigma_P; m_Q, \sigma_Q]\).

**Proof.** See Section III.

It should be noted that the first inequality in \((1)\) gives a nearly-tight bound when \(\frac{(\sigma_P - \sigma_Q)^2}{2(\sigma_P^2 + \sigma_Q^2) + a^2} \ll 1\) by combining \(2(\sigma_P^2 + \sigma_Q^2) = (\sigma_P + \sigma_Q)^2 + (\sigma_P - \sigma_Q)^2\) with the second inequality in \((1)\). For KL-divergence, \(\chi^2\)-divergence and squared Hellinger distance, their divergences between probability measures \(R = (1 - r, r)\) and \(S := (1 - s, s)\) attain lower bounds with given means and variances. However, the TV distance does not have this property for the following reason. By Theorem 2 in \([11]\), we have \(r = s + \frac{a|a|}{v}\), where \(v := \sqrt{(\sigma_Q^2 - \sigma_P^2)^2 + 2a^2(\sigma_P^2 + \sigma_Q^2) + a^4}\). When \(\sigma_P\) and \(\sigma_Q\) are positive, by combining \(d_{TV}(R, S) = |r - s| = \frac{a^2}{v}\) and \((\sigma_P + \sigma_Q)^2 + a^2 > v\) with \((1)\), we have \(d_{TV}(R, S) > \inf d_{TV}(P, Q)\).

We next provide a lower bound for probability measures on \(\mathbb{R}^d\).

**Proposition 1.** Let \(P\) and \(Q\) be probability measures on \(\mathbb{R}^d\), and let \((P, Q) \in \mathcal{P}[m_P, \Sigma_P; m_Q, \Sigma_Q]\). Then,

\[
d_{TV}(P, Q) \geq \frac{\mathbf{a}^T \mathbf{a}}{2(\text{tr}(\Sigma_P) + \text{tr}(\Sigma_Q)) + \mathbf{a}^T \mathbf{a}},
\]

where \(\text{tr}(A)\) denotes the trace of a matrix \(A\), and \(\mathbf{a} := m_P - m_Q\).

**Proof.** See Section III.
A. Proofs of Lemmas

Before proving Theorem 1, we prove the following lemmas.

Lemma 1. Let \( P = (1 - p, p, 0) \) and \( Q = (1 - p, 0, p) \) be probability measures on a common three-element set \( \{x_1, x_2, x_3\} \). Let \( (P, Q) \in \mathcal{P}[m_P, \sigma_P; m_Q, \sigma_Q] \) for \( \sigma_P > 0, \sigma_Q > 0, \) and \( m_P \neq m_Q \). Then, there exist probability measures \( P \) and \( Q \) such that

\[
d_{TV}(P, Q) = \frac{a^2}{(\sigma_P + \sigma_Q)^2 + a^2}. \tag{8}
\]

Proof. The moment constraints reduce to

\[
\begin{align*}
(1 - p)x_1 + px_2 &= m_P, \\
(1 - p)x_1^2 + px_2^2 &= m_P^2 + \sigma_P^2, \\
(1 - p)x_1 + px_3 &= m_Q, \\
(1 - p)x_1^2 + px_3^2 &= m_Q^2 + \sigma_Q^2.
\end{align*} \tag{9}
\]

Subtracting the square of the first equation in (9) from its second equation, we have

\[
x_1 - x_2 = \pm \sigma_P \sqrt{\frac{1}{p(1 - p)}}. \tag{10}
\]

We first consider the first option in (10). Solving simultaneously this relation and the first equation in (9), we obtain

\[
\begin{align*}
x_1 &= m_P + \sigma_P \sqrt{\frac{p}{1 - p}}, \\
x_2 &= m_P - \sigma_P \sqrt{\frac{1 - p}{p}}. \tag{11}
\end{align*}
\]

Similarly, from the third and the forth equation in (9), we obtain

\[
\begin{align*}
x_1 &= m_Q \mp \sigma_Q \sqrt{\frac{p}{1 - p}}, \\
x_3 &= m_Q \pm \sigma_Q \sqrt{\frac{1 - p}{p}}. \tag{14}
\end{align*}
\]

By combining (11) with the first option in (13), we have

\[
- \frac{a}{\sigma_P + \sigma_Q} = \sqrt{\frac{p}{1 - p}}, \quad a < 0. \tag{15}
\]

Solving this equation for \( p \), we have

\[
p = \frac{a^2}{(\sigma_P + \sigma_Q)^2 + a^2} \in (0, 1). \tag{16}
\]

It can be verified that \( x_1, x_2, \) and \( x_3 \) are different from each other from (11)-(14). Since \( d_{TV}(P, Q) = p \), we have

\[
d_{TV}(P, Q) = \frac{a^2}{(\sigma_P + \sigma_Q)^2 + a^2}. \tag{17}
\]

We next consider the second option in (10). By replacing \( \sigma_P \rightarrow -\sigma_P \) in (11), and combining it with the second option in (13), we obtain (16) and (17) for \( a > 0 \). One can verify that (11), (12), (14), and (16) satisfy (9) for \( a < 0 \). Similarly, solutions for \( a > 0 \) satisfy (9) .
**Lemma 2.** Let $P = (1 - p, p)$ and $Q = (1, 0)$ be probability measures on a common two-element set $\{m_Q, x\}$. Let $(P, Q) \in \mathcal{P}[m_P, \sigma_P; m_Q, 0]$ for $\sigma_P > 0$, and $m_P \neq m_Q$. Then, there exist probability measures $P$ and $Q$ such that

$$d_{TV}(P, Q) = \frac{a^2}{\sigma_P^2 + a^2}.$$

(18)

When $\sigma_P = 0$ and $\sigma_Q > 0$, a similar result holds by switching $P$ and $Q$.

**Proof.** The moment constraints reduce to

\[
\begin{cases}
(1 - p)m_Q + px = m_P, \\
(1 - p)m_Q^2 + px^2 = m_P^2 + \sigma_P^2.
\end{cases}
\]

(19)

In the similar way to the proof of Lemma 1 we have

$$m_Q = m_P \pm \sigma_P \sqrt{\frac{p}{1 - p}},$$

(20)

and

$$x = m_P \mp \sigma_P \sqrt{\frac{1 - p}{p}}.$$  

(21)

From the first option in (20), we obtain

$$p = \frac{a^2}{\sigma_P^2 + a^2} \in (0, 1), \quad a < 0.$$  

(22)

The second option in (20) gives the same result for $p$ when $a > 0$. It can be verified that (20)-(22) satisfy (19). Since the TV distance between $P$ and $Q$ is given by $d_{TV}(P, Q) = p$, we obtain (18). Switching $P$ and $Q$ complete the proof.

**B. Proof of Theorem 1**

**Proof.** We first prove Item (a) in Theorem 1. Since the TV distance is invariant under transformation $x \rightarrow x - \frac{m_P + m_Q}{2}$, one can assume $(P, Q) \in \mathcal{P}[\frac{a}{2}, \sigma_P; -\frac{a}{2}, \sigma_Q]$ without any loss of generality. By the Cauchy-Schwarz inequality, we have

$$\left( \int (p - q) |x| d\mu \right)^2 = \left( \int |p - q| \sqrt{\int (p - q)^2 d\mu} \right)^2 
\leq \int |p - q| d\mu \int (p + q) x^2 d\mu
= 2d_{TV}(P, Q) \left( \sigma_P^2 + \sigma_Q^2 + \frac{a^2}{2} \right).$$

(23)

By combining this inequality with $\int |p - q| |x| d\mu \geq \int |(p - q) x d\mu| = |a|$ gives the first inequality in (11). Its second inequality follows from Lemma 1 and Lemma 2 since the case when $\sigma_P = \sigma_Q = 0$ is trivial. We obtain the equation (2) by substituting $\sigma = \sigma_P = \sigma_Q$ into (11). Item (b) also follows from Lemma 1 by substituting $\sigma = \sigma_P = \sigma_Q$ into (11), (12), (14), and (16) when $a < 0$. Similarly, we obtain the result for $a > 0$.

We next prove Item (c) in Theorem 1. Since (6) is symmetric with respect to $P$ and $Q$, it is sufficient to prove for $\sigma_P > \sigma_Q$. Let $m := m_P = m_Q$, and let

$$P_k(x) := \begin{cases}
\frac{1}{2k} - \frac{1}{2k}, & x = m \pm \sigma_Q, \\
\frac{1}{2k}, & x = m \pm \sqrt{(\sigma_P^2 - \sigma_Q^2)k + \sigma_Q^2},
\end{cases}$$

and

$$Q_k(x) := \begin{cases}
\frac{1}{2}, & x = m \pm \sigma_Q, \\
0, & x = m \pm \sqrt{(\sigma_P^2 - \sigma_Q^2)k + \sigma_Q^2}
\end{cases}$$

for sufficiently large $k$. As $k \rightarrow \infty$, we have $d_{TV}(P_k, Q_k) = \frac{1}{k} \rightarrow 0$. \hfill \qed
C. Proof of Proposition 1

By transformation $x \rightarrow x - \frac{m_P + m_Q}{2}$ and using a similar calculation in (23), we have

$$2d_{TV}(P, Q) \left( (\Sigma_P)_{kk}^2 + (\Sigma_Q)_{kk}^2 + \frac{a_k^2}{2} \right) \geq a_k^2. \quad (24)$$

Taking the sum over $1 \leq k \leq d$ for this inequality gives (7).

IV. Conclusion

We provided lower bounds for the TV distance between probability measures on real $d$-space with given means and variances. When $d = 1$ and two probability measures have the same variance, the tight bound was given. We finally conclude with a conjecture.

**Conjecture:** When probability measures on $\mathbb{R}$ have different variances, a tight lower bound is given by $P$ and $Q$ defined on a common $n$-point set $\{x_1, x_2, \cdots, x_n\}$ such that

$$P(x_1) = p_1, \quad Q(x_1) = q_1,$$
$$P(x_2) = p_2, \quad Q(x_2) = q_2,$$
$$P(x_i) = Q(x_i) = p_i, \quad \text{for } 3 \leq i \leq n,$$

where $\sum_{i=3}^{n} p_i = 1 - p_1 - p_2 = 1 - q_1 - q_2$, and $p_i, q_i \in [0, 1]$ for all $1 \leq i \leq n$.

**References**

[1] A. Cichocki and S.-i. Amari. Families of alpha-beta-and gamma-divergences: Flexible and robust measures of similarities. *Entropy*, 12(6):1532–1568, 2010.

[2] S. Davies, A. Mazumdar, S. Pal, and C. Rashtchian. Lower bounds on the total variation distance between mixtures of two gaussians. In *International Conference on Algorithmic Learning Theory*, pages 319–341. PMLR, 2022.

[3] L. Devroye, A. Mehrabian, and T. Reddad. The total variation distance between high-dimensional gaussians. *arXiv preprint arXiv:1810.08693*, 2018.

[4] G. Falasco, M. Esposito, and J.-C. Delvenne. Beyond thermodynamic uncertainty relations: nonlinear response, error-dissipation trade-offs, and speed limits. *Journal of Physics A: Mathematical and Theoretical*, 55(12):124002, 2022.

[5] Y. Hasegawa. Irreversibility, loschmidt echo, and thermodynamic uncertainty relation. *Physical Review Letters*, 127(24):240602, 2021.

[6] E. Hellinger. Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1909(136):210–271, 1909.

[7] S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[8] T. Nishiyama. A tight lower bound for the hellinger distance with given means and variances. *arXiv preprint arXiv:2010.13548*, 2020.

[9] T. Nishiyama. Tight lower bounds for $\alpha$-divergences under moment constraints and relations between different $\alpha$. *arXiv preprint arXiv:2105.12972*, 2021.

[10] T. Nishiyama. On relations between tight bounds for symmetric $f$-divergences and binary divergences. *arXiv preprint arXiv:2210.09571*, 2022.

[11] T. Nishiyama and I. Sason. On relations between the relative entropy and $\chi^2$-divergence, generalizations and applications. *Entropy*, 22(5):563, 2020.

[12] I. Sason and S. Verdú. $f$-divergence inequalities. *IEEE Transactions on Information Theory*, 62(11):5973–6006, 2016.

[13] T. Van Vu, Y. Hasegawa, et al. Unified approach to classical speed limit and thermodynamic uncertainty relation. *Physical Review E*, 102(6):062132, 2020.