Sharp uniform bound between TV and Hellinger distances

Joonhyuk Jung

Oct 6, 2025

THE UNIVERSITY OF CHICAGO

About Me

1. Seoul National University (2017-2023)
   - B.S. in Statistics and B.A. in Economics
2. The University of Chicago (2024-)
   - 2nd-year Ph.D. Student in Statistics
   - Advised by Prof. Chao Gao

Please visit https://joonhyuk.com for more details.

Problem Setup

In general, we have

$$H^2(p,q) \lesssim \mathrm{TV}(p,q) \lesssim H(p,q) \lesssim KL^{1/2}(p||q) \lesssim \chi(p||q),$$

where we define

$$\begin{aligned} H^2(p,q) &:= \frac{1}{2}\int(\sqrt{p}-\sqrt{q})^2, \\ \mathrm{TV}(p,q) &:= \frac{1}{2}\int|p-q|, \\ KL(p||q) &:= \int p\log\frac{p}{q}, \\ \chi^2(p||q) &:= \int\frac{(p-q)^2}{q}. \end{aligned}$$

Gaussian Location Mixture Models

If we assume p and q to be Gaussian location mixtures:

$$p(x) = f_\pi(x) = \int \phi(x-\theta)\, d\pi(\theta),$$
$$q(x) = f_\eta(x) = \int \phi(x-\theta)\, d\eta(\theta),$$

then it was shown in 2023 that

$$H(f_\pi, f_\eta) \asymp KL^{1/2}(f_\pi || f_\eta)$$

provided that π and η are compactly supported [JPW23].

- Here, "$\asymp$" is hiding constants depending on the radius of support, but not depending on π or η.

Question

Under the same circumstances, is it also true that

$$\mathrm{TV}(f_\pi, f_\eta) \asymp H(f_\pi, f_\eta)?$$

- In robust statistics, results on learning in TV are abundant, whereas results on learning in H are rather scarce. Therefore, if one could answer the above question, many results would follow.

Question

Under the same circumstances, is it also true that

$$\mathrm{TV}(f_\pi, f_\eta) \asymp H(f_\pi, f_\eta)?$$

- In robust statistics, results on learning in TV are abundant, whereas results on learning in H are rather scarce. Therefore, if one could answer the above question, many results would follow.
- This has been open at least since 2023.

Question

Under the same circumstances, is it also true that

$$\mathrm{TV}(f_\pi, f_\eta) \asymp H(f_\pi, f_\eta)?$$

- In robust statistics, results on learning in TV are abundant, whereas results on learning in H are rather scarce. Therefore, if one could answer the above question, many results would follow.
- This has been open at least since 2023.
- It turns out that the answer is NO. That is, $H(f_\pi, f_\eta)$ cannot be linearly and uniformly upper bounded by $\mathrm{TV}(f_\pi, f_\eta)$.

$$H(f_\pi, f_\eta) \not\lesssim \mathrm{TV}(f_\pi, f_\eta).$$

Main Result

Theorem (Nonlinear uniform bound of Hellinger distances)

Let π and η be probability measures supported on $[-\frac{M}{2}, \frac{M}{2}]$. Then,

$$H(f_\pi, f_\eta) \lesssim \mathrm{TV}(f_\pi, f_\eta)^{1 - \Omega\left(1/\log\log(1/\mathrm{TV}(f_\pi, f_\eta))\right)}.$$

- The exponent correction of $1/\log\log(1/t)$ from the above bound cannot be improved up to constant factors.

Sharpness

Theorem (Sharpness)

There exist two sequences of probability measures $\{\pi_n\}$ *and* $\{\eta_n\}$ *supported on* $[-\frac{M}{2}, \frac{M}{2}]$ *such that, if we define*

$$H_n := H(f_{\pi_n}, f_{\eta_n}), \qquad \mathrm{TV}_n := \mathrm{TV}(f_{\pi_n}, f_{\eta_n}),$$

then $H_n \downarrow 0$ *as* $n \to \infty$, *and moreover it holds for all* n *that*

$$H_n \geq \mathrm{TV}_n^{1-\epsilon^*(\mathrm{TV}_n)},$$

where we define

$$\epsilon^*(t) := \frac{0.33}{\log\log(1/t)}.$$

References I

Zeyu Jia, Yury Polyanskiy, and Yihong Wu, *Entropic characterization of optimal rates for learning gaussian mixtures*, The Thirty Sixth Annual Conference on Learning Theory, PMLR, 2023, pp. 4296–4335.