Proof of inequalities

Rita Lyu

October 29, 2024

1 Problem 1

Let $\mathcal{X} = \mathbb{R}^d$. Given two distributions $\mathbb{P}$ and $\mathbb{Q}$ with densities p and q, define the Hellinger distance as

$$H(\mathbb{P}, \mathbb{Q}) = \sqrt{\int_{\mathcal{X}} (\sqrt{p(x)} - \sqrt{q(x)})^2 dx}$$

and the total variation distance as

$$TV(\mathbb{P}, \mathbb{Q}) = \sup_{A \subset \mathcal{X}} \int_A |p(x) - q(x)| dx$$

where the supremum is taken over all measurable subsets A. Prove the following relations: 1.

$$\int_{\mathcal{X}} \min(p(x), q(x)) dx \geq \frac{1}{2} \left(\int_{\mathcal{X}} \sqrt{p(x) q(x)} dx \right)^2 = \frac{1}{2} \left(1 - \frac{H^2(\mathbb{P}, \mathbb{Q})}{2} \right)^2.$$

2.

$$\frac{H^2(\mathbb{P}, \mathbb{Q})}{2} \leq TV(\mathbb{P}, \mathbb{Q}) \leq H(\mathbb{P}, \mathbb{Q}) \sqrt{1 - \frac{H^2(\mathbb{P}, \mathbb{Q})}{4}}$$

3. Show that if $\mathbb{P} = \mathbb{P}' \otimes \cdots \otimes \mathbb{P}'$ and $\mathbb{Q} = \mathbb{Q}' \otimes \cdots \otimes \mathbb{Q}'$, both n times, then

$$H^2(\mathbb{P}, \mathbb{Q}) = 2 \left(1 - \left(1 - \frac{H^2(\mathbb{P}', \mathbb{Q}')}{2} \right)^n \right)$$

4.

$$H(\mathbb{P}, \mathbb{Q}) \leq \sqrt{\mathcal{KL}(\mathbb{P}, \mathbb{Q})}, \quad \text{and} \quad TV(\mathbb{P}, \mathbb{Q}) \leq \sqrt{\mathcal{KL}(\mathbb{P}, \mathbb{Q})}$$

1.1

Observe that

$$p(x) q(x) = \min(p(x), q(x)) \max(p(x), q(x)).$$

By Cauchy-Schwartz inequality, we know

$$\left(\int_{\mathcal{X}} \sqrt{\min(p(x), q(x)) \max(p(x), q(x))} \mathrm{d}x \right)^2 \leq \int_{\mathcal{X}} \min(p(x), q(x)) \mathrm{d}x \int_{\mathcal{X}} \max(p(x), q(x)) \mathrm{d}x.$$

Notice

$$\int_{\mathcal{X}} (\max(p(x), q(x)) + \min(p(x), q(x)) \mathrm{d}x = 2.$$

We have

$$\int_{\mathcal{X}} \min(p(x), q(x))\mathrm{d}x \int_{\mathcal{X}} \max(p(x), q(x))\mathrm{d}x = \int_{\mathcal{X}} \min(p(x), q(x))\mathrm{d}x \int_{\mathcal{X}} (2 - \min(p(x), q(x)))\mathrm{d}x$$
$$\frac{\int_{\mathcal{X}} \min(p(x), q(x))\mathrm{d}x \int_{\mathcal{X}}(2 - \min(p(x), q(x)))\mathrm{d}x}{2} = \int_{\mathcal{X}} \min(p(x), q(x))\mathrm{d}x \int_{\mathcal{X}} \left(1 - \frac{\min(p(x), q(x))}{2}\right) \mathrm{d}x$$
$$\leq \int_{\mathcal{X}} \min(p(x), q(x))\mathrm{d}x, \text{ as} \int_{\mathcal{X}} \left(1 - \frac{\min(p(x), q(x))}{2}\right) \mathrm{d}x \leq 1.$$

Thus, we prove the first inequality. For the second equality, by the definition of Hellinger distance, we have

$$\frac{H^2(\mathbb{P}, \mathbb{Q})}{2} = \frac{\int_{\mathcal{X}} \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 \mathrm{d}x}{2} = 1 - \int_{\mathcal{X}} \sqrt{p(x)q(x)}\mathrm{d}x$$
$$\frac{1}{2}\left(\int_{\mathcal{X}} \sqrt{p(x)q(x)}\mathrm{d}x\right)^2 = \frac{1}{2}\left(1 - \frac{H^2(\mathbb{P}, \mathbb{Q})}{2}\right)^2.$$

1.2

We need to first prove an equivalent definition of the total variation distance

$$TV(\mathbb{P}, \mathbb{Q}) = \sup_{A \subset \mathcal{X}} \left| \int_A (p(x) - q(x))\mathrm{d}x \right| = \frac{1}{2} \int_{\mathcal{X}} |p(x) - q(x)|\mathrm{d}x.$$

This could be proved as follows, define $\mathcal{A}_0 = \{x \in \mathcal{X} : q(x) \geq p(x)\}$:

$$TV(\mathbb{P}, \mathbb{Q}) \geq \left| \int_{\mathcal{A}_0} (p(x) - q(x))\mathrm{d}x \right| = \int_{\mathcal{A}_0} (q(x) - p(x))\mathrm{d}x.$$

Note that:

$$\int_{\mathcal{X}} |p(x) - q(x)|\mathrm{d}x = \int_{\mathcal{A}_0} (q(x) - p(x))\mathrm{d}x + \int_{\mathcal{A}_0^c} (p(x) - q(x))\mathrm{d}x$$
$$\int_{\mathcal{X}} q(x)\mathrm{d}x = \int_{\mathcal{X}} p(x)\mathrm{d}x \Rightarrow \int_{\mathcal{A}_0} (q(x) - p(x))\mathrm{d}x = \int_{\mathcal{A}_0^c} (p(x) - q(x))\mathrm{d}x$$
$$\Rightarrow TV(\mathbb{P}, \mathbb{Q}) \geq \frac{1}{2} \int_{\mathcal{X}} |p(x) - q(x)|\mathrm{d}x.$$

And for any $A \subset \mathcal{X}$,

$$\left| \int_A (p(x) - q(x))\mathrm{d}x \right| = \left| \int_{A \cap \mathcal{A}_0} (p(x) - q(x))\mathrm{d}x + \int_{A \cap \mathcal{A}_0^c} (p(x) - q(x))\mathrm{d}x \right|$$
$$\leq \max \left\{ \int_{A \cap \mathcal{A}_0} (q(x) - p(x))\mathrm{d}x, \int_{A \cap \mathcal{A}_0^c} (p(x) - q(x))\mathrm{d}x \right\}$$
$$\leq \max \left\{ \int_{\mathcal{A}_0} (q(x) - p(x))\mathrm{d}x, \int_{\mathcal{A}_0^c} (p(x) - q(x))\mathrm{d}x \right\} = \frac{1}{2} \int_{\mathcal{X}} |p(x) - q(x)|\mathrm{d}x.$$

Thus,

$$TV(\mathbb{P}, \mathbb{Q}) = \sup_{A \subset \mathcal{X}} \left| \int_A (p(x) - q(x))\mathrm{d}x \right| \leq \frac{1}{2} \int_{\mathcal{X}} |p(x) - q(x)|\mathrm{d}x = 1 - \int_{\mathcal{X}} \min(p(x), q(x))\mathrm{d}x,$$

and

$$TV(\mathbb{P},\mathbb{Q}) = \frac{1}{2}\int_{\mathcal{X}} |p(x)-q(x)|\mathrm{d}x = \frac{1}{2}\left(\int_{\mathcal{A}_0}(q(x)-p(x))\mathrm{d}x + \int_{\mathcal{A}_0^c}(p(x)-q(x))\mathrm{d}x\right).$$

We only need to prove

$$1-\int_{\mathcal{X}}\sqrt{p(x)q(x)}\mathrm{d}x \le \frac{1}{2}\int_{\mathcal{X}}|p(x)-q(x)|\mathrm{d}x = 1-\int_{\mathcal{X}}\min(p(x),q(x))\mathrm{d}x.$$

This is obvious because

$$\sqrt{p(x)q(x)} \le \min(p(x),q(x)).$$

We move on to the right-hand side inequality, and we know

$$TV(\mathbb{P},\mathbb{Q}) = 1-\int_{\mathcal{X}}\min(p(x),q(x))\mathrm{d}x \ge 1-\int_{\mathcal{X}}\sqrt{p(x)q(x)}\mathrm{d}x = \frac{H^2(\mathbb{P},\mathbb{Q})}{2}.$$

We also have proven in 1.1

$$\begin{aligned}
&\left(1-\frac{H^2(\mathbb{P},\mathbb{Q})}{2}\right)^2 \le \int_{\mathcal{X}}\min(p(x),q(x))\mathrm{d}x\int_{\mathcal{X}}(2-\min(p(x),q(x)))\mathrm{d}x = (1-TV(\mathbb{P},\mathbb{Q}))(1+TV(\mathbb{P},\mathbb{Q}))\\
&= 1-TV^2(\mathbb{P},\mathbb{Q})\\
&\Rightarrow TV^2(\mathbb{P},\mathbb{Q}) \le \frac{-H^4(\mathbb{P},\mathbb{Q})}{4}+H^2(\mathbb{P},\mathbb{Q}) \quad TV(\mathbb{P},\mathbb{Q}) \le H(\mathbb{P},\mathbb{Q})\sqrt{1-\frac{H^2(\mathbb{P},\mathbb{Q})}{4}}.
\end{aligned}$$

1.3

$$\begin{aligned}
&H^2(\mathbb{P},\mathbb{Q}) = \int_{\mathcal{X}\times\cdots\times\mathcal{X}}\left(\sqrt{p(\mathbf{x})}-\sqrt{q(\mathbf{x})}\right)^2\mathrm{d}\mathbf{x} = \int_{\mathcal{X}\times\cdots\times\mathcal{X}}\left(p(\mathbf{x})+q(\mathbf{x})-2\sqrt{p(\mathbf{x})q(\mathbf{x})}\right)\mathrm{d}\mathbf{x}\\
&= 2-2\int_{\mathcal{X}\times\cdots\times\mathcal{X}}\sqrt{p(\mathbf{x})q(\mathbf{x})}\mathrm{d}\mathbf{x} = 2-2\int_{\mathcal{X}}\sqrt{p'(x_1)q'(x_1)}\mathrm{d}x_1\int_{\mathcal{X}}\sqrt{p'(x_2)q'(x_2)}\mathrm{d}x_2\cdots\int_{\mathcal{X}}\sqrt{p'(x_n)q'(x_n)}\mathrm{d}x_n\\
&= 2-2\left(\int_{\mathcal{X}}\sqrt{p'(x)q'(x)}\mathrm{d}x\right)^n,\\
&\left(1-\frac{H^2(\mathbb{P}',\mathbb{Q}')}{2}\right) = \int_{\mathcal{X}}\sqrt{p(x)'q(x)'}\mathrm{d}x,\\
&\Rightarrow H^2(\mathbb{P},\mathbb{Q}) = 2\left(1-\left(1-\frac{H^2(\mathbb{P}',\mathbb{Q}')}{2}\right)^n\right).
\end{aligned}$$

1.4

$$\begin{aligned}
\mathcal{KL}(\mathbb{P},\mathbb{Q}) &= \int_{\mathcal{X}}\log\left(\frac{p(x)}{q(x)}\right)p(x)\mathrm{d}x = 2\int_{\mathcal{X}}\log\left(\frac{\sqrt{p(x)}}{\sqrt{q(x)}}\right)p(x)\mathrm{d}x\\
&= -2\int_{\mathcal{X}}\log\left(\frac{\sqrt{q(x)}-\sqrt{p(x)}+\sqrt{p(x)}}{\sqrt{p(x)}}\right)p(x)\mathrm{d}x\\
&\ge -2\int_{\mathcal{X}}\frac{\sqrt{q(x)}-\sqrt{p(x)}}{\sqrt{p(x)}}p(x)\mathrm{d}x = -2\left(\int_{\mathcal{X}}\sqrt{p(x)q(x)}\mathrm{d}x-1\right) = H^2(\mathbb{P},\mathbb{Q}).
\end{aligned}$$

In 1.2, we have shown $TV(\mathbb{P},\mathbb{Q})$ could be bounded by Hellinger distance,then

$$TV(\mathbb{P},\mathbb{Q}) \le H(\mathbb{P},\mathbb{Q})\sqrt{1-\frac{H^2(\mathbb{P},\mathbb{Q}}{4}} \le H^2(\mathbb{P},\mathbb{Q}) \le \mathcal{KL}(\mathbb{P},\mathbb{Q}).$$

Then we could prove

$$H(\mathbb{P},\mathbb{Q}) \le \sqrt{\mathcal{KL}(\mathbb{P},\mathbb{Q})},\ TV(\mathbb{P},\mathbb{Q}) \le \sqrt{\mathcal{KL}(\mathbb{P},\mathbb{Q})}.$$