Some Observations about Total Variations

Sihui Wang

1. The Univariate Case:

We can verify that $TV(\mathcal{N}(0,1), \mathcal{N}(0, \sigma^2))$ is equal to $TV(\mathcal{N}(0,1), \mathcal{N}(0, \frac{1}{\sigma^2}))$:

We take $x = \sigma y$ in the following expression:

$$\int_{-\infty}^{\infty} |\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}| dx$$

Then we have:

$$\int_{-\infty}^{\infty} |\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}| dx = \int_{-\infty}^{\infty} |\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} - \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{\sigma^2 y^2}{2}}| dy$$

Which proves that $TV\big(\mathcal{N}(0,1), \mathcal{N}(0, \sigma^2)\big) = TV(\mathcal{N}(0,1), \mathcal{N}(0, \frac{1}{\sigma^2}))$

2. Multivariate Cases:

For example, in the case of $r = 2$, taking certain orthogonal transformation, we can assume that the total variations have the form of:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{x^2}{2\sigma_1^2} - \frac{y^2}{2\sigma_2^2}} - \frac{1}{2\pi} e^{-\frac{x^2}{2} - \frac{y^2}{2}}| dxdy$$

By the inequality $|A_1A_2 - B_1B_2| = |A_1A_2 - B_1A_2 + B_1A_2 - B_1B_2| \le |A_2||A_1 - B_1| + |B_1||A_2 - B_2|$, we have:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left|\frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{x^2}{2\sigma_1^2} - \frac{y^2}{2\sigma_2^2}} - \frac{1}{2\pi} e^{-\frac{x^2}{2} - \frac{y^2}{2}}\right| dxdy$$

$$\le \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{x^2}{2\sigma_1^2}} \left|\frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{y^2}{2\sigma_2^2}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\right| dxdy$$

$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \left|\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{x^2}{2\sigma_1^2}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\right| dxdy$$

$$= \int_{-\infty}^{\infty} \left|\frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{y^2}{2\sigma_2^2}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\right| dy + \int_{-\infty}^{\infty} \left|\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{x^2}{2\sigma_1^2}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\right| dx$$

This boundary is RATHER LOOSE, especially in high dimensional cases. However, based on our previous observations that we can decompose the problems of maximization of the statistical divergences into the sub-problems in each dimension, it might be suggested that we choose the eigenvalues that, in some sense, deviate the most from 1.

If this assumption is true, then based on what we found in the univariate case, we might choose the eigenvalues according to some evaluation functions, such as $\lambda + \frac{1}{\lambda}$.

If this is true, then it is likely that the maximization of Symmetric KL-Divergence, Hellinger distance and total variation will produce similar results, at least under normal distributions.

We have learned that total variation is both upper bounded and lower bounded by some functions of Hellinger distance, so I think it might be plausible that maximization of Hellinger distance and total variation will be similar. Maybe in the numerical section, we could propose several evaluation functions and test if the evaluation function, $\lambda + \frac{1}{\lambda}$, will produce better outcomes for the maximization of total divergence.

3. High Dimensional Cases:

If the dimension is very high, then $X \sim \mathcal{N}(0, I_n)$ will be concentrated in a thin spherical shell around the sphere of radius $\sqrt{n}$, and:[1]

$$\mathbb{P}\left\{\left|\|X\|_2 - \sqrt{n}\right| \geq t\right\} \leq 2e^{-ct^2}$$

It seems that each component of X, x_i, is close to 0, but it is likely that each x_i will more or less deviate a little from 0, and the total deviation, $\sum_{i=1}^{n} \Delta x_i^2$ is likely to be around n. (It seems to suggest that if we evaluate people by n dimensions and n is very large, then virtually everyone will be outstanding in some dimension, and few people will be on the average in every dimension.) So two different distributions $X \sim \mathcal{N}(0, I_n)$ and $Y \sim \mathcal{N}(0, \Sigma)$ will be likely to concentrate in different spherical surfaces or ellipsoid surfaces. This might suggest that the total variation will be close to 1 as dimension grows.

Reference:

[1] R.Vershynin, High-Dimension Probability, An Introduction with Applications in Data Science, pp.50