Exercise 1

Exercise 1. Consider the two sequences of measures $\mu_n = \delta_{1-2^{-n}}$ and $\nu_n = \frac{1}{n}\delta_{-1} + (1-\frac{1}{n})\delta_1$. From the mere look of it one would guess that both of these sequences approach the measure $\pi = \delta_1$ in some way. The KL divergence $KL(\mu|\nu)$ from μ to ν is defined as ∞ if there is an x with $\mu(x) > 0$ and $\nu(x) = 0$ and

$$KL(\mu|\nu) = \sum_{x \in \mathrm{supp}(\mu)} \mu(x) \log\left(\frac{\mu(x)}{\nu(x)}\right)$$

otherwise.

Let's start with exploring the sequences μ_n and ν_n:

$\mu_n = \delta_{1-2^{-n}}$

$n = 0$: $\quad \mu_0 = \delta_0 = \delta(x)$

$n = 1$: $\quad \mu_1 = \delta_{0.5} = \delta(x - 0.5)$

$n = 2$: $\quad \mu_2 = \delta_{0.75} = \delta(x - 0.75)$

$\vdots$

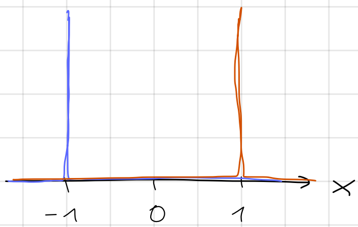
$n = 10$: $\quad \mu_{10} = \delta_{1-2^{-10}} = \delta(x - x_0)$ with $x_0 \approx 0.999$

$\rightarrow$ the x_0 in $\delta(x-x_0)$ converges towards 1

$\nu_n = \frac{1}{n}\delta_{-1} + (1-\frac{1}{n})\delta_1$

$\delta_{-1} = \delta(x+1)$

$\delta_1 = \delta(x-1)$ $\quad$ Sketch:



$\lim_{n\to\infty} \frac{1}{n} = 0$

$\lim_{n\to\infty} (1-\frac{1}{n}) = 1$

$\rightarrow$ first impression: as n increases the second term $(1-\frac{1}{n})\delta_1$ will more and more overpower the first term

There is one important difference between the sequence of μ_n and ν_n:

Regardless of how big n is, the x_0 in $\delta(x-x_0)$ will never reach exactly 1. Therefore, $\mu_n(1) = 0 \quad \forall n \in \mathbb{N}$

In contrast to that, $\nu_n(1) > 0 \quad \forall n \in \mathbb{N}$ because the term δ_1 never vanishes.

Moreover, the following holds:

$$\mathrm{supp}(\mu_n) \cap \mathrm{supp}(\nu_n) = \emptyset \qquad \forall n \in \mathbb{N}$$

and $\left(\bigcup_{n\in\mathbb{N}} \mathrm{supp}(\mu_n)\right) \cap \left(\bigcup_{n\in\mathbb{N}} \mathrm{supp}(\nu_n)\right) = \emptyset$

because $\mathrm{supp}(\mu_n) \subseteq [0,1)$ and $\mathrm{supp}(\nu_n) = \{-1,+1\}$ $\quad \forall n \in \mathbb{N}$

Definition of $KL(\mu|\nu)$:

$$KL(\mu|\nu) = \begin{cases} \infty & \text{if } \exists x: \mu(x) > 0 \text{ and } \nu(x) = 0 \\ \sum_{x \in \mathrm{supp}(\mu)} \mu(x) \log\left(\frac{\mu(x)}{\nu(x)}\right) & \text{otherwise} \end{cases}$$

(i) Calculate $\mathrm{KL}(\pi|\mu_n)$ and $\mathrm{KL}(\pi|\nu_n)$. Do the sequences μ_n and ν_n converge to π w.r.t. the KL-divergence? I.e. do the sequences $\mathrm{KL}(\pi|\mu_n)$ and $\mathrm{KL}(\pi|\nu_n)$ approach 0?

$\pi = \delta_1 = \delta(x-1) \quad \Rightarrow \quad \mathrm{supp}(\pi) = \{+1\}$

$\mathrm{supp}(\mu_n) \subseteq [0,1) \quad \Rightarrow \quad \mathrm{supp}(\pi) \cap \mathrm{supp}(\mu_n) = \emptyset$

$\mathrm{supp}(\nu_n) = \{-1,+1\} \quad \Rightarrow \quad \mathrm{supp}(\pi) \cap \mathrm{supp}(\nu_n) = \{+1\}$

$KL(\pi|\mu_n) = +\infty$ because $\pi(1) > 0$ and $\mu_n(1) = 0$

This holds independent of n, i.e. $\lim_{n\to\infty} KL(\pi|\mu_n) = +\infty$

$$KL(\pi|\nu_n) = \sum_{x \in \mathrm{supp}(\pi)} \pi(x) \log\left(\frac{\pi(x)}{\nu_n(x)}\right)$$

because $\nexists x: \pi(x) > 0$ and $\nu_n(x) = 0$
The only $x \in \mathbb{R}$ with $\pi(x) > 0$ is $x = 1$, but $\nu_n(x) > 0 \;\; \forall n \in \mathbb{N}$

$$= \pi(1) \log\left(\frac{\pi(1)}{\nu_n(1)}\right) = \delta_1(1) \log\left(\frac{\delta_1(1)}{\frac{1}{n}\delta_{-1}(1) + (1-\frac{1}{n})\delta_1(1)}\right)$$

$$= \delta_1(1) \log\left(\frac{\delta_1(1)}{(1-\frac{1}{n})\delta_1(1)}\right) = \delta_1(1) \log\left(\frac{1}{1-1/n}\right) = \delta_1(1) \log\left(\frac{n}{n-1}\right)$$

Note: $\delta_1(1) = 1$

$$\lim_{n\to\infty} KL(\pi|\nu_n) = \lim_{n\to\infty} \delta_1(1) \underbrace{\log\left(\frac{n}{n-1}\right)}_{\to +0} = 0$$

(ii) For each μ_n and ν, find the optimal coupling T between μ_n (resp. ν_n) and π with the procedure from Example 2.29 (by hand). The Wasserstein-2 distance is then given as $\mathcal{W}(T) = \sqrt{\sum_{i,j} t_{ij}|a_i - a_j|^2}$. Do the sequences μ_n and ν_n converge to π in the Wasserstein-2 distance?

a) μ_n and π:

Because the support of μ_n and the support of π only consist of a single element, there is only one possible transport:

$T^* = 1$

Wasserstein-2 distance:

$$W(T^*) = \sqrt{\sum_{i,j} \underbrace{t_{ij}}_{=1} |a_i - a_j|^2} = \sqrt{|1-2^{-n} - 1|^2} = 2^{-n}$$

$$\lim_{n\to\infty} W(T^*) = \lim_{n\to\infty} 2^{-n} = 0$$

$\Rightarrow$ the sequence μ_n converges to π in the Wasserstein-2 distance

b) ν_n and π:

$\text{supp}(\nu_n) = \{-1, +1\}$

To address this problem it helps to interpret the two measures as two discrete random variables (even if formally that is incorrect).

$\nu_n \rightsquigarrow$ X_1 takes on value -1 or $+1$
$\mathbb{P}[X_1 = -1] = \frac{1}{n}$, $\mathbb{P}[X_1 = 1] = 1 - \frac{1}{n}$

$\pi \rightsquigarrow$ X_2 is a constant with $\mathbb{P}[X_2 = 1] = 1$

A coupling between X_1 and X_2 can be described by the vector $T = (t_1 \ t_2)$. T has to fulfill certain conditions to be valid as a coupling:

$$\left.\begin{aligned} &(1)\ t_1 + t_2 = 1 \\ &(2)\ t_1 = 1/n \\ &(3)\ t_2 = 1 - 1/n \end{aligned}\right\}$$

This directly leads to the only possible and therefore optimal coupling:

$$T^* = (\underbrace{\tfrac{1}{n}}_{t_1^*} \quad \underbrace{1-\tfrac{1}{n}}_{t_2^*})$$

Wasserstein-2 distance:

$$W(T^*) = \sqrt{\sum_{i,j} t_{ij} \; |a_i - a_j|^2} =$$

$$= \sqrt{t_1 \, |-1-1|^2 + t_2 \underbrace{|1-1|^2}_{=0}} = \sqrt{4 \frac{1}{n}} = \frac{2}{\sqrt{n}}$$

$$\lim_{n \to \infty} W(T^*) = \lim_{n \to \infty} \frac{2}{\sqrt{n}} = 0$$

$\Rightarrow$ the sequence ν_n converges to π in the Wasserstein-2 distance