EE5143 Problem Set 2: Total Variation Distance and Relative Entropy

Instructor: Dr. Andrew Thangaraj

Jan - May 2024

Total Variation Distance

1. Consider two distributions on $\{1, 2, 3\}$: X or Y. The distribution $X = \{\overset{1/3}{1}, \overset{1/3}{2}, \overset{1/3}{3}\}$, and $Y = \{\overset{1/2}{1}, \overset{1/4}{2}, \overset{1/4}{3}\}$. The total variation distance is given to be $d_{TV} = 1/6$ with $A = \{2, 3\}$ being the subset achieving this maximal value. Find a joint distribution P_{XY} with $d_{TV} = P(X \neq Y)$.

2. Determine the total variation distance between the distributions $P = X$ and $Q = X + c$, where $X \sim Bern(p)$, $p \in (0, 1)$ for the following cases:
   (a) $c \notin \{-1, 0, 1\}$ 1
   (b) $c = 0$ 0
   (c) $c \in \{-1, 1\}$ 1/2

3. Determine an upper bound on the total variation distance between the two distributions P $\sim \mathcal{N}(\mu_1, \sigma^2)$ and Q $\sim \mathcal{N}(\mu_2, \sigma^2)$. Pinsker

Relative Entropy

1. Let the random variable X have five possible outcomes $\{\alpha, \beta, \gamma, \delta, \epsilon\}$. Consider two distributions on this random variable and two functions $L_1(X)$ and $L_2(X)$ for this random variable::

X	$p(x)$	$q(x)$	$L_1(x)$	$L_2(x)$
α	1/2	1/2	1	1
β	1/4	1/8	2	3
γ	1/8	1/8	3	3
δ	1/16	1/8	4	3
ϵ	1/16	1/8	4	3

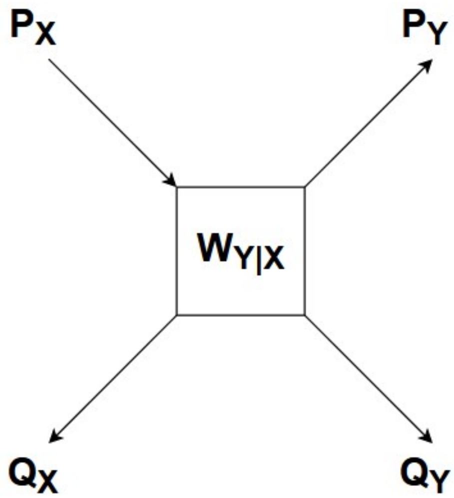
   (a) Calculate $H(p), H(q), D(p||q)$, and $D(q||p)$.
   (b) Find the average of $L_1(X)$ when $X \sim p$ is and similarly for $L_2(X)$ when $X \sim q$. Are they equal to $H(p)$ and $H(q)$, respectively?
   (c) What is the average $L_2(X)$ when $X \sim p$? How much higher is this average when compared to $H(p)$? Verify that the difference is given by $D(p||q)$.

2. Calculate the KL-divergence between the following distributions:
   (a) P $\sim \mathcal{N}(\mu_1, \sigma^2)$ and Q $\sim \mathcal{N}(\mu_2, \sigma^2)$ (Is KL-divergence symmetric?)
   (b) P $\sim Poisson(\lambda_1)$ and Q $\sim Poisson(\lambda_2)$
   (c) P $\sim Geometric(\lambda_1)$ and Q $\sim Geometric(\lambda_2)$

3. Given $x^n = (x_1, \ldots, x_n) \in \mathbb{R}^n$, let $x_{(1)} \leq \ldots \leq x_{(n)}$ denote the ordered entries. Let P, Q be distributions on $\mathbb{R}$ and $P_{X^n} = P^n, Q_{X^n} = Q^n$. Prove or disprove that (where $D(p\|q)$ denotes the *KL-divergence*)

$$D(\text{Binomial}(n, p) \| \text{Binomial}(n, q)) = nD(p\|q)$$

4. *Data processing inequality.* Consider an information processing system with the input being a random variable X and the output being Y. The system is represented as a conditional probability model $W_{Y|X}$. The two distributions from which the input can be generated are P_X, Q_X and the respective output distributions are P_Y, Q_X.



For the setup above, prove the following inequality:

$$\mathrm{D_{KL}}(P_Y, Q_Y) \leq \mathrm{D_{KL}}(P_X, Q_X)$$

Also, derive a sufficient condition on $W_{Y|X}$ for the equality to hold.
Note: This inequality serves as a central inequality in proving Pinkser's inequality for any two general distributions on the same alphabet: https://mathoverflow.net/a/379218