CSC311H1 Assignmnet 1

Xiaoyu Zhou

February 3, 2020

1. (a) From the question, we can know that X,Y are Uni(0, 1). Thus, $f(x) = f(y) = 1$.
For uniform distribution, we can have that:
$E(X^n) = \int_0^1 x^n f(x)dx = \int_0^1 x^n dx = \frac{1}{n+1}|_0^1 = \frac{1}{n+1}$.
So, $E(X) = E(Y) = \frac{1}{2}$, $E(X^2) = E(Y^2) = \frac{1}{3}$.
$E(Z) = E(|X-Y|)^2$
$= E((X-Y))^2$
$= E(X^2 + Y^2 - 2XY)$
$= E(X^2) + E(Y^2) - 2E(XY)$
$= E(X^2) + E(Y^2) - 2E(X)E(Y)$ (since X,Y are independent)
$= \frac{1}{3} + \frac{1}{3} - 2 * \frac{1}{2} * \frac{1}{3}$
$= \frac{1}{6}$
$Var(Z) = E(Z^2) - E(Z)^2$.
$E(Z^2) = E((|X-Y|)^2)^2)$
$= E(|X-Y|)^4$
$= E(X^4 + Y^4 - 4X^3Y + 6X^2Y^2 - 4XY^3)$
$= E(X^4) + E(Y^4) - 4E(X^3Y) + 6E(X^2Y^2) - 4E(XY^3)$
$= E(X^4) + E(Y^4) - 4E(X^3)E(Y) + 6E(X^2)E(Y^2) - 4E(X)E(Y^3)$ (since X,Y are independent)
We proved for uniform distribution, $E(X^n) = \frac{1}{n+1}$. Thus, $E(X^3) = E(Y^3) = \frac{1}{4}$, $E(X^4) = E(Y^4) = \frac{1}{5}$.
So, $E(Z^2) = \frac{1}{5} - 4 * \frac{1}{4} * \frac{1}{2} + 6 * \frac{1}{3} * \frac{1}{3} - 4 * \frac{1}{4} * \frac{1}{2} + \frac{1}{5} = \frac{1}{15}$
So, $Var(Z) = E(Z^2) - E(Z)^2 = \frac{1}{15} - (\frac{1}{6})^2 = \frac{7}{180}$.

(b) Since $x_1 ... x_d$ and $y_1 ... y_d$ are independently and uniformly from [0, 1] and $Z_i = |X_i - Y_i|^2$, from part (a), we can know get that $E(Z_i) = \frac{1}{6}$ and $Var(Z_i) = \frac{7}{180}$.
So $E(R) = E(Z_1 + ... + Z_d)$
$= E(Z_1) + ... + E(Z_d)$
$= \frac{1}{6} + ... + \frac{1}{6}$
$= \frac{d}{6}$
$Var(R) = Var(Z_1 + ... + Z_d)$
$= Var(Z_1) + ... + Var(Z_d)$
$= \frac{7}{180} + ... + \frac{7}{180}$
$= \frac{7d}{180}$

(c) The mean of $\|X - Y\|_2^2 = \frac{d}{6}$

The standard deviation of $\|X - Y\|_2^2 = (\frac{7d}{180})^{\frac{1}{2}}$.
It is obviously that $\frac{1}{6}$ is much bigger than the $(\frac{7}{180})^{\frac{1}{2}}$. So, for different points d, their mean, which represents the distance in 2-dimensions, will have huge differences. However, the standard deviation of different d, which represents the distance in high dimensions, will only change slightly. Thus, in the high dimensions, those points are approximately the same distance.

2. (a) From the properties of probability and because of $x \in \mathcal{X}$, we can know that $0 \leq p(x) \leq 1$. Thus, $\frac{1}{p(x)} \geq 1$.
So $\log_2(\frac{1}{p(x)}) \geq 0$. Also $p(x) \geq 0$.
So $p(x)\log_2(\frac{1}{p(x)}) \geq 0$.
So $\sum_x p(x)\log_2(\frac{1}{p(x)}) \geq 0$.
Thus, the entropy $H(X)$ is non-negative.

(b) $H(X,Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2 p(x,y)$
$= -\sum_{x\in X}\sum_{y\in Y} p(x)p(y)\log_2 p(x)p(y)$ (since X,Y are independent)
$= -\sum_{x\in X}\sum_{y\in Y} p(x)p(y)(\log_2 p(x) + \log_2 p(y))$
$= -\sum_{x\in X}\sum_{y\in Y} p(x)p(y)\log_2 p(x) - \sum_{x\in X}\sum_{y\in Y} p(x)p(y)\log_2 p(y)$
$= -\sum_{x\in X} p(x)\log_2 p(x)\sum_{y\in Y} p(y) - \sum_{y\in Y} p(y)\log_2 p(y)\sum_{x\in X} p(x)$
$= -\sum_{x\in X} p(x)\log_2 p(x) - \sum_{y\in Y} p(y)\log_2 p(y)$ (since $\sum_{x\in X} p(x) = \sum_{y\in Y} p(y) =$ 1)
$= H(X) + H(Y)$

(c) $H(X,Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2 p(x,y)$
$= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2(p(x|y)p(x))$
$= -\sum_{x\in X}\sum_{y\in Y} p(x,y)(\log_2 p(x|y) + \log_2 p(x))$
$= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2 p(x|y) - \sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2 p(x)$
$= -\sum_{x\in X}\sum_{y\in Y} p(x|y)p(y)\log_2 p(x|y) - \sum_{x\in X} p(x)\log_2 p(x)$
$= H(Y|X) + H(X)$

(d) By *Jensen's Inequality*, if $\phi(x)$ is concave, $\phi(E[X]) \geq E[\phi(X)]$. Also we known that $\log_2$ is concave, so, we can have that:
$\mathrm{KL}(p\|q) = \sum_{x\in X} p(x)\log_2\frac{p(x)}{q(x)}$
$= -\sum_{x\in X} p(x)\log_2\frac{q(x)}{p(x)}$
$\geq -\log_2\sum_{x\in X} p(x)\frac{q(x)}{p(x)}$
$= -\log_2\sum_{x\in X} q(x)$
$= -\log_2 1$
$= 0$
So $\mathrm{KL}(p\|q)$ is non-negative.

(e) $\mathrm{KL}(p(x,y)\|p(x)p(y)) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2\frac{p(x,y)}{p(x)p(y)}$
$= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2\frac{p(x)p(y)}{p(x,y)}$
$= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2\frac{p(x)}{p(x,y)} - \sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2 p(y)$

$= \sum_{x\in X} \sum_{y\in Y} p(x,y)\log_2 \frac{p(x,y)}{p(x)} - \sum_{y\in Y} p(y)\log_2 p(y)$
$= \sum_{x\in X} \sum_{y\in Y} p(y|x)p(x)\log_2 p(y|x) - \sum_{y\in Y} p(y)\log_2 p(y)$
$= H(Y) - H(Y|X)$
$= I(Y;X)$