We want to demonstrate equation (5) of [Ferenc Huszr, 2015](https://arxiv.org/abs/1511.05101v1)

$$\begin{aligned} D_{alternative}\left[P \mid\mid Q\right] &= KL\left[P_{x_1} \mid\mid Q_{x_1}\right] + \mathbb{E}_{y\sim P_{x_1}} \mathbb{E}_{z\sim Q_{x_1}} KL\left[P_{x_2|x_1=y} \mid\mid Q_{x_2|x_1=z}\right] \text{ (4)} \\ &= KL\left[P_{x_1} \mid\mid Q_{x_1}\right] + \mathbb{E}_{z\sim Q_{x_1}} KL\left[P_{x_2} \mid\mid Q_{x_2|x_1=z}\right] \text{ (5)} \end{aligned}$$

We study the second right hand side term of the first line (4)

$$\begin{aligned} \mathbb{E}_{y\sim P_{X_1}} \mathbb{E}_{z\sim Q_{X_1}} KL\left[P_{X_2|X_1=y} \mid\mid Q_{X_2|X_1=z}\right] &= \mathbb{E}_{z\sim Q_{X_1}} \sum_{y\sim P_{X_1}} P_{X_1}(X_1=y) \sum_{x_2\sim P_{X_2|X_1=y}(\cdot|X_1=y)} P_{X_2|X_1=y}(X_2=x_2 \mid X_1=y) \log\left[\frac{P_{X_2|X_1=y}(X_2=x_2 \mid X_1=y)}{Q_{X_2=x_2|X_1=z}(X_2=x_2 \mid X_1=z)}\right] \\ &= \mathbb{E}_{z\sim Q_{X_1}} \sum_{(y,x2)\sim P_{X_1,X_2}} P_{X_1,X_2}(X_1=y, X_2=x_2) \log\left[\frac{P_{X_2|X_1=y}(X_2=x_2 \mid X_1=y)}{Q_{X_2|X_1=z}(X_2=x_2 \mid X_1=z)}\right] \\ &= \mathbb{E}_{z\sim Q_{X_1}} \sum_{x_2\sim P_{X_2}} P_{X_2}(X_2=x_2) \sum_{y\sim P_{X_1|X_2=x_2}(\cdot|X_2=x_2)} P_{X_1|X_2=x_2}(X_1=y \mid X_2=x_2) \log\left[\frac{P_{X_2|X_1=y}(X_2=x_2 \mid X_1=y)}{Q_{X_2=x_2|X_1=z}(X_2=x_2 \mid X_1=z)}\right] \end{aligned}$$

Now by Bayes' Rule we know that:

$$P_{X_2|X_1=y}(X_2=x_2 \mid X_1=y) = \frac{P_{X_2}(X_2=x_2)}{P_{X_1}(X_1=y)} P_{X_1|X_2=x_2}(X_1=y \mid X_2=x_2)$$

So we can separate the terms that are only function of x_2 and z as

$$\begin{aligned} = \quad & \mathbb{E}_{z\sim Q_{X_1}} \sum_{x_2\sim P_{X_2}} P_{X_2}(X_2=x_2) \sum_{y\sim P_{X_1|X_2=x_2}(\cdot|X_2=x_2)} P_{X_1|X_2=x_2}(X_1=y \mid X_2=x_2) \left[\log\left(\frac{P_{X_2}(X_2=x_2)}{Q_{X_2|X_1=z}(X_2=x_2 \mid X_1=z)}\right) + \log\left(\frac{P_{X_1|X_2=x_2}(y \mid X_2=x_2)}{P_{X_1}(X_1=y)}\right)\right] \\ = \quad & \mathbb{E}_{z\sim Q_{X_1}} \underbrace{\sum_{x_2\sim P_{X_2}} P_{X_2}(X_2=x_2) \log\left(\frac{P_{X_2}(X_2=x_2)}{Q_{X_2|X_1=z}(X_2=x_2 \mid X_1=z)}\right)}_{KL\left[P_{X_2}||Q_{X_2|X_1=z}\right]} \underbrace{\left(\sum_{y\sim P_{X_1|X_2=x_2}(\cdot|X_2=x_2)} P_{X_1|X_2=x_2}(X_1=y \mid X_2=x_2)\right)}_{=1\text{, we are summing over} P_{X_1|X_2=x_2}} \\ & + \mathbb{E}_{z\sim Q_{X_1}} \sum_{(y,x2)\sim P_{X_1,X_2}} P_{X_1,X_2}(X_1=y, X_2=x_2) \log\left(\frac{P_{X_1|X_2=x_2}(y \mid X_2=x_2)}{P_{X_1}(X_1=y)}\right) \end{aligned}$$

The sum in the last expectation can be rewritten as:

$$\sum_{(y,x2)\sim P_{X_1,X_2}} P_{X_1,X_2}(X_1 = y, X_2 = x_2) \log\left(\frac{P_{X_1|X_2=x_2}(y \mid X_2 = x_2)}{P_{X_1}(X_1 = y)}\right) = \sum_{(y,x2)\sim P_{X_1,X_2}} P_{X_1,X_2}(y, x_2) \log\left(\frac{P_{X_1,X_2}(y, x_2)}{P_{X_1}(y) P_{X_2}(x_2)}\right) = I(X_1; X_2) = 0$$

X_1 and X_2 are independent random variables and their mutual information $I(X_1; X_2)$ is thus equal to zero.
Thus

$$\mathbb{E}_{y\sim P_{X_1}} \mathbb{E}_{z\sim Q_{X_1}} KL\left[P_{X_2|X_1=y} \mid\mid Q_{X_2|X_1=z}\right] = \mathbb{E}_{z\sim Q_{x_1}} KL\left[P_{x_2} \mid\mid Q_{x_2|x_1=z}\right]$$

We have shown equation (5).