In the textbook by Murphy, Jensen's inequality is discussed in the context of Kullback-Leibler (KL) divergence. The Jensen's inequality is used in the context of showing that KL divergence is non-negative. A question that was brought up was with regards to when equality between both sides of the Jensen's inequality can be obtained. To closely emulate the book, we will also use discrete math. However, KL divergence holds for discrete and continuous math.

Before we apply Jensen's inequality, we write out the discrete equation for KL divergence

$$D_{KL}\left(p(x)||q(x)\right) := \sum_{x \in \mathcal{X}} p(x) \log\left(\frac{p(x)}{q(x)}\right) \tag{1}$$

where $0 \cdot \log(\frac{0}{0}) = 0$ by Information Theory convention. Unfortunately, computerized mathematical operations would report that $\frac{0}{0}$ =NaN. To avoid this problem, and to be mathematically correct without the assertion that $0 \cdot \log(\frac{0}{0}) = 0$, we instead restrict the domain of $X \in \mathcal{X}$ to $X \in \mathcal{A}$ where $A = \{x : p(x) > 0\}$. In English, this means that we restrict our computerized math to values x where its probability mass is positive and nonzero. So now we rewrite the definition of discrete KL divergence in terms of the acceptance set, $\mathcal{A}$ as

$$D_{KL}\left(p(x)||q(x)\right) = \sum_{x \in \mathcal{X} \in \mathcal{A}} p(x) \log\left(\frac{p(x)}{q(x)}\right). \tag{2}$$

One thing that we need to understand how to prove that KL divergence is non-negative is that

$$\sum_{x \in \mathcal{A}} p(x)\frac{q(x)}{p(x)} = \sum_{x \in \mathcal{A}} q(x) = 1 \tag{3}$$

Then, by Jensen's inequality,

$$\sum_{x \in \mathcal{A}} p(x) \log\left(\frac{q(x)}{p(x)}\right) \leq \log\left(\sum_{x \in \mathcal{A}} p(x)\frac{q(x)}{p(x)}\right) = \log\left(\sum_{x \in \mathcal{A}} q(x)\right) = \log\left(1\right) = 0. \tag{4}$$

So when do we have the left hand side and right hand side of the Jensen's inequality as equal? The textbook states, without showing, that $p(x) = cq(x)$ for $c \geq 1$ results in equality.

First, we consider the left hand side

$$\sum_{x \in \mathcal{A}} p(x) \log\left(\frac{q(x)}{p(x)}\right) = \sum_{x \in \mathcal{A}} p(x) \log\left(\frac{q(x)}{cq(x)}\right) = \sum_{x \in \mathcal{A}} p(x) \log\left(\frac{1}{c}\right) = \log\left(\frac{1}{c}\right). \tag{5}$$

Let's pause for a moment, **why didn't we end up with the below result**?

$$\sum_{x \in \mathcal{A}} p(x) \log\left(\frac{q(x)}{p(x)}\right) = \sum_{x \in \mathcal{A}} cq(x) \log\left(\frac{q(x)}{cq(x)}\right) = \sum_{x \in \mathcal{A}} cq(x) \log\left(\frac{1}{c}\right) = c\log\left(\frac{1}{c}\right) \tag{6}$$

The answer can be found by revisiting the definition of $\mathcal{A}$.

Now we consider the right hand side of Jensen's inequality

$$\log\left(\sum_{x\in\mathcal{A}} p(x)\frac{q(x)}{p(x)}\right) = \log\left(\sum_{x\in\mathcal{A}} p(x)\frac{q(x)}{cq(x)}\right) = \log\left(\sum_{x\in\mathcal{A}} p(x)\frac{q(x)}{cq(x)}\right) = \log\left(\frac{1}{c.}\right) \tag{7}$$

One question that was not asked, which I now pose to you is why must c be greater than or equal to 1?