Assignment 0: Prerequisites Test

1. Suppose an event consists of tossing a coin three times and each time is independent from all the others. Let $\boldsymbol{X} = (X_1, X_2, X_3)$ be the random variable of this event, where each component X_i ($i = \{1, 2, 3\}$) takes value 1 as head and 0 as tail.

   (1) What is the sample space of $\boldsymbol{X}$? (1 point)

   (2) If $p(X_i = 1) = 0.6\ \forall i$, then what is the probability of the experiment outcome as $\boldsymbol{X} = (0, 0, 0)$; what is the probability if the outcome has two 1's and one 0? (1 point)

2. Fred lives in Los Angeles and commutes 60 miles to work. Whilst at work, he receives an alert from his phone showing that his burglar alarm is ringing. The reliability of his alarm is 97%, which means 97% of burglars trigger the type of alarm. On the other hand, the false alarm rate is 0.2%, which mean the probability of alarm ringing but no burglar is 0.2%. The average burglary rate in that area is about 0.5% in a year, he decided to go back his home and check.

   - What is the probability that there was a burglar in his house? (1 point)
   - While driving back home to investigate, Fred hears on the radio that there was a small earthquake that day near his home. Usually, an earthquake has 1% of chance to set off an alarm, as mentioned in the instruction manual and in the area of Los Angeles, the probability of having an earthquake is 0.1%. Assume that earthquakes are independent of burglars, with the new information, what is the probability that there was a burglar? (1 point)

3. Consider a categorical dsitribution over the random variable X. The sample space of X is $\mathcal{K} = \{1, 2, \ldots, K\}$. For every $k \in \mathcal{K}$, the probability is defined as $P(X = k) = p_k$ with p_k to be estimated. If we want to estimate $\{p_k\}_{k=1}^{K}$ with N observations of X as $\{x^{(1)}, x^{(2)}, \ldots, x^{(N)}\}$, we can use the maximum likelihood estimation by written the likelihood function as

$$f(p) = \prod_{n=1}^{N} P(x^{(n)}), \tag{.1}$$

   or the log-likelihood function as

$$\log f(p) = \sum_{n=1}^{N} \log p(x^{(n)}). \tag{.2}$$

   Please show that the solution of maximizing the (log-)likelihood function is

$$p_k = \frac{\sum_{n=1}^{N} \delta(x^{(n)}, k)}{N} \tag{.3}$$

where $\delta(x^{(n)}, k)$ is the delta function with $\delta(x^{(n)}, k) = 1$ if $x^{(n)} = k$ and 0 otherwise. [**Hint**: $\sum_{k=1}^{K} p_k = 1$] (2 points)

4. Consider a Bernoulli distribution $P(X)$ with $P(X = 1) = 0.7$ and $P(X = 0) = 0.3$

   (1) Compute the expectation of this distribution $\mathbb{E}_P(X)$; (0.5 point)

   (2) What is the entropy of this random variable X under the distribution P, $H_P(X)$? (0.5 point)

   (3) Given another Bernoulli distribution $Q(X)$ with $Q(X = 1) = 0.4$ and $Q(X = 0) = 0.6$, what is the KL divergence of these two distribution, $\mathrm{KL}(P\|Q)$? (1 point)

   (4) Prove that, no matter how $Q(X)$ changes, $\mathrm{KL}(P\|Q) \geq 0$. Furthermore, $\mathrm{KL}(P\|Q) = 0$ if and only if $P = Q$. (2 points)