Question 1: As a consultant to a local plant that manufactures usb drives you need to estimate the failure probability of their usb drives. In order to do that, a series of experiments is performed; in each experiment several usb drives are tried until a good one is found. The number of failures, k, is recorded.

- Given that p is the failure probability, what is the probability of k failures before a good usb drive is found?
  $p^k(1-p)$

- You have performed m independent experiments of this form, recording $k_1, k_2, \ldots, k_m$. Estimate the most likely value of p as a function of m and $k_1, k_2, \ldots, k_m$.
  $P(D|h) = \Pi_{i=1}^{m} p^{k_i}(1-p)$
  $LL(p(D|h)) = \sum_{i=1}^{m} \ln(p^{k_i}) + m \ln(1-p)$
  $mle = \frac{\sum_{i=1}^{m} k_i}{p} - \frac{m}{1-p} = 0$
  $\equiv \sum_{i=1}^{m} k_i = mp + \sum_{i=1}^{m} k_i p$
  $\equiv \sum_{i=1}^{m} k_i = p \left(m + \sum_{i=1}^{m} k_i\right)$
  $mle_p = \frac{\sum_{i=1}^{m} k_i}{m + \sum_{i=1}^{m} k_i}$

Question 2: Let θ be the probability that "Thumbtack 1" (we will abbreviate it as T1) shows heads and 2θ be the probability that "Thumbtack 2" (we will abbreviate it as T2) shows heads. You are given the following Dataset (6 examples).

T1	T2	T1	T1	T2	T2
Tails	Heads	Tails	Tails	Heads	Heads

- What is the likelihood of the data given θ.
  $p(D|\theta) = (1-\theta)^3(2\theta)^3$

- What is the maximum likelihood estimate of θ?
  $LL(p(D|\theta)) = 3\ln(1-\theta) + 3\ln(2) + 3\ln(\theta)$
  $mle = \frac{(-1)(3)}{1-\theta} + \frac{3}{\theta} = 0 \equiv$
  $\frac{3}{\theta} = \frac{3}{1-\theta} \equiv$
  $3 - 3\theta = 3\theta$
  $\equiv 3 = 6\theta$
  $\equiv \theta = .5$

Question 3: Write down the expression for the log-likelihood of the data and then derive the expression for the maximum likelihood estimates for the following two distributions.

- Categorical distributions defined over the support $\{1, \ldots, k\}$ (namely x is a discrete variable and can take k values); find an expression for the MLE of the parameters $\theta_1, \ldots, \theta_k$ s.t. $\theta_i \geq 0$ and $\sum_{i=1}^{k} \theta_i = 1$.
  $p(D|\theta_1, \theta_2, \ldots, \theta_k = \Pi_{j=1}^{m} \theta_1^{[x=1]} \times \theta_2^{[x=2]}, \times \cdots \times \theta_k^{[x=k]})$
  $LL = \sum_{j=1}^{m} \sum_{i=1}^{k} [x_j = i] \ln(\theta_i)$
  $mle_{\theta_i} = \frac{\sum_{j=1}^{m} \sum_{i=1}^{k} [x_j = i]}{\sum_{j=1}^{m}}$

- The probabilistic model is given by: $P_\theta(x) = \theta e^{-\theta x}$ where $\theta \in \mathbb{R}$ is a parameter and $x \in [0, \infty]$.
  $P(D|\theta) = \Pi_{i=1}^{m} \theta e^{-\theta x_i}$
  $LL = m \ln(\theta) - \theta \sum_{i=1}^{m} x_i$
  $\frac{d}{d\theta} = \frac{m}{\theta} - \sum_{i=1}^{m} x_i$
  $mle_\theta = \frac{m}{\sum_{i=1}^{m} x_i}$

Question 4: You are given a coin and a thumbtack and you put Beta priors $\beta(5,5)$ and $\beta(20,20)$ on the coin and thumbtack respectively. You perform the following experiment: toss both the thumbtack and the coin 100 times. To your surprise, you get 20 heads and 80 tails for both the coin and the thumbtack. Are each of the following two statements true or false:

* The MLE estimate of both the coin and the thumbtack is the same.

This statement is true, the reason for this is that we are using a Bernoulli likelihood. Our MLE is trying to maximize our likelihood. The maximum likelihood estimate of our data is the average number of times we flipped heads. Because this average is the same for both experiments, the MLE of both the coin and the thumbtack is the same.

* The MAP estimate of the parameter θ (probability of landing heads) for the coin is greater than the MAP estimate of θ for the thumbtack.

The $MAP = argmax_h(p(D|\theta)p(\theta))$ which we can calculate by taking the derivative of $\theta^{N_0+\gamma_1}(1-\theta)^{N_1+\gamma_2}$, setting it to 0 and solving.
$$\frac{N_0+\gamma_1}{\theta} - \frac{N_1+\gamma_2}{1-\theta}$$

The MAP for the coin is:
$$\frac{20+4}{\theta} = \frac{20+4}{1-\theta}$$
$$24 = 48\theta$$
$$\theta = .5$$

The MAP for the thumbtack is:
$$\frac{39}{\theta} = \frac{39}{1-\theta}$$
$$39 = 78\theta$$
$$\theta = .5$$

This statement is false, the MAP estimate is also the same. This is because our likelihood is the same, and our hyperparameters in the beta priors are the same for both the coin and the thumbtack.

Question 5: Let p be the probability of a coin landing heads up when tossed. You flip the coin 8 time and observe 5 tails and 3 heads. Suppose p can only take two values: 0.3 or 0.6. Find the Maximum likelihood estimate of p over the set of possible values $\{0.3, 0.6\}$.

$$MLE = argmax_{h\in H}(P(D|h))$$
$$MLE(\{0.3, 0.6\}) = argmax((.7^5.3^3),(.4^5.6^3))$$
$$\equiv argmax_p(.004537, .002211)$$
$$\equiv 0.3$$

Question 6: Suppose that you have the following prior on the parameter $p : P(p = 0.3) = 0.2$ and $P(p = .6) = .8$. Given that you flipped the coin and observed 5 tails and 3 heads, find the MAP estimate of p over the set $\{0.3, 0.6\}$, using the prior.

$$MAP = argmax_p(P(D|h)p(h))$$
$$\equiv argmax_p(.004537(.2), .002211(.8))$$
$$\equiv argmax_p(.000907, .00176)$$
$$\equiv 0.6$$