Problem 3 Maximum Likelihood Estimation [40 pts]

This problem explores maximum likelihood estimation (MLE), which is a technique for estimating an unknown parameter of a probability distribution based on observed samples. Suppose we observe the values of n i.i.d.[1] random variables $X_1, \ldots, X_n$ drawn from a single Bernoulli distribution with parameter θ. In other words, for each X_i, we know that:

$$P(X_i = 1) = \theta \qquad \text{and} \qquad P(X_i = 0) = 1 - \theta$$

Our goal is to estimate the value of θ from these observed values of X_1 through X_n.

For any hypothetical value $\hat{\theta}$, we can compute the probability of observing the outcome $X_1, \ldots, X_n$ if the true parameter value θ were equal to $\hat{\theta}$. This probability of the observed data is often called the data likelihood, and the function $L(\hat{\theta}) = P(X_1, . , X_n | \hat{\theta})$ that maps each $\hat{\theta}$ to the corresponding likelihood is called the likelihood function. A natural way to estimate the unknown parameter θ is to choose the $\hat{\theta}$ that maximizes the likelihood function. Formally,

$$\hat{\theta}^{MLE} = argmaxL(\hat{\theta})$$

Often it is more convenient to work with the log likelihood function $l'(\hat{\theta}) = logL(\hat{\theta})$. Since the log function is increasing, we also have:

$$\hat{\theta}^{MLE} = argmaxl(\hat{\theta})$$

3.1 **[8 Points]** Write a formula for the log likelihood function, $l(\hat{\theta})$. Your function should depend on the random variables $X_1, \ldots, X_n$, the hypothetical parameter $\hat{\theta}$ and should be simplified as far as possible (i.e., don't just write the definition of the log likelihood function). Does the log likelihood function depend on the order of the random variables?

$$L(\hat{\theta}) = (P(X_1,\ldots,X_n | \hat{\theta})$$

Bernoulli:

$$L(\hat{\theta}) = \prod_{i=1}^{n} \hat{\theta}^{x_i}(1-\hat{\theta})^{(1-x_i)}$$

Log likelihood:

$$\log\left(\prod_{i=1}^{n} \hat{\theta}^{x_i}(1-\hat{\theta})^{(1-x_i)}\right)$$

$$= \sum_{i=1}^{n} \log\left(\hat{\theta}^{x_i} + (1-\hat{\theta})^{(1-x_i)}\right) = \sum_{i=1}^{n} (x_i \log(\hat{\theta}) + (1-x_i)\log(1-\hat{\theta}))$$

Since the variables are added together, the order of random variables will not matter.

3.2 **[8 Points]** Consider the following sequence of 10 samples: $X = (0,1,0,1,1,0,0,1,1,1)$.

Compute the maximum likelihood estimate for the 10 samples. Show all of your work (hint: recall that if x^* maximizes $f(x)$, then $f'(x^*) = 0$).

The likelihood is

$$L(\hat{\theta}) = \prod_{i=1}^{n} \hat{\theta}^{x_1}(1-\hat{\theta})^{(1-x_1)}$$

First count the occurrences of 0 and 1 in X, which has n = 10,

$L(\hat{\theta}) = (P(X_1,\ldots,X_n \,|\, \hat{\theta})$

$= P(x=0)P(X=1)\ldots P(x=1)$

$= P(x=0)^4 P(x=1)^6$

$= (1-\hat{\theta})^4 \hat{\theta}^6$

Using the log likelihood function:

$$\iota(\hat{\theta}) = \sum_{i=1}^{n} log(x_i|\theta)$$

$$\sum_{i=1}^{n} x_i \log(\hat{\theta}) + (1-x_i)\log(1-\hat{\theta})$$

$$6\log(\hat{\theta}) + 4\log(1-\hat{\theta}) + C$$

Where constant C does not depend on $\hat{\theta}$

Let the derivative of $l(\hat{\theta})$ with respect to $\hat{\theta}$ be zero:

$$\frac{dl}{d\hat{\theta}} = \frac{6}{\hat{\theta}} - \frac{4}{1-\hat{\theta}} = 0$$

By solving for , $\hat{\theta}$ we get the MLE, which is $\hat{\theta}$ = 2.5

3.3 **[8 Points]** Now we will consider a related distribution. Suppose we observe the values of m iid random variables $Y_1,.\ ,Y_m$ drawn from a single Binomial distribution $B(n,\theta)$. A Binomial
distribution models the number of 1's from a sequence of n independent Bernoulli variables with parameter. In other words,

$$P(Y_i = k) = \binom{n}{k}\theta^k(1-\theta)^{n-k} = \frac{n!}{k!(n-k)!}\cdot\theta^k(1-\theta)^{n-k}$$

Write a formula for the log likelihood function, $l(\hat{\theta})$ Your function should depend on the random variables $Y_1,\ldots,Y_m$ and the hypothetical parameter $\hat{\theta}$.

$l(\hat{\theta}) = \log P(Y_i = k)$

$$= \log(\left(\frac{n!}{Y_m!(n-Y_m)!}\right)\theta^{Y_m}(1-\theta)^{n-Y_m})$$

$$= \log\left(\frac{n!}{Y_m!(n-Y_m)!}\right) + Y_m log(\theta) + (n-Y_m)\log(1-\theta)$$

3.4 **[8 Points]** Consider two Binomial random variables Y_1 and Y_2 with the same parameters, $n = 5$ and θ. The Bernoulli variables for Y_1 and Y_2 resulted in (0,1,0,1,1) and (0,0,1,1,1), respectively.

Therefore, $Y_1 = 3$ and $Y_2 = 3$. Compute the maximum likelihood estimate for the 2 samples. Show your work.

For Y_1:

$$= \log\left(\frac{n!}{Y_m!(n-Y_m)!}\right) + Y_m log(\theta) + (n-Y_m)\log(1-\theta)$$

$$= \log\left(\frac{5*4*3*2*1}{3!(5-3)!}\right) + 3\log(\theta) + (5-3)\log(1-\theta)$$

Let the derivative of $l(\theta)$ with respect to θ be zero:

$$= \frac{dl}{d\theta} = \frac{3}{\theta} - \frac{2}{1-\theta} = 0$$

By solving for θ, we get the MLE, which is θ

= 2.5

For Y_2:

$$= \log\left(\frac{n!}{Y_m!(n-Y_m)!}\right) + Y_m log(\theta) +$$
$$(n - Y_m)\log(1-\theta)$$

$$=\log\left(\frac{5*4*3*2*1}{3!(5-3)!}\right) + 3\log(\theta) +$$
$$(5-3)\log(1-\theta)$$

Let the derivative of $l(\theta)$ with respect to θ be zero:

$$\frac{dl}{d\theta} = \frac{3}{\theta} + \frac{2}{1-\theta} = 0$$

By solving for θ, we get the MLE, which is θ

= 2.5