DS 122 Homework 4

Xiang Fu
xfu@bu.edu
Boston University Faculty of Computing & Data Sciences

Contents

1 Q1 Problem A

1.1 Q1.1 Part 1

Differentiate the expression $y(x) = \log(x^3) + \log(4x) - \log(3x + 2)$

1.2 Answer

Given:

$$y(x) = \log(x^3) + \log(4x) - \log(3x + 2)$$

For $\log(x^3)$

Let $u = x^3$. Then:

$$u' = 3x^2$$

The derivative of $\log(x^3)$ is:

$$\frac{3x^2}{x^3} = \frac{3}{x}$$

For $\log(4x)$

Let $u' = 4x$. Then:

$$u' = 4$$

Thus, the derivative of $\log(4x)$ is:

$$\frac{4}{4x} = \frac{1}{x}$$

For $\log(3x + 2)$

Let $u = 3x + 2$. Then:

$$u' = 3$$

Thus, the derivative of $\log(3x + 2)$ is:

$$\frac{3}{3x + 2}$$

Combining the results:

The overall derivative, $y'(x)$, is the sum of the derivatives of each item:

$$y'(x) = \frac{3}{x} + \frac{1}{x} - \frac{3}{3x + 2}$$

1.3 Q1.2 Part 2

Given the functions $u(x) = e^x$ and $v(x) = \log(x + 1)$, determine the derivative of $y(x) = u(x)v(x)$

1.4 Answer

To find the derivative of $y(x) = u(x)v(x)$, we'll need to make use of the product rule for differentiation. The product rule states:

If $y(x) = u(x)v(x)$

$$y'(x) = u'(x)v(x) + u(x)v'(x)$$

Given:

$$u(x) = e^x$$

$$v(x) = \log(x + 1)$$

Differentiating $u(x)$:

$$u'(x) = \frac{d}{dx}e^x = e^x$$

Differentiating $v(x)$:

Using the chain rule:

$$v'(x) = \frac{d}{dx}\log(u) = \frac{u'}{u}$$

Let $u = x + 1$. Then:

$$u' = 1$$

Thus, the derivative of $\log(x + 1)$ is:

$$v'(x) = \frac{1}{x + 1}$$

Applying the Product Rule

$$y'(x) = u'(x)v(x) + u(x)v'(x)$$

$$y'(x) = e^x \log(x + 1) + e^x \frac{1}{x + 1}$$

1.5 Q1.3 Part 3

Using the logarithmic properties to simplify, find the derivative of $y(x) = \log\left(x^4 \cdot (x + 1)^3\right)$

1.6 Answer

Given: $y(x) = \log\left(x^4 \cdot (x+1)^3\right)$

Using the Properties of Logarithms:

Using the logarithmic property $\log(a \cdot b) = \log(a) + \log(b)$

$$y(x) = \log(x^4) + \log\left((x+1)^3\right)$$

Using the property $\log\left(a^b\right) = b \cdot \log(a)$:

$$y(x) = 4\log(x) + 3\log(x+1)$$

Differentiating the Expression:

Differentiating $4\log(x)$

$$\frac{d}{dx}4\log(x) = 4 \cdot \frac{1}{x} = \frac{4}{x}$$

Differentiating $3\log(x+1)$:

Using the chain rule:

$$\frac{d}{dx}3\log(x+1) = 3 \cdot \frac{1}{x+1} = \frac{3}{x+1}$$

Combining the Results:

The overall derivative, $y'(x)$, is the sum of the derivatives of each term:

$$y'(x) = \frac{4}{x} + \frac{3}{x+1}$$

1.7 Q1.4 Part 4

For the function $y(x) = x^2\log(x^3+1)$, apply the chain rule and logarithmic differentitation to determine $\frac{dy}{dx}$.

1.8 Answer

Given the function:

$$y(x) = x^2\log(x^3+1)$$

We can use both the product rule and the chain rule combined with logarithmic differentiation

Let's call:

$$u(x) = x^2$$

$v(x) = \log(x^3 + 1)$

Using the product rule:

$$\frac{d}{dx} = u'(x)v(x) + u(x)v'(x)$$

Differentiating $u(x)$:

$$u'(x) = \frac{d}{dx}x^2 = 2x$$

Differentiating $v(x)$:

$$v'(x) = \frac{d}{dx}\log(x^3 + 1)$$

$$v'(x) = \frac{1}{x^3 + 1} * \frac{d}{dx}(x^3 + 1)$$

$$v'(x) = \frac{1}{x^3 + 1} * 3x^2$$

$$v'(x) = \frac{3x^2}{x^3 + 1}$$

Applying the product rule:

$$\frac{dy}{dx} = 2x\log(x^3 + 1) + x^2\frac{3x^2}{x^3 + 1}$$

1.9 Q1.5 Part 5

If $y(x) = x^x$, take the natural logarithm of both sides and then differentiate implicitly to directly determine $\frac{dy}{dx}$.

1.10 Answer

Given $y(x) = x^x$

Taking the Natural Logarithm of Both Sides

$$\ln(y) = \ln(x^x)$$

Using the property of logarithms $\ln\left(a^b\right) = b\ln(a)$:

$$\ln(y) = x\ln(x)$$

Differentiating the left side with respect to x using the chain rule:

$$\frac{1}{y} * \frac{dy}{dx}$$

Differentiating the right side:

The term $x\ln(x)$ is a product of two functions, so we use the product rule.

Let $u(x)$ and $v(x) = \ln(x)$

$$u'(x) = 1$$

$$v'(x) = \frac{1}{x}$$

Using the product rule:

$$\frac{d}{dx}[x\ln(x)] = x * \frac{1}{x} + \ln(x) * 1 = 1 + \ln(x)$$

Equating both derivatives:

$$\frac{1}{y} * \frac{dy}{dx} = 1 + \ln(x)$$

Solving for $\frac{dy}{dx}$

$$\frac{dy}{dx} = y(1 + \ln(x))$$

We know from the original equation that $y = x^x$. Plugging this in:

$$\frac{dy}{dx} = x^x(1 + \ln(x))$$

2 Q2 Problem B

Consider the histograms below which depict the sampling distributions of four different estimators for a population parameter. The true population parameter value is 7.

Based on the histograms:

2.1 Q2.1

Find the estimators with the lowest bias.

- Both 1 and 2 have lowest bias
- Choice 2 of 4: Both 1 and 3 have lowest bias
- Choice 3 of 4: Only 3 has lowest bias
- Choice 4 of 4: Both 2 and 4 have lowest bias

2.2 Q2.2

Find the estimator(s) with the highest variance.

- Choice 1 of 4: Both 1 and 4 have highest variance
- Choice 2 of 4: Both 2 and 4 have highest variance
- Choice 3 of 4: Only 1 has highest variance
- Choice 4 of 4: Only 2 has highest variance

2.3 Q2.3

Considering the fact that Mean Squared Error (MSE) is a combination of variance and squared bias, which estimator(s) likely has the highest MSE?

- Choice 1 of 4: Estimator 1
- Choice 2 of 4: Estimator 2
- Choice 3 of 4: Estimator 1 and 2
- Choice 4 of 4: Estimator 4

3 Q3 Problem C

Suppose $\left\{y^{(1)}, y^{(2)}, ..., y^{(n)}\right\}$ is an i.i.d. sample from the continuous uniform distribution with parameters 0 and θ.

Let $\hat{\theta} = \frac{2}{n} * \left(y^{(1)}, y^{(2)}, ..., y^{(n)}\right)$be an estimator of θ

Hint: You can use the following facts about the continuous uniform distribution with parameters a and b:

- its means is equal to $\frac{a+b}{2}$
- its variance is equal to $\frac{(b-a)^2}{12}$

3.1 Q3.1 Part 1

Determine the bias of $\widehat{\theta}$.

3.2 Answer

Given:

$\{y^{(1)}, y^{(2)}, ..., y^{(n)}\}$ is an i.i.d. sample from the continuous uniform distribution with parameters 0 and θ.

Estimator:

$$\hat{\theta} = \frac{2}{n} \sum_{i=1}^{n} y^{(i)}$$

We want to determine the bias of $\hat{\theta}$.

Determine $E(\hat{\theta})$

Given $y^{(i)}$ is from a uniform distribution with parameters 0 and θ, the mean $E(y^{(i)})$ is:

$$E(y^{(i)}) = \frac{0 + \theta}{2} = \frac{\theta}{2}$$

Now, the expected value of the sum of the $y((i))$ values is:

$$E\left(\sum_{i=1}^{n} y((i))\right) = nE(y^{(i)}) = n * \frac{\theta}{2} = \frac{n\theta}{2}$$

Given the estimator:

$$\hat{\theta} = \frac{2}{n} \sum_{i=1}^{n} y^{(i)}$$

The expected value $E(\hat{\theta})$ is:

$$E(\hat{\theta}) = \frac{2}{n} * E\left(\sum_{i=1}^{n} y((i))\right)$$

$$E(\hat{\theta}) = \frac{2}{n} * \frac{n\theta}{2} = \theta$$

Compute the Bias

Now that we have $E(\hat{\theta}) = \theta$:

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta = \theta - \theta = 0$$

Therefore, the estimator $\hat{\theta}$ is unbiased for θ.

3.3 Q3.2 Part 2

Determine the mean of $\widehat{\hat{\theta}}$.

3.4 Answer

To determine the mean of θ, we need to find its expected value, $E\left(\hat{\theta}\right)$.

Recall from the previous part, the estimator is:

$$\hat{\theta} = \frac{2}{n}\sum()$$

Given that $y^{(i)}$ is from a uniform distribution with parameters 0 and θ, the mean $E(y^{(i)}$ is:

$$E\left(y^{(i)}\right) = \frac{0}{\theta}, 2) = \frac{\theta}{2}$$

Now, the expected value of the sum of the $y^{(i)}$ values is:

$$E(\sum_{i=1}^{n} y^{(i)} nE\left(y^{(i)}\right) = n \cdot \frac{\theta}{2} = \frac{(n \cdot \theta)}{2}$$

Given the estimator:

$$\hat{\theta} = \frac{2}{n} \sum_{(i=1\}})^{n} y^{(i)}$$

The expected value $E(\hat{\theta}$ is:

$$E\left(\hat{\theta}\right) = \frac{2}{n} \cdot E\left(\sum_{i=1}^{n} y^{(i)}\right)$$

$$E\left(\hat{\theta}\right) = \frac{2}{n} \cdot \frac{n\theta}{\theta} = \theta$$

Thus, the mean of $\hat{\theta}$ is θ.

3.5 Q3.3 Part 3

Determine the variance of $\hat{\theta}$.

3.6 Answer

To determine the variance of $\hat{\theta}$, we will start by calculating the variance of a single observation $y^{(i)}$ and then extend it to be the variance of the estimator.

Given:

$y^{(i)}$ are i.i.d samples from the continuous uniform distribution with parameters 0 and θ

The variance $\mathrm{Var}\left(y^{(i)}\right)$ is:

$$\mathrm{Var}\left(y^{(i)}\right) = \frac{(\theta - 0)^2}{12} = \frac{\theta^2}{12}$$

Since the observations are independent, the variance of the sum is the sum of the variances:

$$\mathrm{Var}\left(\sum_{i=1}^{n} y((i))\right) = n \cdot \frac{\theta^2}{12} = \frac{n\theta^2}{12}$$

Now, let's consider the estimator:

$$\hat{\theta} = \frac{2}{n}\sum_{i=1}^{n} y((i))$$

Using the properties of variance:

$$\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X)$$

where c is a constant.

The variance $\mathrm{Var}\left(\hat{\theta}\right)$ is:

$$\mathrm{Var}\left(\hat{\theta}\right) = \left(\frac{2}{n}\right)^2 \mathrm{Var}(X)$$

$$\mathrm{Var}\left(\hat{\theta}\right) = \frac{4}{n^2} \cdot \frac{(n\theta)^2}{12}$$

$$\mathrm{Var}\left(\hat{\theta}\right) = \frac{4\theta^2}{12n}$$

$$\mathrm{Var}\left(\hat{\theta}\right) = \frac{\theta^2}{3n}$$

Thus, the varaince of $\hat{\theta}$ is $\frac{\theta^2}{3n}$.

4 Q4 Problem D

Suppose y is a discrete random variable with the following probability mass function (pmf):

y	**0**	**1**	**2**	**3**
p(y;θ)	$\frac{2}{3}\theta$	$\frac{1}{3}\theta$	$\frac{2}{3}(1-\theta)$	$\frac{1}{3}(1-\theta)$

Where $0 \leq \theta \leq 1$ is a parameter. Given the following 10 independent observations from this distribution:

$$Y = \{1, 1, 0, 2, 2, 1, 3, 2, 0, 3\}$$

4.1 Q4.1

Compute the likelihood function $p(Y;\theta)$.

4.2 Answer

We can multiply together the probabilities assigned by the pmf to each observed va;ie. according to the principle of likelihood.

Given observations $Y = \{1,1,0,2,2,1,3,2,0,3\}$, let's define: n_0, n_1, n_2, n_3 as the number of times 0, 1, 2, 3 appear in Y respectively.

The likelihood function, $L(\theta;Y)$, is computed as:

$$L(\theta;Y) = p(Y;\theta) = p(y_1;\theta)p(y_2;\theta)...p(y_{10};\theta)$$

Since the observations are independent, we multiply the probabilities. Given the provided pmf:

$$p(y;\theta) =$$

We can compute the likelihood function as:

$$L(\theta;Y) = \left(\frac{2}{3}\theta\right)^{n_0}\left(\frac{1}{3}\theta\right)^{n_1}\left(\frac{2}{3}\theta(1-\theta)\right)^{n_2}\left(\frac{1}{3}\theta(1-\theta)\right)^{n_3}$$

Where:

$$n_0 = \text{number of 0s in Y}$$

$$n_1 = \text{number of 1s in Y}$$

$$n_2 = \text{number of 2s in Y}$$

$$n_3 = \text{number of 3s in Y}$$

Given the observations $Y = \{1,1,0,2,2,1,3,2,0,3\}$, we can substitute:

$$n_0 = 2$$

$$n_1 = 3$$

$$n_2 = 3$$

$$n_3 = 2$$

Now, substituting these counts into the likelihood function we can get:

$$L(\theta;Y) = \left(\frac{2}{3}\theta\right)^{2}\left(\frac{1}{3}\theta\right)^{3}\left(\frac{2}{3}\theta(1-\theta)\right)^{3}\left(\frac{1}{3}\theta(1-\theta)\right)^{2}$$

4.3 Q4.2

Compute the log-likelihood function.

4.4 Answer

The log-likelihood function is the natural logarithm of the likelihood function. Given our likelihood function from Q4.1:

$$L(\theta;Y) = \left(\frac{2}{3}\theta\right)^2 \left(\frac{1}{3}\theta\right)^3 \left(\frac{2}{3}\theta(1-\theta)\right)^3 \left(\frac{1}{3}\theta(1-\theta)\right)^2$$

We can take the natural logarithm of both sides:

$$l(\theta;Y) = \ln[L(\theta;Y)]$$

Applying the properties of logarithms:

$$l(\theta;Y) = 2\ln\left(\frac{2}{3}\theta\right) + 3\ln\left(\frac{1}{3}\theta\right) + 3\ln\left(\frac{2}{3}(1-\theta)\right) = 2\ln\left(\frac{1}{3}(1-\theta)\right)$$

Break it down further:

$$l(\theta;Y) = 2\left[\ln\left(\frac{2}{3}\right) + \ln(\theta)\right] + 3\left[\ln\left(\frac{1}{3}\right) + \ln(\theta)\right] + 3\left[\ln\left(\frac{2}{3}\right) + \ln(1-\theta)\right] + 2\left[\ln\left(\frac{1}{3}\right) + \ln(1-\theta)\right]$$

$$l(\theta;Y) = 2\ln\left(\frac{2}{3}\right) + 2\ln(\theta) + 3\ln\left(\frac{1}{3}\right) + 3\ln(\theta) + 3\ln\left(\frac{2}{3}\right) + 3\ln(1-\theta) + 2\ln\left(\frac{1}{3}\right) + 2\ln(1-\theta)$$

4.5 Q4.3

Find the potential Maximum Likelihood Estimator (MLE) for θ.

4.6 Answer

We will need to maximize the log-likelihood function iwth respect to θ to find the Maximum Likelihood Estimator (MLE) for θ.

From Q4.2, our log-likelihood function is:

$$l(\theta;Y) = 2\ln\left(\frac{2}{3}\right) + 2\ln(\theta) + 3\ln\left(\frac{1}{3}\right) + 3\ln(\theta) + 3\ln\left(\frac{2}{3}\right) + 3\ln(1-\theta) + 2\ln\left(\frac{1}{3}\right) + 2\ln(1-\theta)$$

To find the MLE, we will differentiate $l(\theta;Y)$ with respect to θ and set the result eqaul to zero:

$$\frac{dl(\theta;Y)}{d\theta} = 0$$

Differentiating:

1. Derivative of $\ln(\theta)$ is $\frac{1}{\theta}$
2. Derivative of $\ln(1-\theta)$ is $-\frac{1}{1-\theta}$
3. Constants and terms that don't include θ will have derivatives of zero

$$\therefore \frac{dl(\theta; Y)}{d\theta} = \frac{2}{\theta} + \frac{3}{\theta} - \frac{3}{1-\theta} - \frac{2}{1-\theta}$$

Setting all this to zero:

$$\frac{5}{\theta} - \frac{5}{1-\theta} = 0$$

$$5(1-\theta) - 5\theta = 0$$

$$\theta = \frac{1}{2}$$

So, the Maximum Likelihood Estimator (MLE) for θ is θ = 0.5.

4.7 Q4.4

Explain in words how to confirm that the computed point for part c is indeed a maximum. Note: no computations are required for this part.

4.8 Answer

1. Second Derivative Test
   - By computing the second derivative of the log-likelihood function with respect to θ and evaluating it at the computed point, we can determine the nature of the point:
     - If the second derivative is negative, it indicates that the function is concave down at that point, which means our computed point is a local maximum.
     - If the second derivative is positive, it indicates that the function is concave up, which would mean our computed point is a local minimum.
     - If the second derivative is zero, the test is inconclusive.
2. Log-Likelihood Surface
   - If we were to plot the log-likelihood function, the computed point would be a peak on this surface if it is a maximum.
   - We cam look for the highest point in this plot.
3. Sign Change of First Derivative
   - If the first derivative changes sign (from positive to negative) as θ crosses the computed value, this indicates that the function increases up to our point and then decreases after, suggesting it's a maximum.

5 Q5 Problem E

In a clinical trial for a new drug, 30 patients were treated. Out of them, 24 patients showed improvement in their conditions.

Assuming the patients' responses follow a binomial distribution, find the Maximum Likelihood Estimator (MLE) for the probability $\theta = p$ that a randomly chosen patient shows improvement after being treated with the drug.

Hint: The likelihood function for binomial distribution is given by:

$$p(x; \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

where:

- n is the number of trials
- x is the number of successes
- θ is the parameter, that represents the probability of success on a single trial.

5.1 Answer

The goal here is to find the value of θ (the probability of a patient showing improvement) that maximizes the likelihood of observing the given data, which follows a binomial distribution.

Given:
- $n = 30$ (total number of patients)
- $x = 24$ (number of patients who showed improvement)

The likelihood function for the binomial distribution is:

$$L(\theta) = p(x; \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

The Maximum Likelihood Estimator (MLE) of θ is the value of θ that maximizes this likelihood function.

To find the MLE, we'll differentiate the log likelihood with respect to θ and set it to zero. This is because the log function is a strictly increasing function, so where the log likelihood achieves its maximum, the likelihood also achieves its maximum.

Taking the natural logarithm of both sides:

$$\ln(L(\theta)) = \ln\left(\binom{n}{x}\theta^x(1-\theta)^{n-x}\right)$$

$$\ln(L(\theta)) = \ln\left(\binom{n}{x}\right) + x\ln(\theta) + (n-x)\ln(1-\theta)$$

Now, differentiate with respect to θ:

$$\frac{d}{d\theta}\ln(L(\theta)) = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

To find the MLE, set the derivative to zero and solve for θ:

$$\frac{x}{\theta} - \frac{n-x}{1-\theta} = 0$$

Multiplying through by $\theta(1-\theta)$ and rearranging:

$$x(1-\theta) = (n-x)\theta$$

$$x - x\theta = n\theta - x\theta$$

$$x = n\theta$$

Thus, the MLE for θ is:

$$\hat{\theta} = \frac{x}{n}$$

Given $x = 24$ and $n = 30$:

$$\hat{\theta} = \frac{24}{30} = \frac{4}{5} = 0.8$$

Therefore, the Maximum Likelihood Estimator (MLE) for the probability θ that a randomly chosen patient shows improvement after being treated with the drug is 0.8 or 80%.

6 Q6 Computational

See the Jupyter Notebook for details.