STA4020 Problem Set 3

Maximum Likelihood Estimation

Fall 2024

General Instructions:

This assignment contains 6 questions. Please ensure that all work is clearly shown and that final answers are boxed or highlighted. This assignment is due by **5 PM on October 18, 2024**. Late submissions will incur a penalty unless prior arrangements have been made.

Ethical and Responsible Usage of AI

While you are encouraged to explore modern tools such as AI and machine learning platforms to enhance your learning, it is essential to use these resources responsibly. All solutions must be your original work. You may use AI to assist with understanding concepts or to guide you, but copying AI-generated answers directly without comprehension is not permitted. Always strive to understand the underlying methods and derivations behind your solutions.

Additional Reminders:

1. Independent Work: Each student must submit their own work. Direct copying from others is considered academic misconduct.
2. Formatting: Please submit your answers as a typed document (preferably in PDF format) or as clearly scanned handwritten work. Both electronic submissions and hard copies are acceptable.
3. References: If you use any external sources, including textbooks, articles, or online resources, make sure to cite them appropriately.

If you have any questions or need clarification on any of the questions, feel free to reach out before the due date.

Good luck!

Problem 1 (20 Points)

Consider a random variable $X \sim \text{Poisson}(\lambda)$, where λ is the rate parameter. The probability mass function (PMF) is given by:

$$P(X = x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots.$$

- **(6 Points)** Derive the log-likelihood function $\log L(\lambda; x)$ for the Poisson distribution.
- **(6 Points)** Compute the score function (the derivative of the log-likelihood function) with respect to λ.
- **(8 Points)** Show that the expectation of the score function is zero.

Problem 2 (15 Points)

A biased coin is tossed n times and lands heads k times. Assume the probability of getting heads is θ. Derive the maximum likelihood estimator for the probability of heads θ.

Problem 3 (15 Points)

You want to estimate the number of registered participants in a CUHK, Shenzhen seminar. You know that participants are numbered from 1 to n, where n is the total number of participants. You randomly select three participants from the seminar and ask for their registration numbers, which turn out to be 5, 19, and 31. Find the maximum likelihood estimate for n. (The registration numbers are drawn from a discrete uniform distribution.)

Hint: Consider the behavior of the likelihood function when n is smaller than 31 and when n is greater than or equal to 31.

Problem 4 (15 Points)

Suppose we have data $x_1, x_2, \ldots, x_n$ that are independently drawn from a uniform distribution $U(a, b)$, where a and b are unknown parameters.

- **(10 Points)** Derive the maximum likelihood estimators (MLE) for a and b.
- **(5 Points)** Provide an intuitive explanation about your MLE for a and b.

Problem 5 (15 Points)

Let $X_1, X_2, \ldots, X_n$ be an independent and identically distributed (i.i.d.) sample from an exponential distribution with parameter λ. The probability density function of the exponential distribution is:

$$f(x|\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

- **(5 Points)** Derive the MLE of λ.
- **(10 Points)** Suppose you are interested in the quantity $\theta = \log(\lambda)$. Use the Delta Method to find the asymptotic distribution of $\hat{\theta} = \log(\hat{\lambda})$, where $\hat{\lambda}$ is the MLE of λ that you derived in part 1.

Problem 6 (20 Points)

This is a coding exercise. You are given a dataset `data.txt` containing n independent observations from an exponential distribution with an unknown parameter λ. The probability density function for the exponential distribution is:

$$f(x|\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

Tasks:

1. Load the dataset from the provided 'data.txt' file and visualize the data using a histogram. (Your submission should contain such a historgram).
2. Estimate the MLE for λ using the following steps:
   - Derive the log-likelihood function for the exponential distribution.
   - Write a Python function to compute the negative log-likelihood.
   - Use Python's `scipy.optimize.minimize` function to find the MLE for λ.
3. Visualize the fit by plotting the histogram of the data along with the estimated probability density function (PDF) for the exponential distribution using your estimated $\hat{\lambda}$.
4. Submit your code used for the analysis.

Problem 1.

$$L(\lambda, x) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{(x_i)!}$$

$$= \frac{\lambda^{\sum^n x_i} \cdot e^{-n\lambda}}{\prod_{i=1}^{n} (x_i)!}$$

$$\log L(\lambda, x) = \log\left[\lambda^{\sum^n x_i} \cdot e^{-n\lambda}\right] - \log\left[\prod_{i=1}^{n} (x_i)!\right]$$

$$= \sum^n_i x_i \log(\lambda) - n\lambda - \log\left(\prod_{i=1}^{n} x_i!\right)$$

Score function:
$$l(\lambda, x) = \log L(\lambda, x)$$

$$\frac{\partial l(\lambda, x)}{\partial \lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n$$

$$\mathbb{E}\left[\frac{\partial l(\lambda, x)}{\partial \lambda}\right] = \mathbb{E}\left[\frac{\sum_{i=1}^{n} x_i}{\lambda} - n\right] = \frac{\sum_{i=1}^{n} \mathbb{E}(x_i)}{\lambda} - n$$

$$= \frac{n(\lambda)}{\lambda} - n$$

$$= n - n$$

$$= 0 \quad \text{(shown)}$$

Problem 2 :

$$P(X = x ; \theta) = \theta \cdot (1-\theta)$$

where x denotes the outcome of getting heads

$$L(\theta, x) = \binom{n}{k} \theta^{k} \cdot (1-\theta)^{n-k}$$

$$l(\theta, x) = \log\binom{n}{k} + k \log \theta + (n-k) \log(1-\theta)$$

$$\frac{\partial l}{\partial \theta} = 0 + \frac{k}{\theta} + \frac{-(n-k)}{1-\theta} = 0$$

$$\therefore \quad \frac{k}{\theta} = \frac{n-k}{1-\theta}$$

$$k - \theta k = \theta n - \theta k$$

$$k = \theta n$$

$$\theta = \frac{k}{n}$$

$\therefore$ the MLE $\hat{\theta} = \frac{k}{n}$

Problem 3 :

let sample size be m, $m = 3$

$$L(n) = \prod_{i=1}^{m} \left(\frac{1}{n}\right) = \frac{1}{n^3}$$

Case ① : $n < 31$,

if $n < 31$, we should not have a sample with registration number 31.
i.e. the likelihood of observing 31 is zero as it lies outside the range of registration numbers.

$\Rightarrow L(n) = 0$

Case 2 : $n \geqslant 31$,

our observations $(5, 19, 31)$ are valid
and the likelihood function is as above : $L(n) = \frac{1}{n^3}$

$$\therefore \arg\max \hat{n} = \max_n \left(\frac{1}{n^3}\right) = \min_n (n)$$

To obtain MLE $\hat{n}$ is to maximise $L(n)$, which is to minimize n.

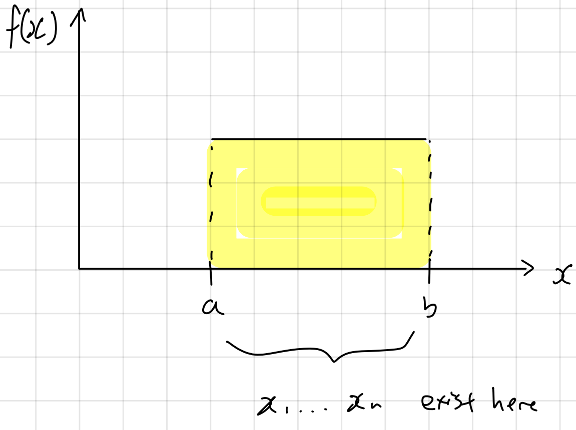
However, n must be at least 31 (i.e. $n \geqslant 31$)

$\therefore \hat{n} = 31.$

Problem 4:

$x_1, x_2, \dots x_n \overset{iid}{\sim} Unif(a,b)$, a, b unknown

$$f(x_i) = \frac{1}{b-a}$$



$$L(a,b) = \prod_{i=1}^{n} f(x_i) = \frac{1}{(b-a)^n}$$

$$\ell(a,b) = -n \log(b-a)$$

$$\hat{a}: \quad \frac{\partial \ell}{\partial a} = -n \cdot \frac{1}{b-a} \cdot -1 = \frac{n}{b-a} \quad \text{— } ①$$

$$\hat{b}: \quad \frac{\partial \ell}{\partial b} = -n \cdot \frac{1}{b-a} = -\frac{n}{b-a} \quad \text{— } ②$$

Since $a \leq \min(X_1, \dots X_n)$ and $b \geq \max(X_1, \dots, X_n)$ by definition of $Unif(a,b)$,

and ① is increasing, the largest possible $\hat{a} = \min(X_1, \dots X_n)$

and similarly, ② is decreasing, the smallest possible $\hat{b} = \max(X_1, \dots, X_n)$

Explanation: ① Since all x_i's $\sim Unif(a,b)$, all values of x_i are equally likely i.e. each event has the same probability of happening.
To fit all x_i's (ideally), 'a' should be as large as possible, while still smaller than or equal to the smallest observation (i.e. $a \leq \min\{x_i \mid i \in [1,n]\}$).
Similarly, 'b' should be as small as possible but still larger then or equal to the largest observation of the data (i.e. $b \geq \max\{x_i \mid i \in [1,n]\}$).

② Since we want the MLE estimates $\hat{a}, \hat{b}$, assigning $\hat{a}$ to the smallest observed datapoint and $\hat{b}$ to the largest observed datapoint ensures that we fully capture all / full range of data.

⇒ this maximises our likelihood of observations

⇒ other choices for $\hat{a}$ and $\hat{b}$ will increase $(b-a)$ and decrease likelihood since:

$$L(a,b) \propto \frac{1}{(b-a)}$$

Problem 5:

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n \cdot e^{-\lambda \sum_{i=1}^{n} x_i}$$

$$\ell(\lambda) = n \log(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

$$n\lambda^{-1} = n(-1)(\lambda)^{-2}$$
$$= -\frac{n}{\lambda^2}$$

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\therefore \quad \frac{n}{\lambda} = \sum_{i=1}^{n} x_i$$

and $\hat{\lambda} = \dfrac{n}{\sum_{i=1}^{n} x_i} = \dfrac{1}{\bar{x}}$

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

$\mathbb{E}(X_i) = \frac{1}{\lambda}$, $Var(X_i) = \frac{1}{\lambda^2}$

$\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \frac{n}{\lambda}$, $Var\left(\sum_{i=1}^{n} X_i\right) = \frac{n}{\lambda^2}$

Since $\theta = \log(\lambda)$, $\frac{\partial\theta}{\partial\lambda} = \frac{1}{\lambda}$

$$I_{x_1 \ldots x_n}(\lambda) = \mathbb{E}\left[-\frac{\partial^2}{\partial\lambda^2}\ell(\lambda)\right] = \mathbb{E}\left[-\left(-\frac{n}{\lambda^2}\right)\right] = \mathbb{E}\left[\frac{n}{\lambda^2}\right]$$
$$= \frac{n}{\lambda^2}$$

For a single observation, $I_1(\lambda) = \frac{1}{\lambda^2} \Rightarrow I_1(\lambda)^{-1} = \lambda^2$

By the delta method, $\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}(\log(\hat{\lambda}) - \log(\lambda)) \sim N\left(0, \left(\frac{1}{\lambda}\right)^2 \cdot \lambda^2\right) = N(0, 1)$

$\therefore$ Asymp distribution: $\hat{\theta} \sim N\left(\log(\lambda), \frac{1}{n}\right)$