CGS698C Assignment 1

Jiyanshu Dhaka, 220481

Part 1: Probability

1.1 Two-Coin Tossing Experiment

In this experiment, two coins are tossed at the same time. Each coin toss results in either heads (H) or tails (T).

(a) Sample Space

The sample space Ω of the two-coin-tossing experiment is the set of all possible outcomes. Since each coin can result in either H or T, the sample space is:

$$\Omega = \{(HH), (HT), (TH), (TT)\}$$

(b) Event Space

The event space is the set of all possible events (subsets of the sample space) that can occur. For the sample space Ω, the event space includes:

$\{\}$,
$\{(HH)\}, \{(HT)\}, \{(TH)\}, \{(TT)\}$,
$\{(HH), (HT)\}, \{(HH), (TH)\}, \{(HH), (TT)\}, \{(HT), (TH)\}, \{(HT), (TT)\}, \{(TH), (TT)\}$,
$\{(HH), (HT), (TH)\}, \{(HH), (HT), (TT)\}, \{(HH), (TH), (TT)\}, \{(HT), (TH), (TT)\}$,
$\{(HH), (HT), (TH), (TT)\}$

(c) Probabilities of Events

Assume that all the outcomes in the sample space have equal probabilities.

i. Probability of Each Outcome There are four possible outcomes in the sample space Ω. Since all outcomes are equally likely, the sum of their probabilities must equal 1:

$$P(\{(HH)\}) + P(\{(HT)\}) + P(\{(TH)\}) + P(\{(TT)\}) = 1$$

Let $P(\{(HH)\}) = P(\{(HT)\}) = P(\{(TH)\}) = P(\{(TT)\}) = p$. Then:

$$4p = 1$$

Solving for p:

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

Thus, the probability of each outcome is:

$$P(\{(HH)\}) = P(\{(HT)\}) = P(\{(TH)\}) = P(\{(TT)\}) = \frac{1}{4}$$

ii. Probability of the Event that at Least One Head Appears The event that at least one head appears is given by the set:

$$A = \{(HH), (HT), (TH)\}$$

The probability of this event is the sum of the probabilities of the outcomes in A:

$$P(A) = P(\{(HH)\}) + P(\{(HT)\}) + P(\{(TH)\}) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

iii. Probability of the Event that Exactly One Head Appears The event that exactly one head appears is given by the set:

$$B = \{(HT), (TH)\}$$

The probability of this event is the sum of the probabilities of the outcomes in B:

$$P(B) = P(\{(HT)\}) + P(\{(TH)\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

Part 2: Discrete Random Variables

2.1 Word Recognition Experiment

The probability assigner function is given by:

$$f(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

where k is the number of correctly recognized words, n is the total number of words, and p is the underlying probability of correctly recognizing a word.

Given:

$$n = 50, \quad k = 45, \quad p = 0.9$$

We need to find $f(45, 50, 0.9)$:

$$f(45, 50, 0.9) = \binom{50}{45} (0.9)^{45} (0.1)^5$$

The binomial coefficient is calculated as:

$$\binom{50}{45} = \frac{50!}{45! \cdot 5!}$$

Therefore, the probability is:

$$f(45, 50, 0.9) = \frac{50!}{45! \cdot 5!}(0.9)^{45}(0.1)^5$$

2.2 Road Accidents in a Small City

Suppose in a small city, 10 road accidents happen on average in a single day. The probability of k number of road accidents in a day is given by the probability mass function:

$$f(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where $\lambda = 10$.

(a) Probability of Zero Road Accidents in a Day

The probability that zero road accidents happen in a day is:

$$f(0, 10) = \frac{10^0 e^{-10}}{0!} = e^{-10}$$

(b) Probability of More than 7 but Less than 10 Road Accidents in a Day

The probability of occurrence of more than 7 but less than 10 road accidents in a day is the sum of probabilities from 8 to 9:

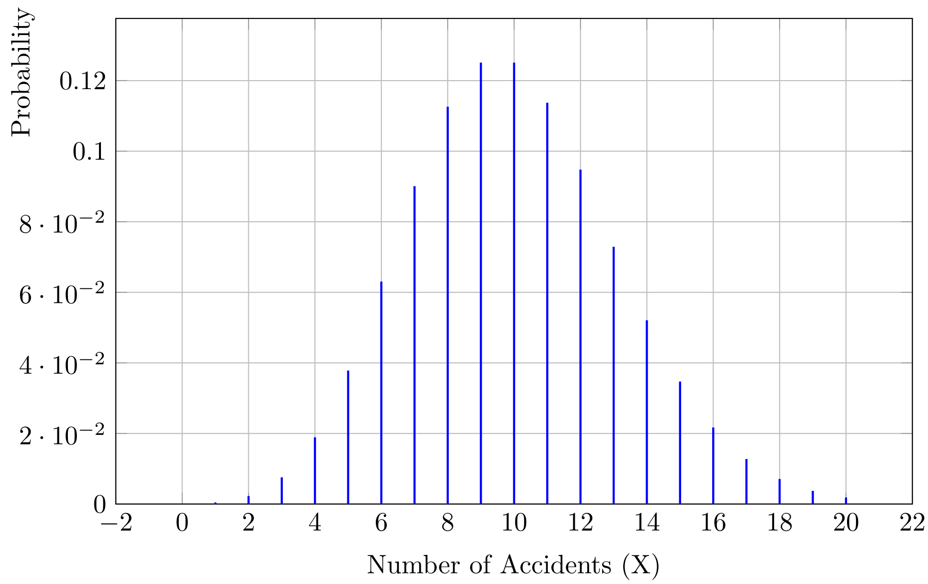
$$P(8 \leq k \leq 9) = f(8, 10) + f(9, 10)$$

$$f(8, 10) = \frac{10^8 e^{-10}}{8!}, \quad f(9, 10) = \frac{10^9 e^{-10}}{9!}$$

$$P(8 \leq k \leq 9) = \frac{10^8 e^{-10}}{8!} + \frac{10^9 e^{-10}}{9!}$$

(c) Graph of the Probability Mass Function

The graph of the probability mass function $f(x)$ for $X = 0, 1, 2, \ldots, 20$ is shown below:



3.1 Normal Distribution

Suppose a random variable X is normally distributed. The probability density function of the normal distribution is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(a) Find the probability density of obtaining $x = 0$, given that $\mu = 1$, $\sigma = 1$.

Given $\mu = 1$ and $\sigma = 1$:

$$f(0) = \frac{1}{1 \cdot \sqrt{2\pi}} e^{-\frac{(0-1)^2}{2 \cdot 1^2}}$$

$$f(0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}}$$

$$f(0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}}$$

(b) Find the probability density of obtaining $x = 1$, given that $\mu = 0$, $\sigma = 1$.

Given $\mu = 0$ and $\sigma = 1$:

$$f(1) = \frac{1}{1 \cdot \sqrt{2\pi}} e^{-\frac{(1-0)^2}{2 \cdot 1^2}}$$

$$f(1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}}$$

$$f(1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}}$$

(c) Finding the probability that the outcome occurs between x_2 and x_3.

You are given:

$$P(x_1 \leq X \leq x_2) = 0.3$$

$$P(x_1 \leq X \leq x_3) = 0.45$$

To find $P(x_2 \leq X \leq x_3)$:

$$P(x_2 \leq X \leq x_3) = P(x_1 \leq X \leq x_3) - P(x_1 \leq X \leq x_2)$$

$$P(x_2 \leq X \leq x_3) = 0.45 - 0.3 = 0.15$$

Part 4: Likelihood Function

Suppose a random variable X has the probability density function $f(x, \theta)$ where θ is a parameter of the probability density function and x is a value of the random variable X. The PDF tells you the probability density of generating an outcome x when the value of θ is known or assumed. For example, if I know or assume $\theta = 2$, I can calculate the probability density for different values of the random variable such as $X = 5$, $X = 3$, or $X = 100$. Basically, the PDF is viewed as a function of x when θ is fixed.

However, I can also view the PDF in a different way. I can calculate the probability density of obtaining a given, fixed outcome x for different values of θ. That is, the PDF can be viewed as a function of θ when x is fixed. This alternative view of the PDF is called the likelihood function.

The likelihood function maps the values of the parameter θ to probability densities when the sample x is taken as a fixed, observed quantity.

In summary, the PDF is a function of x where θ is assumed to be fixed; the likelihood function is a function of the parameter θ when the sample x is fixed or known.

The likelihood function is often represented by $L(\theta|x)$:

$$L(\theta|x) = f(x, \theta) \text{ when } x \text{ is fixed}$$

4.1 Visual Word Recognition Experiment

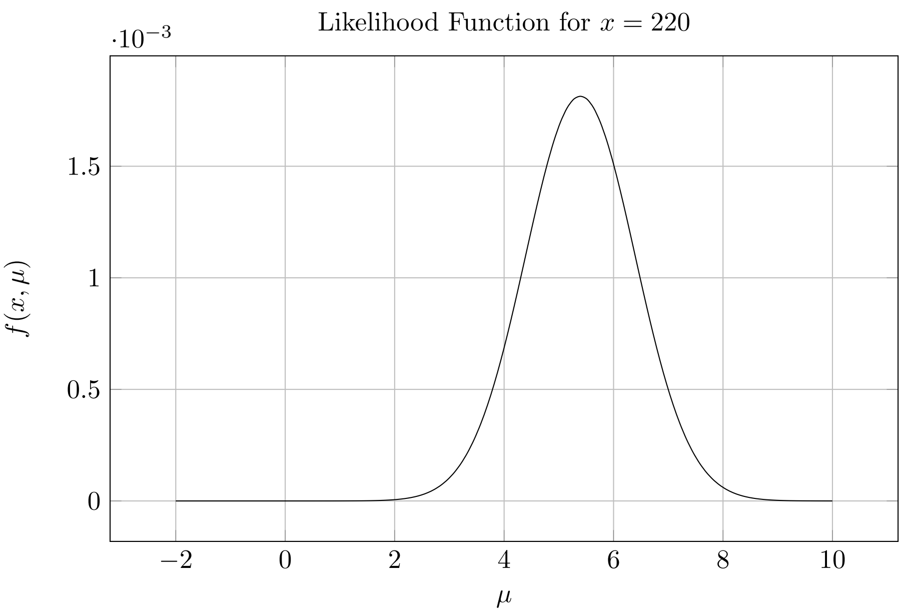
In a visual word recognition experiment, a participant has to recognize whether a string shown on the screen is a meaningful word (e.g., "book") or a non-word (e.g., "bktr"). The participant is asked to answer "yes" if the shown string is a meaningful word, and "no" if it is a meaningless non-word. Suppose a participant is shown 5 strings on the screen one by one. The time taken by the participant to recognize each string is shown below (in milliseconds):

Recognition time for 5 strings: 303, 443, 220, 560, 880

Suppose the random variable X represents the string recognition times. A researcher proposes a hypothesis that the string recognition times are generated by the probability density function $f(x, \mu)$:

$$f(x, \mu) = \frac{1}{x\sqrt{2\pi}} e^{-\frac{(\log x - \mu)^2}{2}}$$

1 (a) Plot the graph of the likelihood function with respect to values of μ, assuming that x is fixed to 220



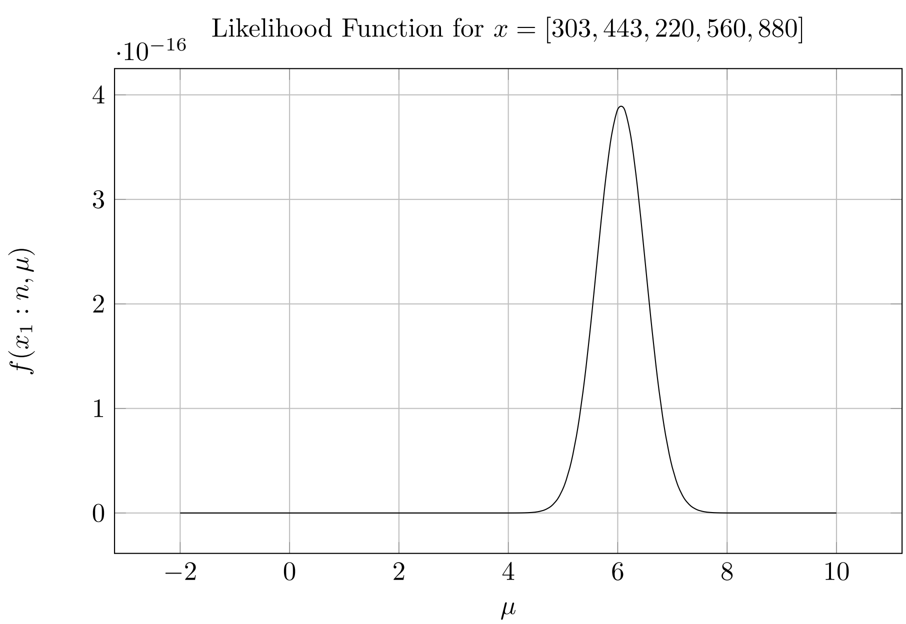
Below is a table showing the likelihood values at different points of μ for $x = 220$:

μ	Likelihood (L)
-2.0	2.44×10^{-75}
-0.67	1.91×10^{-11}
0.67	2.54×10^{-8}
2.0	5.72×10^{-6}
3.3	0.22×10^{-3}
4.67	0.14×10^{-2}
5.39	0.18×10^{-2}
6.00	0.15×10^{-2}
7.33	0.27×10^{-3}
8.66	0.85×10^{-5}
10.0	0.95×10^{-7}

2 (b) Graph the likelihood function when x is the observed sample of recognition times i.e., 303, 443, 220, 560, 880

The likelihood function for a sample of recognition times $x = [x_1, x_2, x_3, \ldots, x_n]$ is given by:

$$f(x_1 : n, \mu) = \left(\prod_{i=1}^{n} \frac{1}{x_i \sqrt{2\pi}} \right) e^{-\sum_{i=1}^{n} \frac{(\log x_i - \mu)^2}{2}}$$



Below is a table showing the likelihood values at different points of μ for the observed sample $303, 443, 220, 560, 880$:

μ	Likelihood (L)
-2.0	1.23×10^{-92}
-0.67	3.76×10^{-39}
0.67	7.95×10^{-38}
2.0	6.42×10^{-32}
3.3	1.72×10^{-27}
4.67	2.98×10^{-25}
5.39	2.95×10^{-24}
6.06	3.89×10^{-16}
7.33	1.27×10^{-18}
8.66	4.56×10^{-21}
10.0	1.26×10^{-23}

3 (c) Maximum Likelihood Estimate for μ

The value of μ that maximizes the likelihood function for the observed sample $303, 443, 220, 560, 880$ is approximately 6.06 .

From the analysis, we see that the maximum likelihood estimate for μ is approximately 6.06, indicating that this value of μ best explains the observed sample of recognition times.