Assignmnet 2

Problem 1

Bayes' Formula:

$$P(\text{Disease}|\text{Positive}) = \frac{P(\text{Positive}|\text{Disease}) \cdot P(\text{Disease})}{P(\text{Positive}|\text{Disease}) \cdot P(\text{Disease}) + P(\text{Positive}|\text{No Disease}) \cdot P(\text{No Disease})}$$

Substituting Values:

$$P(\text{Disease}|\text{Positive}) = \frac{0.95 \cdot 0.001}{(0.95 \cdot 0.001) + (0.001 \cdot 0.999)}$$

Calculation:

1. **Numerator**: $0.95 \cdot 0.001 = 0.00095$
2. **Denominator**: $0.00095 + (0.001 \cdot 0.999) = 0.00095 + 0.000999 = 0.001949$
3. **Result**:

$$P(\text{Disease}|\text{Positive}) = \frac{0.00095}{0.001949} \approx 0.4874$$

Final Answer:

$$\boxed{0.487}$$

Interpretation: There is approximately a **48.7% chance** that you actually have the disease if you have the symptom and test positive.

Problem 2

1. **Define the events**:
   - A: The family has two boys (BB).
   - B: You met a boy.
2. **Compute** $P(A|B)$ using Bayes' theorem:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}.$$

3. **Calculate each component**:
   - $P(A) = \frac{1}{4}$ (probability of BB).
   - $P(B|A) = 1$ (if the family is BB, you always meet a boy).
   - $P(B)$: Total probability of meeting a boy across all family types: $$ P(B) = \underbrace{P(B | BB) \cdot P(BB)}*{BB} + \underbrace{P(B | BG) \cdot P(BG)}*{BG} + \underbrace{P(B | GB) \cdot P(GB)}*{GB} + \underbrace{P(B | GG) \cdot P(GG)}*{GG}.

$$Substitutingvalues:$$

P(B) = \left(1 \cdot \frac{1}{4}\right) + \left(\frac{1}{2} \cdot \frac{1}{4}\right) + \left(\frac{1}{2} \cdot \frac{1}{4}\right) + \left(0 \cdot \frac{1}{4}\right) = \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}. $$

4. **Final probability**:

$$P(A|B) = \frac{1 \cdot \frac{1}{4}}{\frac{1}{2}} = \frac{1/4}{1/2} = \frac{1}{2}.$$

Why Opinion (2) is Incorrect

Opinion (2) assumes the sample space reduces to BB, BG, GB after observing a boy, treating all three cases as equally likely. However, this **ignores the selection mechanism** of randomly meeting one child:

- In BB, you **always** meet a boy.
- In BG or GB, you have only a 50% chance of meeting a boy.

Thus, BG and GB are **underrepresented** in the observed data (meeting a boy). The correct calculation weights BB twice as heavily as BG or GB, leading to a probability of $\frac{1}{2}$.

Conclusion

- **Correct answer**: $\boxed{\frac{1}{2}}$.
- **Opinion (2) is wrong** because it fails to account for the unequal likelihood of observing a boy in BB vs. BG/GB. The selection process biases the sample toward families with more boys.

Problem 3

Answer: No, the independence of A and B, and B and C, does **not** guarantee that A and C are independent. Here is a counterexample:

Counterexample

Consider a sample space with 4 equally likely outcomes: $1, 2, 3, 4$, each with probability $\frac{1}{4}$. Define:

- $A = 1, 2$,
- $B = 2, 3$,
- $C = 3, 4$.

Step 1: Verify A and B are independent

- $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$,
- $P(A \cap B) = P(2) = \frac{1}{4}$,
- $P(A)P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Since $P(A \cap B) = P(A)P(B)$, A and B are independent.

Step 2: Verify B and C are independent

- $P(B) = \frac{1}{2}$, $P(C) = \frac{1}{2}$,
- $P(B \cap C) = P(3) = \frac{1}{4}$,
- $P(B)P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Since $P(B \cap C) = P(B)P(C)$, B and C are independent.

Step 3: Check A and C

- $P(A) = \frac{1}{2}$, $P(C) = \frac{1}{2}$,
- $P(A \cap C) = P(\emptyset) = 0$,
- $P(A)P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Since $P(A \cap C) \neq P(A)P(C)$, A and C are **dependent**.

Conclusion: Independence is **not transitive**. A counterexample demonstrates that A and C can be dependent even if $A \perp B$ and $B \perp C$.

Problem 4

Solution:

Assumption (1): Marking keys after unsuccessful trials (without replacement)

- **PMF**: Uniform distribution. Each trial is equally likely to be the successful one.

$$P(X = k) = \frac{1}{5}, \quad k = 1, 2, 3, 4, 5.$$

- **Reasoning**: Since keys are marked and never reused, the correct key is equally likely to be in any of the 5 positions. For example:
    - $P(X = 1) = \frac{1}{5}$,
    - $P(X = 2) = \frac{4}{5} \cdot \frac{1}{4} = \frac{1}{5}$,
    - $P(X = 3) = \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{5}$, etc.

Assumption (2): Choosing any key equally likely (with replacement)

- **PMF**: Geometric distribution. Trials are independent, with success probability $p = \frac{1}{5}$.

$$P(X = k) = \left(\frac{4}{5}\right)^{k-1} \cdot \frac{1}{5}, \quad k = 1, 2, 3, \ldots.$$

- **Reasoning**: Each trial has a fixed probability $\frac{1}{5}$ of success, independent of previous trials. For example:
    - $P(X = 1) = \frac{1}{5}$,
    - $P(X = 2) = \frac{4}{5} \cdot \frac{1}{5}$,
    - $P(X = 3) = \left(\frac{4}{5}\right)^2 \cdot \frac{1}{5}$, etc.

Final Answer

1. **For Assumption (1)**:

$$P(X = k) = \frac{1}{5}, \quad \boxed{k = 1, 2, 3, 4, 5}.$$

2. **For Assumption (2)**:

$$P(X = k) = \left(\frac{4}{5}\right)^{k-1} \cdot \frac{1}{5}, \quad \boxed{k = 1, 2, 3, \ldots}.$$

Problem 5

Answer: The best-fitting Poisson distribution is **Poisson(3)**. The predicted numbers of hours are calculated as follows:

Step 1: Calculate the Sample Mean

The observed data gives a sample mean of:

$$\bar{k} = \frac{0 \cdot 45 + 1 \cdot 152 + 2 \cdot 215 + 3 \cdot 232 + 4 \cdot 168 + 5 \cdot 96 + 6 \cdot 55 + 7 \cdot 37}{1000} \approx 3.02$$

This suggests **Poisson(3)** is the most appropriate model.

Step 2: Compute Poisson Probabilities for $\lambda = 3$

Using the Poisson PMF $P(k; \lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$, we calculate:

- $P(0; 3) = 0.0498$
- $P(1; 3) = 0.1494$
- $P(2; 3) = 0.2240$
- $P(3; 3) = 0.2240$
- $P(4; 3) = 0.1680$
- $P(5; 3) = 0.1008$
- $P(6; 3) = 0.0504$
- $P(k \geq 7; 3) = 1 - \sum_{k=0}^{6} P(k; 3) = 0.0336$

Step 3: Predicted Hours for $\lambda = 3$

Multiply probabilities by 1000 (total hours):

Number of emails, k	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7 or more**
Predicted Hours	49.79	149.36	224.04	224.04	168.03	100.82	50.41	33.60

Step 4: Why Poisson(3) Fits Best

- **Poisson(2)** underestimates higher k values (e.g., $k = 3$ observed: 232 vs. predicted: 180.4).
- **Poisson(4)** overestimates higher k values (e.g., $k \geq 7$ observed: 37 vs. predicted: 110.6).
- **Poisson(3)** aligns closely with the observed distribution (e.g., $k = 3$ observed: 232 vs. predicted: 224.04).

Final Answer

The best-fitting distribution is **Poisson(3)**. The completed table is:

Number of emails, k	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7 or more**
Predicted Hours	49.79	149.36	224.04	224.04	168.03	100.82	50.41	33.60

Problem 6

Solution:

Step 1: Define the Problem

- **Premium per participant**: $12.
- **Compensation per death**: $2,000.
- **Number of participants**: 2,500.
- **Probability of death per participant**: 0.002.

Let X be the number of deaths in a year. Then:

- **Total revenue**: $12 \times 2500 = 30,000$.
- **Total cost**: $2000 \times X$.
- **Profit**: $\text{Profit} = 30,000 - 2000X$.

We want to find $P(\text{Profit} < 0)$, i.e., $P(30,000 - 2000X < 0)$.

Step 2: Simplify the Inequality

$$30,000 - 2000X < 0 \implies X > 15.$$

Thus, the insurance company loses money if $X > 15$.

Step 3: Model X

X follows a **Binomial distribution**:

$$X \sim \text{Binomial}(n = 2500, p = 0.002).$$

Since n is large and p is small, we approximate X using a **Poisson distribution** with $\lambda = np = 2500 \times 0.002 = 5$:

$$X \sim \text{Poisson}(\lambda = 5).$$

Step 4: Calculate $P(X > 15)$

Using the Poisson PMF $P(k;\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$, we compute:

$$P(X > 15) = 1 - P(X \leq 15).$$

Using a Poisson calculator or table for $\lambda = 5$:

$$P(X \leq 15) \approx 0.99999.$$

Thus:

$$P(X > 15) = 1 - 0.99999 = 0.00001.$$

Final Answer

The probability that the insurance company loses money is:

$$\boxed{0.00001}.$$

Problem 7

Solution:

---

(1) Distribution of T

Jobs are sent independently at a constant rate of 3 jobs per hour. This implies:

- The time interval T between two consecutive jobs follows an **exponential distribution**.
- The rate parameter $\lambda = 3$ jobs/hour.

Thus:

$$T \sim \text{Exponential}(\lambda = 3).$$

---

(2) Probability the Next Job Arrives Within 5 Minutes

First, convert 5 minutes to hours:

$$5 \text{ minutes} = \frac{5}{60} \text{ hours} = \frac{1}{12} \text{ hours}.$$

The CDF of an exponential distribution is:

$$P(T \leq t) = 1 - e^{-\lambda t}.$$

Substitute $\lambda = 3$ and $t = \frac{1}{12}$:

$$P\left(T \leq \frac{1}{12}\right) = 1 - e^{-3\cdot\frac{1}{12}} = 1 - e^{-\frac{1}{4}}.$$

Calculate $e^{-\frac{1}{4}}$:

$$e^{-\frac{1}{4}} \approx 0.7788.$$

Thus:

$$P\left(T \leq \frac{1}{12}\right) = 1 - 0.7788 = 0.2212.$$

Final Answer

1. The distribution of T is:

$$\boxed{T \sim \text{Exponential}(\lambda = 3)}.$$

2. The probability that the next job arrives within 5 minutes is:

$$\boxed{0.2212}.$$

Problem 8

Group Members:

1. 12213031 田源坤
2. 12312735 何捷
3. 12312735 何捷