DSW_ASSIGNMENT-2

Untitled Attachment

1. An Earphone manufacturing company claims that the average life of its product is 2.1 years. If the standard deviation is 0.17 and the significance level is set to 0.05. Do hypothesis testing to see if company's claim is right or wrong? We have the following 10 samples.

Assignment - 2

1. H_0 mean of product $= \mu = 2.1$ yrs $\quad H_1 : \mu \neq 2.1$
   std $= 0.17$

Sample mean $= \bar{X} = \dfrac{1.9 + 2.3 + 2.1 + 2.2 + 1.9 + 2.4 + 2.1 + 2.3 + 2.2 + 2.0}{10}$

$= 2.14$

$$\text{Std } (s) = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{0.04 + 0.16 + 0.00 + 0.04 + 0.04 + 0.16 + 0.00 + 0.16 + 0.04 + 0.04}{9}}$$

$= 0.173$

$$t = \frac{(\bar{X} - \mu)}{(s/\sqrt{n})} = \frac{(2.14 - 2.1)}{(0.173/\sqrt{10})} = \frac{0.4}{0.0547} = 7.31$$

Degree of freedom (df) $= n - 1 = 10 - 1 = 9$

Cricital t-value $= \pm 2.26$

df $= 9$
significane level $= 0.005$

The t-staticatic 7.31 is greater then the cricital t-value (2.26) ~~and~~ is greater the signifiance value 0.05

~~Hypothesis testing is for~~ company claim is wrong.

Avg. battery life $\neq$ 2.1 yrs.

2. What can you say if the company had claimed that the average life of it's batteries are 4.1 with a standard deviation of 0.1

2. $H_0: \mu = 4.1$

$H_1: \mu \neq 4.1$

$$t = \frac{(\bar{x} - \mu)}{(s/\sqrt{n})} = \frac{(2.14 - 4.1)}{(0.173/\sqrt{10})} = \frac{-1.96}{0.0547} = -35.83$$

$df = n - 1 = 10 - 1 = 9$

t-table ($df = 9$, significance level $= 0.05$)

cricital t-value $= \pm 2.26$

~~t-st~~ t-statistic $(-35.83) <$ t-critical (-2.26)

∴ We can say the avg. battery life is 4.1 yrs is incorrect

3.What are random variables? Discuss about continuous and discrete random variables with examples.

Ans-

Random variable

Answer-A random variable in statistics is a function that assigns a real value to an outcome in the sample space of a random experiment.
Random variables can have specific values or any value in a range.
There are two basic types of random variables,

1.Discrete Random Variables

2.Continuous Random Variables

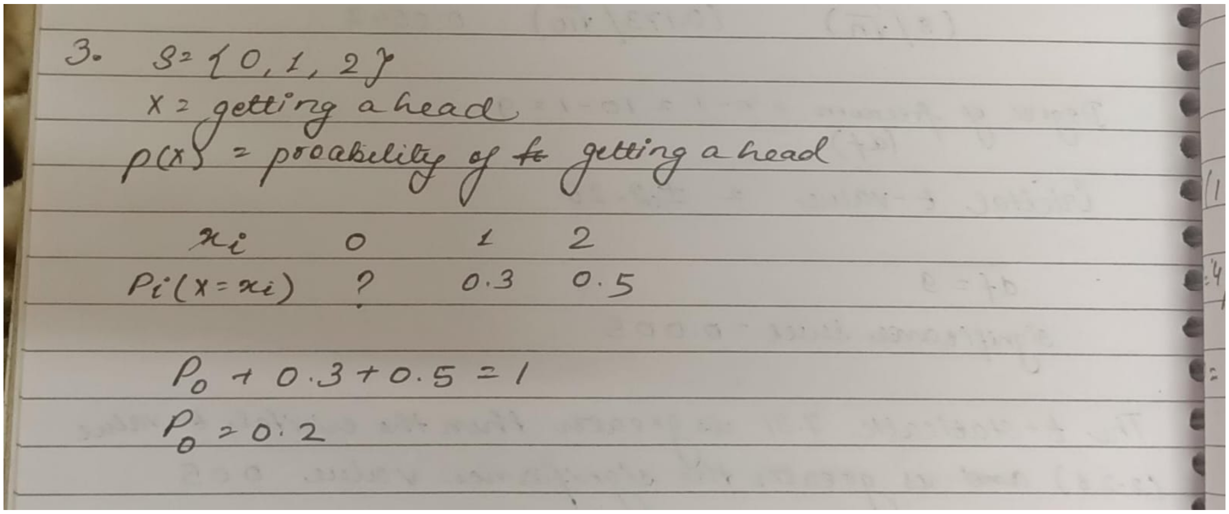
Discrete Random Variables

A discrete random variable is a type of random variable that can take on a finite or countable number of distinct values
If X is a discrete random variable and the PMF of X is P(xi), then

- $0 \le p_i \le 1$
- $\sum p_i = 1$ where the sum is taken over all possible values of x

Example-tossing of two coins
p(x)=probability getting a head
s={0,1,2}
x=getting a head

xi	0	1	2
Pi(X=xi)	?	0.3	0.5



3. $S = \{0, 1, 2\}$

X = getting a head

$p(x)$ = probability of ~~to~~ getting a head

x_i	0	1	2
$P_i(X = x_i)$	?	0.3	0.5

$P_0 + 0.3 + 0.5 = 1$

$P_0 = 0.2$

Continuous Random Variable

Continuous Random Variable takes on an infinite number of values.

If X is a continuous random variable. P (x < X < x + dx) = f(x)dx then,

- $0 \leq f(x) \leq 1$; for all x
- $\int f(x)\,dx = 1$ over all values of x

Then P (X) is said to be a PDF of the distribution.

Example-

Find the value of P (1 < X < 2)

Such that,

$f(x) = kx^3$; $0 \leq x \leq 3 = 0$

Otherwise f(x) is a density function.

$P(1 < X < 2)$

$f(x) = kx^3;\ 0 \leq x \leq 3 = 0$

$\int f(x)\,dx = 1$

$\int kx^3\,dx = 1$

$\frac{kx^4}{4} = 1$

$0 \leq x \leq 3 = 0$

$\frac{k\,[3^4 - 0^4]}{4} = 1$

$k\left(\frac{81}{4}\right) = 1$

$\boxed{k = \frac{4}{81}}$

$P(1 < X < 2) = k \times [X^4]/4$

$P = 4/81 \times \frac{[16-1]}{4}$

$P = \frac{15}{81}$

4. Write a program

a. To compute the mean, median and mode of a python list.

1. To compute the variance and standard deviation of a vector(list).
2. To compute the covariance and correlation between two vectors(list).

Untitled Attachment

5. State central limit theorem. Demonstrate CLT with an example. Use python programming for the following steps. Consider a population which satisfies binomial distribution. Draw random samples from the population, find each sample mean. What can you say about the list containing sample means.

Ans-**Central Limit Theorem** states that when large samples usually greater than thirty are taken into consideration then the distribution of sample arithmetic mean approaches the normal distribution irrespective of the fact that random variables were originally distributed normally or not.

Example 1. The male population’s weight data follows a normal distribution. It has a mean of 70 kg and a standard deviation of 15 kg. What would the mean and standard deviation of a sample of 50 guys be if a researcher looked at their records?

7. A: the event that a person has the disease
B: the event that a person tests positive for the disease

$$P(A) = \frac{1}{500} = 0.002$$

$$P(B/A) = 0.99$$

$$P\left(\frac{B'}{A'}\right) = 0.95$$

$$P(A/B) = P(B/A) * \frac{P(A)}{P(B)}$$

$$P(A') = 1 - P(A)$$
$$P(B/A') = 1 - 0.95 = 0.05$$

$$P(B) = 0.99 \times 0.002 + 0.05(1 - 0.002)$$
$$= 0.00198 + 0.04995$$
$$= 0.05193$$

$$P(A/B) = 0.99 * \frac{0.002}{0.05193} = 0.0364$$

The probability that a person who tests +ve for the disease actually has it is approx. 0.0364 or 3.64%.

6. What is Baye’s theorem. What are some real-world applications of Bayes’ theorem in fields such as medicine and machine learning?

Bayes theorem (also known as the Bayes Rule or Bayes Law) is used to determine the conditional probability of event A when event B has already occurred.

The general statement of Bayes’ theorem is “**The conditional probability of an event A, given the occurrence of another event B, is equal to the product of the event of B, given A and the probability of A divided by the probability of event B.”**i.e.

$$P(A|B) = P(B|A)P(A) / P(B)$$

where,

- ***P(A)****and* ***P(B)****are the probabilities of events A and B*
- ***P(A|B)****is the probability of event A when event B happens*
- ***P(B|A)****is the probability of event B when A happens*

Example are taken for www.geeksforgeeks.org

Bayes’ Theorem has a wide range of applications in various fields. Let’s explore some real-world scenarios where Bayes’ Theorem plays a crucial role:

1. **Spam Filtering**:
    - **Application**: Email systems use Bayes’ Theorem to distinguish between legitimate emails and spam.

- **How It Works**: By analyzing the frequency of certain words or phrases in both spam and non-spam emails, the filter assigns probabilities to incoming messages being spam. For example, if an email contains words commonly found in spam messages (e.g., "free" or "discount"), the filter calculates the likelihood that it is spam and routes it accordingly[1].

2. **Weather Forecasting**:
   - **Application**: Meteorologists apply Bayes' Theorem to improve the accuracy of weather forecasts.
   - **How It Works**: By incorporating data from various sources (satellite imagery, weather stations, historical patterns), forecast models calculate the probability of different weather outcomes. For instance, if a high-pressure system is moving in from the west, Bayes' Theorem helps assess the likelihood of clear skies versus rain in a specific region[1].
3. **DNA Testing (Forensic Science)**:
   - **Application**: In criminal investigations, Bayes' Theorem is used to interpret DNA evidence.

- **How It Works**: By comparing DNA samples collected from crime scenes to databases of known DNA profiles, analysts calculate the probability that a suspect's DNA matches the evidence. This aids in making informed decisions during investigations[1].

4. **Medical Diagnosis**:
    - **Application**: While not a replacement for medical expertise, Bayes' Theorem can be used in conjunction with patient data and medical history to calculate the probability of a specific disease.
    - **How It Works**: Given a patient's symptoms and test results, Bayes' Theorem helps assess the likelihood of a particular disease. It assists doctors in prioritizing further tests and making informed decisions[2].
5. **Machine Learning (Classification)**:
    - **Application**: In machine learning, Bayes' Theorem is used for classification tasks.
    - **How It Works**: For instance, in spam detection, natural language processing (NLP), and image recognition, Bayes' Theorem helps identify objects in photographs or classify text based on word frequencies. Bayesian methods are also used in probabilistic models like Naive Bayes classifiers[3].

6. **Financial Forecasting**:
   - **Application**: Bayes' Theorem can be applied to financial markets and risk assessment.
   - **How It Works**: By incorporating new information (e.g., economic indicators, market trends), investors can update their beliefs about future stock prices or other financial variables[1].
7. **Fault Diagnosis in Engineering**:
   - **Application**: Engineers use Bayes' Theorem to diagnose faults in complex systems (e.g., machinery, electronics).
   - **How It Works**: By combining prior knowledge, sensor data, and statistical models, Bayes' Theorem helps identify faulty components or predict system failures[1].

7. A certain disease affects 1 in every 500 people. A diagnostic test for the disease is 99% accurate when a person has the disease and 95% accurate when a person does not have the disease. If a person tests positive for the disease, what is the probability that they actually have it?

7. A: the event that a person has the disease
B: the event that a person tests positive for the disease

$P(A) = \frac{1}{500} = 0.002$

$P(B/A) = 0.99$

$P\left(\frac{B'}{A'}\right) = 0.95$

$P(A/B) = P(B/A) * \frac{P(A)}{P(B)}$

$P(A') = 1 - P(A)$
$P(B/A') = 1 - 0.95 = 0.05$

$P(B) = 0.99 \times 0.002 + 0.05(1 - 0.002)$
$= 0.00198 + 0.04995$
$= 0.05193$

$P(A/B) = 0.99 * \frac{0.002}{0.05193} = 0.0364$

The probability that a person who tests +ve for the disease actually has it is approx. 0.0364 or 3.64%.

8. In a spam detection system, 1% of emails are actually spam. The spam filter correctly identifies 95% of spam emails but also incorrectly marks 3% of legitimate emails as spam. If an email is flagged as spam by the filter, what is the probability that it is actually spam?

FREEMIND

Date

Page

8. A: the event that a ~~person~~ email is span

B: " " " " email is flagged as spam

$P(A) = 0.01$

$P(B/A) = 0.95$

$P(B'/A') = 0.03$

$$P(A/B) = P\left(\frac{B}{A}\right) * \frac{P(A)}{P(B)}$$

$$P(B) = P\left(\frac{B}{A}\right) * P(A) + P\left(\frac{B}{A'}\right) * P(A')$$

$P(A') = 1 - P(A)$

$P(B/A') = 1 - P(B'/A') = 1 - 0.03 = 0.97$

$$P(B) = 0.95 \times 0.01 + 0.97 \times (1 - 0.01).$$
$$= 0.0095 + 0.9603$$
$$= 0.9698$$

$$P(A/B) = \frac{0.95 \times 0.01}{0.9698} \approx 0.0098$$

Probability 0.98%.

9. Consider two vectors of length n(input from the user). Take the vectors as input from the user. Write a python

program to find the

a. It’s dot product

b. Distance between two vectors.

c. Sum of the two vectors.

d. Normalize a vector

10. Take two 2 × 2 matrices as user input. Write a python program to find the sum of the two matrices.

Untitled Attachment

Untitled Attachment