1. **Basic probability theory notation and terms**

   a)

- **Probability** - a measurement of the likelihood that an event will occur and is in range between 0 (impossibility) and 1 (certainty).
- **Probability Mass** - the probability associated with discrete random variables.
- **Probability Density** - the probability associated with continuous random variables.
- **Probability Mass Function** - a function that gives us the probability associated with discrete random variables.
- **Probability Density Function** - a function that gives us the probability associated with continuous random variables.
- **Probability Distribution** - a function that gives us the probabilities of occurrence of all the possible outcomes in an experiment.
- **Discrete Probability Distribution** - the probability of occurrence of each value of a discrete random variable.
- **Continuous Probability Distribution** - the probabilities of the possible values of a continuous random variable.
- **Cumulative Distribution Function** - a function that calculates the probability that a random variable X takes on a value less than or equal to x.

- **Likelihood** - the probability of a model parameter value, given specific observed data.

b)

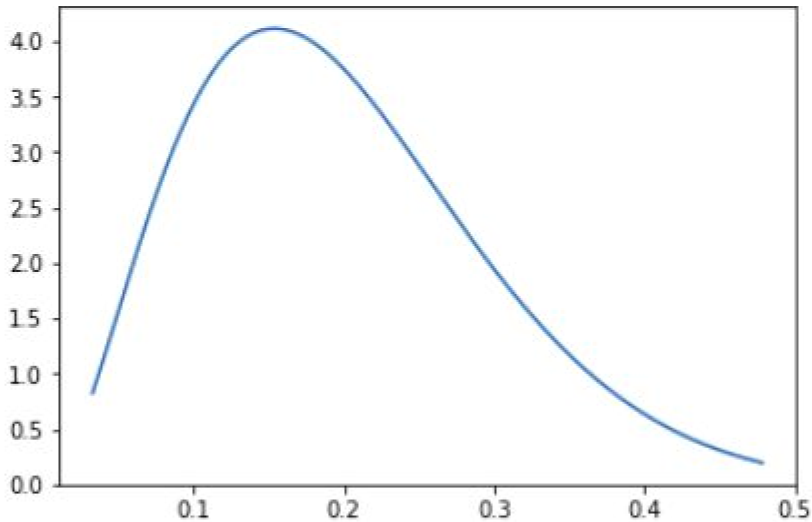
- **What is observation model?**
    - The observation model is the expression that relates model parameters to the observations.
      The observation is the value at a particular period of a particular variable.

- **What is statistical model?**
    - A statistical model is a model that implements statistical assumption regarding the generation of sample data from a larger population. The assumptions describe a set of probability distributions.

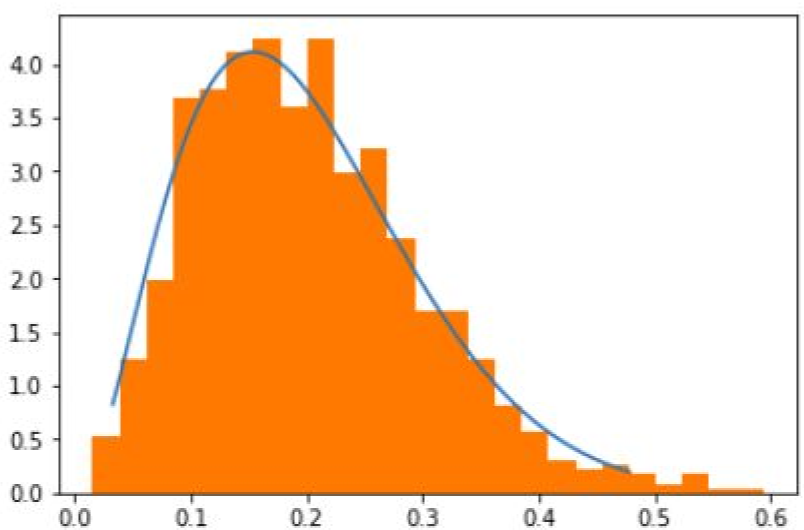
- **What is the difference between mass and density?**
    - Mass is used for describing discrete random variables and density is used for describing continuous random variables. Mass function assigns probability to each point in the sample space, whereas a density function gives a probability that a random variable falls within some interval.

2. Basic Computer Skills

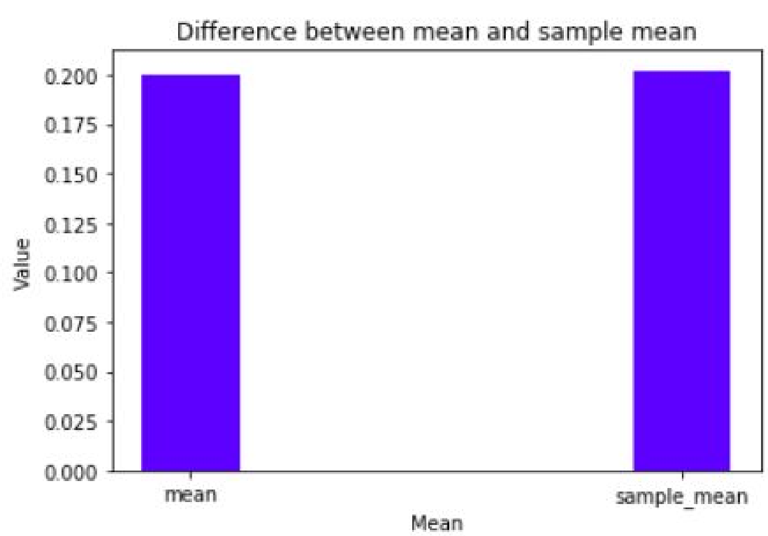
```
import numpy as np
from scipy.stats import beta

import matplotlib
import matplotlib.pyplot as plt


mean = 0.2
variance = 0.01

def calc_alfa(mean, variance):
    return mean * (mean * (1 - mean) / variance - 1)

a = calc_alfa(mean, variance)

def calc_beta(mean, a):
    return a * (1 - mean) / mean

b = calc_beta(mean, a)

x = np.linspace(beta.ppf(0.01, a, b), beta.ppf(0.99, a, b), num=10000)
y = beta.pdf(x, a, b)

plt.plot(x, y)
plt.show()
```



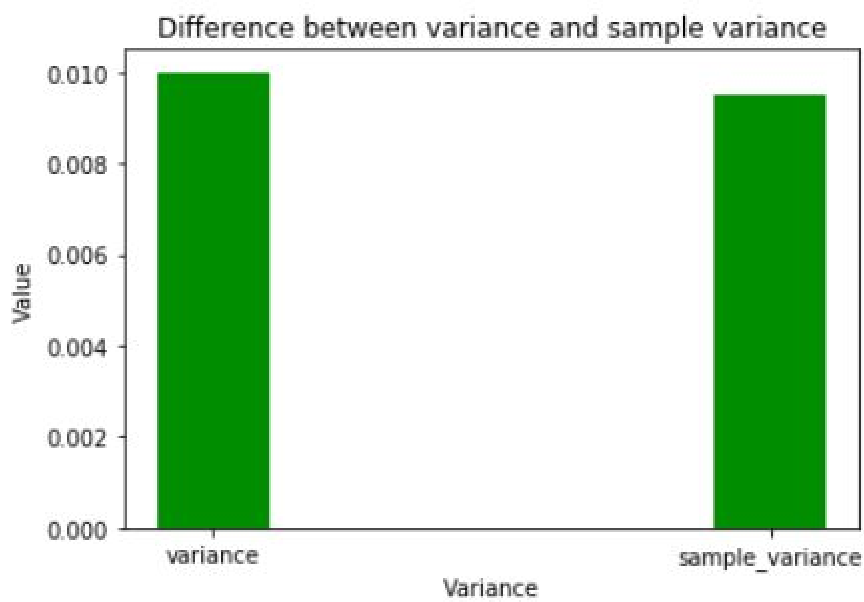

```
# Take 1000 random samples
sample = beta.rvs(a, b, size = 1000)
plt.plot(x, y)
plt.hist(sample, density=True, bins = 25)
plt.show()
```



```
# Calculate the mean and variance of the sample data
sample_mean = np.mean(sample)
sample_variance = np.var(sample)

print('The sample mean is: {0} . The sample variance is:
{1}'.format(sample_mean, sample_variance))

# The sample mean is: 0.20214093998293126 . The sample variance is:
0.009487014906088123

# Compare the mean and the sample mean
mean_plot = plt.bar([0, 1], [mean, sample_mean], 0.2,
                color='b',
                label='mean')

plt.xlabel('Mean')
plt.ylabel('Value')
plt.title('Difference between mean and sample mean')
plt.xticks([0, 1], ('mean', 'sample_mean'))

plt.show()
```



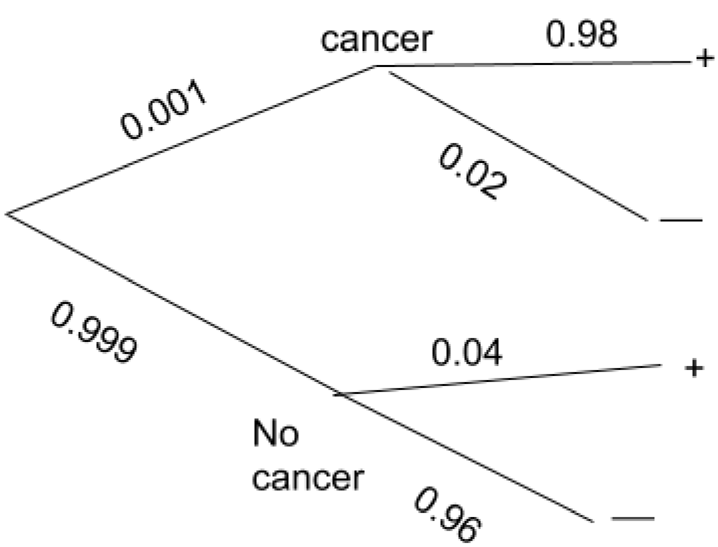
```
# Compare the variance and the sample variance
mean_plot = plt.bar([0, 1], [variance, sample_variance], 0.2,
              color='g',
              label='Variance')

plt.xlabel('Variance')
plt.ylabel('Value')
plt.title('Difference between variance and sample variance')
plt.xticks([0, 1], ('variance', 'sample_variance'))

plt.show()
```




```
confidence = np.percentile(sample, 95)
print('The central 95% interval of the sample distribution is:
{0}'.format(confidence))

# The central 95% interval of the sample distribution is: 0.3744919405953623
```

3. Bayes' Theorem

Given:
A: Subject has lung cancer
B: Test gives positive result
P(B | A) = 0.98
P(~B | ~A) = 0.96
P(A) = 0.001



Solution

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)} = \frac{0.98 * 0.001}{(0.98 * 0.001) + (0.999 * 0.04)} = \frac{0.00098}{0.00098 + 0.03996} = 0.02$$

If the test gives positive result, the chances of patient having a lung cancer are **2%.**

$$P(A^{\subset} \mid B^{\subset}) = \frac{P(B^{\subset} \mid A^{\subset}) * P(A^{\subset})}{P(B^{\subset})} = \frac{0.96 * 0.999}{(0.96 * 0.999) + (0.001 * 0.02)} = \frac{0.959}{0.959 + 0.00002} = 0.99$$

If the test gives negative result, the chances of patient not having a lung cancer are **99%.**

$$P(A^{\subset} \mid B) = \frac{P(B \mid A^{\subset}) * P(A^{\subset})}{P(B)} = \frac{0.04 * 0.999}{(0.04 * 0.999) + (0.98 * 0.001)} = \frac{0.03966}{0.03996 + 0.00098} = 0.97$$

If the test gives positive result, the chances of patient not having a lung cancer are **97%.**

$$P(A \mid B^{\subset}) = \frac{P(B^{\subset} \mid A) * P(A)}{P(A)} = \frac{0.02 * 0.001}{0.001} = \frac{0.0002}{0.001} = 0.02$$

If the test gives negative result, the chances of patient having a lung cancer are **2%.**

I would advise the researchers to improve their test.
Even if the test gives positive result, the chances of patient having a lung cancer are 2%.

4. Bayes' Theorem

Given:

Box A: 2 red and 5 white balls
Box B: 4 red and 1 white ball
Box C: 1 red and 3 white balls

Probability of picking a ball from A: 40% = 0.4
Probability of picking a ball from B: 10% = 0.1
Probability of picking a ball from C: 50% = 0.5

Solution:

Probability of picking a red ball from A: $\frac{2}{7} * 0.4 = 0.285 * 0.4 = 0.114$
Probability of picking a red ball from B: $\frac{4}{5} * 0.1 = 0.8 * 0.1 = 0.08$
Probability of picking a red ball from C: $\frac{1}{4} * 0.5 = 0.25 * 0.5 = 0.125$

Probability of picking a red ball: $0.114 + 0.08 + 0.125 = 0.319$ or ~32%

If the ball that was picked was a red ball, it was most likely that is was picked from the **box C** because box C has the highest probability of picking a red ball from all the boxes.

5. Bayes' Theorem

F - fraternal twins
I - identical twins
M - male
FE - female
MM - male-male
FF - female-female

$P(F) = \frac{1}{125} = 0.008$
$P(I) = \frac{1}{300} = 0.003$

Identical twins can be either two males or two females
Fraternal twins can be: male-male, male-female, female-male, female-female.

$P(MM \mid I) = 0.5$ (half of the population are male and half are female)
$P(MM \mid F) = 0.25$ (there are 4 combinations [male-male, male-female, female-male, female-female])

$$P(MM) = (0.5 * 0.003) + (0.25 * 0.003) = 0.002$$

$$P(I \mid MM) = \frac{P(MM \mid I) * P(I)}{P(MM)} = \frac{0.5 * 0.003}{0.002} = 0.75$$

The chances of Elvis Presley having an identical twin brother are **0.75 or 75%.**