Benha University
Faculty Of Engineering (Shoubra)
Communications And Computer Engineering

Course: EMP301 (Engineering Mathematics (5))
Term: 241
Course Project cover page

S#	Student Name	Edu Email	Student ID	Marks		
				Report & Slides (35)	Implementation (45)	Presentation (20)
1	Abdelrahman Salah El-dein Abdelaziz	abdelrahman402720@feng.bu.edu.eg	221903085			
2	Abdullah Mohamed Mohamed Galal	abdalah.mohamed21@feng.bu.edu.eg	221903118			
3	Farida Waheed Abd El Bary	farida414766@feng.bu.edu.eg	221903168			
4	Mohamed Ahmed Mohamed Hassan	mohammed405815@feng.bu.edu.eg	221903120			
5	Nour Hesham El Sayed	nor.hesham21@feng.bu.edu.eg	221902960			
6	Omar Sami Mohamed Ahmed	omar413298@feng.bu.edu.eg	221903198			
7	Razan Ahmed Fawzy	razan403035@feng.bu.edu.eg	221903165			

Date handed in: 19 / 12 / 2023.

Report & Slides (35)		Points
1. Content	Project includes all material needed to give a good understanding of the topic.	10
	All content throughout the paper & presentation is accurate. There are no factual errors.	10
2. Format	Sequencing of Information: Information is organized in a clear, logical way. It is easy to anticipate the next slide.	5
	All graphics are attractive (size and colors) and support the topic of the presentation.	5
	Text-Font Choice & Formatting: Font formats (color, bold, italic) have been carefully planned to enhance readability and content.	5
Presentation (20)		
1. Delivery	Members understand presented material and can clearly answer all questions about it.	10
	Members spoke at a good rate, volume and with good grammar. They maintained eye-contact while using, but not reading their notes.	5
2. Cooperation	Group's members share tasks, and all performed responsibly all of the time.	5
Implementation (45)		
1. Code	Members understand the code and can clearly answer all questions about it.	20
	The code is completely functional and correctly producing the outputs.	25
Total Score		**100**

Benha University
Faculty Of Engineering (Shoubra)
Communications And Computer Engineering

RANDOM VARIABLES

ENGINEERING MATHEMATICS (5)
EMP301

UNDER THE SUPERVISION OF:
DR. LAMIA ALREFAAI

Table of Contents:

Summary

This in-depth research offers a thorough examination of discrete and continuous random variables, diving into the complex worlds of probability and statistics. The conditions, probability functions, cumulative distributions, and practical applications of discrete variables—such as the Bernoulli, Binomial, Geometric, and Poisson distributions—are examined in detail. The investigation also includes continuous random variables, such as the Gaussian (Normal), exponential, and uniform distributions, all of which are carefully analysed for their mathematical characteristics and applications. The study emphasizes how important these distributions are for modelling variability and uncertainty in a variety of domains, offering a strong basis for statistical analysis and practical applications.

Abstract

This paper explores the complex relationship between probability and statistics, revealing the fundamental role that discrete and continuous random variables play in describing variability and uncertainty. The contrast between these variables is examined, illuminating their unique traits and underlying principles. Continuous random variables, like the Gaussian, exponential, and uniform distributions, use probability density functions to express measurements, whereas discrete random variables, like the Binomial and Bernoulli distributions, etc……, record countable outcomes.

Introduction

The fundamental contrast between discrete and continuous random variables is the basis of this paper's investigation into the broad field of probability and statistics. Probability mass functions are used to express discrete random variables, which are best illustrated by situations such as coin flips and represent countable outcomes. However, probability density functions are used for continuous random variables, which are represented by measures such as height distributions. To shed light on how these tools function as potent models for comprehending and navigating uncertainty and variability in a variety of real-world scenarios, this study aims to untangle the complex mathematical features and practical applications inherent in both categories.

Discrete Random Variables

1.Bernoulli Distribution

1.1. Definition

Bernoulli distribution is a special kind of distribution, and it is random experiment that can only have an outcome of either 1 or 0 is known as a Bernoulli trial.

$$P_X(x) = \begin{cases} 1 & \text{for } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

then X is a discrete random variable that can only take one value, i.e., X=1 with a probability of one and any other value will be zero.

1.2. Conditions

- There should be only two possible outcomes of your trial.
- Each of the two outcomes should have a fixed probability of occurrence.
- The trials should be independent of each other.

1.3. Parameter

The distribution is characterized by a single parameter (p), representing the probability of success. This parameter determines the shape and behaviour of the distribution.

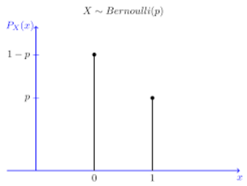
1.4. Probability Mass Function (PMF)

The probability mass function (PMF) for a discrete random variable assigns a probability to each value of the variable. The probability mass function for a Bernoulli distribution equals either p (the probability of success), or 1-p (the probability of failure).

$$\text{PMF} = f(x, p) = \begin{cases} p & \text{if } x = 1 \\ q = 1 - p & \text{if } x = 0 \end{cases}$$

We can also express this formula as,

$$f(x, p) = p^x (1 - p)^{1-x}, x \in \{0, 1\}$$

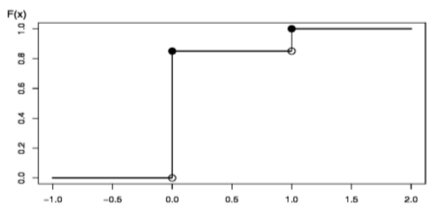


1.5. Cumulative Distribution Function (CDF)

The cumulative distribution function of a Bernoulli random variable X when evaluated at x is defined as the probability that X will take a value lesser than or equal to x. The formula is given as follows:

$$\text{CDF} = F(x, p) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - p & \text{if } 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

Stepwise increase is observed in the CDF. It represents cumulative probability for values 0 and 1, starting at 0 and rising to 1−p at 0 and abruptly reaching 1 at 1.



1.6. Mean and Variance

1.5.1. Mean

The arithmetic mean of many independent realizations of the random variable X gives us the expected value or mean. The expected value can also be thought of as the weighted average.

$$\mu = p$$

- Proof:

We know that for X,

P(X = 1) = p

P(X = 0) = q

E[X] = P(X = 1) . 1 + P(X = 0) . 0

E[X] = p . 1 + q . 0

E[X] = p

Thus, the mean or expected value of a Bernoulli distribution is given by E[X] = p.

1.5.2. Variance

The variance can be defined as the difference of the mean of X^2 and the square of the mean of X.

$$\sigma^2 = p(1-p)$$

- Proof:

$Var[X] = E[X^2] - (E[X])^2$

Using the properties of $E[X^2]$, we get,

$E[X^2] = \sum x^2 \, P(X = x)$

$E[X^2] = 1^2 . p + 0^2 . q = p$

Substituting this value in $Var[X] = E[X^2] - (E[X])^2$ we have

$Var[X] = p - p^2$

$= p(1 - p)$

$= p . q$

Hence, the variance of a Bernoulli distribution is $Var[X] = p(1 - p) = p . q$

1.7. Applications

1. In medicine, Bernoulli distributions are used to model the events experienced by a single patient. These events could be disease, death, and so on.
2. Logistic regressions use Bernoulli distribution to model the occurrence of certain events such as the specific outcome of a dice roll.
3. Bernoulli distribution is also used as a basis to derive several other probability distributions that have applications in the engineering, aerospace, and medical industries.

Also in real-life applications:

1. The success or failure of a medical treatment
2. Transmission or non-transmission of a disease

1.8. Code

```
def generate_bernoulli_randvar(n,p):
    return random.choices([0,1],weights=[(1-p),p],k=n)
```

```python
def bernoulli_pmf(probability,number):
    if number == 1:
        return probability
    elif number == 0:
        return 1-probability
def bernoulli_cdf(probability,number):
    if number == 1:
        return 1
    elif number == 0:
        return 1-probability
def bernoulli(probability, size):
    x = [0,1] #failure or success.
    #probability of 0 and1
    pmf = [bernoulli_pmf(probability, i) for i in x]
    # cummulative probability of 0 and 1
    cdf = [bernoulli_cdf(probability, i) for i in x]
    # generates size no. of bernoulli random variables.
    histogram = generate_bernoulli_randvar(size, probability)
    mean = probability
    variance = probability * (1 - probability)
    #plots the pmf, cdf and histogram graphs.
    plt.vlines(x, 0, pmf, lw=40, alpha=1)
    plt.text(0.5, 0.05, f'E: {mean}\nvar: {variance}', fontsize=12)
    plt.title("bernoulli PMF")
    plt.xlabel("X")
    plt.ylabel("probability")
    plt.show()
    plt.step(x,cdf,lw=10,where='post')
    plt.title("bernoulli CDF")
    plt.xlabel("X")
    plt.ylabel("cummulative probability")
    plt.show()
    plt.hist(histogram, bins='auto')
    plt.title("bernoulli histogram")
    plt.xlabel("X")
    plt.ylabel("samples")
    plt.show()
```

- **generate_bernoulli_randvar (n, p)**

This function generates **n** random variables following a Bernoulli distribution. Each variable is either 0 or 1, with the probability of being 1 equal to the top. The **random. choice's** function is used to perform this sampling, where 0 is chosen with probability **(1-p)** and 1 with probability **p**.

- **bernoulli_pmf (probability, number)**

This function calculates the Probability Mass Function (PMF) of the Bernoulli distribution. The Bernoulli distribution is characterized by two possible outcomes (0 or 1) with probabilities **(1–p)** and **p** respectively.

- **probability**: Probability of success (1).
- **number**: The random variable for which the PMF is calculated.

The function returns the probability of observing the specified **number** according to the Bernoulli distribution.

- **bernoulli_cdf (probability, number)**

This function calculates the Cumulative Distribution Function (CDF) of the Bernoulli distribution. The CDF gives the probability that a random variable is less than or equal to a specified value.

- **probability**: Probability of success (1).
- **number**: The random variable for which the CDF is calculated.

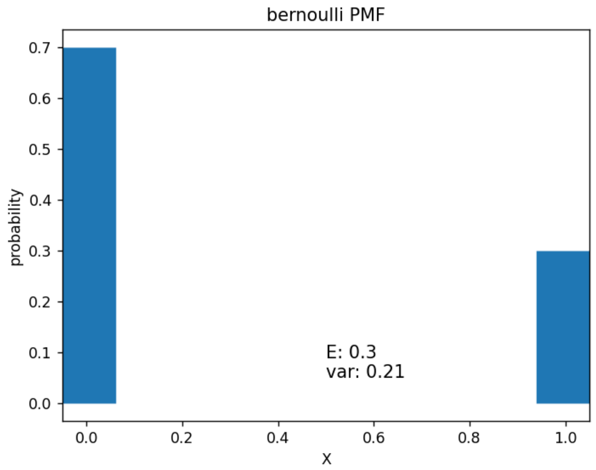
The function returns the cumulative probability up to the specified **number** in the Bernoulli distribution.

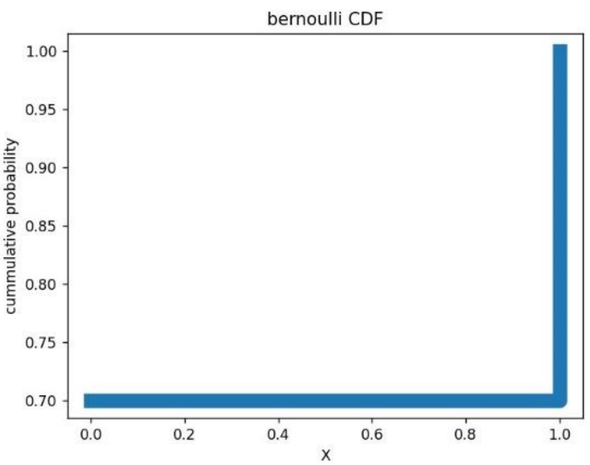
- **bernoulli (probability, size)**

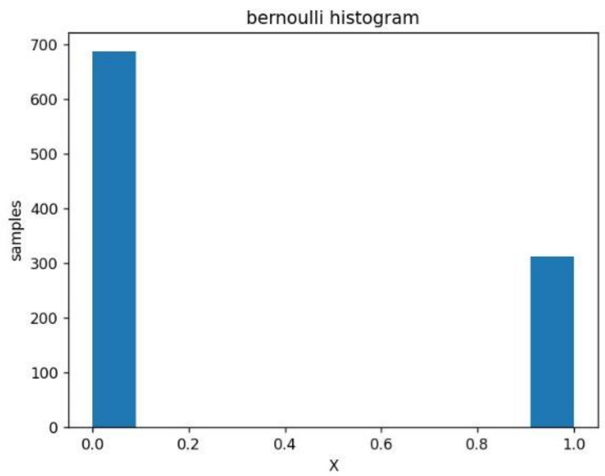
This function visualizes various aspects of the Bernoulli distribution:

- **Random Variable Values (x list):** It creates a list **[0, 1]** to represent the possible outcomes of a Bernoulli distribution.
- **PMF Calculation (pmf list):** It calculates the PMF for each possible value in **x** using the **bernoulli_pmf** function.
- **CDF Calculation (cdf list):** It calculates the CDF for each possible value in **x** using the **bernoulli_cdf** function.
- **Histogram of Random Samples (histogram list):** It generates random samples using the **generate_bernoulli_randvar** function and plots a histogram of the results.
- **Mean and Variance Calculation:** It calculates the mean and variance of the Bernoulli distribution.
- **Graph Plots Using matplotlib:**
    1. **PMF Plot:** It uses **plt. vlines** to plot vertical lines representing the PMF at each point.
    2. **CDF Plot:** It uses **plt. step** to plot the step function representing the CDF.
    3. **Histogram Plot:** It uses **plt. hist** to plot a histogram of the random samples.
- **Display:** It uses **plt. show ()** to display each plot separately.

```
Choose which type of random variables that you will use:
1. Binomial
2. Bernoulli
3. Geometric
4. Poisson
5. Uniform Discrete
6. Continuous Uniform
7. Exponential
8. Gaussian
Enter the corresponding number (1-8): 2
Enter the probability of success (p): 0.3
Enter the number 0 or 1 (n): 1
The PMF of Bernoulli distribution:  0.3
The CDF of Bernoulli distribution:  1
```







NOTE:

- *One of characteristics of Bernoulli distribution is that probabilities are not affected by the outcomes of other trials which means the trials are independent.*
- *Bernoulli distribution is a special case of binomial distribution when only 1 trial is conducted.*

2. Binomial Random Variable

2.1. Definition

The binomial distribution represents the probability for 'x' successes of an experiment in 'n' trials, given a success probability 'p' for each trial at the experiment, where there are two distinct complementary outcomes, a “success” and a “failure”.

2.2. Conditions

All conditions must be satisfied:

1. The experiment consists of an identical trial.
2. Each trial results in one of the two outcomes, called success and failure.
3. The probability of success, denoted p, remains the same from trial to trial.
4. The n trials are independent. That is, the outcome of any trial does not affect the outcome of the others.

2.3. Parameters

- p is the probability of success on an individual trial.
- q is the probability of failure on an individual trial (q=1−p).
- n is the total number of trials.
- k is the number of successes.

2.4. Probability Mass Function (PMF)

The Probability Mass Function (PMF) of a binomial distribution describes the probability of obtaining exactly k successes in an independent Bernoulli trial. The PMF is denoted as P(X=k), where X is the random variable representing the number of successes.

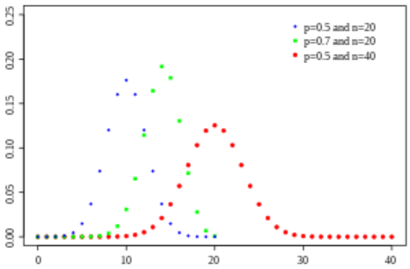
$$P(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

where
n = the number of trials (or the number being sampled)
x = the number of successes desired
p = probability of getting a success in one trial
$q = 1 - p$ = the probability of getting a failure in one trial

2.5. Cumulative Distribution Function (CDF)

The Cumulative Distribution Function (CDF) of a binomial distribution gives the probability that the random variable X, representing the number of successes in an independent Bernoulli trial, is less than or equal to a specific value k. The CDF is denoted as P(X≤k) and is calculated by summing the probabilities from k=0 to k:

$$P(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i q^{n-i}$$



2.6. Mean and Variance

2.6.1. Mean

The probability function for a binomial random variable is b (x; n, p) = n*p. This is the probability of having x successes in a series of n independent trials when the probability of success in any one of the trials is p.

$$\mu = np$$

➢Proof:

$$\mathsf{E}(X) = \sum_{x \in \Omega_X} x \Pr(X = x)$$

Thus:

$$\begin{aligned} \mathsf{E}(X) &= \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k q^{n-k} \\ &= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)} \\ &= np \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j} \\ &= np \end{aligned}$$

2.6.2. Variance

Variance of the binomial distribution is a measure of the dispersion of the probabilities with respect to the mean value. The variance of the binomial distribution is.

$$\sigma^2 = np(1-p)$$

where n is the number of trials, p is the probability of success, and q is the probability of failure.

➢Proof:

Now, $Var(X) = \mathbb{E}[X^2] - [\mathbb{E}[X]]^2$

$\mathbb{E}[X^2] = \sum_{x=0}^{n} x^2 \cdot \binom{n}{x} p^x q^{n-x}$

$= \sum_{x=0}^{n} [x(x-1)+x] \cdot \binom{n}{x} p^x q^{n-x} + \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x q^{n-x}$

$= \sum_{x=2}^{n} \frac{x(x-1)n!}{(n-x)!x(x-1)(x-2)!} p^x q^{n-x} + np$

$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(n-x)!(x-2)!} p^{x-2} q^{n-x}$

$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{[(n-2)-(x-2)]!(x-2)!} p^{x-2} q^{n-x}$

$= n(n-1)p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} q^{n-x}$

$= n(n-1)p^2 [{}^{n-2}C_0 q^{n-2} + {}^{n-2}C_1 p q^{n-3} + {}^{n-2}C_2 p^2 q^{n-4} + \ldots + {}^{n-2}C_{n-2} p^{n-2}] + np$

$= n(n-1)p^2 [(p+q)^{n-2}] + np$

Since p + q = 1, we have

$\mathbb{E}[X^2] = n(n-1)p^2 + np$

Using this,

$Var(X) = n(n-1)p^2 + np - (np)^2$

$= n^2p^2 - np^2 + np - n^2p^2$

$= np(1-p)$

$= npq$

2.7. Applications

1. **Biomedical Research:** Diagnostic Tests: Binomial distribution is used to model the number of correct diagnoses or false positives/negatives in medical tests.
2. **Economics:** Market Share Analysis: It can be used to model the number of successes (e.g., customers choosing a particular brand) in a fixed number of trials.
3. **Marketing:** Response Rates: Marketers often use the binomial distribution to model the number of responses (e.g., clicks on an ad) in a fixed number of advertising exposures.
4. **Educational Testing:** Exam Scores: It can be used to model the number of correct answers in a multiple-choice exam.
5. **Project Management:** Project Success: In project management, the binomial distribution can model the probability of successfully completing a certain number of tasks within a project.

2.8. Real-life Applications

we get this data from research where people aged 30 to 80 faced the possibility of breast cancer at a rate of one in eight, statisticians employed the binomial distribution to unravel the mystery.
The mean revealed that, on average, one person out of eight would experience breast cancer.
Crafting a Probability Mass Function (PMF), the statisticians illustrated the likelihood of different outcomes within the group.

we get this data from research where people aged 30 to 80 faced the possibility of breast cancer at a rate of one in eight, statisticians employed the binomial distribution to unravel the mystery.
The mean revealed that, on average, one person out of eight would experience breast cancer.
Crafting a Probability Mass Function (PMF), the statisticians illustrated the likelihood of different outcomes within the group.

2.9. Code

```python
def binomial_pmf(n,r,p):
    return comb(n,r)*p**r*(1-p)**(n-r)
def binomial(n, p,size):
    pmf = [binomial_pmf(n,i,p) for i in range(n+1)] #calculate the pmf of 0 to n numbers
    x = [i for i in range(n+1)] # list of random variable values from 0 to n
    cdf = [sum(pmf[:i+1])for i in range(n+1)] # calculates the cdf at each point
    # generates the random variables,
    # which are the no. of successes in n bernoulli experiments over size no. of samples.
    histogram = [generate_bernoulli_randvar(n,p).count(1) for _ in range(size)]
    mean = n*p
    variance = n*p*(1-p)
    #plots the pmf, cdf and histogram graphs.
    plt.vlines(x, 0, pmf, lw=20, alpha=1)
    plt.text(0, 0.27, f'E: {mean}\nvar: {variance}', fontsize=12)
    plt.title("binomial PMF")
    plt.xlabel("X")
    plt.ylabel("probability")
    plt.show()
    plt.step(x,cdf, lw=5,where="post")
    plt.title("binomial CDF")
    plt.xlabel("X")
    plt.ylabel("cummulative probability")
    plt.show()
```

```
plt.hist(histogram,bins=n+1,align='mid',edgecolor='black')
plt.title("binomial histogram")
plt.xlabel("X")
plt.ylabel("samples")
plt.show()
```

- **binomial_pmf (n, r, p)**

This function calculates the probability mass function (PMF) of the binomial distribution. The binomial distribution describes the number of successes in a fixed number of independent Bernoulli.

trials, each with the same probability of success.

- **n**: Number of trials.
- **r**: Number of successes.
- **p**: Probability of success in a single trial.

The function uses the binomial coefficient (**comb (n, r)**) from the **scipy. special** module to calculate the number of combinations of **n** items taken **r** at a time. The formula for the binomial PMF is then applied:

$P(X=r) = \binom{n}{r} \cdot p^{r} \cdot (1-p)^{n-r}$

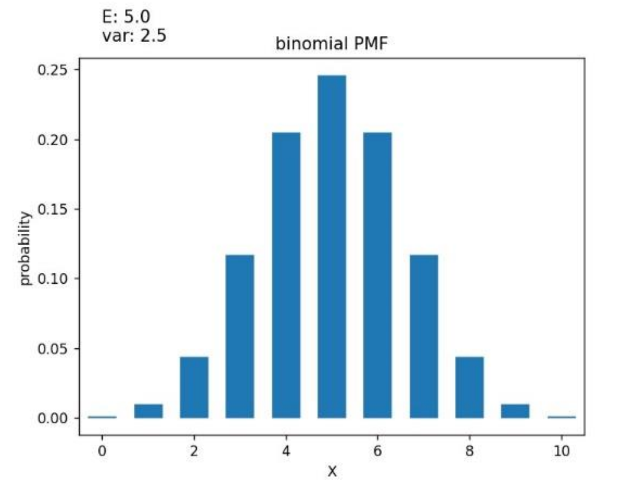
This result is returned as the probability of observing **r** successes in **n** trials.

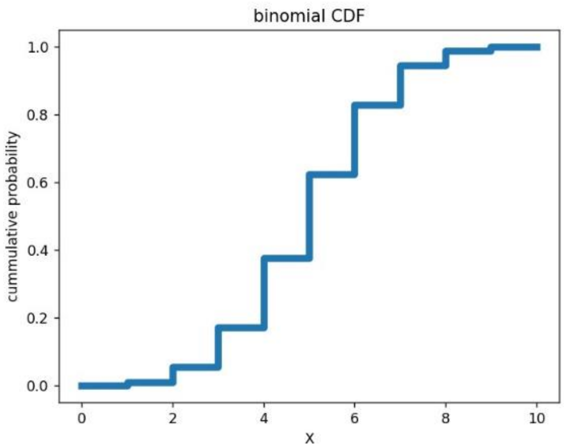
- **binomial (n, p, size)**

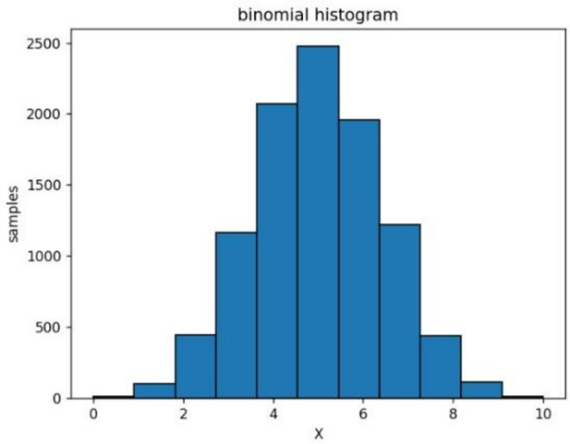
This function generates and visualizes various aspects of the binomial distribution:

- **PMF Calculation (pmf list):** It calculates the PMF for each possible number of successes from 0 to **n** using the **binomial_pmf** function.
- **Random Variable Values (x list):** It creates a list of random variable values from 0 to **n** to represent the possible outcomes.
- **CDF Calculation (cdf list):** It calculates the cumulative distribution function (CDF) at each point by summing up the probabilities in the PMF list up to that point.
- **Histogram of Random Samples (histogram list):** It generates random samples using the **generate_bernoulli_randvar** function (not provided) and counts the number of successes in each sample.
- **Mean and Variance Calculation:** It calculates the mean and variance of the binomial distribution.
- **Graph Plots Using matplotlib:**
    1. **PMF Plot:** It uses **plt. vlines** to plot vertical lines representing the PMF at each point.
    2. **CDF Plot:** It uses **plt. step** to plot the step function representing the CDF.
    3. **Histogram Plot:** It uses **plt. hist** to plot a histogram of the random samples.
- **Display:** It uses **plt. show ()** to display each plot separately.

```
Choose which type of random variables that you will use:
1. Binomial
2. Bernoulli
3. Geometric
4. Poisson
5. Uniform Discrete
6. Continuous Uniform
7. Exponential
8. Gaussian
Enter the corresponding number (1-8): 2
Enter the probability of success (p): 0.3
Enter the number 0 or 1 (n): 1
The PMF of Bernoulli distribution:  0.3
The CDF of Bernoulli distribution:  1
```







3. Geometric Random Variable

3.1. Definition

It variable is a discrete random variable that models the number of trials needed to observe the first success in a sequence of independent Bernoulli trials. Each trial has two possible outcomes: success or failure. The geometric random variable represents the number of trials required until the first success occurs.

3.2. Conditions

- A phenomenon that has a series of trials
- Each trial has only two possible outcomes – either success or failure
- The probability of success is the same for each trial.

3.3. Parameters

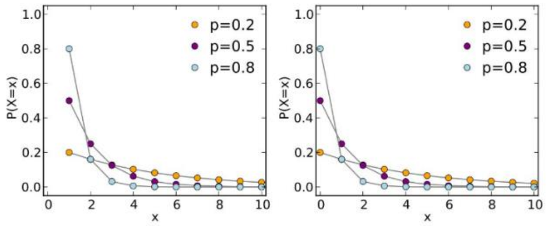
The parameters of a geometric random variable are as follows:
1. Probability of success (p): This parameter represents the probability of success in each trial. It must be a value between 0 and 1, inclusive. It determines the likelihood of observing the first success in each trial.
2. Number of trials (x): This is the random variable itself and represents the number of trials needed to observe the first success. It can take on integer values starting from 1 and so on.

3.4. Probability Mass Function (PMF)

The probability mass function can be defined as the probability that a discrete random variable, X, will be exactly equal to some value, x.

$$P(X = x) = (1 - p)^{x - 1}p$$

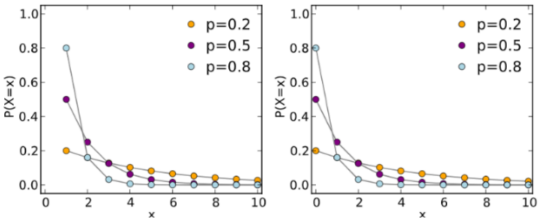
where, $0 < p \leq 1$.



3.5. Cumulative Distribution Function (CDF)

The cumulative distribution function of a random variable, X, that is evaluated at a point, x, can be defined as the probability that X will take a value that is lesser than or equal to x.

$$P(X \le x) = 1 - (1 - p)^x$$



➢Proof:

$$\begin{aligned} F(x) &= P(X \le x) \\ &= 1 - P(X > x) \\ &= 1 - \sum_{w=x+1}^{\infty} f(w) \\ &= 1 - \sum_{w=x+1}^{\infty} p(1-p)^w \\ &= 1 - p(1-p)^{x+1}\left[1 + (1-p) + (1-p)^2 + \cdots\right] \\ &= 1 - p(1-p)^{x+1}\left[\frac{1}{1-(1-p)}\right] \\ &= 1 - (1-p)^{x+1} \qquad x = 0, 1, 2, \ldots \end{aligned}$$

3.6. Mean and Variance

3.6.1. Mean

The expectation or mean of a geometric random variable X is given by:

$$E(X) = 1 / p$$

➢Proof:

Let $q = 1 - p$,

$$\begin{aligned} E[X] &= \sum_{i=1}^{\infty} i q^{i-1} p \\ &= \sum_{i=1}^{\infty} (i-1+1) q^{i-1} p \\ &= \sum_{i=1}^{\infty} (i-1) q^{i-1} p + \sum_{i=1}^{\infty} q^{i-1} p \\ &= \sum_{j=0}^{\infty} j q^{j} p + 1 \\ &= q \sum_{j=1}^{\infty} j q^{j-1} p + 1 \\ &= qE[X] + 1 \end{aligned}$$

$$pE[X] = 1$$

$$E[X] = 1/p$$

3.6.2. Variance

The variance of a geometric random variable X is given by:

$$Var(X) = (1-p)/p^2$$

➢ Proof:

$$\begin{aligned}
\mathrm{Var}[X] &= \mathrm{E}[X^2] - \mathrm{E}[X]^2 \\
&= \frac{2-3p+p^2}{p^2} - \left(\frac{1-p}{p}\right)^2 \\
&= \frac{2-3p+p^2-(1-2p+p^2)}{p^2} \\
&= \frac{2-3p+p^2-1+2p-p^2}{p^2} \\
&= \frac{1-p}{p^2}
\end{aligned}$$

3.7. Applications

1. **Quality Control:** In manufacturing processes, the geometric random variable can be used to analyse the number of defective items produced before the first non-defective item is observed. This information can help identify potential issues in the production process and improve quality control measures.
2. **Finance:** In investment analysis, the geometric random variable can be used to model the number of unsuccessful investments before a successful one is made. This can provide insights into the risk and return characteristics of investment portfolios.
3. **Sports Analytics:** In sports analytics, the geometric random variable can be used to analyse the number of games a team needs to win before winning a championship. This information can help teams strategize and make informed decisions based on their performance in previous games.
4. **Investment Returns:** When analysing the performance of investments over multiple periods, the geometric mean can provide a more accurate representation of the average return. This is because investment returns are often compounded over time, and the geometric mean considers the compounding effect.
5. **Environmental Studies:** In ecology and environmental studies, the geometric mean is used to calculate average growth rates or population changes. This is because population growth tends to follow an exponential pattern, and the geometric mean captures this trend more effectively than the arithmetic mean.

3.8. Code

```
def geometric_pmf(probability,numer_of_trials):
        return ((1-probability)**(numer_of_trials-1))*probability
def geometric_cdf(probability,no_of_trials):
    return 1-((1-probability)**no_of_trials)
def genetrate_geometric_randvar(probability):
    count = 0
    while True:
        result = generate_bernoulli_randvar(1, probability)
        count = count + 1
        if result == [1]:
            break
    return count
```

```python
    pmf = [geometric_pmf(probability, i) for i in x]
    # calculates cdf values for each element in x.
    cdf = [geometric_cdf(probability, i) for i in x]
    #generates size no. of random variables
    histogram = [genetrate_geometric_randvar(probability) for _ in range(size)]
    mean = 1 / probability
    variance = (1 - probability) / (probability ** 2)
    #plots the pmf, cdf and histogram graphs.
    plt.vlines(x, 0, pmf, lw=20, alpha=1)
    plt.text(7, 0.05, f'E: {mean}\nvar: {variance}', fontsize=12)
    plt.title("geometric PMF")
    plt.xlabel("X")
    plt.ylabel("probability")
    plt.show()
    plt.step(x,cdf, lw=3,where='post')
    plt.title("geometric CDF")
    plt.xlabel("X")
    plt.ylabel("cummulative probability")
    plt.show()
    plt.hist(histogram, bins=max(histogram)+1,align="mid",edgecolor='black')
    plt.title("geometric histogram")
    plt.xlabel("X")
    plt.ylabel("samples")
    plt.show()
```

- **geometric_pmf (probability, number_of_trials)**

This function calculates the Probability Mass Function (PMF) of the geometric distribution. The geometric distribution models the number of Bernoulli trials required for the first success.

- **probability**: Probability of success in a single trial.
- **number_of_trials**: The number of trials until the first success.

The PMF for a geometric distribution is given by $(X=k) = (1-p)^{k-1} \cdot p$, where k is the number of trials until the first success.

- **geometric_cdf (probability, number_of_trials)**

This function calculates the Cumulative Distribution Function (CDF) of the geometric distribution. The CDF gives the probability that a geometric random variable is less than or equal to a specified value.

- **probability**: Probability of success in a single trial.
- **number_of_trials**: The number of trials.

The CDF for a geometric distribution is given by $F(X \le k) = 1-(1-p)^{k}$, where k is the number of trials.

- **generate_geometric_randvar(probability)**

This function generates a random variable following a geometric distribution. It simulates Bernoulli trials until the first success occurs. The function returns the count of trials needed to achieve the first success.

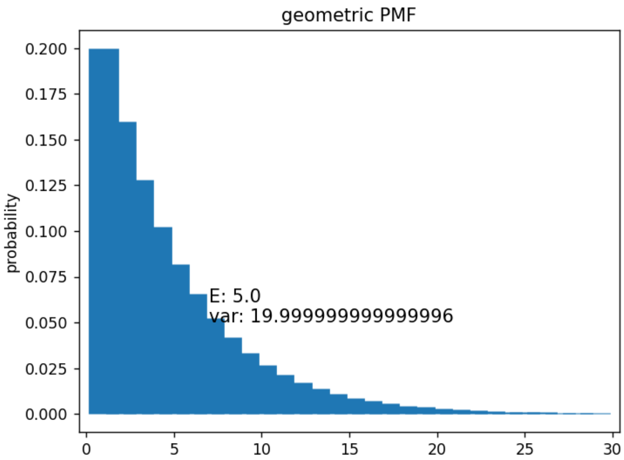
- **geometric (probability, size)**

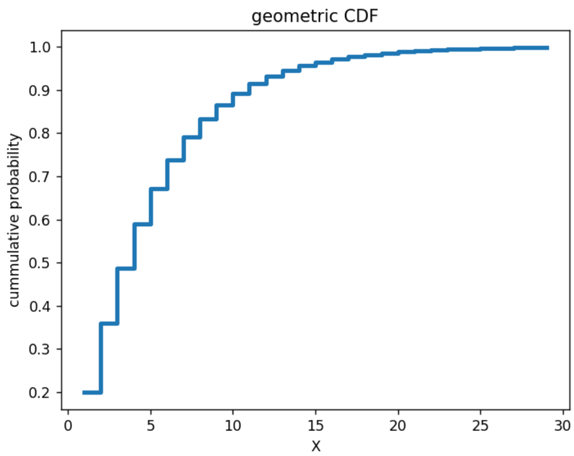
This function visualizes various aspects of the geometric distribution:

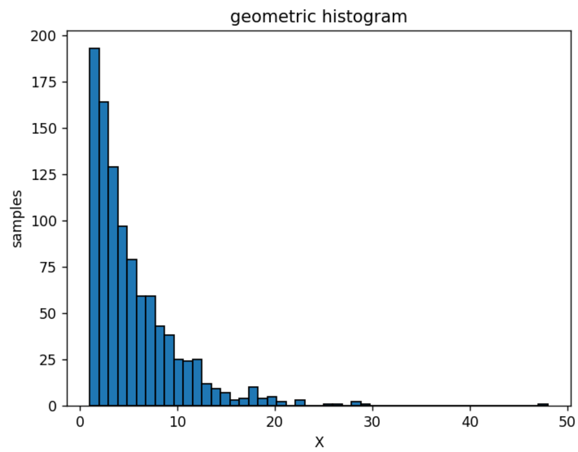
- **Random Variable Values (x list):** It creates a list of random variables from 1 to 29 (sample limit) to represent the possible number of trials until the first success.
- **PMF Calculation (pmf list):** It calculates the PMF for each possible value in **x** using the **geometric_pmf** function.
- **CDF Calculation (cdf list):** It calculates the CDF for each possible value in **x** using the **geometric_cdf** function.

- **Histogram of Random Samples (histogram list):** It generates random samples using the **generate_geometric_randvar** function and plots a histogram of the results.
- **Mean and Variance Calculation:** It calculates the mean and variance of the geometric distribution.
- **Graph Plots Using matplotlib:**
  1. **PMF Plot:** It uses **plt. vlines** to plot vertical lines representing the PMF at each point.
  2. **CDF Plot:** It uses **plt. step** to plot the step function representing the CDF.
  3. **Histogram Plot:** It uses **plt. hist** to plot a histogram of the random samples.
- **Display:** It uses **plt. show ()** to display each plot separately.

```
Choose which type of random variables that you will use:
1. Binomial
2. Bernoulli
3. Geometric
4. Poisson
5. Uniform Discrete
6. Continuous Uniform
7. Exponential
8. Gaussian
Enter the corresponding number (1-8): 3
Enter the probability of success (p): 0.2
Enter the number of trials (n): 10
The PMF of Geomatric distribution:  0.026843545600000015
The CDF of Geomatric distribution:  0.8926258175999999
```







4. Uniform Random Variable

4.1. Definition

A uniform discrete random variable is a type whose outcomes are equally likely over a finite set of values. In other words, each possible outcome has the same probability of occurring. The uniform discrete distribution is also characterized by its constant probability mass function.

4.2. Characteristics

- Finite Set of Values: The outcomes of a uniform discrete random variable must be finite and well-defined as it takes on values from a specific, finite set.
- The probability mass function is constant for all possible values.
- Mutual Exclusivity: This means that only one outcome can occur at a time.

- Collectively Exhaustive: The set of possible outcomes must be collectively exhaustive, meaning that one of the outcomes must occur. The sum of the probabilities of all possible outcomes is 1.

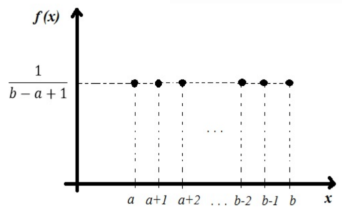
4.3. Parameters

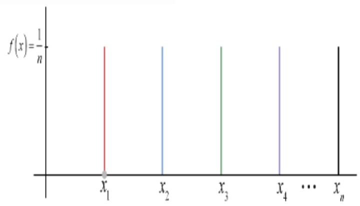
- Lower Bound (a): The smallest value in the set of possible outcomes. The uniform distribution is defined over the integers starting from this lower bound.
- Upper Bound (b): The largest value in the set of possible outcomes. The uniform distribution is defined over the integers up to and including this upper bound.

4.4. Probability Mass Function (PMF)

The Probability mass function of a uniform discrete random variable is a constant for each possible outcome, for each possible outcome xi, the probability of X taking on the value xi is 1/n. The PMF ensures that each outcome has an equal probability of occurring, which is a special characteristic of a uniform discrete random variable.

$$f_X(x) = \frac{1}{b-a+1} \quad \text{where} \quad x \in \{a, a+1, \dots, b-1, b\} \ .$$





➢Proof:

The probability mass function (PMF) of a uniform discrete random variable is:

$$P(X = x_i) = \frac{1}{k}$$

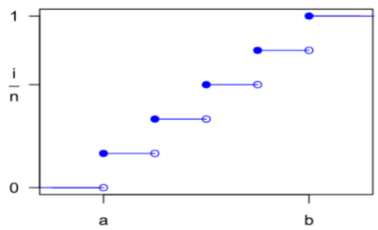
for $i = 1, 2, \dots, k$, where k is the number of distinct values in the set.

$$\sum_{i=1}^{k} P(X = x_i) = \sum_{i=1}^{k} \frac{1}{k} = 1$$

4.5. Cumulative Distribution Function (CDF)

The cumulative distribution function (CDF) of a discrete uniform random variable is a step function that increases by a constant amount at each possible outcome. where each outcome has an equal probability, and it is given by:

$$F_X(x) = \begin{cases} 0 \ , & \text{if } x < a \\ \frac{\lfloor x \rfloor - a + 1}{b - a + 1} \ , & \text{if } a \le x \le b \\ 1 \ , & \text{if } x > b \ . \end{cases}$$



➢Proof

$$F(x) = P(X \le x) = \sum_{k=a}^{x} P(X = k)$$

$$F(x) = \sum_{k=a}^{x} \frac{1}{b-a+1}$$

$$F(x) = (x - a + 1) \cdot \frac{1}{b-a+1}$$

$$F(x) = \frac{x-a+1}{b-a+1}$$

$$F_X(x) = \begin{cases} 0 \ , & \text{if } x < a \\ \frac{\lfloor x \rfloor - a + 1}{b - a + 1} \ , & \text{if } a \le x \le b \\ 1 \ , & \text{if } x > b \ . \end{cases}$$

4.6. Mean and Variance

4.6.1. Mean

The mean or expected value of a uniform random variable defined over a<x<b is:

$$\mu = E(X) = \frac{a+b}{2}$$

➢Proof:

$$\begin{aligned} E(Y) &= E(kX + (a-k)) \\ &= kE(X) + (a-k) \\ &= k\left(\frac{N+1}{2}\right) + (a-k) \\ &= \frac{k}{2}\left(\frac{b-a+k}{k} + 1\right) + a - k \\ &= \frac{a+b}{2} \end{aligned}$$

4.6.2. Variance

The variance σ2 for a discrete uniform distribution is given by:

$$\sigma^2 = \frac{(b-a+1)^2 - 1}{12} \quad \text{which is equal to} \quad \boxed{\frac{n^2-1}{12}}$$

➢Proof:

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}(X-\mu)^2 \\ &= \mathbb{E}(X^2) - \mu^2 \\ &= \frac{1}{b-a+1}\sum_{x=a}^{b} x^2 - \frac{(a+b)^2}{4} \\ &= \frac{1}{b-a+1}\left(\frac{(b)(b+1)(2b+1) - (a-1)((a-1)+1)(2(a-1)+1)}{6}\right) - \frac{(a+b)^2}{4} \\ &= \frac{1}{b-a+1}\left(\frac{(b)(b+1)(2b+1) - (a-1)(a)(2a-1)}{6}\right) - \frac{(a+b)^2}{4} \\ &= \frac{1}{b-a+1}\left(\frac{(2b^3+3b^2+b) - (2a^3-3a^2+a)}{6}\right) - \frac{(a+b)^2}{4} \\ &= \frac{1}{b-a+1}\frac{(4b^3+6b^2+2b) - (4a^3-6a^2+2a) - (3a^2+6ab+3b^2)}{12} \\ &= \frac{1}{b-a+1}\frac{4b^3+3b^2+2b-4a^3+3a^2-2a-6ab}{12} \\ &= \ldots \\ &= \frac{(b-a+1)^2-1}{12} \end{aligned}$$

4.7. Applications

1. **Rolling a Fair Die:** When you roll a fair six-sided die, the outcome can be any of the numbers 1, 2, 3, 4, 5, or with equal probability.
2. **Choosing a Random Card from a Well-Shuffled Deck:** If you have a well-shuffled standard deck of 52 playing cards, and you randomly select one card, each of the 52 cards has an equal probability of being chosen.
3. **Selecting a Random Day of the Week:** If you randomly select a day of the week, Monday, Tuesday, Wednesday, and so on, each has an equal probability of being chosen.
4. **Picking a Number from a Hat:** If you write numbers 1 to 10 on pieces of paper, place them in a hat, and randomly pick one, the probability of selecting any specific number is the same for each.
5. **Queueing Theory:** In queueing models, if customers arrive at a service point randomly and independently with a constant arrival rate, the number of arrivals in each time may follow a discrete uniform distribution.

4.8. code

```
def discrete_uniform_pmf(upper_bound, lower_bound, point):
    if upper_bound>=point>=lower_bound:
        return 1/(upper_bound-lower_bound+1)
    else:
        return 0
def discrete_uniform_cdf(lower_bound,upper_bound,point):
    if lower_bound <= point <= upper_bound:
        return (point - lower_bound+1) / (upper_bound - lower_bound+1)
    elif point > upper_bound:
        return 1
    else:
        return 0
def generate_uniform_randvar(a,b,size):
    #chooses a piont randomly size number of times and store it in a list
    return [random.randrange(a,b+1) for _ in range(size)]
def uniform_discrete(a,b,size):
    x = [i for i in range(a,b+1)] #each point within the given interval
    # sets the probability and duplicate it for the sixe of the interval.
    pmf = [discrete_uniform_pmf(b,a,i) for i in x]
    # since all the pmf values are equal all
    # we need is to multiply the probability by itt's index +1 ie:1,2,3...
    cdf = [discrete_uniform_cdf(a,b,i) for i in x]
    # generates all the random variables needed for the histogram.
    histogram = generate_uniform_randvar(a,b,size)
    mean = (b-a)/2
    variance = round((((b-a+1)**2)-1)/12,1)
    # calculates pmf values for each element in x.
    plt.vlines(x, 0, pmf, lw=20, alpha=1)
    plt.text(a, 1/(b-a), f'E: {mean}\nvar: {variance}', fontsize=12)
    plt.title("uniform PMF")
    plt.xlabel("n")
    plt.ylabel("probability")
    plt.show()
    plt.step(x,cdf, lw=5)
    plt.title("uniform CDF")
    plt.xlabel("X")
    plt.ylabel("probability")
    plt.show()
    plt.hist(histogram, bins=len(x),edgecolor='black',align='mid')
    plt.xticks(x)
    plt.title("uniform histogram")
    plt.xlabel("X")
    plt.ylabel("samples")
    plt.show()
```

- **discrete_uniform_pmf (upper_bound, lower_bound, point)**

This function calculates the Probability Mass Function (PMF) for a discrete uniform distribution. The discrete uniform distribution assumes that each outcome in the specified range has an equal probability.

- **upper_bound**: The upper bound of the distribution.
- **lower_bound**: The lower bound of the distribution.
- **point**: The specific point for which the PMF is calculated.

The function checks whether the given **point** is within the specified range. If it is, the PMF is calculated using the formula $P(X=k) = 1/(b-a+1)$, and if not, the PMF is 0.

- **discrete_uniform_cdf (lower_bound, upper_bound, point)**

This function calculates the Cumulative Distribution Function (CDF) for a discrete uniform distribution. The CDF gives the probability that a random variable is less than or equal to a specified value.

- **lower_bound**: The lower bound of the distribution.
- **upper_bound**: The upper bound of the distribution.
- **point**: The specific point for which the CDF is calculated.

The function checks whether the given **point** is within the specified range. If it is, the CDF is calculated using the formula $F(X\leq k) = (k-a+1)/(b-a+1)$. If the point is greater than the upper bound, the CDF is 1. If the point is outside the range, the CDF is 0.

- **generate_uniform_randvar (a, b, size)**

This function generates random variables following a discrete uniform distribution within the specified range.

- **a**: The lower bound of the distribution.
- **b**: The upper bound of the distribution.
- **size**: The number of random variables to generate.

The function uses **random. randrange (a, b+1)** to randomly choose a value within the specified range and repeat this process **size** times, storing the results in a list.

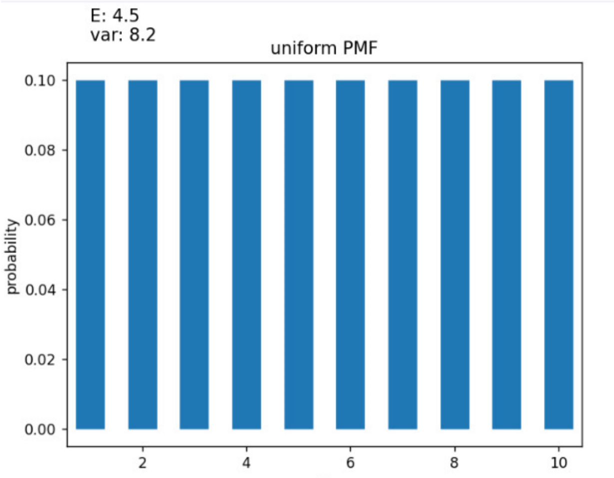
- **uniform_discrete (a, b, size)**

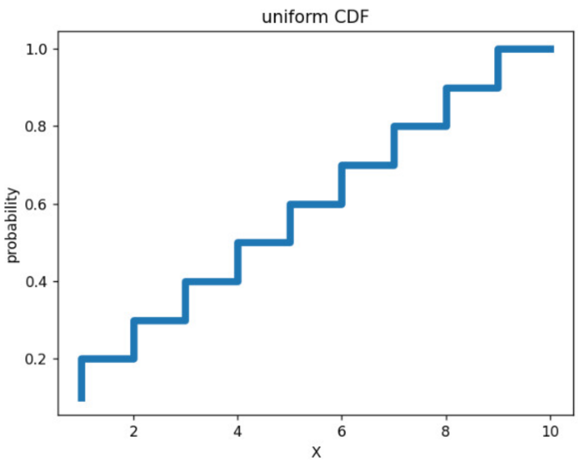
This function generates and visualizes various aspects of the discrete uniform distribution:

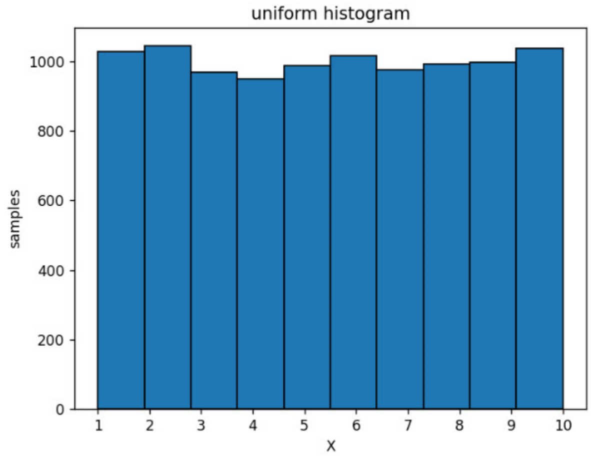
- **Random Variable Values (x list):** It creates a list of values within the given interval.
- **PMF Calculation (pmf list):** It calculates the PMF for each value in **x** using the **discrete_uniform_pmf** function.
- **CDF Calculation (cdf list):** It calculates the CDF for each value in **x** using the **discrete_uniform_cdf** function.
- **Histogram of Random Samples (histogram list):** It generates random samples using the **generate_uniform_randvar** function and plots a histogram of the results.
- **Mean and Variance Calculation:** It calculates the mean and variance of the discrete uniform distribution.

- **Graph Plots Using matplotlib:**
  1. **PMF Plot:** It uses **plt. vlines** to plot vertical lines representing the PMF at each point.
  2. **CDF Plot:** It uses **plt. step** to plot the step function representing the CDF.
  3. **Histogram Plot:** It uses **plt. hist** to plot a histogram of the random samples.
- **Display:** It uses **plt. show ()** to display each plot separately.

```
Choose which type of random variables that you will use:
1. Binomial
2. Bernoulli
3. Geometric
4. Poisson
5. Uniform Discrete
6. Continuous Uniform
7. Exponential
8. Gaussian
Enter the corresponding number (1-8): 5
Enter lower bound for Uniform distribution: 1
Enter upper bound for Uniform distribution: 10
Enter a point for the distribution: 5
The PMF of discrete uniform distribution:  0
The CDF of discrete uniform distribution:  0.5
```







5. Poisson Random Variable

5.1. Definition

A Poisson random variable is a discrete probability distribution that stands for the number of events that occur within a fixed interval of time or space with a known constant mean rate and independently of the time since the last event. It is a limited form of binomial distribution.

5.2. Conditions

- Intervals must be fixed and constant.
- Independence: The occurrence of one event should not affect the occurrence of another. events are assumed to be independent of each other.
- Constant Rate: The average rate of events (λ) is assumed to be constant across the entire interval. This means that the probability of an event occurring in an exceedingly small subinterval is proportional to the length of that subinterval.
- Discreteness: The number of events in each interval is a non-negative integer.
- Rare Events: The probability of more than one event occurring in an infinitesimally small interval is negligible. This is often expressed mathematically as $\lim_{\Delta t \to 0} P(X>1) = 0$.

5.3. Parameters

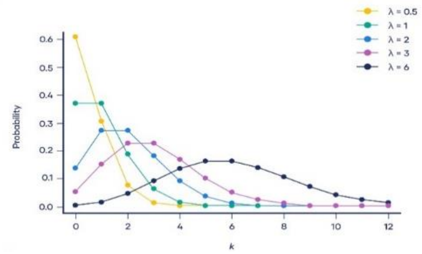
- -The Poisson distribution is characterized by a single parameter denoted as λ (lambda).
- -This parameter is the average rate of events occurring in a fixed interval of time or space

- -The value of λ is crucial in defining the distribution and deciding the likelihood of seeing a specific number of events.

5.4. Probability Mass Function (PMF)

This function is used for discrete random variables. The PMF gives the probability of the random variable taking on a specific value. The sum of the probabilities over all possible values is equal to 1. where λ > 0 (average rate of events), k = 1, 2, . . .

$$P(X=k)=\frac{e^{-\lambda}.\lambda^{k}}{k!}$$



The Poisson distribution has only one parameter, λ (lambda), which is the mean number of events. The graph above shows examples of Poisson distributions with different values of λ.

➢Proof:

$$P_\lambda(x)=\frac{\lambda^x e^{-\lambda}}{x!}$$

$$P_\lambda(t,x)=\frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

$$\lambda t=np \Rightarrow \lambda=\frac{np}{t}$$

$$\Delta n=\frac{t}{\Delta t}\cdot n$$

$$P(x,t+dt)=P(x,t)P(0,dt)+P(x-1,t)P(1,dt)$$

$$P(x,t+dt)=P(x,t)(1-\lambda dt)+P(x-1,t)\lambda dt$$

$$\frac{dP(x,t)}{dt}+\lambda P(x,t)=\lambda P(x-1,t)$$

$$e^{\lambda t}\frac{dP(x,t)}{dt}+e^{\lambda t}\lambda P(x,t)=e^{\lambda t}\lambda P(x-1,t)$$

$$\frac{d}{dt}[e^{\lambda t}P(x,t)]=e^{\lambda t}\lambda P(x-1,t)$$

$$\frac{d}{dt}[e^{\lambda t}P(1,t)]=e^{\lambda t}\lambda P(0,t)=e^{\lambda t}\lambda e^{-\lambda t}=\lambda$$

$$e^{\lambda t}P(1,t)=\int \lambda dt=\lambda t+C=\lambda t$$

$$P(1,t)=\lambda t e^{-\lambda t}$$

$$P(x,t)=\frac{(\lambda t)^x}{x!}e^{-\lambda t}$$

5.5. Cumulative Distribution Function (CDF)

The cumulative distribution function (CDF) of a Poisson random variable X is the probability that the random variable takes a value less than or equal to a specified value k. The CDF is denoted by FX (k) and is given by the sum of the probabilities from 0 to k using the Poisson probability mass function (PMF).

$$P(X\le k)=\sum_{i=0}^{k}\frac{e^{-\lambda}.\lambda^{i}}{i!}$$

➢Proof:

$$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$$

$$P(X\le k)=\sum_{i=0}^{k}P(X=i)$$

$$P(X\le k)=\sum_{i=0}^{k}\frac{e^{-\lambda}\lambda^i}{i!}$$

5.6. Mean and Variance

5.6.1. Variance

The mean of a Poisson distribution is denoted by λ, which stands for the average rate of events occurring in a fixed interval of time or space.

$$\mu = \lambda$$

➢ Proof:

$$\begin{aligned}
&\mathrm{E}[X] \\
&= \sum_{x \in R_X} x p_X(x) \\
&= \sum_{x=0}^{\infty} x \exp(-\lambda) \frac{1}{x!} \lambda^x \\
&= 0 + \sum_{x=1}^{\infty} x \exp(-\lambda) \frac{1}{x!} \lambda^x && \text{(the first term of the sum is zero since } x = 0\text{)} \\
&= \sum_{y=0}^{\infty} (y+1) \exp(-\lambda) \frac{1}{(y+1)!} \lambda^{y+1} && \text{(by changing variable: } y = x - 1\text{)} \\
&= \sum_{y=0}^{\infty} (y+1) \exp(-\lambda) \frac{1}{(y+1)y!} \lambda \lambda^{y} && \text{(since } (y+1)! = (y+1)y!\text{)} \\
&= \lambda \sum_{y=0}^{\infty} \exp(-\lambda) \frac{1}{y!} \lambda^{y} \\
&= \lambda \sum_{y=0}^{\infty} p_Y(y) && (p_Y \text{ is the pmf of a Poisson r.v. with parameter } \lambda) \\
&= \lambda && \text{(the sum of a pmf over its support is 1)}
\end{aligned}$$

5.6.2. Variance

Variance (σ 2): The variance is a measure of how much the values of a random variable deviate from its mean. It is calculated as the average of the squared differences between each value and the mean.

$$\sigma^2 = \lambda$$

➢ Proof:

$$\begin{aligned}
&= \lambda \left\{ \sum_{y=0}^{\infty} y p_Y(y) + \sum_{y=0}^{\infty} p_Y(y) \right\} \\
&= \lambda \{ \mathrm{E}[Y] + 1 \} && \text{(the sum of a pmf over its support is 1)} \\
&= \lambda \{ \lambda + 1 \} && \text{(expected value of a Poisson r.v. with parameter } \lambda \text{ is } \lambda\text{)} \\
&= \lambda^2 + \lambda \\
&\mathrm{E}[X]^2 = \lambda^2 \\
&\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda \\
&= \sum_{y=0} (y+1)^2 \exp(-\lambda) \frac{1}{(y+1)!} \lambda^{y+1} && \text{(by changing variable: } y = x - 1\text{)} \\
&= \sum_{y=0}^{\infty} (y+1)^2 \exp(-\lambda) \frac{1}{(y+1)y!} \lambda \lambda^{y} && \text{(since } (y+1)! = (y+1)y!\text{)} \\
&= \lambda \sum_{y=0}^{\infty} (y+1) \exp(-\lambda) \frac{1}{y!} \lambda^{y} \\
&= \lambda \sum_{y=0}^{\infty} (y+1) p_Y(y) && (p_Y \text{ is the pmf of a Poisson r.v. with parameter } \lambda)
\end{aligned}$$

5.7. Applications

1. **Traffic Engineering:** Predicting the number of vehicles passing through a toll booth or a traffic signal in a specific period.
2. **Insurance:** Estimating the number of insurance claims filed within a certain period, assuming these events are rare.

3. **Medicine:** Examining the occurrence of rare diseases or adverse reactions to a medication in a clinical trial.
4. **Physics:** Describing the distribution of particle hits on a detector in particle physics experiments.
5. **Environmental Science:** Modeling the number of earthquakes in a region or the number of forest fires in each area.

5.8. Real-life Applications

There there was a fabric producer in the busy town of Textilia who was well-known for creating rolls of beautiful fabrics. Their primary focus was on fabric quality control; thus, they were quite interested in figuring out how likely it was that their rolls would have flaws.
Stella was a dedicated statistician who lived in the center of the manufacturing plant. Stella was given a challenge one day by the production head: figure out how likely it was that she would discover precisely ten flaws in two rolls of cloth that were consecutive. Equipped with her skills in statistics, Stella explored the field of Poisson distribution.
calculating the probability of of finding a certain number of defects can help manufacturers determine the extra fabric they may need to counteract the defects.

Assume the average number of defects in a fabric roll = 6 (lambda)
Let's say a manufacturer wants to calculate the probability of finding 10 defects in two rolls.
Then the new lambda equals 2*6 = 12 defects
Then according to the PMF formula of poission random variable: lambda^k*e^-lambda / k!
The probability of finding 8 defects = 12^8 * e^-12 / 12! = 0.105 (10.5%)

5.9. Code

```
def poisson_pmf(lambda_,k):
    return (math.exp(-lambda_) * (lambda_**k) / math.factorial(k))
def poisson_cdf(lambda_, k):
    # calculate the pmf values up to k and sum them.
    return sum(poisson_pmf(lambda_,i) for i in range(k+1))
def inverse_transform_sampling_poisson(lambda_):
    u = random.uniform(0, 1)
    k = 0
    cumulative_prob = poisson_cdf(lambda_, k)
    while u > cumulative_prob:
        k += 1
        cumulative_prob += poisson_pmf(lambda_, k)
    return k
def generate_poisson_randvar(lambda_,size):
    return [inverse_transform_sampling_poisson(lambda_) for _ in range(size)]
def poisson(lambda_,lower_bound,upper_bound,size):
    #create list of x values max 20.
    x = [i for i in range(lower_bound,upper_bound+1)]
    #calculate the pmf value for each x value.
    pmf = [poisson_pmf(lambda_,i) for i in x]
    #calculate the cdf at each x value.
    cdf = [poisson_cdf(lambda_,i) for i in x]
    #calculate mean and variance.
    mean = variance = lambda
```

```
    #generate the random variables for the histogram.
    histogram = generate_poisson_randvar(lambda_,size)
    #plots the pmf, cdf and histogram graphs.
    plt.vlines(x, 0, pmf, lw=12)
    plt.text(10, 0.1, f'E: {mean}\nvar: {variance}', fontsize=12)
    plt.title("poisson PMF")
    plt.xlabel("X")
    plt.ylabel("probability")
    plt.show()
    plt.step(x,cdf, lw=5,where="post")
    plt.title("poisson CDF")
    plt.xlabel("X")
    plt.ylabel("probability")
    plt.show()
    plt.hist(histogram, bins=max(histogram)+1,edgecolor='black')
    plt.title("poisson histogram")
    plt.xlabel("X")
    plt.ylabel("samples")
    plt.show()
```

- **poisson_pmf (lambda_, k)**

This function calculates the Probability Mass Function (PMF) for the Poisson distribution, which models the number of events occurring in a fixed interval of time or space.

- **lambda_**: The average rate of events per unit interval.
- **k**: The specific number of events.

The PMF for a Poisson distribution is given by the formula $P(X=k) = (e^{-\lambda} \lambda^{k})/k!$ It returns the probability of observing **k** events in a given interval.

- **poisson_cdf (lambda_, k)**

This function calculates the Cumulative Distribution Function (CDF) for the Poisson distribution.

- **lambda_**: The average rate of events per unit interval.
- **k**: The specific number of events.

The CDF for a Poisson distribution is calculated by summing the PMF values up to the specified **k**. It uses the **poisson_pmf** function to calculate individual PMF values for each **i** from 0 to **k**.

- **inverse_transform_sampling_poisson(lambda_)**

This function performs inverse transform sampling to generate a random variable following a Poisson distribution.

- **lambda_**: The average rate of events per unit interval.

It generates a random variable using the inverse transform method. It repeatedly increments the count **k** and calculates the cumulative probability until it exceeds a randomly generated value **u**. The final count **k** is then returned.

- **generate_poisson_randvar (lambda_, size)**

This function generates a list of random variables following a Poisson distribution using the inverse transform sampling method.

- **lambda_**: The average rate of events per unit interval.
- **size**: The number of random variables to generate.

It calls **inverse_transform_sampling_poisson size** times and stores the results in a list.

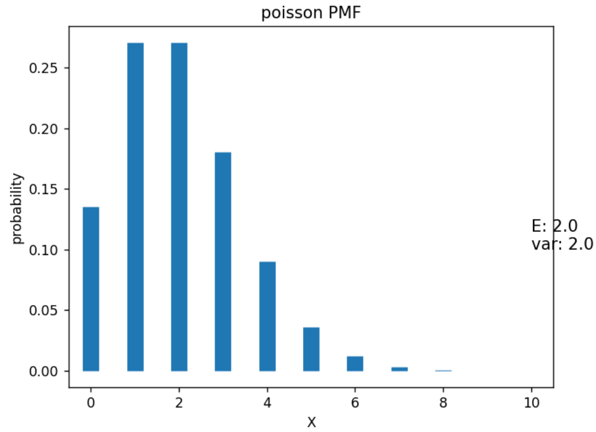
- **poisson (lambda_, lower_bound, upper_bound, size)**

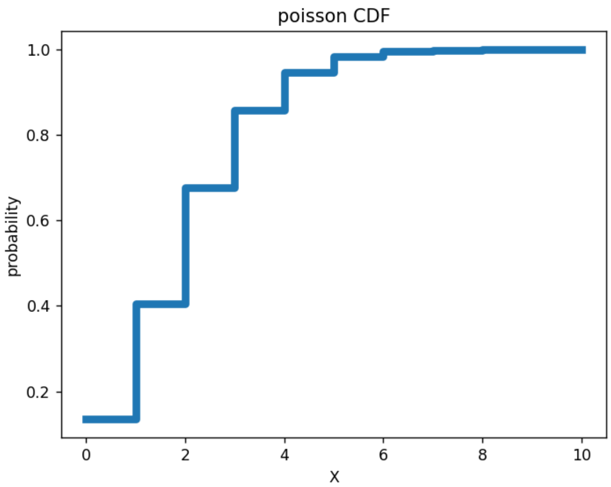
This function generates and visualizes various aspects of the Poisson distribution:

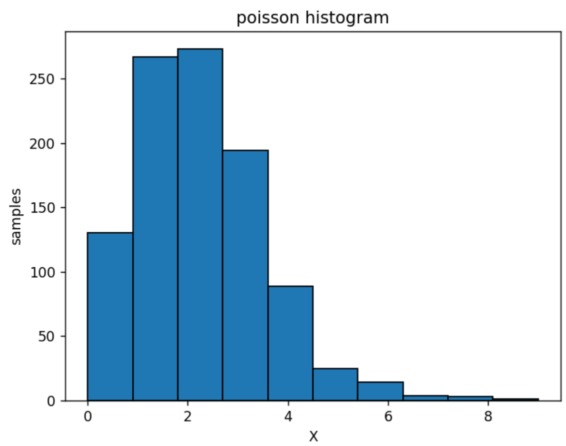
- **Random Variable Values (x list):** It creates a list of values within the specified interval.

- **PMF Calculation (pmf list):** It calculates the PMF for each value in **x** using the **poisson_pmf** function.
- **CDF Calculation (cdf list):** It calculates the CDF for each value in **x** using the **poisson_cdf** function.
- **Mean and Variance Calculation:** Both the mean and variance of a Poisson distribution are equal to **lambda_**.
- **Histogram of Random Samples (histogram list):** It generates random samples using the **generate_poisson_randvar** function and plots a histogram of the results.
- **Graph Plots Using matplotlib:**
  1. **PMF Plot:** It uses **plt. vlines** to plot vertical lines representing the PMF at each point.
  2. **CDF Plot:** It uses **plt. step** to plot the step function representing the CDF.
  3. **Histogram Plot:** It uses **plt. hist** to plot a histogram of the random samples.
- **Display:** It uses **plt. show ()** to display each plot separately.

```
Choose which type of random variables that you will use:
1. Binomial
2. Bernoulli
3. Geometric
4. Poisson
5. Uniform Discrete
6. Continuous Uniform
7. Exponential
8. Gaussian
Enter the corresponding number (1-8): 4
Enter lambda for Poisson distribution: 2
Enter the number of events for Poisson distribution: 5
Enter lower bound for Uniform distribution: 0
Enter upper bound for Uniform distribution: 10
The PMF of Poisson distribution:  0.03608940886309672
The CDF of Poisson distribution:  0.9834363915193857
```







NOTE:

- *The Poisson approximation to binomial: The Poisson approximation is a mathematical approximation, and it is a key in the limit theorem that is often used when certain conditions are met within the context of the binomial distribution.*
- *The conditions for a valid Poisson approximation are:*
  Small Probability of Success (p): The probability of success in each trial (p) should be small. This ensures that the events are rare.
  Large Number of Trials (n)
- *Rule:*
  $(n/k).p^k (1-p)^{(n-k)} \approx (\lambda^k/k!).\ e^{-\lambda}$, *where* $\lambda = np$.

- *Relation between Poisson and binomial distribuation*

  In some situations, a binomial distribution approaches a Poisson distribution as the number of trials increases and the success probability decreases.

  - *Conditions:*

    1. Rare occurrences: When modeling rare occurrences, when the average rate of recurrence (λ) is low, the Poisson distribution is frequently employed. "Rare events" in the context of the binomial distribution suggest a small probability of success (p) in each trial.
    2. Large Number of Trials: The binomial distribution (n) has a large number of trials.
    3. Small Success Probability: There is a low chance of success (p) for every trial.
    The product (np) stays small or in the moderate

  - *Relationship:*

    Take into consideration a binomial distribution with the following parameters: (np = λ) = number of trials (n) and (p) = probability of success in each trial. The binomial distribution approaches a Poisson distribution with parameter λ as (n) grows large and (p) decreases while (np) stays constant.

Continuous Random Variables

1. Uniform Random Variable

1.1. Definition

A continuous uniform distribution is a type of symmetric probability distribution that describes an experiment in which the outcomes of the random variable have equally likely probabilities of occurring within an interval [a, b].

1.2. Conditions

- Uniformity: The probability density function is constant within the specified interval.
- Normalization: The total area under the probability density function over the entire range is equal to 1. This ensures that the probabilities sum to 1.

1.3. Parameters

- a and b are known as the parameters of continuous uniform distribution. We cannot have an outcome of either less than a or greater than b
- the random variable X takes values between a (lower limit) and b (upper limit).

1.4. Probability Density Function (PDF)

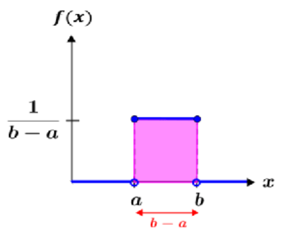
The Probability Density Function (PDF) quantifies the likelihood of values for a continuous random variable. In a uniform distribution, it stays constant over an interval, signifying equal probability. within that range.

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b, \\ 0, & \text{otherwise}, \end{cases}$$

where $[a, b]$ is the interval on which X is defined. We write

$$X \sim \text{Uniform}(a, b)$$

to say that X is drawn from a uniform distribution on an interval $[a, b]$.



➢Proof:

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_a^b \frac{1}{b-a}\,dx$$

$$\int_a^b \frac{1}{b-a}\,dx = \frac{1}{b-a}\int_a^b dx$$

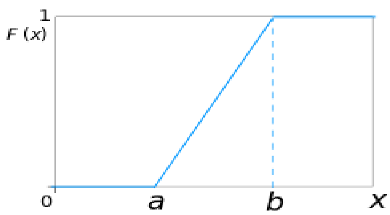
$$\frac{1}{b-a}[x]_a^b = \frac{1}{b-a}\cdot(b-a) = 1$$

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

1.5. Cumulative Distribution Function (CDF)

The Cumulative Distribution Function (CDF) gives the probability that a random variable is less than or equal to a specific value. It supplies a cumulative view of the likelihood of outcomes in a continuous or discrete probability distribution.

$$F(x) = \begin{cases} 0 & \text{for } x < a, \\ \frac{x-a}{b-a} & \text{for } a \le x \le b, \\ 1 & \text{for } x > b. \end{cases}$$



➢Proof:

$$F(x) = \int_a^x f(t)\,dt$$

$$F(x) = \int_a^x \frac{1}{b-a}\,dt$$

$$F(x) = \left[\frac{t}{b-a}\right]_a^x$$

$$F(x) = \frac{x-a}{b-a} - \frac{a-a}{b-a}$$

$$F(x) = \frac{x-a}{b-a}$$

1.6. Mean and Variance

1.6.1. Mean

The mean, or expected value, of a random variable represents its average value. It's a measure central to probability distributions, indicating the balance point or centre of the distribution.

$$\mu = E(x) = \frac{a+b}{2}$$

➢Proof:

$$\begin{aligned} \mathrm{E}(X) &= \int_a^b x \cdot \frac{1}{b-a}\,\mathrm{d}x \\ &= \left[\frac{1}{2}\frac{x^2}{b-a}\right]_a^b \\ &= \frac{1}{2}\frac{b^2-a^2}{b-a} \\ &= \frac{1}{2}\frac{(b+a)(b-a)}{b-a} \\ &= \frac{1}{2}(a+b)\,. \end{aligned}$$

1.6.2. Variance

Variance quantifies the spread or dispersion of a random variable's values around its mean. It provides insht into the variability within a probability distribution.

$$\sigma^2 = \frac{(b-a)^2}{12}$$

➢Proof:

$$\sigma^2 = \int_a^b (x-\mu)^2 \cdot f(x)\,dx$$

Substitute the expressions for μ and $f(x)$:

$$\sigma^2 = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \cdot \frac{1}{b-a}\,dx$$

Expand and simplify:

$$\sigma^2 = \frac{1}{b-a}\int_a^b \left(x^2 - 2x\frac{a+b}{2} + \left(\frac{a+b}{2}\right)^2\right)\,dx$$

Use the formulas for the sum of consecutive integers and the sum of consecutive squares:

$$\sigma^2 = \frac{1}{b-a} \cdot \frac{(b-a)^3}{3}$$

Simplify:

$$\sigma^2 = \frac{(b-a)^2}{12}$$

1.7. Applications

1. **Waiting Time at a Bus Stop:** Suppose the time it takes for a bus to arrive at a bus stop follows a continuous uniform distribution between 0 and 10 minutes. Each minute within that range is equally likely, and the distribution is uniform.
2. **Height of a Perfectly Mixed Sample:** In chemistry, when a sample is perfectly mixed, the concentration of a substance in the sample may follow a continuous uniform distribution over a certain range.
3. **Arrival Time of Customers at a Service Centre:** The time at which customers arrive at a service centre during business hours may follow a continuous uniform distribution, assuming they are equally likely to arrive at any moment within the operating hours.
4. **Lifetime of a Battery:** The lifetime of a certain type of battery might be modelled using a continuous uniform distribution if all possible lifetimes within a specified range are equally likely.

1.8. Code

```
    # calculates the pdf value for each x value.
    cdf = [continous_uniform_cdf(lower_bound,upper_bound,i) for i in x]
    mean = (upper_bound - lower_bound) / 2
    variance = round((((upper_bound - lower_bound + 1) ** 2) - 1) / 12, 1)
    #plots the PDF, CDF and histogram.
    plt.plot(x, pdf, label='PDF')
    plt.text(lower_bound, 1 / (upper_bound - lower_bound), f'E: {mean}\nvar: {variance}', fontsize=12)
    plt.title('Continuous Uniform PDF')
    plt.xlabel('Value')
    plt.ylabel('PDF')
    plt.legend()
    plt.show()
    plt.plot(x, cdf, label='CDF')
    plt.title('Continuous Uniform CDF')
    plt.xlabel('Value')
    plt.ylabel('CDF')
    plt.legend()
    plt.show()
    plt.hist(histogram, bins='auto', edgecolor='black',align='mid')
    plt.title("uniform histogram")
    plt.xlabel("samples")
    plt.ylabel("probability")
    plt.show()
```

➢ **continuous_uniform_pmf (upper_bound, lower_bound, point)**

This function calculates the Probability Density Function (PDF) for a continuous uniform distribution. It determines the likelihood of a random variable taking a specific value within a given range.

- **upper_bound**: The upper bound of the distribution.
- **lower_bound**: The lower bound of the distribution.
- **point**: The specific point for which the PDF is calculated.

The function checks if the given point is within the specified range. If it is, it returns $1/(b-a)$, indicating a uniform distribution. Otherwise, it returns 0.

➢ **continuous_uniform_cdf (lower_bound, upper_bound, point)**

This function calculates the Cumulative Distribution Function (CDF) for a continuous uniform distribution. It represents the probability that a random variable is less than or equal to a specific value.

- **lower_bound**: The lower bound of the distribution.

- **upper_bound**: The upper bound of the distribution.
- **point**: The specific point for which the CDF is calculated.

The function checks if the given point is within the specified range. If it is, it returns *(x-a)/(b-a)*, which represents the cumulative probability up to that point. If the point is greater than the upper bound, it returns 1, indicating that all values are covered. If the point is outside the range, it returns 0.

➢ **generate_continuous_uniform_randvar (upper_bound, lower_bound, size)**

This function generates random variables following a continuous uniform distribution within the specified range.

- **upper_bound**: The upper bound of the distribution.
- **lower_bound**: The lower bound of the distribution.
- **size**: The number of random variables to generate.

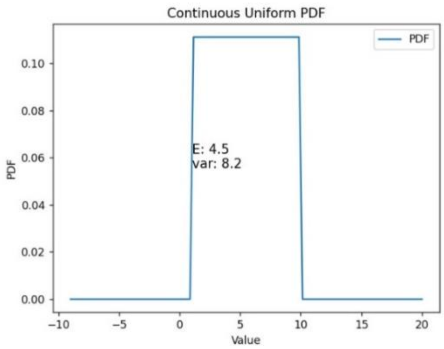
The function uses **random. uniform (lower_bound, upper_bound)** to generate random values within the specified range and repeats this process **size** times, storing the results in a list.

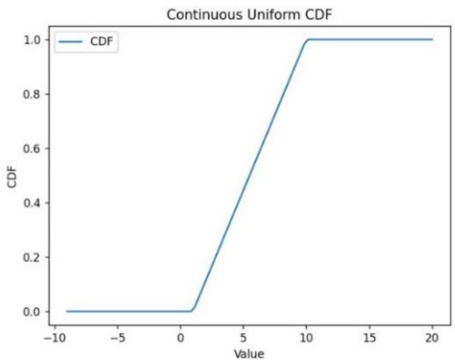
➢ **continuous_uniform (lower_bound, upper_bound, npoints, size)**

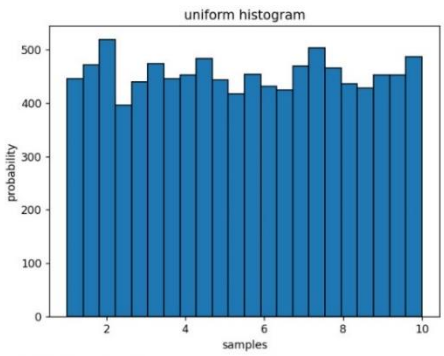
This function generates and visualizes various aspects of the continuous uniform distribution:

- **Points (x list):** It creates a list of points within the specified interval, obtained using the **line_space** function (not defined in the provided code).
- **PDF Calculation (pdf list):** It calculates the PDF for each point in **x** using the **continuous_uniform_pmf** function.
- **CDF Calculation (cdf list):** It calculates the CDF for each point in **x** using the **continuous_uniform_cdf** function.
- **Mean and Variance Calculation:** It calculates the mean and variance of the continuous uniform distribution.
- **Histogram of Random Samples (histogram list):** It generates random samples using the **generate_continuous_uniform_randvar** function and plots a histogram of the results.
- **Graph Plots Using matplotlib:**
    1. **PDF Plot:** It uses **plt. plot** to plot the PDF.
    2. **CDF Plot:** It uses **plt. plot** to plot the CDF.
    3. **Histogram Plot:** It uses **plt. hist** to plot a histogram of the random samples.
- **Display:** It uses **plt. show ()** to display each plot separately.

```
Choose which type of random variables that you will use:
1. Binomial
2. Bernoulli
3. Geometric
4. Poisson
5. Uniform Discrete
6. Continuous Uniform
7. Exponential
8. Gaussian
Enter the corresponding number (1-8): 6
Enter lower bound for Uniform distribution: 1
Enter upper bound for Uniform distribution: 10
Enter a point for the distribution: 5
Enter the number of points for the distribution: 100
The PMF of continous uniform distribution:  0
The CDF of continous uniform distribution:  0.4444444444444444
```







2. Exponential Distribution

2.1. Definition

An exponential continuous random variable is a type of probability distribution that models the time until an event occurs in a (Poisson process) process that evolves continuously and independently over time.

2.2. Conditions

The exponential distribution is used to model the time until an event occurs, assuming continuous and independent events at a constant rate (λ). Notable characteristics include its memoryless property, no upper time limit, and a relationship with the Poisson process. The mean and variance are both (1/ λ), and the survival function is S(x) = e^λx. This distribution is often applied in scenarios involving waiting times or inter-arrival times in queuing theory.

2.3. Parameters

- x is the random variable representing the time between events
- λ is the rate parameter, indicating the average number of events per unit time.

2.4. Probability Density Function (PDF)

The probability density function of the exponential distribution is given by:

$$f_X(x) = \lambda e^{-\lambda x}$$

➢Proof:

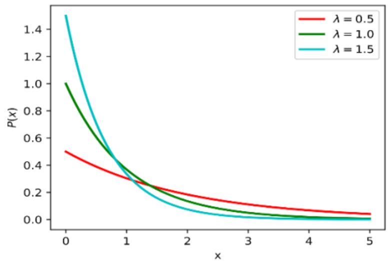
The cumulative distribution function (CDF) of the exponential distribution is given by:

$$F(x;\lambda) = \int_0^x \lambda e^{-\lambda t}\, dt$$

Now, let's differentiate the CDF to get the PDF:

$$f(x;\lambda) = \frac{d}{dx} F(x;\lambda)$$

$$f(x;\lambda) = \frac{d}{dx}\left(\int_0^x \lambda e^{-\lambda t}\, dt\right)$$



Performing the differentiation step by step:

$$f(x;\lambda) = \frac{d}{dx}\left(-e^{-\lambda t}\Big|_0^x\right)$$

$$f(x;\lambda) = \frac{d}{dx}\left(-e^{-\lambda x} + e^0\right)$$

$$f(x;\lambda) = \lambda e^{-\lambda x}$$

So, the final probability density function (PDF) is:

$$f(x;\lambda) = \lambda e^{-\lambda x}$$

2.5. Cumulative Distribution Function (CDF)

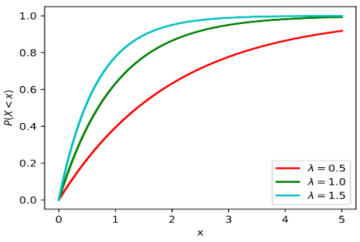
The cumulative distribution function is expressed as:

$$F_X(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$$

➢Proof:

$$F_X(x) = \begin{cases} 0\,, & \text{if } x < 0 \\ 1 - \exp[-\lambda x]\,, & \text{if } x \geq 0\,. \end{cases}$$

$$\mathrm{Exp}(x;\lambda) = \begin{cases} 0\,, & \text{if } x < 0 \\ \lambda \exp[-\lambda x]\,, & \text{if } x \geq 0\,. \end{cases}$$



$$F_X(x) = \int_{-\infty}^{x} \mathrm{Exp}(z;\lambda)\,\mathrm{d}z\,.$$

$$F_X(x) = \int_{-\infty}^{x} 0\,\mathrm{d}z = 0\,.$$

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{0} \mathrm{Exp}(z;\lambda)\,\mathrm{d}z + \int_{0}^{x} \mathrm{Exp}(z;\lambda)\,\mathrm{d}z \\ &= \int_{-\infty}^{0} 0\,\mathrm{d}z + \int_{0}^{x} \lambda \exp[-\lambda z]\,\mathrm{d}z \\ &= 0 + \lambda \left[-\frac{1}{\lambda}\exp[-\lambda z]\right]_0^x \\ &= \lambda \left[\left(-\frac{1}{\lambda}\exp[-\lambda x]\right) - \left(-\frac{1}{\lambda}\exp[-\lambda \cdot 0]\right)\right] \\ &= 1 - \exp[-\lambda x]\,. \end{aligned}$$

2.6. Mean and Variance

2.6.1.Mean

The mean of the exponential distribution is calculated using the integration by parts.

$$E[X] = \frac{1}{\lambda}$$

➢Proof:

$$Mean = E[X] = \int_0^\infty x\lambda e^{-\lambda x} dx$$

$$= \lambda \left[\left| \frac{-xe^{-\lambda x}}{\lambda} \right|_0^\infty + \frac{1}{\lambda} \int_0^\infty e^{-\lambda x} dx \right]$$

$$= \lambda \left[0 + \frac{1}{\lambda} \frac{-e^{-\lambda x}}{\lambda} \right]_0^\infty$$

$$= \lambda \frac{1}{\lambda^2}$$

$$= \frac{1}{\lambda}$$

2.6.2.Variance

Find the variance of the exponential distribution, we need to find the second moment of the exponential distribution, and it is given by:

$$E[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x} = \frac{2}{\lambda^2}$$

We can calculate the variance using this formula: Var (X) = E(X2) - E(X)2

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

➢Proof:

the second moment becomes

$$\begin{aligned}
\mathrm{E}[X^2] &= \lambda\left[\left(-\frac{x^2}{\lambda}-\frac{2x}{\lambda^2}-\frac{2}{\lambda^3}\right)\exp(-\lambda x)\right]_0^{+\infty} \\
&= \lambda\left[\lim_{x\to\infty}\left(-\frac{x^2}{\lambda}-\frac{2x}{\lambda^2}-\frac{2}{\lambda^3}\right)\exp(-\lambda x)-\left(0-0-\frac{2}{\lambda^3}\right)\exp(-\lambda\cdot 0)\right] \\
&= \lambda\left[0+\frac{2}{\lambda^3}\right] \\
&= \frac{2}{\lambda^2}\,.
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\left[X^2\right]-\mathrm{E}[X]^2 \\
&= \frac{2}{\lambda^2}-\left(\frac{1}{\lambda}\right)^2 \\
&= \frac{2}{\lambda^2}-\frac{1}{\lambda^2} \\
&= \frac{1}{\lambda^2}\,.
\end{aligned}$$

2.7. Applications

1. **Reliability Engineering:** Modelling time until the failure of a system or component.
2. **Queuing Theory:** Analysing waiting times in queues or lines.
3. **Telecommunications:** Representing the time between phone calls or message arrivals.
4. **Finance:** Modelling the time until a financial event occurs

2.8. Real-life Application

There once was a busy day at the airport, with ninety-two planes waiting impatiently to take off. Every plane was getting ready to take off into the endless sky, and there was an air of ordered chaos. Equipped with statistical knowledge, the air traffic controllers determined that the mean interval between departures was a swift 10.8 minutes. Because of this smooth flow, a flight would depart the ground with grace every 10.8 minutes on average.
Still, there was a lingering curiosity about the possibility of delays. The mean was transformed with a mathematical wand into a lambda of 0.09 flights per minute, which gave rise to a key that unlocked the likelihood of delays.
One question emerged as the timer ran out: What was the likelihood of a delay longer than five minutes? With careful consideration, the astute statisticians unlocked the potential of probability theory. They deducted from the enchanted number 1 the cumulative probability of departures within five minutes.
The answer then became clear: there was a 90.5% chance that the farewell dance would require a little longer than the customary five minutes. The airport, with its rhythms of comings and goings, went on its choreographed symphony in the air, now and then broken up by the tardy elegance of a plane taking its sweet time to enter the heavenly ballet.

Sum of time between departures = 996 mins
The total number of flights = 92
Mean = 996/92 = 10.8 mins
Lambda = 1/mean = 0.09 flights/mins

What's the Probability of delay between departures being more than 5 mins:
p(x>5) = 1-P(x<=5)
P(x<=5) = 0.095
p(x>5) = 1-0.095 = 0.905

2.9. Code

```
def generate_exponential_random_variables(lambda_, size=1000):
    return [(-1 / lambda_) * math.log(random.uniform(0,1)) for _ in range(size)]
def exponential_pdf(lambda_,point):
    return lambda_*math.exp(-lambda_*point) # returns the pdf at point
def exponential_cdf(lambda_,point):
    return 1-math.exp(-lambda_*point) # returns the cdf at point
def exponential(lambda_,lower_bound,upper_bound,npoints,size):
    # creates the interval of values of the random variables
    points = line_space(lower_bound,upper_bound,npoints)
    # generates the random variables for the histogram.
    histogram = generate_exponential_random_variables(lambda_,size)
    #calculates the pdf for each value of x
    ypdf = [exponential_pdf(lambda_,x) for x in points]
     # calculates the cdf for each value of x
    ycdf = [exponential_cdf(lambda_,x) for x in points]
    mean = 1/lambda_
    variance = 1/lambda_**2
    # plots the PDF, CDF and histogram.
    plt.plot(points, ypdf, label=f'λ = {lambda_}')
    plt.text(lower_bound, 1 / (upper_bound - lower_bound), f'E: {mean}\nvar: {variance}', fontsize=12)
    plt.title('Exponential PDF')
    plt.xlabel('Value')
    plt.ylabel('PDF')
    plt.legend()
    plt.show()
    plt.plot(points, ycdf, label=f'λ = {lambda_}')
    plt.title('Exponential CDF')
    plt.xlabel('Value')
    plt.ylabel('CDF')
    plt.legend()
    plt.show()
    plt.hist(histogram, bins=round(max(histogram)), edgecolor='black')
    plt.title("Exponential histogram")
    plt.xlabel("X")
    plt.ylabel("probability")
    plt.show()
```

- **generate_exponential_random_variables (lambda_, size=1000)**

This function generates random variables following an exponential distribution using the inverse transform sampling method.

- **lambda_**: The rate parameter of the exponential distribution.
- **size**: The number of random variables to generate (default is 1000).

The function uses the inverse transform sampling technique. It generates random variables by applying the inverse of the cumulative distribution function (CDF) to uniformly distributed random variables between 0 and 1. In the case of the exponential distribution, the inverse CDF is calculated as $X=-(1/\lambda)\ln(U)$, where U is a uniformly distributed random variable. The generated random variables are stored in a list.

- **exponential_pdf (lambda_, point)**

This function calculates the Probability Density Function (PDF) of the exponential distribution at a specific point.

- **lambda_**: The rate parameter of the exponential distribution.
- **point**: The specific point for which the PDF is calculated.

The PDF formula for the exponential distribution is $f(x; \lambda) = \lambda e^{-\lambda x}$.

➢ **exponential_cdf (lambda_, point)**

This function calculates the Cumulative Distribution Function (CDF) of the exponential distribution at a specific point.

- **lambda_**: The rate parameter of the exponential distribution.
- **point**: The specific point for which the CDF is calculated.

The CDF formula for the exponential distribution is $F(x; \lambda) = 1 - e^{-\lambda x}$.

➢ **exponential (lambda_, lower_bound, upper_bound, npoints, size)**

This function generates and visualizes various aspects of the exponential distribution:

- **lambda_**: The rate parameter of the exponential distribution.
- **lower_bound**: The lower bound of the x-axis for plotting.
- **upper_bound**: The upper bound of the x-axis for plotting.
- **npoints**: The number of points to generate between **lower_bound** and **upper_bound**.
- **size**: The number of random variables to generate for the histogram.

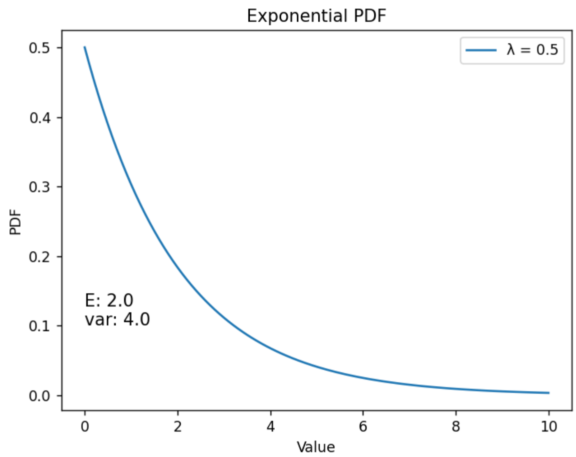
The function performs the following steps:

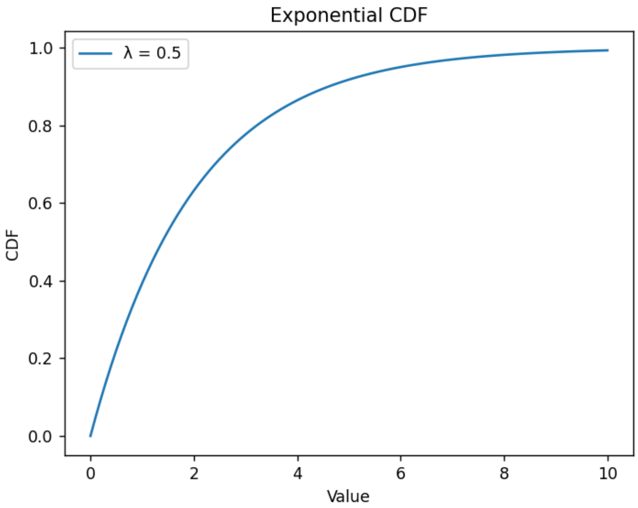
1. **Generate Points:**
    - Creates a list of points (**points**) between **lower_bound** and **upper_bound** to represent the x-axis.
2. **Generate Random Variables:**
    - Generates random variables using the **generate_exponential_random_variables** function for the histogram.
3. **Calculate PDF Values:**
    - Calculates the PDF values (**ypdf**) for each point using the **exponential_pdf** function.
4. **Calculate CDF Values:**
    - Calculates the CDF values (**ycdf**) for each point using the **exponential_cdf** function.
5. **Calculate Mean and Variance:**
    - Computes the mean and variance of the distribution.
6. **Plot PDF:**
    - Plots the PDF of the exponential distribution using **plt. plot**.
7. **Plot CDF:**
    - Plots the CDF of the exponential distribution using **plt. plot**.
8. **Plot Histogram:**
    - Generates and plots a histogram of the random variables using **plt.hist**.

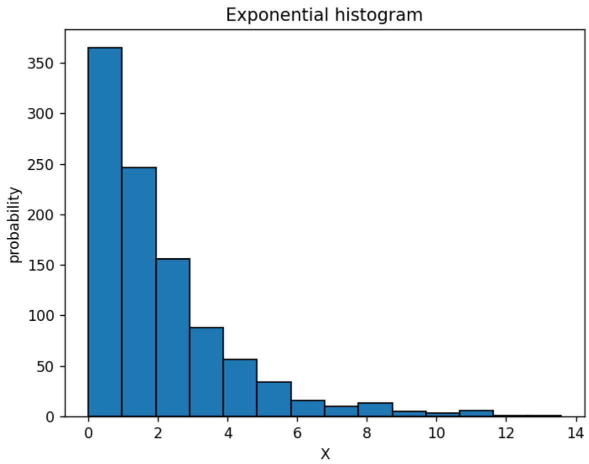
Visualization:

- The code produces three plots: PDF, CDF, and a histogram of random variables.
- Text annotations provide information about the mean and variance.

```
Choose which type of random variables that you will use:
1. Binomial
2. Bernoulli
3. Geometric
4. Poisson
5. Uniform Discrete
6. Continuous Uniform
7. Exponential
8. Gaussian
Enter the corresponding number (1-8): 7
Enter lambda for Exponential distribution: 0.5
Enter a value at which to calculate the PDF: 5
Enter lower bound for Uniform distribution: 0
Enter upper bound for Uniform distribution: 10
Enter the number of points for the distribution: 100
The PMF of Exponential distribution:  0.0410424993119494
The CDF of Exponential distribution:  0.9179150013761012
```

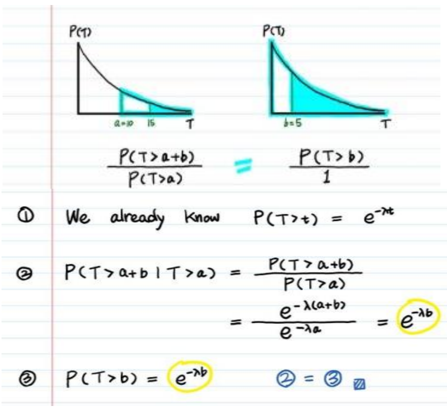






NOTE:

- The exponential distribution is closely associated with the Poisson process, which is a stochastic process that models the number of events occurring in fixed intervals of time or space. Key features of the Poisson process include independence of events and a constant average rate of occurrence.
- Exponential Decay: The distribution showcases exponential decay, indicating a constant hazard rate. The probability of an event occurring in the next instant remains constant, regardless of how much time has elapsed.
- Memory-lessness: The memoryless property of the exponential distribution can be defined as: $P(T > a + b \mid T > a) = P(T > b)$ This means that given an exponential random variable T, the probability that T exceeds a sum of two time periods (a + b) given that it has already exceeded the first period a, is equal to the probability that T exceeds just the second period b.



3. Gaussian (Normal) Random Variable

3.1. Definition

The Gaussian distribution, sometimes referred to as the normal distribution, is the most important continuous probability distribution in statistics and probability theory. It also goes by the name "bell curve" occasionally.

3.2. Parameters

Gaussian random variables have two parameters (μ, σ2). It is noteworthy that the mean is μ and the variance is σ 2 — these two parameters are exactly the first moment and the second central moment of the random variable. Most other random variables do not have this property.

3.3. Probability Density Function (PMF)

The probability density function or probability distribution function is the same. PDF can be considered as a function which maps each value of the random variable to its frequency.

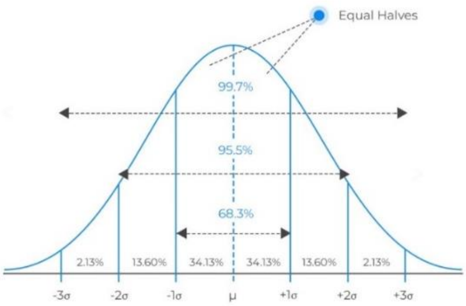
$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Where,

- x is the variable
- μ is the mean
- σ is the standard deviation

Probabilities follow the empirical rule:

- Approximately 68.3% of the data falls within one standard deviation of the mean.
- Approximately 95.5% of the data falls within two standard deviations of the mean.
- Approximately 99.7% of the data fall within three standard deviations of the mean.



3.3.1. Standard Normal Distribution

The standard normal distribution has the PDF:

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

Where mean μ=0 and standard deviation σ=1

➢Proof:

$$\int_{-\infty}^{\infty} f_X(x)dx$$
$$= (2\pi)^{-1/2} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}x^2\right)dx$$
$$= (2\pi)^{-1/2} 2 \int_{0}^{\infty} \exp\left(-\frac{1}{2}x^2\right)dx \qquad \text{(since the integrand is even)}$$
$$= (2\pi)^{-1/2} 2 \left(\int_{0}^{\infty} \exp\left(-\frac{1}{2}x^2\right)dx \int_{0}^{\infty} \exp\left(-\frac{1}{2}x^2\right)dx\right)^{1/2}$$
$$= (2\pi)^{-1/2} 2 \left(\int_{0}^{\infty} \exp\left(-\frac{1}{2}x^2\right)dx \int_{0}^{\infty} \exp\left(-\frac{1}{2}y^2\right)dy\right)^{1/2}$$
$$= (2\pi)^{-1/2} 2 \left(\int_{0}^{\infty}\int_{0}^{\infty} \exp\left(-\frac{1}{2}(x^2+y^2)\right)dydx\right)^{1/2}$$
$$= (2\pi)^{-1/2} 2 \left(\int_{0}^{\infty}\int_{0}^{\infty} \exp\left(-\frac{1}{2}(x^2+s^2x^2)\right)xdsdx\right)^{1/2} \qquad \text{(changing variable } y = xs\text{)}$$
$$= (2\pi)^{-1/2} 2 \left(\int_{0}^{\infty}\int_{0}^{\infty} \exp\left(-\frac{1}{2}x^2(1+s^2)\right)xdxds\right)^{1/2}$$

$$= (2\pi)^{-1/2} 2\left(\int_0^\infty \left[-\frac{1}{1+s^2}\exp\left(-\frac{1}{2}x^2(1+s^2)\right)\right]_0^\infty ds\right)^{1/2}$$
$$= (2\pi)^{-1/2} 2\left(\int_0^\infty \left(0+\frac{1}{1+s^2}\right)ds\right)^{1/2}$$
$$= (2\pi)^{-1/2} 2\left(\int_0^\infty \frac{1}{1+s^2}ds\right)^{1/2}$$
$$= (2\pi)^{-1/2} 2([\arctan(s)]_0^\infty)^{1/2}$$
$$= (2\pi)^{-1/2} 2(\arctan(\infty) - \arctan(0))^{1/2}$$
$$= (2\pi)^{-1/2} 2\left(\frac{\pi}{2} - 0\right)^{1/2}$$
$$= 2^{-1/2}\pi^{-1/2} 2\pi^{1/2} 2^{-1/2} = 1$$

3.3.2. General Gaussian Distribution

The General normal distribution has the PDF:

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

➢Proof:

$\sigma > 0$

Prove: $\int_{-\infty}^{\infty} f(x \mid \mu, \sigma^2)\, dx = 1$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx$$

$$z = \frac{x-\mu}{\sigma} \;;\; x = \sigma z + \mu \;;\; dx = \sigma\, dz$$

$$= \frac{1}{\sqrt{2\pi}\,\cancel{\sigma}}\int_{-\infty}^{\infty} e^{-(\sigma z + \cancel{\mu} - \cancel{\mu})^2/2\sigma^2} \cancel{\sigma}\, dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\cancel{\sigma^2} z^2/2\cancel{\sigma^2}} dz$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-z^2/2} dz \qquad \int_{-\infty}^{\infty} e^{-y^2} dy \quad \text{(Gaussian Integral)}$$

$$y^2 = z^2/2 \qquad = \sqrt{\pi}$$

$$y = \frac{z}{\sqrt{2}} \;;\; z = \sqrt{2}\, y \;;\; dz = \sqrt{2}\, dy$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2} \cancel{\sqrt{2}}\, dy = \frac{1}{\sqrt{\pi}} \underbrace{\int_{-\infty}^{\infty} e^{-y^2} dy}_{\sqrt{\pi}}$$

$$= \frac{1}{\cancel{\sqrt{\pi}}}\cancel{\sqrt{\pi}} = \underline{\underline{1}}$$

3.4. Cumulative distribution function (CDF)

It is a function derived from the probability density function for a continuous random variable. It gives the probability of finding the random variable at a value less than or equal to a given cut-off. It is calculating area under the curve in the distribution curve.

$F_X(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right]$ Where erf(x) is the error function defined as $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)\,dt$.

3.4.1. Standard Normal Distribution

PDF of the Standard Normal Distribution:

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

➢Proof:

$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} f(t)\, dt$

$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt$

The integral does not have a closed-form solution, but it is often expressed in terms of the error function (erf):

$$\Phi(z) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right]$$

3.4.2. General Gaussian Distribution

PDF of the General Gaussian distribution:

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

➢Proof:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}\,dt$$

This integral is solved in terms of the error function (erf):

$$F(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

3.5. Mean and Variance

3.5.1. Mean

The expected value (or mean) of a Gaussian distribution is equal to its mean parameter μ. In mathematical notation, if X is a random variable following a Gaussian distribution with mean μ and standard deviation σ, the expected value is given by:

3.5.1.1. Standard Normal Distribution

The expected value of a standard normal random variable X is:

$$E[X] = 0$$

➢Proof:

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f_X(x)dx \\
&= (2\pi)^{-1/2} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2}x^2\right)dx \\
&= (2\pi)^{-1/2} \int_{-\infty}^{0} x \exp\left(-\frac{1}{2}x^2\right)dx + (2\pi)^{-1/2} \int_{0}^{\infty} x \exp\left(-\frac{1}{2}x^2\right)dx \\
&= (2\pi)^{-1/2}\left[-\exp\left(-\frac{1}{2}x^2\right)\right]_{-\infty}^{0} + (2\pi)^{-1/2}\left[-\exp\left(-\frac{1}{2}x^2\right)\right]_{0}^{\infty} \\
&= (2\pi)^{-1/2}[-1+0] + (2\pi)^{-1/2}[0+1] \\
&= (2\pi)^{-1/2} - (2\pi)^{-1/2} \\
&= 0
\end{aligned}$$

3.5.1.2. General Gaussian Distribution

The expected value of a normal random variable X is:

$$E[X] = \mu$$

➢Proof:

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x$$

So:

$$\begin{aligned}
\mathrm{E}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\mathrm{d}x \\
&= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu)\exp(-t^2)\,\mathrm{d}t \\
&= \frac{1}{\sqrt{\pi}}\left(\sqrt{2}\sigma \int_{-\infty}^{\infty} t\exp(-t^2)\,\mathrm{d}t + \mu \int_{-\infty}^{\infty} \exp(-t^2)\,\mathrm{d}t\right) \\
&= \frac{1}{\sqrt{\pi}}\left(\sqrt{2}\sigma\left[-\frac{1}{2}\exp(-t^2)\right]_{-\infty}^{\infty} + \mu\sqrt{\pi}\right) \\
&= \frac{\mu\sqrt{\pi}}{\sqrt{\pi}} \\
&= \mu
\end{aligned}$$

3.5.2. Variance

The variance of a Gaussian distribution is denoted as Var[X] or σ 2, where X is a random variable following a Gaussian distribution with mean μ and standard deviation σ. The formula for the variance is given by:

3.5.2.1. Standard Normal Distribution Variance:

The variance of a standard normal random variable X is:

$$\operatorname{Var}[X] = \sigma^2 = 1^2 = 1$$

➢Proof:

$$\begin{aligned}
\mathrm{E}[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x)dx \\
&= (2\pi)^{-1/2} \int_{-\infty}^{\infty} x^2 \exp\left(-\tfrac{1}{2}x^2\right)dx \\
&= (2\pi)^{-1/2} \left\{ \int_{-\infty}^{0} x\left(x\exp\left(-\tfrac{1}{2}x^2\right)\right)dx + \int_{0}^{\infty} x\left(x\exp\left(-\tfrac{1}{2}x^2\right)\right)dx \right\} \\
&= (2\pi)^{-1/2} \left\{ \left[-x\exp\left(-\tfrac{1}{2}x^2\right)\right]_{-\infty}^{0} + \int_{-\infty}^{0} \exp\left(-\tfrac{1}{2}x^2\right)dx + \left[-x\exp\left(-\tfrac{1}{2}x^2\right)\right]_{0}^{\infty} \right. \\
&\quad \left. + \int_{0}^{\infty} \exp\left(-\tfrac{1}{2}x^2\right)dx \right\} \quad \text{(integrating by parts)} \\
&= (2\pi)^{-1/2} \left\{ (0-0) + (0-0) + \int_{-\infty}^{0} \exp\left(-\tfrac{1}{2}x^2\right)dx + \int_{0}^{\infty} \exp\left(-\tfrac{1}{2}x^2\right)dx \right\} \\
&= (2\pi)^{-1/2} \int_{-\infty}^{\infty} \exp\left(-\tfrac{1}{2}x^2\right)dx \\
&= \int_{-\infty}^{\infty} f_X(x)dx = 1 \quad \text{(the integral of a pdf over its support is equal to 1)}
\end{aligned}$$

$$\mathrm{E}[X]^2 = 0^2 = 0$$

$$\operatorname{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = 1 - 0 = 1$$

3.5.2.2. General Normal Distribution Variance:

The variance of a standard normal random variable X is:

$$\operatorname{Var}[X] = \sigma^2$$

➢Proof:

$$\operatorname{var}(X) = \int_{-\infty}^{\infty} x^2 f_X(x)\,\mathrm{d}x - (\mathrm{E}(X))^2$$

So:

$$\begin{aligned}
\operatorname{var}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\mathrm{d}x - \mu^2 \\
&= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu)^2 \exp(-t^2)\,\mathrm{d}t - \mu^2 \\
&= \frac{1}{\sqrt{\pi}} \left(2\sigma^2 \int_{-\infty}^{\infty} t^2 \exp(-t^2)\,\mathrm{d}t + 2\sqrt{2}\sigma\mu \int_{-\infty}^{\infty} t\exp(-t^2)\,\mathrm{d}t + \mu^2 \int_{-\infty}^{\infty} \exp(-t^2)\,\mathrm{d}t\right) - \mu^2 \\
&= \frac{1}{\sqrt{\pi}} \left(2\sigma^2 \int_{-\infty}^{\infty} t^2 \exp(-t^2)\,\mathrm{d}t + 2\sqrt{2}\sigma\mu \left[-\frac{1}{2}\exp(-t^2)\right]_{-\infty}^{\infty} + \mu^2\sqrt{\pi}\right) - \mu^2 \\
&= \frac{1}{\sqrt{\pi}} \left(2\sigma^2 \int_{-\infty}^{\infty} t^2 \exp(-t^2)\,\mathrm{d}t + 2\sqrt{2}\sigma\mu \cdot 0\right) + \mu^2 - \mu^2 \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 \exp(-t^2)\,\mathrm{d}t \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \left(\left[-\frac{t}{2}\exp(-t^2)\right]_{-\infty}^{\infty} + \frac{1}{2}\int_{-\infty}^{\infty} \exp(-t^2)\,\mathrm{d}t\right) \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2} \int_{-\infty}^{\infty} \exp(-t^2)\,\mathrm{d}t \\
&= \frac{2\sigma^2\sqrt{\pi}}{2\sqrt{\pi}} \\
&= \sigma^2
\end{aligned}$$

3.6. Applications

1. **Statistical Analysis and Hypothesis Testing:** Many statistical tests and methods assume normality, making it a crucial tool in hypothesis testing. It is used in constructing confidence intervals and determining statistical significance.

2. **Financial Modelling and Risk Management:** Stock prices, returns, and financial market indices often follow a normal distribution. Tools like the Black-Scholes model for option pricing assume normality in the distribution of asset returns.
3. **Quality Control and Manufacturing Processes:** In manufacturing, product characteristics like length, weight, and strength often exhibit a normal distribution. Control charts and process capability analysis rely on the assumption of normality.
4. **Biostatistics and Medical Research:** Many biological and physiological measurements in medicine, such as blood pressure, height, and weight, are often normally distributed. Tools like linear regression and analysis of covariance often assume normality.
5. **Machine Learning and Data Science:** Gaussian distribution is a key concept in machine learning, particularly in algorithms like Gaussian Naive Bayes. It is used in clustering algorithms. Gaussian processes are employed in regression and optimization tasks.

3.7. Real-life Applications

The story of retirement ages became a tale of resilience and unpredictability, celebrated through the lens of statistical exploration. And so, armed with their newfound insights, the group eagerly continued their data-driven adventure into the heart of football analytics.
Ultimately, the retirement age story was reinterpreted through statistical exploration as a story of tenacity and unpredictable outcomes.
Equipped with their recently acquired knowledge, the team enthusiastically proceeded with their data-driven journey into the core of football analytics.and the number of players is 212.

Let’s say we have research about 212 players retirement The mean is 35 and standard deviation is 3.5 If we want to calculate the probability of a player retiring after the age of 38.
So, the $p(x>38) = 1\text{-}P(x<=38)$
The cdf formula is 0.5 * (1 + erf ((x- mean) / std_dev * sqrt (2)) So the $P(x=38) = 0.78$ So $p(x>38) = 1\text{-}0.78 = 0.22$
(22%)

3.8. Code

```
def generate_gaussian_random_variables(mean, std_dev, size=1000):
    return [random.gauss(mean, std_dev) for _ in range(size)]
def gaussian_pdf(mean,std_dev,variance,point):
    return (1/(std_dev*math.sqrt(2*math.pi)))*math.exp((-(point-mean)**2)/(2*std_dev**2))
def gaussian_cdf(mean,std_dev,point):
    return 0.5 * ((1 + math.erf((point - mean) / (std_dev * math.sqrt(2)))))
def gaussian(mean, std_dev,size):
    variance = std_dev**2
    # generate the random variables for the histogram.
    histogram = generate_gaussian_random_variables(mean,std_dev,size)
    #generates all the random variable values
    points = line_space(mean-6*std_dev,mean+6*std_dev,100)
    # calculates thepdf for each point.
    pdf = [gaussian_pdf(mean,std_dev,variance,x)for x in points]
    #calculates the cdf value for each point.
    cdf = [gaussian_cdf(mean,std_dev,x) for x in points]
    # plots the PDF, CDF and histogram.
    plt.plot(points, pdf, label=f'μ: {mean}')
    plt.text(mean-10,0.4, f'var: {variance}', fontsize=12)
    plt.title('Gaussian PDF')
    plt.xlabel('X')
```

```python
    plt.ylabel('probability')
    plt.legend()
    plt.show()
    plt.plot(points, cdf, label=f'μ: {mean}')
    plt.title('Gaussian CDF')
    plt.xlabel('X')
    plt.ylabel('CDF')
    plt.legend()
    plt.show()
    plt.hist(histogram, bins='auto', edgecolor='black')
    plt.title("Gaussian histogram")
    plt.xlabel("X")
    plt.ylabel("samples")
    plt.show()
```

- **generate_gaussian_random_variables (mean, std_dev, size=1000)**

This function generates random variables following a Gaussian (normal) distribution.

- **mean**: The meaning of the Gaussian distribution.
- **std_dev**: The standard deviation of the Gaussian distribution.
- **size**: The number of random variables to generate (default is 1000).

It utilizes **random. gauss (mean, std_dev)** to generate random variables based on the specified mean and standard deviation. The generated random variables are stored in a list.

- **gaussian_pdf (mean, std_dev, variance, point)**

This function calculates the Probability Density Function (PDF) of the Gaussian distribution at a specific point.

- **mean**: The meaning of the Gaussian distribution.
- **std_dev**: The standard deviation of the Gaussian distribution.
- **variance**: The variance of the Gaussian distribution.
- **point**: The specific point for which the PDF is calculated.

- **gaussian_cdf (mean, std_dev, point)**

This function calculates the Cumulative Distribution Function (CDF) of the Gaussian distribution at a specific point.

- **mean**: The meaning of the Gaussian distribution.
- **std_dev**: The standard deviation of the Gaussian distribution.
- **point**: The specific point for which the CDF is calculated.

- **gaussian (mean, std_dev, size)**

This function generates and visualizes various aspects of the Gaussian distribution:

- **mean**: The meaning of the Gaussian distribution.
- **std_dev**: The standard deviation of the Gaussian distribution.
- **size**: The number of random variables to generate for the histogram.

The function performs the following steps:

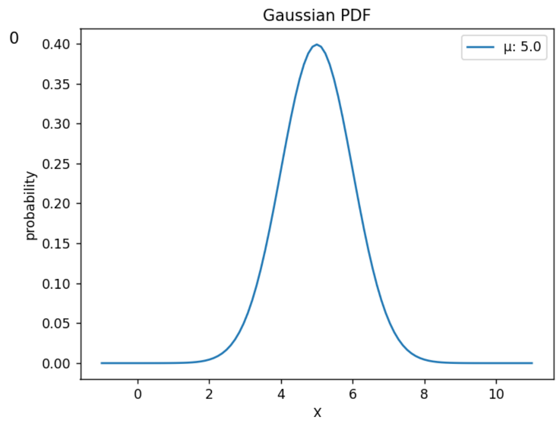
1. **Generate Points:**
    - Creates a list of points (**points**) for the x-axis within a specified range around the mean.
2. **Generate Random Variables:**
    - Generates random variables using the **generate_gaussian_random_variables** function for the histogram.
3. **Calculate PDF Values:**
    - Calculates the PDF values (**pdf**) for each point using the **gaussian_pdf** function.

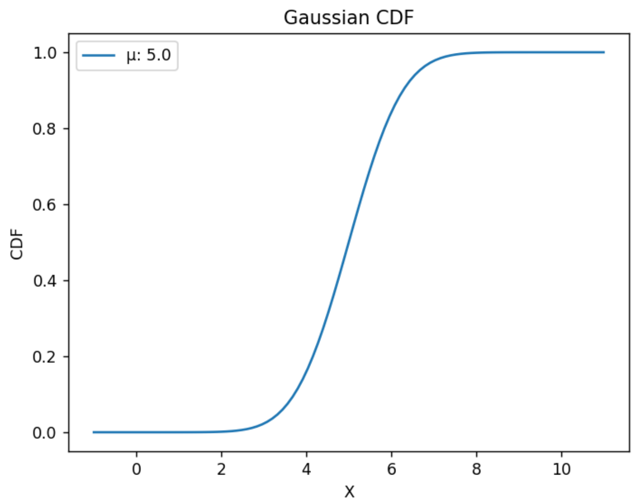
4. **Calculate CDF Values:**
   - Calculates the CDF values (**cdf**) for each point using the **gaussian_cdf** function.
5. **Calculate Mean and Variance:**
   - Computes the variance of the distribution.
6. **Plot PDF:**
   - Plots the PDF of the Gaussian distribution using **plt. plot**.
7. **Plot CDF:**
   - Plots the CDF of the Gaussian distribution using **plt. plot**.
8. **Plot Histogram:**
   - Generates and plots a histogram of the random variables using **plt.hist**.

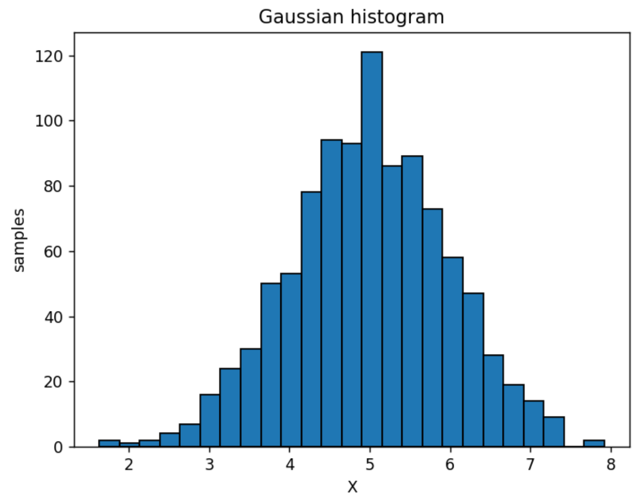
Visualization:

- The code produces three plots: Gaussian PDF, CDF, and a histogram of random variables.
- Text annotations provide information about the mean and variance.

```
Choose which type of random variables that you will use:
1. Binomial
2. Bernoulli
3. Geometric
4. Poisson
5. Uniform Discrete
6. Continuous Uniform
7. Exponential
8. Gaussian
Enter the corresponding number (1-8): 8
Enter mean for Gaussian distribution: 5
Enter standard deviation for the Gaussian distribution: 1
Enter variance for Gaussian distribution: 1
Enter a value at which to calculate the PDF: 7
The PMF of Gaussian(Normal) distribution:  0.05399096651318806
The CDF of Gaussian(Normal) distribution:  0.9772498680518209
```







```python
def calculate_mean(data):
    return sum(data) / len(data)
def calculate_variance(data, mean):
    return sum((x - mean) ** 2 for x in data) / len(data)
def calculate_standard_deviation(data):
    mean = calculate_mean(data)
    variance = calculate_variance(data, mean)
    return variance ** 0.5
def football_players_histogram():
    retirment_ages = [24,33,34,33,36,26,28,31,33,32,32,27,39,30,37,30,35,
                      34,31,29,36,35,33,32,35,32,35,35,34,32,36,31,33,33,29,32,33,35,
                      36,33,35,35,35,34,31,31,36,35,33,34,35,33,31,28,35,33,34,39,35,
                      41,35,37,40,36,32,36,36,34,31,35,37,39,36,37,36,36,35,36,36,35,
                      36,40,31,41,32,32,40,34,35,32,36,36,37,44,33,37,34,41,34,36,33,
                      35,35,29,32 ,31,30,27,47,35,35,37,38,36,45,44,45,30,28,34,39,32,
                      37,35,37,41,36,37,34,36,37,37,35,36,34,38,33,32,40,34,32,36,32,
                      38,35,35,35,41,30,35,37,31,36,38,39,34,39,37,43,41,38,35,32,36,
                      40,37,37,35,40,37,35,40,38,38,36,38,38,36,38,38,36,37,32,33,35,
                      37,33,38,36,33,39,36,38,35,38,35,39,41,41,37,35,35,36,35,39,37,
                      38,40,38,36,37,46,35]
    mean = calculate_mean(retirment_ages)
    std_dev = calculate_standard_deviation(retirment_ages)
    variance = calculate_variance(retirment_ages,mean)
    gaussian(mean,std_dev,1000)
    print("The mean:",mean)
    print("variance:",variance)
    print("Standard deviation:",std_dev)
```

➢ **Statistical Functions**

1. **calculate_mean(data)**
    - Calculates the meaning of a given dataset.
    - Formula:
      mean=sum of datanumber of data pointsmean=number of data pointssum of data
2. **calculate_variance (data, mean)**
    - Calculates the variance of a given dataset.

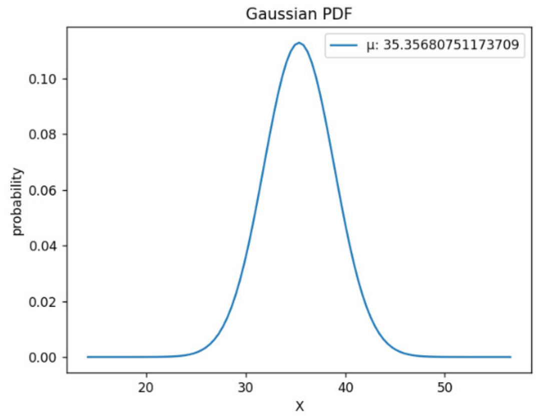
3. **calculate_standard_deviation(data)**
    - Calculates the standard deviation of a given dataset.
    - Uses the mean and variance calculated by the above functions.
    - Formula: standard deviation=variancestandard deviation=variance

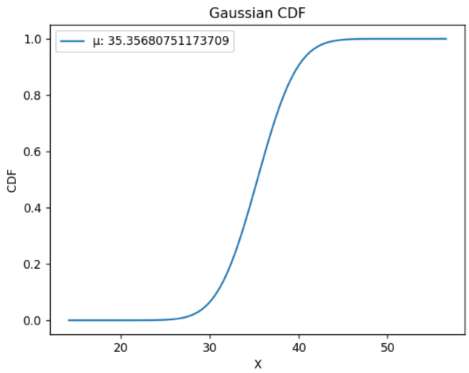
➢ **football_players_histogram ()**

This function specifically deals with retirement ages of football players:

- **retirement_ages**
    - A list containing retirement ages of football players.
- **Calculate Mean, Variance, and Standard Deviation:**
    - Calls the statistical functions to compute mean, variance, and standard deviation of retirement ages.
- **Visualize the Gaussian Distribution:**
    - Calls the **gaussian** function (assumed to be defined elsewhere) to visualize the Gaussian distribution of retirement ages.

- **Print Summary Statistics:**
  - Prints the mean, variance, and standard deviation.





Note:

- The probability density function (pdf) is used to describe probabilities for continuous random variables while for the discrete random variable, probability mass function (pmf) will be used.
- The central limit theorem is the basis for how normal distributions work in statistics.
- In probability theory, (CLT) states that the distribution of a sample variable approximates a normal distribution as the sample size becomes larger, assuming that all samples are identical in size.
- Put another way, CLT is a statistical premise that, given a sufficiently large sample size from a population with a finite level of variance, the mean of all sampled variables from the same population will be approximately equal to the mean of the whole population.

Task Assignment

	Name	Task	%
1	Abdelrahman Salah El-dein Abdelaziz	• Testing and modifying all random Variable codes. • Writing and coordinating the whole report • Searching for Information about all Random Variables. • Writing and coordinating the report • Searched for the real-life examples of all Random variables. • Tested the random variables python codes. • Review the random variable. • Creating the PowerPoint slides.	14.3
2	Abdullah Mohamed Mohamed Galal		14.3
3	Farida Waheed Abd El Bary		14.3
4	Mohamed Ahmed Mohamed Hassan		14.3
5	Nour Hesham El Sayed		14.3
6	Omar Sami Mohamed Ahmed		14.3
7	Razan Ahmed Fawzy		14.3

List Of Reference

1. https://analystprep.com/cfa-level-1-exam/quantitative-methods/key-properties-normal-distribution/
2. https://medium.com/@ashok.1055/gaussian-normal-distribution-and-its-pdf-probability-density-function-f1e4f187308b
3. https://mathworld.wolfram.com/NormalDistribution.html#:~:text=de%20Moivre%
4. https://www.investopedia.com/terms/c/central_limit_theorem.asp
5. https://www.statlect.com/probability-distributions/normal-distribution
6. https://www.linkedin.com/pulse/normal-distribution-its-applications-quality-control-process-shaikh
7. https://proofwiki.org/wiki/Variance_of_Gaussian_Distribution
8. https://proofwiki.org/wiki/Expectation_of_Gaussian_Distribution
9. https://www.cuemath.com/binomial-distribution-formula/
10. https://www.probabilitycourse.com/chapter3/3_1_5_special_discrete_distr.php#:~:text=A%20Bernoulli%20random%20variable%20is,take%20a%20pass%2Dfail%20exam.
11. https://online.stat.psu.edu/stat500/lesson/3/3.2/3.2.2
12. https://careerfoundry.com/en/blog/data-analytics/what-is-bernoulli-distribution/#what-is-bernoulli-distribution
13. https://www.simplilearn.com/tutorials/data-science-tutorial/bernoulli-distribution
14. https://www.cuemath.com/data/variance-of-binomial-distribution/
15. https://testbook.com/maths/mean-and-variance-of-binomial-distribution
16. https://byjus.com/maths/exponential-distribution/
17. https://en.wikipedia.org/wiki/Exponential_distribution#/media/File:Exponential_distribution_cdf_-
18. _public_domain.svg
19. https://towardsdatascience.com/what-is-exponential-distribution-7bdd08590e2a
20. Bibliography
21. Chan, S. H. (2021). PROBABILITY for DATA SCIENCE. Michigan: Michigan Publishing.
22. HAQUE, D. M. (2022). Geometric Random Variable: 7 Important Characteristics. Retrieved from Lambda Geeks: https://lambdageeks.com/geometric-random-variable/
23. Eberly college of science. (2021). Introduction to Probability Theory. Retrieved from Eberly college of science: https://online.stat.psu.edu/stat414/lesson/11/11.2#:~:text=The%20mean%20of%20a%20geometric%20random%20variable%20%28X%29,of%20a%20geometric%20random%20variable%20%28X%29%20is%3A%20%28sigma%5E2%3DVar%28X%29%3Ddfrac%7B1-p%7D%7Bp%5E2%7D%29
24. https://en.wikipedia.org/wiki/Geometric_distribution
25. https://www.cuemath.com/geometric-distribution-formula/
26. https://byjus.com/maths/geometric-distribution/
27. https://www.cuemath.com/data/poisson-distribution/
28. https://en.wikipedia.org/wiki/Poisson_distribution
29. https://www.scribbr.com/statistics/poisson-distribution/
30. https://builtin.com/data-science/poisson-process
31. https://byjus.com/maths/exponential-distribution/