Handouts of Probability 02

Dr. Brahim Taoussi

MCB

PhD in Applied Statistics

b_taoussi@esc-alger.dz

+213 797 457 060

2024 - 2025 / 2^{nd} Year

Contents

Chapter 1

Random Variables and Probability Distributions

1.1 The Concept of Random Variables

1.1.1 General Introduction

A random variable X is a function from the sample space Ω to the set of real numbers $\mathbb{R}$.

$$X : \Omega \to \mathbb{R}$$

Example 1.1.1
Suppose we toss a fair coin twice and record the sequence of heads (H) and tails (T). A possible outcome of this experiment is then given by $\omega_1 = (HT)$.

1. Ω is the set of all possible outcomes that corresponds the random ζ experiment. it is:

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

   Note that Ω has $2^2 = 4$ elements.

2. The Power set $(P(\Omega))$ is the set of all subsets of Ω. It has $2^{|\Omega|}$ elements. In our case it has $2^4 = 16$ elements.

$$P(\Omega) = \{\emptyset, \{(HH)\}, \ldots, \Omega\}$$

   A set $\mathcal{F}$, that is contained in the power set, $\mathcal{F} \subseteq P(\Omega)$, is called $\sigma -$ Algebra when it satisfies certain properties. In this context $\mathcal{F}$ is also known as the Event Space.

   For example $\mathcal{F} = \{\emptyset, \{(HH)\}, \{(HT), (TH), (TT)\}, \Omega\}$ is a $\sigma -$ Algebra.

3. The Probability Measure $\mathbb{P}$ is a function from $\mathcal{F}$ to $[0, 1]$ with $\mathbb{P}(\Omega) = 1$, and such that if the events $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint:

$$\mathbb{P}\left[\cup_{j=1}^{\infty} A_j\right] = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$$

   Assuming that the coin is fair (50/50 chance of getting heads/tails), then the probability function for the $\sigma -$ Algebra consisting of all subsets of Ω is:

Event A_j	(HH)	(HT)	(TH)	(TT)	$\emptyset$	$\{(HT), (TH), (TT)\}$	$\ldots$
$\mathbb{P}(A_j)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{3}{4}$	$\ldots$

   In other words, the probability measure can be expressed as:

$$\mathbb{P} : \mathcal{F} \to [0, 1]$$
$$\forall A \in \Omega \mapsto P[A] = \frac{|A|}{|\Omega|}$$

   It should be mentioned that the triplet $(\Omega, \mathcal{F}, \mathbb{P})$ is known as the Probability Space.

1.1.2 Definitions

Random variable

Let the random experiment ζ and $(\Omega, \mathcal{F}, \mathbb{P})$ is the corresponding probability space where $\mathcal{F} \subseteq P(\Omega)$. A random variable is a function from Ω to $\mathbb{R}$. where:

$$\forall x \in R_X : X^{-1}(x) = \Omega_x. \quad \text{The event } \Omega_x \in \mathcal{F}$$

Note

We denote random variables with capital letters.

Example 1.1.2

Suppose we are only interested in tosses that result in heads in Example 1.1.1. We can define a random variable X that tracks the number of heads obtained in an outcome. So, if outcome (HT) is obtained, then X will equal 1.

Formally, we denote this as follows:

$$\begin{aligned} X : \Omega &\to \mathbb{R} \\ \omega &\mapsto \text{Number of heads in } \omega \end{aligned}$$

Since there are only four outcomes in Ω, we can list the value of X for each outcome individually:

$$\begin{gathered} \text{Inputs: } \Omega \xrightarrow{\text{function:}X} \text{Outputs: } \mathbb{R} \\ (HH) \overset{X}{\longmapsto} 2 \\ (HT) \overset{X}{\longmapsto} 1 \\ (TH) \overset{X}{\longmapsto} 1 \\ (TT) \overset{X}{\longmapsto} 0 \end{gathered}$$

We can also write the above as follows:

$$X(HH) = 2,\ X(HT) = 1,\ X(TH) = 1,\ X(TT) = 0.$$

The advantage of defining the random variable X in this context is that the two outcomes (HT) and (TH) are both assigned a value of 1, meaning we are not focused on the actual sequence of heads and tails that resulted in obtaining one head.
Note that the inverse of the random variable X maps values from R_X back to Ω:

$$\begin{aligned} X^{-1}(0) &= (TT) \\ X^{-1}(1) &= \{(HT), (TH)\} \\ X^{-1}(2) &= (HH) \\ X^{-1}(3) &= \emptyset \\ X^{-1}\{1, 2\} &= \{(HT), (TH), (HH)\} \\ X^{-1}[0, 1] &= \{(TT), (HT), (TH)\} \end{aligned}$$

The range of a random variable

The range of a random variable X (R_X), or Range(X), is the set of possible values of X. R_X can be finite, countable infinite, or uncountable infinite.

Note that the random variable we defined in Example 1.1.2 only equals one of three possible values: 0, 1, 2. Meaning that $R_X = \{0, 1, 2\}$

Types of random variables

The random variable X is:

- Discrete, if R_X is finite or countable infinite.
- Continuous, if R_X is uncountable infinite.

In Example 1.1.2, X is clearly a discrete random variable since R_X is finite.

Further Examples

- Rolling a six-sided die results in a finite set of potential outcomes.
- Continuously tossing a six-sided die until a one is rolled yields a countably infinite set of potential outcomes.
- Measuring temperature leads to an uncountably infinite set of possible values.

1.1.3 The Indicator Function

The indicator function of an event $I_A(\omega)$ takes on only two values: 0 or 1. It is typically used to indicate whether a specific condition or event has occurred. The indicator function of event A is:

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{otherwise} \end{cases}$$

In Example 1.1.1, let's consider the event:

$$A = \{\text{Both tosses have the same output}\} \Rightarrow A = \{(HH), (TT)\}$$

- $P[I_A(\omega) = 1] = P[\omega \in A] = P[\{(HH), (TT)\}] = \frac{2}{4} = P[A]$
- $P[I_A(\omega) = 0] = 1 - P[I_A(\omega) = 1] = \frac{2}{4} = P[\bar{A}]$

1.1.4 The Cumulative Distribution Function

Let X be a random variable and $(\Omega, \mathcal{F}, \mathbb{P})$ is the corresponding probability space. The Cumulative Distribution Function (CDF) of X, denoted by $F_X(x)$, is defined to be the function with domain the real line and range the interval [0,1], which satisfies

$$\forall x \in \mathbb{R} : F_X(x) = P[X \leq x] = P[\{\omega : X(\omega) \leq x\}]$$

For every real number x. F has the following properties:
Let X be a random variable with a CDF F. Then F satisfies the following properties:

1. F is non-decreasing, i.e., F may be constant, but otherwise, it is increasing.
2. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.
3. $\lim_{x\to a^+} F_X(x) = F_X(a)$ continuous from the right.

In Example 1.1.2, Let's start by calculating the probability of the values of the range of X.

$$P_X(0) = P(X = 0) = \frac{1}{4}$$
$$P_X(1) = P(X = 1) = \frac{2}{4}$$
$$P_X(2) = P(X = 2) = \frac{1}{4}$$

To find the CDF of X $F_X(x)$, we argue as follows.
First, note that if $x < 0$, then

$$F_X(x) = P(X \leq x) = 0, \text{ for } x < 0.$$

Next, if $x \geq 2$,

$$F_X(x) = P(X \leq x) = 1, \text{ for } x \geq 2.$$

Next, if $0 \leq x < 1$

$$F_X(x) = P(X \leq x) = P(X = 0) = \frac{1}{4}, \text{ for } 0 \leq X < 1.$$

Finally, if $1 \leq x < 2$,

$$F_X(x) = P(X \leq x) = P(X = 0) + P(X = 1) = \frac{1}{4} + \frac{2}{4} = \frac{3}{4}, \text{ for } 1 \leq x < 2.$$

Thus, to summarize, we have

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ \frac{1}{4}, & \text{if } 0 \leq x < 1 \\ \frac{3}{4}, & \text{if } 1 \leq x < 2 \\ 1, & \text{if } x \geq 2 \end{cases}$$

1.2 Discrete Random Variables

Now that we've established the concept of a random variable, let's discuss the two main categories: discrete and continuous random variables. We'll begin by introducing the discrete random variable, providing its definition, and subsequently exploring the concept of the Probability Mass Function (PMF).

1.2.1 Definition

A discrete random variable is a random variable that has only a finite or countably infinite (think integers or whole numbers) number of possible values.

Examples:

- Tossing a coin twice and record the output.
- Rolling a die three times and record the output.
- Number of children in a family.
- Number of successes in a series of trials.

1.2.2 Probability Mass Function

The PMF (or frequency function) of a discrete random variable X assigns probabilities to the possible values of the random variable. More specifically, if $x_1, x_2, \ldots$ denote the possible values of a random variable X, then the PMF of X $p_X(x)$ is:

$$p_X(x_i) = P(X = x_i) = P(\{\omega \in \Omega / X(\omega) = x_i\}).$$

Example 1.2.1
Continuing in the context of Example 1.1.2, we compute the probability that the random variable X equals 1. There are two outcomes that lead to X taking the value 1, namely HT and TH. So, the probability that $X = 1$ is given by

$$P_X(1) = P(X = 1) = P(\{(HT), (TH)\} = P(\{(HT)\} + P(\{(TH)\} = \frac{2}{4}.$$

The PMF of X is

x	0	1	2
$p_X(x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

Properties of PMFs

Let X be a discrete random variable with possible values denoted $x_1, x_2, \ldots$. The PMF of X, denoted p_X, must satisfy the following:

- $\sum_{x_i} p_X(x_i) = p_X(x_1) + p_X(x_2) + \ldots = 1$

- $p_X(x_i) \geq 0$ for all x_i
- Furthermore, if A is a subset of Ω, then the probability that X takes a value in A is given by

$$P(X \in A) = \sum_{xi \in A} p_X(x_i)$$

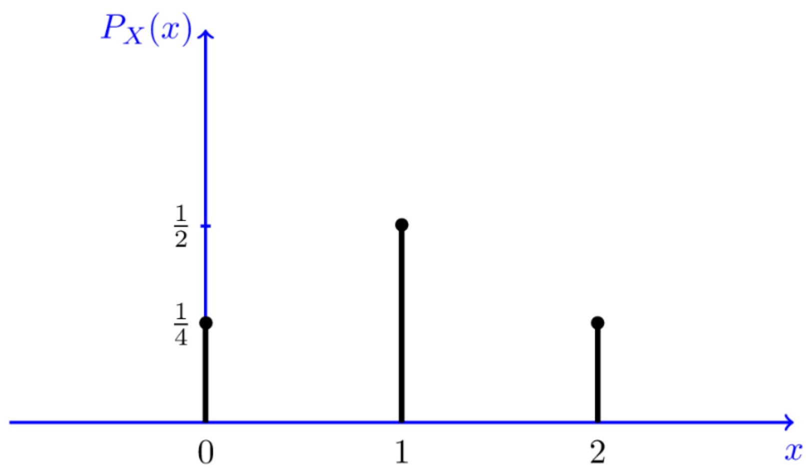


Figure 1.1: The PMF for random variable X in Example 1.1.2

Example 1.2.2
Let's consider the following pairs (x, p_x), where $\forall x \in R_x, p_x = P[X = x]$ such that the table represents a PMF of X?

x	-2	-1	0	1	2	3
$p_X(x)$	c	0.2	0.3	0.2	0.1	c

1.2.3 Cumulative Distribution Function

Definition

X is a discrete random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where R_X is the range of X and $p_X(x)$ is the PMF of X.
The CDF of X is defined as

$$\forall x \in \mathbb{R}, F_X(x) = P[X \leq x] = \sum_{x_i \leq x} P[X = x_i] = \sum_{xi_{\leq} x} p_X(x_i)$$

In Example 1.1.2, the CDF of X is

x	$x < 0$	0	1	2	$x > 2$
$p_X(x)$	0	0.25	0.5	0.25	0
$F_X(x)$	0	0.25	0.75	1	1

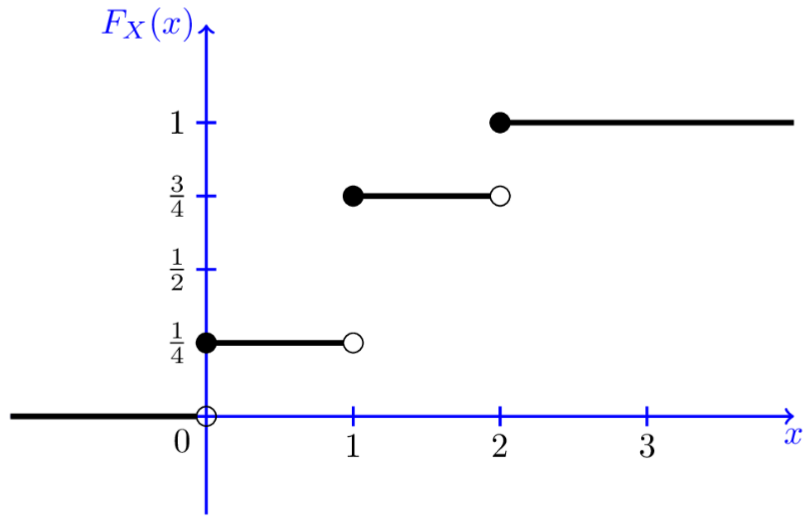


Figure 1.2: The CDF for random variable X in Example 1.1.2

Remarks

In addition to the aforementioned characteristics of CDF, we have:

- $F_X(x)$ is defined for all $\mathbb{R}$, not just R_X.
- If $x < x_1 \Rightarrow F_X(x) = 0$, and if $x \geq x_n \Rightarrow F_X(x) = 1$
- If $x_i \leq x < x_{i+1} \Rightarrow F_X(x) = F_X(x_i)$, i.e. $F(.)$ is constant in the interval $[x_i, x_{i+1}[$
- $\forall x_i \in R_X$, $p_X(x) = F_X(x_i) - F_X(x_{i-1})$

These characteristics are better illustrated in the Figure 1.3.

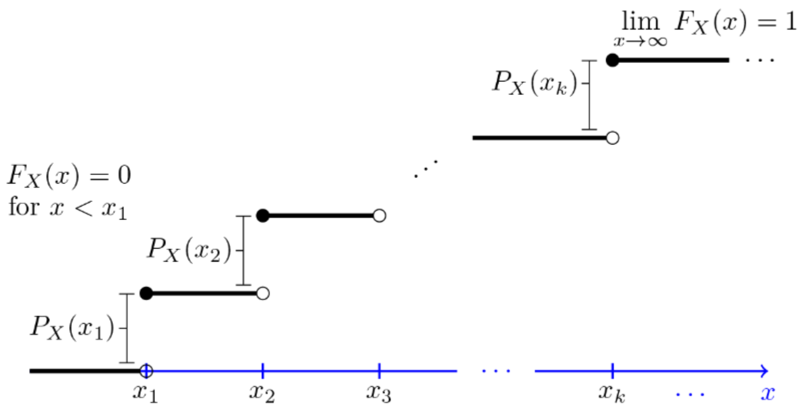


Figure 1.3: CDF of a discrete random Variable

Additional Remarks

- $P[a < X \leq b] = F_X(b) - F_X(a)$
- $P[a < X < b] = F_X(b) - F_X(a) - P[X = b]$
- $P[a \leq X \leq b] = F_X(b) - F_X(a) + P[X = a]$
- $P[a \leq X < b] = F_X(b) - F_X(a) + P[X = a] - P[X = b]$

Example 1.2.3
In Example 1.1.2, calculate the following probabilities:

1. $P[X \leq -2]$
2. $F_X(0.99)$
3. $P[1.5 < X \leq 2.5]$

1.2.4 Exercises

Exercise 1.2.1
The sample space of a random experiment is $\Omega = \{a, b, c, d, e, f\}$, and each outcome is equally likely. A random variable X is defined as follows:

ω	a	b	c	d	e	f
x	0	0	1.5	1.5	2	3

Determine the PMF of X.

Exercise 1.2.2
Let X be a discrete random variable with the following PMF:

$$p_X(x) = \begin{cases} \frac{3}{6} & \text{for} \quad x = 0 \\ \frac{2}{6} & \text{for} \quad x = 1 \\ \frac{1}{6} & \text{for} \quad x = 2 \\ 0 & \text{otherwise} \end{cases}$$

1. Find R_X, the range of the random variable X.
2. Find $P(X \geq 1.5)$.
3. Find $P(0 < X < 2)$.
4. Find $P(X = 0|X < 2)$.

Exercise 1.2.3
A coin is tossed three times. Let the random variable X be the number of tails.

1. Find the probability distribution of X.
2. Give the CDF of X.

1.3 Continuous Random Variables

We say that a random variable is continuous if the probability of observing any particular value in the real line is zero $\forall x \in \mathbb{R}, P(X = x) = 0$. Accordingly, the notion of probability distribution function is meaningless and we have to come up with something a bit different, though with a similar interpretation. The analogue of the probability distribution function for continuous random variables is the Probability Density Function (PDF). To understand the latter, we first note that, within the context of continuous random variables, it only makes sense to talk about the probability of observing a value within a given interval. The PDF then measures the mass of probability of an infinitesimal interval, that is to say, $Pr(x < X < x + \Delta x)$ for a very small $\Delta x > 0$. In the following definition, we formalize such a notion.

- **Example**
  The chances that a student hight is exactly 176.00 cm are very slim.

Definition

X is a random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $R_X \subseteq \mathbb{R}$. Let X be a random variable with a CDF $F_X(x)$. If $F_X(x)$ is continuous and also has a derivative $f_X(x) = \dfrac{dF_X(x)}{dx}$ which exists everywhere except at possibly a finite number of points and is piecewise continuous, then X is called a continuous random variable.

1.3.1 Probability Density Function

The $f_X(x)$ is such that

1. $f_X(x) \geq 0, \forall x \in \mathbb{R}$
2. $\int_{R_X} f_X(x)\, dx = 1$
3. $P[a \leq X \leq b] = \int_a^b f_X(x)\, dx$ for $-\infty < a < b < \infty$.

Important Remarks

- $\forall x \in \mathbb{R}, P(X = x) = 0$
- $f_X(x)$ does not express probabilities. $f_X(x) \neq P[X = x]$
- $f_X(x)$ is always non-negative and can exceed one.
- $f_X(x) = \dfrac{dF_X(x)}{dx}$
- $F_X(x_0) = \int_{-\infty}^{x_0} f_X(x)\, dx, \quad \forall x_0 \in \mathbb{R}$
- $\forall a, b \in \mathbb{R}, P[a < X < b] = P[a \leq X < b] = P[a < X \leq b] = P[a \leq X \leq b] = P[X \leq b] - P[X \leq a] = F_X(b) - F_X(a)$

Example 1.3.1 1. Let a random variable satisfy the following PDF

$$f_X(x) = \begin{cases} 2x, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

- Verify that $f_X(x)$ is a valid PDF.
- Compute the following probabilities $P\left(X \leq \frac{1}{2}\right)$ and $P\left(X \leq \frac{1}{2} / \frac{1}{3} < X < \frac{2}{3}\right)$.

2. Let X denote a random variable with a CDF

$$F_X(x) = \begin{cases} 1 - e^{-x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

- Find the PDF $f_X(x)$

1.3.2 Exercises

Exercise 1.3.1
Let X be a continuous random variable with the following PDF:

$$f_X(x) = \begin{cases} ce^{-x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

1. Find c.
2. Find the CDF of X, $F_X(x)$.
3. Find $P(1 < X < 3)$.

Exercise 1.3.2
Suppose that $f_X(x) = \frac{x}{8}$ for $3 < x < 5$. Determine the following probabilities:

1. is $f_X(x)$ a valid PDF?
2. $P(X < 4)$
3. $P(X > 3.5)$
4. $P(4 < X < 5)$
5. $P(X < 3.5 \text{ or } X > 4.5)$

Exercise 1.3.3
The daily expenditure X, in euros, spent by a certain customer in a department store is distributed according to the following density function:

$$f(x) = \begin{cases} \frac{x^2}{9000}, & \text{if } 0 < x < 30, \\ 0, & \text{otherwise.} \end{cases}$$

1. Is this a valid PDF?
2. Calculate the probability that the client has a daily spending between 15 and 20 euros.
3. What is his expected daily expenditure? Calculate the variance of spending.
4. Next winter, the department store is granting the following "discount vouchers," whose amount depends on the client's expenditure:

Expenditure Range (Euros)	Discount Voucher (Euros)
Between 10 and 20	1
Between 20 and 25	1.5
Exceeds 25	3

Derive the distribution and the expected value of the discounts obtained by the client.

5. The customer has paid for his purchases in the store with a credit card, which applies a 4% charge. What is the probability that the customer has paid (with charges included) between 10 and 15 euros?

1.4 Probabilistic Characteristics of Random Variables

In this section, we look at various numerical characteristics of discrete and continuous random variables. This give us a way of classifying and comparing random variables.

1.4.1 Expected Value

The Case of Discrete Random Variables

If X is a discrete random variable with R_X, and a PMF $p_X(x)$, then the expected value (or mean) is given by

$$E[X] = \sum_{x \in R_X} x\, p_X(x)$$

The expected value of X may also be denoted as μ_X or simply μ.

Notes

- $E(X)$ is a weighted average where the probabilities $p_X(x)$ are the weights.
- $E(X)$ is bordered by $\text{Min}(X)$ and $\text{Max}(X)$.

Example

Find the expected value in Example 1.1.1.

x	0	1	2	$\sum$
$p_X(x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	1
$x * p_X(x)$	0	$\frac{2}{4}$	$\frac{2}{4}$	1

The Case of Continuous Random Variables

If X is a continuous random variable with a PDF $f_X(x)$, then the expected value (or mean) of X is given by

$$\mu = \mu_X = E(X) = \int_{R_X} x\, f_X(x) \quad dx$$

The formula for the expected value of a continuous random variable is the continuous analogue of the expected value of a discrete random variable, where instead of summing over all possible values we integrate.

Example 1.4.1
Let the random variable X denote the time a person waits for an elevator to arrive. Suppose the longest one would need to wait for the elevator is 2 minutes, so that the possible values of X (in minutes) are

given by the interval $[0,2]$. The PDF for X is given by

$$f_X(x) = \begin{cases} x, & \text{for } 0 \le x \le 1 \\ 2-x, & \text{for } 1 < x \le 2 \\ 0, & \text{otherwise} \end{cases}$$

- Verify that $f_X(x)$ is a viable PDF.
- Calculate the probability that a person waits less than 30 seconds.
- Find the CDF of X.
- Compute the mean.

Expected Value of Functions of a Random Variable

Let X be a random variable, and let $Y = g(X)$. Suppose that we are interested in finding $E(Y)$. One way to find $E(Y)$ is to first find the distribution of Y and then use the expectation formula. But there is another way which is usually easier. It is called the law of the unconscious statistician.

The law of the unconscious statistician states

$$E(g(X)) = \begin{cases} \sum_{R_X} g(x) * p_X(x), & \text{if X is a discrete R.V.} \\ \int_{R_X} g(x) * f_X(x)\, dx, & \text{if X is a continuous R.V.} \end{cases}$$

Example 1.4.2
Find the expected value of $Y = g(X) = X^2$ in examples 1.1.1 and **??**.

Characteristics of The Expected Value

let X be a R.V. and $\forall a, b \in \mathbb{R}$:

1. $E[a] = a$
2. $E[a\,X + b] = a\,E[X] + b$
3. $E[a\,g(X) + b\,h(X)] = a\,E[g(X)] + b\,E[h(X)]$

1.4.2 Variance

Let X be a R.V. and R_X is its range of values. The variance of a random variable X is given by

$$\sigma^2 = V(X) = Var(X) = E\left[(X - E(X))^2\right] = \begin{cases} \sum_{R_X} (X - E(X))^2 * p_X(x), & \text{if X is a discrete R.V.} \\ \int_{R_X} (X - E(X))^2 * f_X(x)\, dx, & \text{if X is a continuous R.V.} \end{cases}$$

In other words, the variance of a random variable is the average of the squared deviations of the random variable from its mean (expected value).

Example 1.4.3
Find the variance of X in examples 1.1.1 and **??**.

Koenig-Huyghens Formula for the Variance

To compute the variance of a random variable X, we usually employ the following formula

$$V(X) = E\left[X^2\right] - E\left[X\right]^2$$

Standard Deviation

Notice that the variance of a random variable will result in a number with units squared. On the other hand, the standard deviation that has the same units as the random variable X. Thus, the standard deviation is easier to interpret, which is why we make a point to define it.

$$\sigma = \sqrt{V(X)}$$

The variance and standard deviation give us a measure of spread (dispersion or variability) for random variables. The standard deviation is interpreted as a measure of how "spread out" the possible values of X are with respect to the mean of X, $\mu = E[X]$.

Characteristics of the Variance

let X be a R.V. and $\forall a, b \in \mathbb{R}$:

1. $V[a] = 0$
2. $V[a\,X + b] = a^2\,V[X]$

1.4.3 Moments

As with expected value and variance, the moments of a random variable are used to characterize the distribution of the random variable and to compare the distribution to that of other random variables.

Non-Central Moments

The r^th raw moment $(\forall r \in \mathbb{N}^*)$ of a random variable X, it is also called non-central moment or moment about the origin, denoted by $\mu_r^{'}$ is defined as

$$\mu_r^{'} = E\left[X^r\right]$$

$$\mu_r^{'} = \begin{cases} \sum_{R_X} x^r * p_X(x), & \text{if X is a discrete R.V.} \\ \int_{R_X} x^r * f_X(x)\,dx, & \text{if X is a continuous R.V.} \end{cases}$$

- The expected value of X, $E[X]$, is the first raw moment.
- When X is continuous. The r^th moment about the origin is only defined if $E[X^r]$ exists.

Central Moments

The r^th central moment ($\forall r \in \mathbb{N}^*$) of a random variable X, it is also called moment about the mean, denoted by μ_r is defined as

$$\mu_r = E\left[(X - E(X))^r\right]$$

$$\mu_r = \begin{cases} \sum_{R_X} \left[X - E(X)\right]^r * p_X(x), & \text{if X is a discrete R.V.} \\ \int_{R_X} \left[X - E(X)\right]^r * f_X(x)\, dx, & \text{if X is a continuous R.V.} \end{cases}$$

- Note that $\mu_1 = 0$.
- The variance of X, $V[X]$, is the second central moment.
- When X is continuous. The r^th moment about the mean is only defined if $E\left[(X - E(X))^r\right]$ exists.

Factorial Moments

The r^th factorial moment ($\forall r \in \mathbb{N}^*$) of a random variable X, denoted by $\mu_{[r]}$ is defined as

$$\mu_{[r]} = E\left[X(X-1)(X-2)\ldots(X-r+1)\right]$$

$$\mu_{[r]} = \begin{cases} \sum_{R_X} \left[X(X-1)(X-2)\ldots(X-r+1)\right] * p_X(x), & \text{if X is a discrete R.V.} \\ \int_{R_X} \left[X(X-1)(X-2)\ldots(X-r+1)\right] * f_X(x)\, dx, & \text{if X is a continuous R.V.} \end{cases}$$

- Note that $\mu_{[1]}$ is the expected value of X.
- Note that $\mu_{[2]}$ is the variance of X.

Example 1.4.4
Find $\mu_3^{'}$, μ_3, and $\mu_{[3]}$ in Example 1.1.1.

1.4.4 Moment-Generating Function

Moments can be calculated directly from the definition, but, even for moderate values of r, this approach becomes cumbersome. The next definition and theorem provide an easier way to generate moments.

The moment-generating function (MGF) of a random variable X is given by

$$M_X(t) = E\left[e^{tx}\right], \quad \text{for } t \in \mathbb{R}$$

$$M_X(t) = \begin{cases} \sum_{R_X} e^{tx} * p_X(x), & \text{if X is a discrete R.V.} \\ \int_{R_X} e^{tx} * f_X(x)\, dx, & \text{if X is a continuous R.V.} \end{cases}$$

MGF and raw moments

If random variable X has MGF $M_X(t)$, then

$$M_X^{(r)}(0) = \frac{d^r}{dt^r}\left[M_X(t)\right]_{t=0} = E\left[X^r\right]$$

Example 1.4.5
Find the MGF of the following probability distribution, then obtain $E\left[X^r\right]$ for $r = 1$ and $r = 2$.

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

1.4.5 Mode, Median, Quartiles, and Other Characteristics

Mode

The mode M_o is the most likely possible value. For a random variable, it is the value for which the probability mass function (PMF) or probability density function (PDF) attains its maximum.

$$M_O = \begin{cases} x, x \in R_X / Max\left[p_X(x)\right], & \text{if X is a discrete R.V.} \\ x, x \in R_X / Max\left[f_X(x)\right], & \text{if X is a continuous R.V.} \end{cases}$$

Median

The median of a set of numbers is that number (or numbers) which lies in the middle of the set once the numbers are arranged into ascending or descending order. The median of a random variable X is any number M_e that satisfies

$$P\left[X \leq M_e\right] \geq \frac{1}{2} \quad P\left[X \geq M_e\right] \geq \frac{1}{2}$$

- The previous condition can be rewritten as

$$M_e = Inf\left\{x, x \in R_X \quad | \quad F_X(x) \geq \frac{1}{2}\right\}$$

- If X is a continuous R.V. then

$$F(M_e) = \frac{1}{2} \Leftrightarrow M_e = F^{-1}(\frac{1}{2})$$

Example 1.4.6
Find the mode and median in examples 1.1.1, **??**, and 1.4.1.

Quartiles

Let X be a random variable with a CDF $F_X(x)$. Let $0 < p < 1$. The p^{th} quantile or p^{th} percentile of X is defined to be the first $x_{(p)}$ such that $F_X(x_{(p)}) \geq p$.

$$x_{(p)} = Inf\left\{x, x \in R_X \quad | \quad F_X(x) \geq p\right\}$$

If X is a continuous R.V. then

$$F(x_{(p)}) = p \Leftrightarrow x_{(p)} = F^{-1}(p)$$

- $p = 0.5 \Rightarrow x_{(p)} = M_e$.
- $p \in \{0.25, 0.5, 0.75\} \Rightarrow x_{(p)}$ represents the quantiles Q_1, Q_2, and Q_3.

- $p \in \{0.1, 0.2, 0.3, \ldots, 0.9\} \Rightarrow x_{(p)}$ represents the deciles D_1, D_2, …, and D_9.
- $p \in \{0.01, 0.02, 0.03, \ldots, 0.99\} \Rightarrow x_{(p)}$ represents the percentiles P_1, P_2, …, and P_{99}.

Skewness and Kurtosis

As discussed in descriptive statistics, the expected value, median, and mode are considered measures of central tendency. The variance and standard deviation shows the dispersion of the distribution of the random variable X around its mean. Moreover, the quartiles are known as measures of position. Skewness and Kurtosis are measures of shape. Skewness (γ_1)measures the lack of symmetry in the distribution of X, whereas kurtosis (γ_2)measures the tailedness of a distribution of X. These statistical tools collectively contribute to a comprehensive understanding of the features of the studied random variable.

- Skewness

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{E(X-\mu)^3}{\sigma^3}$$

- Kurtosis

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{E(X-\mu)^4}{\sigma^4} - 3$$

1.4.6 Chebyshev's Inequality

Chebyshev's inequality expresses the relationship between $\mu = E(X)$ and $\sigma^2 = V(X)$. Chebyshev's theorem is given as

$$P\left[\mu - k\,\sigma \leq X \leq \mu + k\,\sigma\right] \geq 1 - 1/k^2$$

In other words,

$$P\left[|X-\mu| \leq k\,\sigma\right] \geq 1 - 1/k^2$$

and consequently,

$$P\left[|X-\mu| > k\,\sigma\right] < 1/k^2$$

Chebyshev's inequality states that the probability that X will fall within k standard deviation of the mean is at least $1 - 1/k^2$, where $k > 1$.

Example 1.4.7
A sample of size $n = 50$ has a mean $\bar{x} = 28$ and a standard deviation $s = 3$. Without knowing anything else about the sample, we want to determine the number of observations that lie within the interval $(22, 34)$, and those that lie outside it.

Solution

The interval $(22, 34)$ is formed by adding and subtracting two standard deviations from the mean:

$$28 - 2(3) = 22 \quad \text{and} \quad 28 + 2(3) = 34.$$

By Chebyshev's Theorem, at least $\frac{3}{4}$ of the data must lie within two standard deviations of the mean. Therefore, at least $\frac{3}{4}$ of the 50 observations are in the interval.

$$\frac{3}{4} \times 50 = 37.5.$$

Since we cannot have a fractional observation, we conclude that at least 38 observations lie within the interval $(22, 34)$.

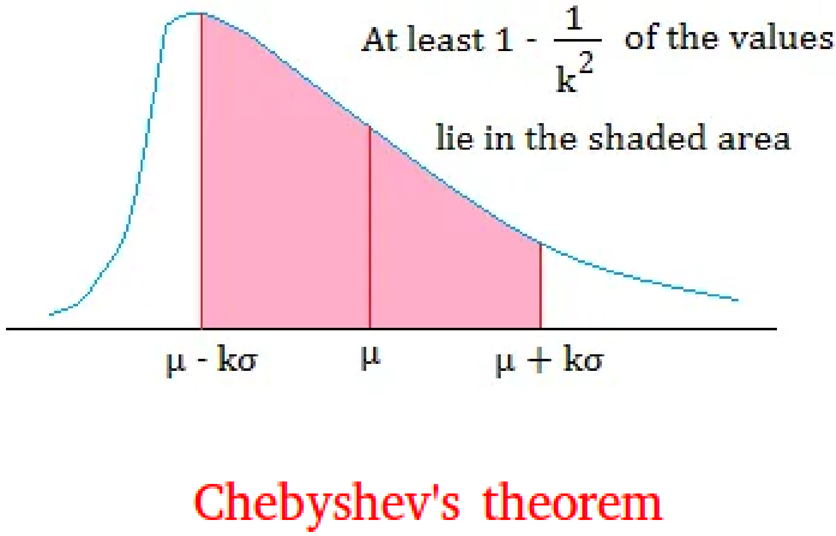


Figure 1.4: Chebyshev's theorem.

Now, if at least $\frac{3}{4}$ of the observations are within the interval, at most $\frac{1}{4}$ of the observations are outside the interval. Therefore:

$$\frac{1}{4} \times 50 = 12.5.$$

Since fractional observations are not possible, at most 12 observations lie outside the interval.

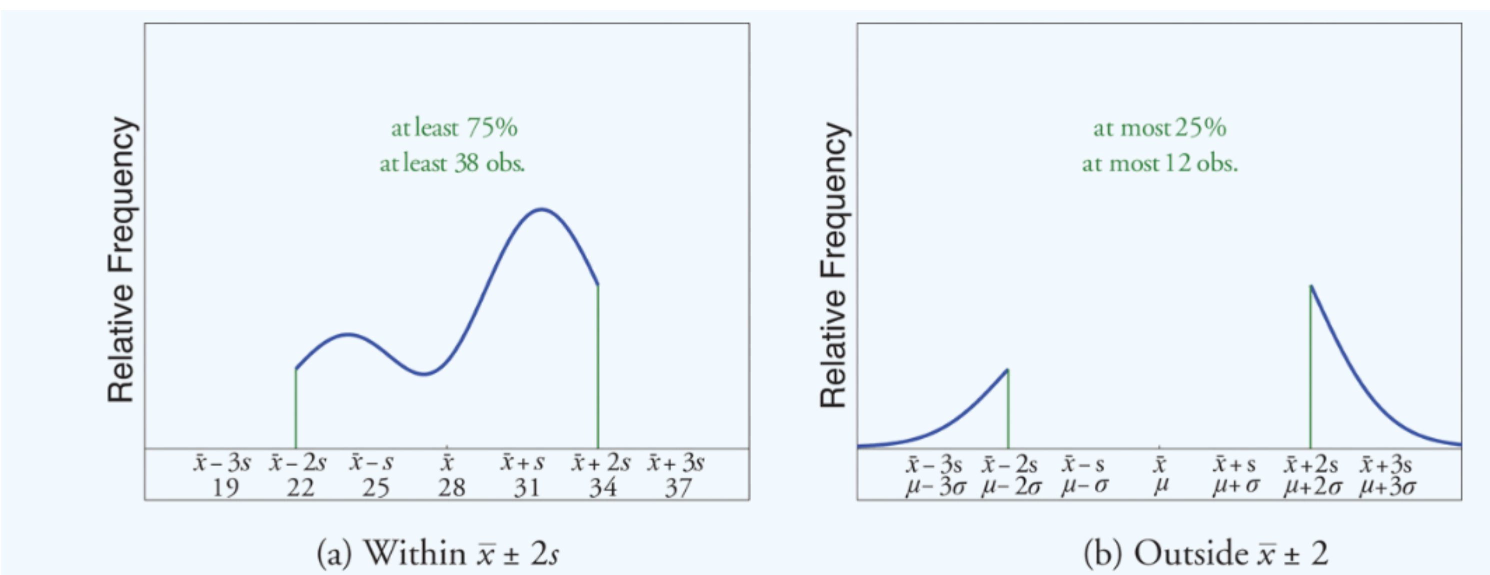


Figure 1.5: Within and outside observations according to Chebyshev theorem.

Example 1.4.8
Let X be the GPA of a freshman student at ESC. The mean and variance of X are 8.4 and 1.44 respectively.

- What is the probability that X is between 6 and 10.8.
- What is the probability that X is between 7 and 10.

Example 1.4.9
Suppose X is the sum of two rolls of a fair die.

- Find the PMF and CDF of X.
- Compute μ and σ.
- Using Chebyshev's theorem, estimate the probability that X falls within 2.17 and 11.83.

Markov's Inequality

Suppose that the R.V. X takes only non-negative values and that $\mu = E(X)$ is assumed to be finite. Let c be any positive number. Then,

$$P[X \geq c] \leq \frac{\mu}{c}$$

Example 1.4.10
Suppose we have a random variable X representing income in a population, with an expected value $\mathbb{E}[X] = 50000\,DA$ (the average income).
What is the probability that someone earns at least 200000 DA?

By Markov's Inequality, for any $c > 0$, we have:

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Setting $a = 200000$, we find:

$$P(X \geq 200000) \leq \frac{50000}{200000} = 0.25$$

Thus, the probability that someone earns at least \$200,000 is at most 25%.
The virtue of Chebyshev's and Markov's inequalities is that they make no restrictive assumptions on the random variable X. Whenever μ and σ are finite, Chebyshev's inequality is applicable, and whenever μ is finite, Markov's inequality applies, provided the random variable is non-negative.

1.4.7 Exercises

Exercise 1.4.1
Let X be a discrete random variable with the following PMF:

$$P_X(x) = \begin{cases} 0.1, & \text{for } x = 0.2 \\ 0.2, & \text{for } x = 0.4 \\ 0.2, & \text{for } x = 0.5 \\ 0.3, & \text{for } x = 0.8 \\ 0.2, & \text{for } x = 1 \\ 0, & \text{otherwise} \end{cases}$$

1. Find R_X, the range of the random variable X.
2. Determine the CDF of X.
3. Find $P(X \leq 0.5)$.
4. Find $P(0.25 < X < 0.75)$.

5. Find $P(X = 0.2 \mid X < 0.6)$.

6. Find the MFG of X.

Exercise 1.4.2
Let X be a continuous random variable with PDF:

$$f_X(x) = \begin{cases} x + \frac{1}{2}, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

1. Is $f_X(x)$ a valid PDF.

2. Find $E(X^n)$, where $n \in \mathbb{N}$.

3. Determine $E(X)$, σ, M_e, and M_o.

Exercise 1.4.3
Let X be a random variable with the PDF given by:

$$f_X(x) = \begin{cases} cx^2, & |x| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

1. Find the constant c.

2. Determine the CDF of X.

3. Find $P(X \geq \frac{1}{2})$.

4. Find $E(X)$, $V(X)$, M_e, and M_o.

Exercise 1.4.4
A casino advertises that the expected prize won by its clients is 600 dinars, with risk $\sigma = 360$.

1. Calculate the probability that a player wins between 150 and 1050 euros.

2. How would the above probability change if $\sigma = 420$?

3. What is the probability that the prize won by a player deviates at least 504 dinars from its expectation?

Usual Probability Distributions

In the next two section, we introduce families of common probability distributions, i.e., probability distributions for discrete and continuous random variables. We refer to these as ”families” of distributions because in each case we will define a PMF or PDF by specifying an explicit formula, and that formula will incorporate a constant (or set of constants) that are referred to us parameters. By specifying values for the parameter(s) in the PMF/PDF, we define a specific probability distribution for a specific random variable. For each family of distributions introduced, we will list a set of defining characteristics that will help determine when to use a certain distribution in a given context.

Chapter 2

Standard Discrete Probability distributions

2.1 Discrete Uniform Distribution

2.1.1 Definition

Let X be a discrete R.V. that takes the values x_i; where $x_i \in R_X$ and $R_X = \{x_1, x_2, \dots, x_n\}$. X is said to have a discrete uniform distribution if and only if it has the following PMF:

$$X \sim \mathrm{U}\left[|R_X|\right] \Leftrightarrow \forall x_i \in R_X, \quad p_X(x) = \frac{1}{|R_X|} = \frac{1}{n}$$

Example

One of the simplest probability distributions is that of a digit taken at random. In this case the integers $0, 1, 2, \dots, 9$ are to be equally likely. The probabilities are

$$P\left[\text{integer chosen} = i\right] = \frac{1}{10}; \quad i = 0, 1, 2, \dots, 9.$$

2.1.2 Special case

The discrete uniform distribution on $R_X = \{1, 2, \dots, n\}$ is defined by the PMF:

$$X \sim \mathrm{U}[n] \Leftrightarrow \forall x \in R_X, \quad p_X(x) = \frac{1}{n}$$

Clearly for any given integer x where $1 \leq x \leq n$ the CDF of X is

$$F_X(x) = \frac{x}{n}$$

Example

X is the obtained outcome of rolling a fair six-sided die. $X \sim \mathrm{U}[n = 6]$. The PMF of of rolling a fair six-sided die once is illustrated in the following figure.

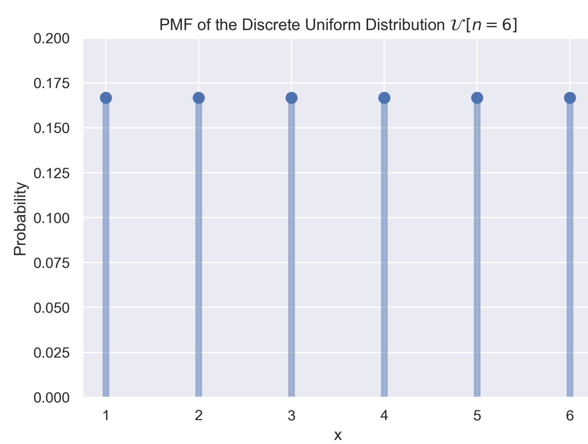


Figure 2.1: The PMF of $X \sim \mathrm{U}[n = 6]$

2.1.3 Moments

In the **special case**, the first few moments are found easily. For example,

$$\mu = \sum_{x=1}^{n} x\, p(x) = \frac{n+1}{2}$$

Similarly,

$$E(X^2) = \sum_{x=1}^{n} x^2\, p(x) = \frac{(n+1)(2n+1)}{6}$$

Therefore,

$$\sigma^2 = \frac{n^2-1}{12}$$

It follows from the trivial symmetric nature of the discrete uniform distribution that

$$\mu_3 = E\left[(X-\mu)^3\right] = 0$$

We can also find

$$\mu_4 = E\left[(X-\mu)^4\right] = \frac{(3n^2-7)(n^2-1)}{240}$$

Thus, the skewness and the kurtosis of the discrete uniform distribution are

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = 0$$

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{-6\left(n^2+1\right)}{5\left(n^2-1\right)}$$

In addition, the MGF is expressed as

$$M_X(t) = E[e^{tx}] = \frac{1}{n}\sum_{x=1}^{n} e^{tx}$$

Remember that,

$$\sum_{x=1}^{n} x^2 = \frac{n\left(n+1\right)\left(2n+1\right)}{6}$$

$$\sum_{x=1}^{n} x^3 = \frac{n^2\left(n+1\right)^2}{4}$$

$$\sum_{x=1}^{n} x^4 = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}$$

To sum up, the probabilistic characteristics of the special case of the discrete uniform distribution are given in the following table.

Table 2.1: Probabilistic characteristics of the special case of the uniform distribution

X	$\boldsymbol{\mathcal{U}\,[n]}$
Range of X	$\{1, 2, 3, \ldots, n\}$
PMF	$p_X(x) = \dfrac{1}{n}; \quad \text{if} \quad x \in R_X$
CDF	$F_X(x) = \dfrac{x}{n}; \quad \text{if} \quad 1 \leq x \leq n$
MGF	$M_X(t) = \dfrac{1}{n} \sum_{x=1}^{n} e^{tx}$
Expected Value	$\mu = \dfrac{n+1}{2}$
Variance	$\sigma^2 = \dfrac{n^2-1}{12}$
Third Central Moment	$\mu_3 = 0$
Fourth Central Moment	$\mu_4 = \dfrac{(3n^2-7)(n^2-1)}{240}$
Skewness	$\gamma_1 = 0$
Kurtosis	$\gamma_2 = \dfrac{-6\,(n^2+1)}{5\,(n^2-1)}$

Example 2.1.1
Let the random variable X have a discrete uniform distribution on the integers $1 \leq x \leq 3$.

1. Determine μ, σ, γ_1 and γ_2 of X.

2. Answer the same question, assuming that $0 \leq x \leq 100$.

2.1.4 Exercises

Exercise 2.1.1
A deck of traditional cards has 52 cards of four suits, hearts, diamonds, clubs, and spades. A card is selected at random.

1. Express this selection using the discrete uniform distribution identifying its parameter.

Exercise 2.1.2
10 ping pong balls are numbered 1 through 10 and placed in a bag. One ping pong ball is removed from the bag randomly.

1. Find the expected value, variance, median, and mode.

2. Find the probability that the number on the drawn ping pong ball is between 7 and 10.

2.1.5 Python Implementation

Contents

Usual Probability Distributions

In the next two section, we introduce families of common probability distributions, i.e., probability distributions for discrete and continuous random variables. We refer to these as "families" of distributions because in each case we will define a PMF or PDF by specifying an explicit formula, and that formula will incorporate a constant (or set of constants) that are referred to us parameters. By specifying values for the parameter(s) in the PMF/PDF, we define a specific probability distribution for a specific random variable. For each family of distributions introduced, we will list a set of defining characteristics that will help determine when to use a certain distribution in a given context.

Chapter 1

Standard Discrete Probability distributions

1.1 Discrete Uniform Distribution

1.1.1 Definition

Let X be a discrete R.V. that takes the values x_i; where $x_i \in R_X$ and $R_X = \{x_1, x_2, \ldots, x_n\}$. X is said to have a discrete uniform distribution if and only if it has the following PMF:

$$X \sim \mathrm{U}\left[|R_X|\right] \Leftrightarrow \forall x_i \in R_X, \quad p_X(x) = \frac{1}{|R_X|} = \frac{1}{n}$$

Example

One of the simplest probability distributions is that of a digit taken at random. In this case the integers $0, 1, 2, \ldots, 9$ are to be equally likely. The probabilities are

$$P\left[\text{integer chosen} = i\right] = \frac{1}{10}; \quad i = 0, 1, 2, \ldots, 9.$$

1.1.2 Special case

The discrete uniform distribution on $R_X = \{1, 2, \ldots, n\}$ is defined by the PMF:

$$X \sim \mathrm{U}[n] \Leftrightarrow \forall x \in R_X, \quad p_X(x) = \frac{1}{n}$$

Clearly for any given integer x where $1 \leq x \leq n$ the CDF of X is

$$F_X(x) = \frac{x}{n}$$

Example

X is the obtained outcome of rolling a fair six-sided die. $X \sim \mathrm{U}[n = 6]$. The PMF of of rolling a fair six-sided die once is illustrated in the following figure.

1.1.3 Moments

In the **special case**, the first few moments are found easily. For example,

$$\mu = \sum_{x=1}^{n} x\, p(x) = \frac{n+1}{2}$$

Similarly,

$$E(X^2) = \sum_{x=1}^{n} x^2\, p(x) = \frac{(n+1)(2n+1)}{6}$$

Therefore,

$$\sigma^2 = \frac{n^2 - 1}{12}$$

It follows from the trivial symmetric nature of the discrete uniform distribution that

$$\mu_3 = E\left[(X - \mu)^3\right] = 0$$

2.2 Bernoulli Distribution

Several of the discrete probability distributions discussed in this chapter are based on experiments or processes in which a sequence of trials, called Bernoulli trials, are conducted.

2.2.1 Bernoulli Trials

A Bernoulli trial results in one of two mutually exclusive outcomes, typically denoted S (for Success) and F (for Failure).

For example, tossing a coin is a Bernoulli trial since only one of two different outcomes can occur, head (H) or tail (T).

Characteristics

The characteristics of a Bernoulli trial are:

- The trial results in one of two mutually exclusive outcomes. (We denote one outcome by S and the other by F).
- The outcomes are exhaustive, i.e., no other outcomes are possible.
- The probabilities of S and F are denoted by p and q, respectively. That is, $P(S) = p$ and $P(F) = q$. Note that $p + q = 1$.

2.2.2 The Indicator Function and Bernoulli Distribution

Let A be an event in a sample space Ω. Suppose we are only interested in whether or not the outcome of the underlying probability experiment is in the specified event A. To track this we can define an indicator random variable, denoted I_A, given by

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \in \bar{A} \end{cases}$$

In other words, the random variable I_A will equal 1 if the resulting outcome is in event A, and I_A equals 0 if the outcome is not in A. Thus, I_A is a discrete random variable. We can state the probability mass function of I_A in terms of the probability that the resulting outcome is in event A, i.e., the probability that event A occurs, $P(A)$:

$$p(1) = P(I_A = 1) = P(A)$$

$$p(0) = P(I_A = 0) = P(\bar{A}) = 1 - p(A)$$

In the above Example, the random variable I_A is a Bernoulli random variable because its PMF has the form of the Bernoulli probability distribution, which we define next.

Definition

X is defined as the numerical outcome of a Bernoulli trial, where $X = 1$ if a success occurs and $X = 0$ if a failure occurs. Consequently, the probability distribution for $X = x$ is shown in the following table:

Outcome	x	$p_X(x)$
Success	1	p
Failure	0	q

In other words,

$$X \sim \mathcal{B}(p) \Leftrightarrow P\left[X = x\right] = p^x q^{1-x}; \quad x \in \{0, 1\}$$

Which is the same as,

$$P(X = x) = \begin{cases} p, & \text{if } x = 1 \\ q, & \text{if } x = 0 \end{cases}$$

The graphical representation of the PMF of a Bernoulli distribution is given in Figure 2.2.

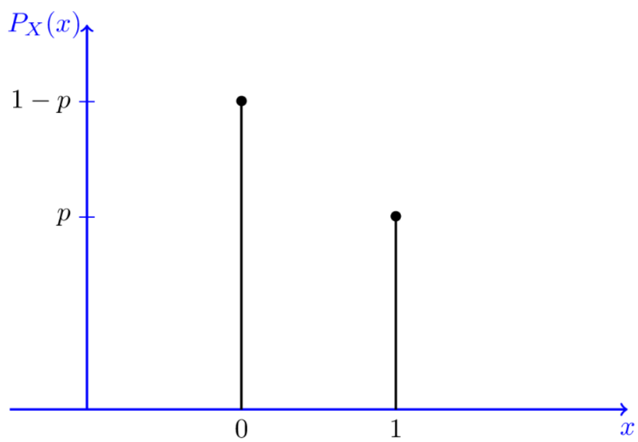


Figure 2.2: The PMF of $X \sim \mathrm{B}[p]$

2.2.3 Moments

The mean and variance of the Bernoulli random variable are, respectively,

- $\mu = p$
- $\sigma^2 = p\,q$

Example 2.2.1
In the Bernoulli coin-tossing experiment, define H as a success and T as a failure. Then $X = 1$ if H occurs and $X = 0$ if T occurs. Since $P(H) = P(T) = 0.5$ if the coin is balanced, the probability distribution for X is

$$P(X = x) = \begin{cases} 0.5, & \text{if } x = 1 \\ 0.5, & \text{if } x = 0 \end{cases}$$

- $\mu = 0.5$
- $\sigma^2 = 0.25$

The probabilistic characteristics of the Bernoulli distribution are given in the following table.

Table 2.2: Probabilistic characteristics of the Bernoulli distribution

X	$\boldsymbol{\mathcal{B}\ [p]}$
Range of X	$R_X = \{0, 1\}$
Parameters	$0 < p < 1; q = 1 - p$
PMF	$p_X(x) = p^x q^{1-x}; \quad \text{if} \quad x \in R_X$
MGF	$M_X(t) = q + p\, e^t$
Expected Value	$\mu = p$
Variance	$\sigma^2 = p\, q$
Third Central Moment	$\mu_3 = p\, q\, (q - p)$
Fourth Central Moment	$\mu_4 = p\, q\, (1 - 3\, p\, q)$
Skewness	$\gamma_1 = \dfrac{q - p}{\sqrt{p\, q}}$
Kurtosis	$\gamma_2 = \dfrac{1 - 6\, p\, q}{p\, q}$

A Bernoulli random variable, by itself, is of little interest in engineering and science applications. Conducting a series of Bernoulli trials, however, leads to some well-known and useful discrete probability distributions. One of these is described in the next section.

Example 2.2.2
Using the table above, compare the characteristics of the random variables X and Y, where

- X represents the number of heads when a fair coin is rolled once.
- Y represents the number of heads when an unfair coin is flipped once. (the probability of getting a tail is 0.6).

2.2.4 Python Implementation

Bernoulli

[1]:
```
import numpy as np
from scipy.stats import bernoulli
from IPython.display import display, Math
import matplotlib.pyplot as plt
```

1 Parameters

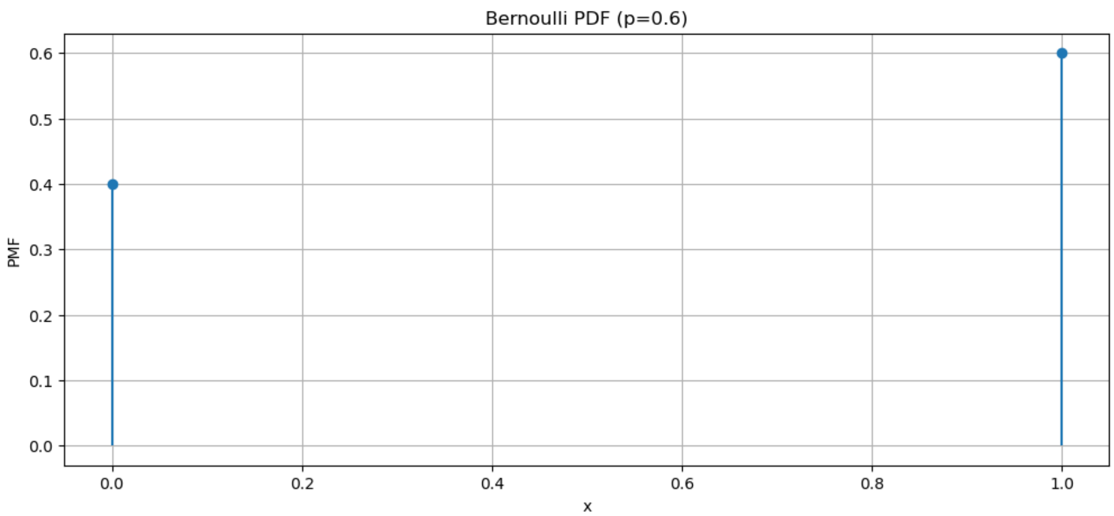
[2]:
```
# Set probability of success (p)
p = 0.6
```

2 Probability Mass Function (PMF)

[3]:
```
PMF = bernoulli.pmf([0, 1], p)
print("PMF values:", PMF)
```

```
PMF values: [0.4 0.6]
```

[4]:
```
plt.figure(figsize=(12, 5))
x_values = range(0, 2)
plt.stem(x_values, PMF, basefmt=" ")
plt.xlabel('x')
plt.ylabel('PMF')
plt.title(f'Bernoulli PDF (p={p})')
plt.grid(True)
plt.show()
```

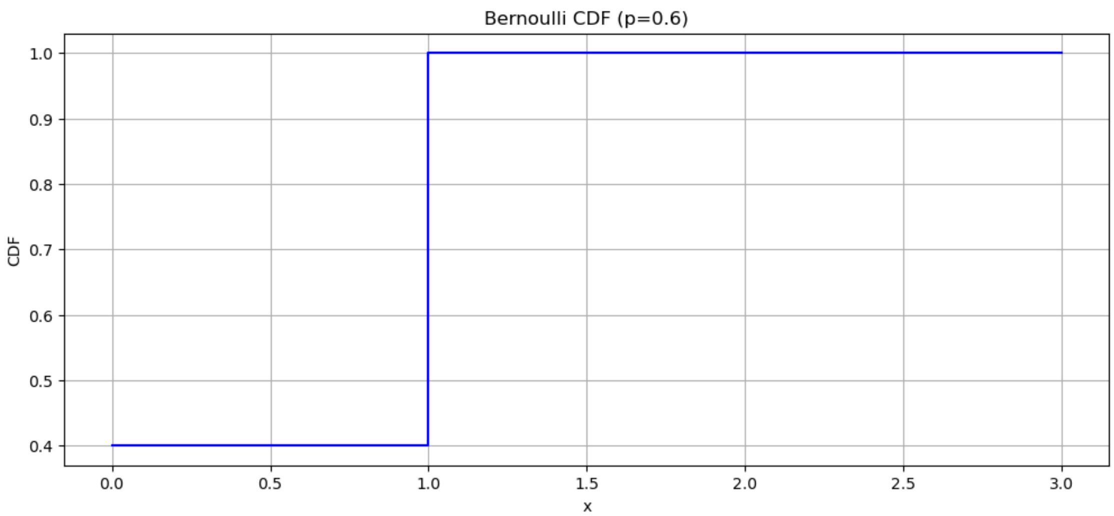


3 Cumulative Distribution Function (CDF)

[5]:
```
CDF = bernoulli.cdf(range(0, 4), p)
print("CDF values:", CDF)
```

```
CDF values: [0.4 1.  1.  1. ]
```

[6]:
```
plt.figure(figsize=(12, 5))
x_values = range(0, 4)
plt.step(x_values, CDF, where='post',  color='blue')
plt.xlabel('x')
plt.ylabel('CDF')
plt.title(f'Bernoulli CDF (p={p})')
plt.grid(True)
plt.show()
```



4 Expected value or mean (μ)

[7]:
```
mean = bernoulli.mean(p)
display(Math(r"\mu = " + str(mean)))
```

$\mu = 0.6$

5 Variance (σ^2)

[8]:
```
variance = bernoulli.var(p)
display(Math(r"\sigma^2 = " + str(variance)))
```

$\sigma^2 = 0.24$

6 Skewness(γ_1)

[9]:
```
skewness = bernoulli.stats(p, moments='s')
display(Math(r"\gamma_1 = " + str(skewness)))
```

$\gamma_1 = -0.4082482904638634$

7 Excess Kurtosis (γ_2)

[10]:
```
excess_kurtosis = bernoulli.stats(p, moments='k')
display(Math(r"\gamma_2 = " + str(excess_kurtosis)))
```

$\gamma_2 = -1.8333333333333308$

8 Median M_e

[11]:

```
median = bernoulli.median(p)
display(Math(r"M_e = " + str(median)))
```

$M_e = 1.0$

9 Mode M_o

[12]:

```
mode = np.argmax(PMF)
display(Math(r"$M_o$ = " + str(mode)))
```

$M_o = 1$

2.3 Binomial Distribution

2.3.1 Preamble

The binomial probability distribution is derived as follows. A simple event for a binomial experiment consisting of n Bernoulli trials can be represented by the symbol

$$SSFSFSSSF\dots FFS$$

where the letter in the i^{th} position, proceeding from left to right, denotes the outcome of the i^{th} trial (success or failure).

Let Y **be a random variable that represents the number of successes** (S) **in the** n **trails**.

Since we want to find the probability $p(Y = y)$ of observing y successes in the n trials, we will need to sum the probabilities of all simple events that contain y successes and $(n - y)$ failures. Such simple events would appear symbolically as

$$\overbrace{SSSS\dots S}^{y}\ \overbrace{FF\dots F}^{(n-y)}$$

Given that the trials are independent (the selection is made with replacement), the probability of a particular simple event implying y successes is

$$\overbrace{P(SSS\dots S}^{y}\ \overbrace{FF\dots F)}^{(n-y)} = p^{y}q^{n-y}$$

The number of these equiprobable simple events is equal to the number of ways we can arrange the y successes and the $(n - y)$ failures in n positions corresponding to the n trials. This is equal to the number of ways of selecting y positions (trials) for the y successes from a total of n positions. This number is

$$C_n^y = \frac{n!}{y!(n-y)!}$$

We have determined the probability of each simple event that results in y successes, as well as the number of such events. We now sum the probabilities of these simple events to obtain

$$p(Y = y) =$$
$$\text{Number of simple events implying } y \text{ successes} \times \text{Probability of one of these equiprobable simple events}$$

$$p(y) = C_n^y\, p^y\, q^{n-y}$$

Example 2.3.1
Suppose we toss a coin three times and record the sequence of heads H and tails T. Let X be a discrete random variable that represents the number of heads in an outcome. What is the distribution of X?

2.3.2 Characteristics of the Binomial Distribution

1. The experiment consists of n identical Bernoulli trials.

2. There are only two possible outcomes on each trial: S (for Success) and F (for Failure).

3. $P(S) = p$ and $P(F) = q$ remain the same from trial to trial. (Note that $p + q = 1$.)

4. **The trials are independent.**

5. The binomial random variable X is the number of S's in n trials.

2.3.3 Definition

Suppose that n independent trials of the same probability experiment are performed, where each trial results in either a "success" (with probability p), or a "failure" (with probability $1 - p$). If the random variable X denotes the total number of successes in the n trials, then X has a binomial distribution with parameters n and p, which we write

$$X \sim \mathcal{B}(n, p) \Leftrightarrow P[X = x] = C_n^x p^x q^{n-x}; \quad x \in \{0, 1, 2, \ldots, n\}$$

The graphical representation of the PMF of a Binomial(n = 10, p = 0.3) distribution is given in Figure 2.3.

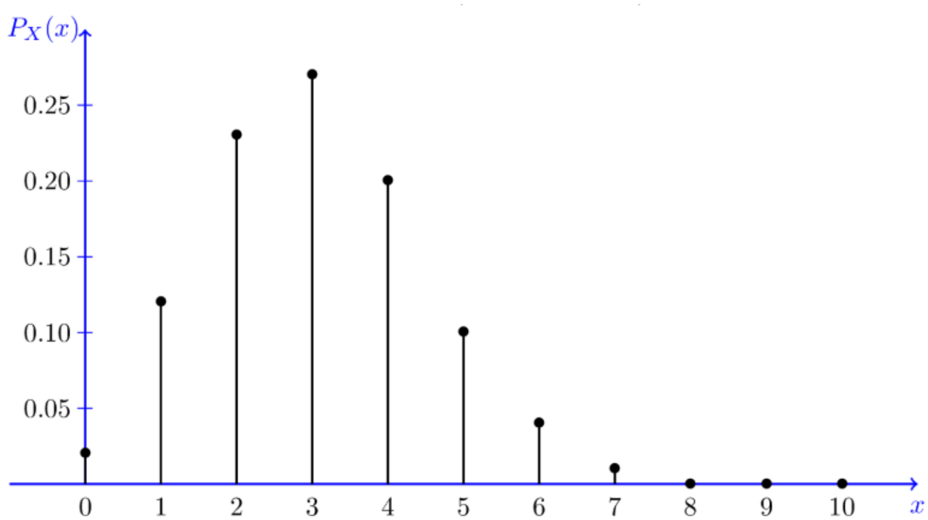


Figure 2.3: The PMF of $X \sim$ B[10, 3]

2.3.4 Moments

The probabilistic characteristics of the Binomial distribution are summarized in the following table.

$\boldsymbol{X}$	$\boldsymbol{\mathcal{B}(n,p)}$
Parameters	$n \in \mathbb{N},\ 0 < p < 1$
$\boldsymbol{R_X}$	$\{0, 1, 2, 3, \ldots, n\}$
PMF	$p_X(x) = C_n^x\, p^x q^{n-x}; \quad \text{if} \quad x \in R_X$
MGF	$M_X(t) = (p\, e^t + q)^n$
Expected Value	$\mu = n\, p$
Variance	$\sigma^2 = n\, p\, q$
Third Central Moment	$\mu_3 = n\, p\, q\, (q - p)$
Fourth Central Moment	$\mu_4 = n\, p\, q\, (1 - 6\, p\, q + 3n\, p\, q)$
Skewness	$\gamma_1 = \dfrac{q-p}{\sqrt{np\, q}}$
Kurtosis	$\gamma_2 = \dfrac{1 - 6\, p\, q}{np\, q}$

Example 2.3.2 1. Prove that the $\mathcal{B}\,(n,\, p)$ is a valid PMF.

2. Find the MGF of the Binomial distribution.

2.3.5 Exercises

Exercise 2.3.1
Suppose that I have a coin for which $P(H) = p$ and $P(T) = 1 - p$. I toss the coin 5 times.

1. What is the probability that the outcome is THHHH?
2. What is the probability that the outcome is HTHHH?
3. What is the probability that the outcome is HHTHH?
4. What is the probability that I will observe exactly four heads and one tail?
5. What is the probability that I will observe exactly three heads and two tails?
6. If I toss the coin n times, what is the probability that I observe exactly k heads and $n - k$ tails?

Exercise 2.3.2
An urn contains 30 red balls and 70 green balls.
What is the probability of getting exactly x red balls in a sample of size 20 if the sampling is done with replacement?

Exercise 2.3.3
You take an exam that contains 20 multiple-choice questions. Each question has 4 possible options. You know

the answer to 10 questions, but you have no idea about the other 10 questions, so you choose answers randomly. Your score X on the exam is the total number of correct answers.

1. Find the PMF of X.
2. What is $P(X > 15)$?
3. What is the expected score?

2.3.6 Python Implementation

Binomial

[2]:
```
from scipy.stats import binom
from IPython.display import display, Math
import numpy as np
import matplotlib.pyplot as plt
```

1 Parameters

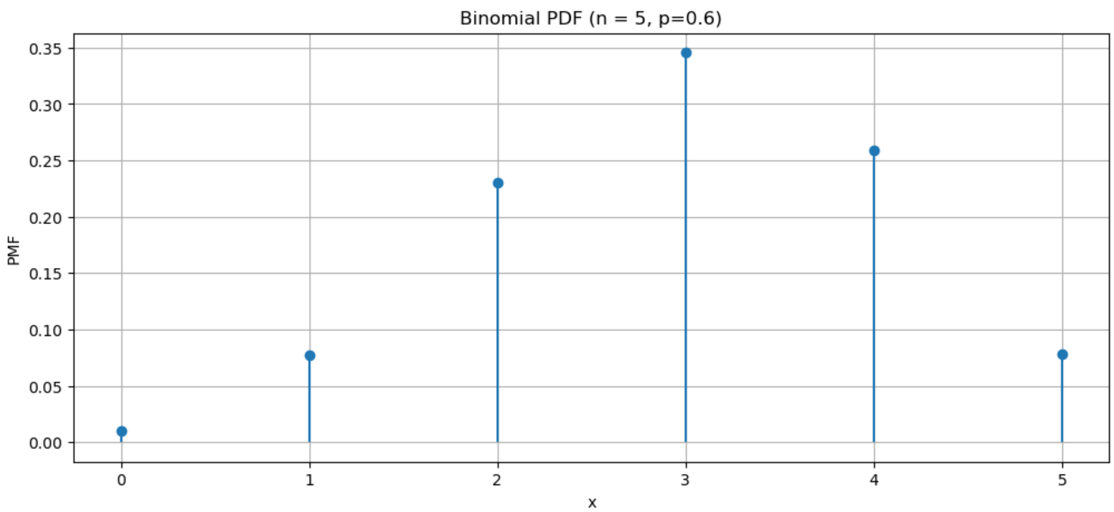
[3]:
```
n = 5  # number of trials
p = 0.6  # probability of success
```

2 Probability Mass Function (PMF)

[4]:
```
PMF = binom.pmf(range(0, n+1), n, p)
print("PMF values:", PMF)
```

```
PMF values: [0.01024 0.0768  0.2304  0.3456  0.2592  0.07776]
```

[5]:
```
plt.figure(figsize=(12, 5))
x_values = range(0, 6)
plt.stem(x_values, PMF, basefmt=" ")
plt.xlabel('x')
plt.ylabel('PMF')
plt.title(f'Binomial PDF (n = {n}, p={p})')
plt.grid(True)
plt.show()
```



3 Cumulative Distribution Function (CDF)

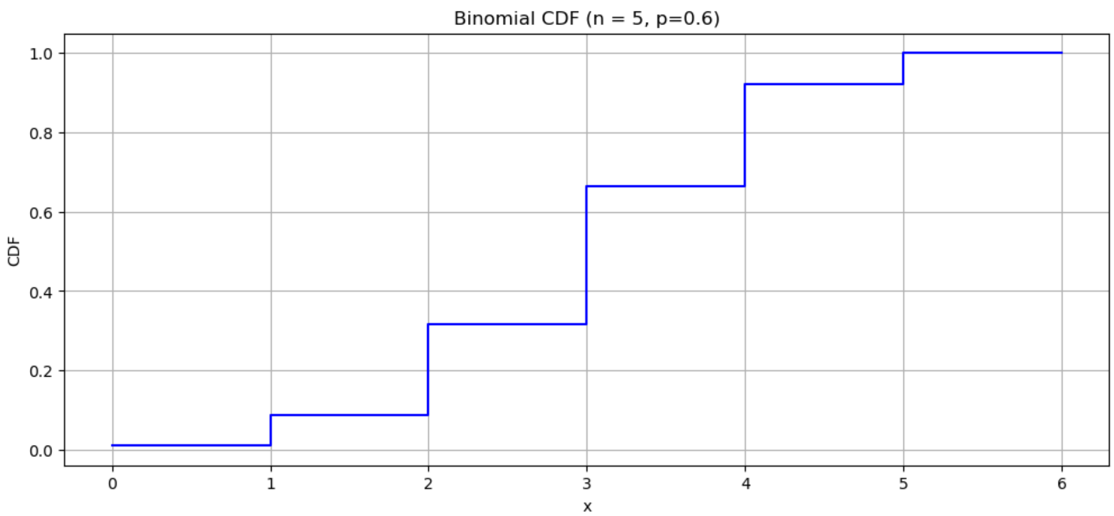
[6]:

```
CDF = binom.cdf(range(0, n+2), n, p)
print("CDF values:", CDF)
```

```
CDF values: [0.01024 0.08704 0.31744 0.66304 0.92224 1.      1.     ]
```

[7]:

```
plt.figure(figsize=(12, 5))
x_values = range(0, 7)
plt.step(x_values, CDF, where='post',  color='blue')
plt.xlabel('x')
plt.ylabel('CDF')
plt.title(f'Binomial CDF (n = {n}, p={p})')
plt.grid(True)
plt.show()
```



4 Expected value or mean (μ)

[8]:
```
mean = binom.mean(n, p)
display(Math(r"\mu = " + str(mean)))
```

$\mu = 3.0$

5 Variance (σ^2)

[9]:
```
variance = binom.var(n, p)
display(Math(r"\sigma^2 = " + str(variance)))
```

$\sigma^2 = 1.2000000000000002$

6 Skewness(γ_1)

[10]:
```
skewness = binom.stats(n, p, moments='s')
display(Math(r"\gamma_1 = " + str(skewness)))
```

$\gamma_1 = -0.1825741858350553$

7 Excess Kurtosis (γ_2)

[11]:
```
excess_kurtosis = binom.stats(n, p, moments='k')
display(Math(r"\gamma_2 = " + str(excess_kurtosis)))
```

$\gamma_2 = -0.3666666666666666$

8 Median M_e

[12]:
```
median = binom.median(n, p)
display(Math(r"M_e = " + str(median)))
```

$M_e = 3.0$

9 Mode M_o

[13]:
```
mode = np.argmax(PMF)
display(Math(r"$M_o$ = " + str(mode)))
```

$M_o = 3$

2.4 Hypergeometric Distribution

2.4.1 Illustrative Example

An urn contains a total of 10 balls, where 4 of the balls are orange and the remaining 6 are grey. Suppose we draw 3 balls from the urn without replacement, meaning once we select a ball we do not place it back in the urn before drawing out the next one. Then some of the balls in our selection may be orange and some may be grey.

We can define the discrete random variable X to give the number of orange balls in our selection. In other words, obtaining an orange ball is a success (S) and consequently getting a grey ball is (F).

$$\begin{aligned}
R_X &= \{0, 1, 2, 3\} \\
P[X=0] &= P[(FFF)] = \frac{C_4^0\,C_6^3}{C_{10}^3} \\
P[X=1] &= P[(FFS),(FSF),(SFF)] = \frac{C_4^1\,C_6^2}{C_{10}^3} \\
P[X=2] &= P[(FSS),(SSF),(SFS)] = \frac{C_4^2\,C_6^1}{C_{10}^3} \\
P(X=3) &= P[(SSS)] = \frac{C_4^3\,C_6^0}{C_{10}^3}
\end{aligned}$$

The probability distribution of X is referred to us as the Hypergeometric distribution, which we define next.

2.4.2 Definition

Suppose in a collection of N objects, N_p are of type 1 (successes) and $N_q = N - N_p$ are of another type 2 (failures). Furthermore, suppose that n objects are randomly selected from the collection **without replacement**. Define the discrete random variable X to give the number of successes. Then X has a Hypergeometric distribution with parameters N, n, p.

$$X \sim \mathcal{H}(N,\, n,\, p) \Leftrightarrow P\left[X = x\right] = \frac{C_{N_p}^{x}\,C_{N_q}^{n-x}}{C_N^n}; \quad \text{Max}(0, n - N_q) \leq x \leq \text{Min}(n, N_p)$$

The graphical representation of the PMF of a Hypergeomtric(N = 10, n = 3, p = 0.4) distribution is given in Figure 2.4.

2.4.3 Moments

The moments of the Hypergeometric distribution are given in the following table.

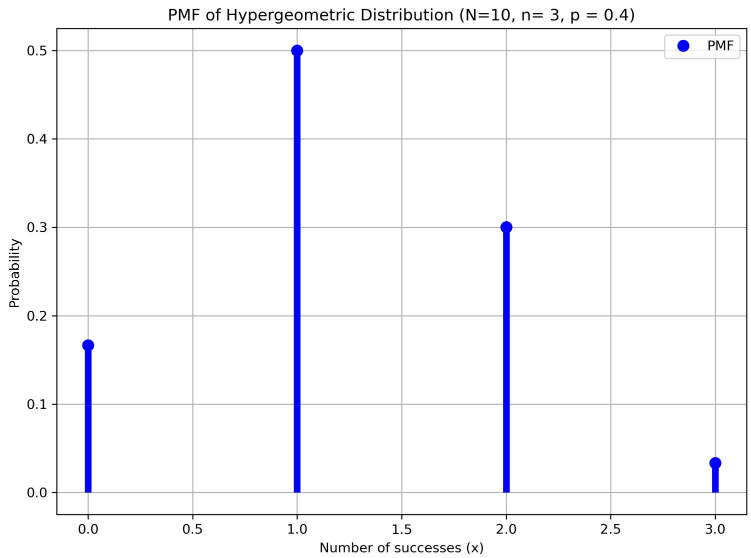


Figure 2.4: The PMF of $X \sim \mathcal{H}[10, 3, 0.4]$

Random Variable X	$\boldsymbol{\mathcal{H}\ [N, n, p]}$
Parameters	$N \in \mathbb{N}$; $n \in \mathbb{N}, n \leq N$; $0 \leq p \leq 1$
Range of X	$R_X = \text{Max}(0, n - N_q) \leq x \leq \text{Min}(n, N_p)$ $\quad N_p = p\,N$ $\quad N_q = (1-p)\,N$
PMF	$\dfrac{C_{N_p}^{x}\,C_{N_q}^{n-x}}{C_N^n}$
Expected Value	$\mu = n\,p$
Variance	$\sigma^2 = \dfrac{N-n}{N-1} n\,p\,q$

2.4.4 Approximation to the Binomial Distribution

Let X follows a hypergeometric distribution $\mathcal{H}\ [N, n, p]$. When N is too large $N \to \infty$, X is approximated to the Binomial distribution $\mathcal{B}\ [n, p]$.

$$X \sim \mathcal{H}\ [N, n, p]_{N \to +\infty} \approx X \sim \mathcal{B}\ [n, p]$$

- The Hypergeometric distribution is approximated to the Binomial distribution when

$$\frac{n}{N} < 10\%$$

Example 2.4.1
Let $X \sim \mathcal{H}\ [N = 5000, n = 100, p = 0.1]$. Compute $P[X = 20]$

$$P[X=20] = \frac{C_{500}^{20}\, C_{4500}^{500-20}}{C_{5000}^{100}} = 0.001065$$

Since $\dfrac{n}{N} = \dfrac{100}{5000} = 0.02 < 0.1$ then $X \approx \mathcal{B}\left[n=100, p=0.1\right]$

$$P[X=20] = C_{100}^{20}\, 0.1^{20} 0.9^{100-20} = 0.00117$$

Example 2.4.2
An urn contains a total of 100 balls, with one-fourth of them colored red and the remaining balls in various other colors. A random selection of 30 balls is made from the urn.
Let X represents the number of red balls in the sample.

1. If the selection is conducted with replacement. Answer the following
   (a) Determine the probability of selecting non-red balls.
   (b) Calculate the probability of selecting all red balls.
   (c) Compute the probability of selecting at least 10 red balls.
   (d) Find the probability that the number of selected red balls is less than the number of selected non-red balls.
   (e) Calculate the average number of selected red balls and the standard deviation of selected red balls.
2. Answer the previous questions if the selection is conducted without replacement.

2.4.5 Exercises

Exercise 2.4.1
An urn contains five chips, two red and three white. Suppose that two are drawn out simultaneously. Let X denote the number of red chips in the sample.
Find $V(X)$.

Exercise 2.4.2
Suppose X has a hypergeometric distribution with $N = 100$, $n = 4$, and $N_p = 20$. Determine the following:

1. $P(X = 1)$
2. $P(X = 4)$
3. $P(X \geq 18)$
4. Determine the mean and variance of X.

Exercise 2.4.3
A shipment of 10 boxes of meat contains 2 boxes of contaminated goods. An inspector randomly selects 4 boxes; let Z be the number of boxes of contaminated meat among the selected 4 boxes.

(a) What is the PMF of Z?

(b) What is the probability that at least one of the four boxes is contaminated?

(c) How many boxes must be selected so that the probability of having at least one contaminated box is larger than 0.75?

2.4.6 Python Implementation

Hypergeometric

November 14, 2024

[1]:
```
import numpy as np
from scipy.stats import hypergeom
from IPython.display import display, Math
import matplotlib.pyplot as plt
```

1 Parameters

[2]:
```
N = 10  # population size
N_p = 4  # Total number of successes in the population
N_q = N - N_p # Total number of failuers in the population
n = 3  # Sample size
```

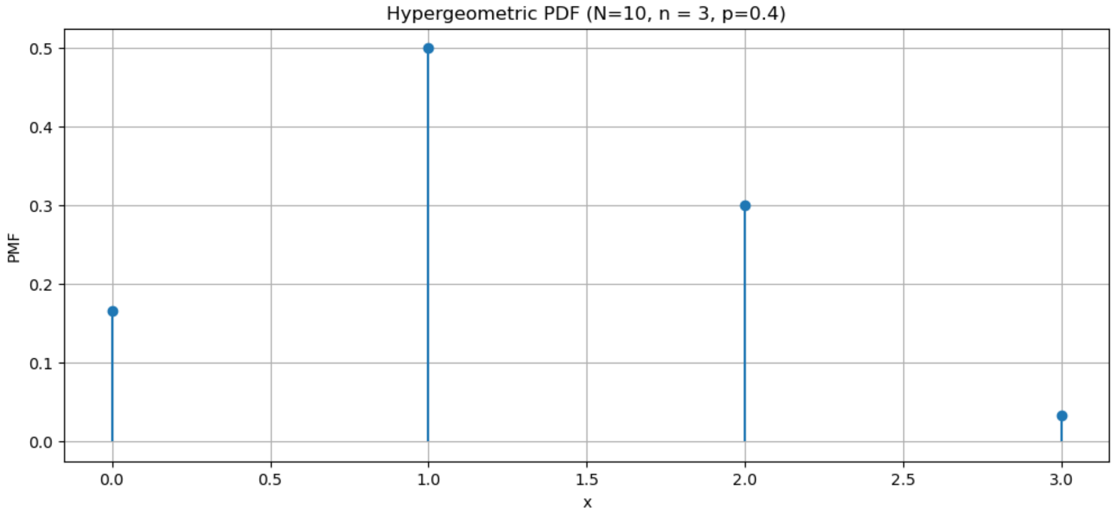
[4]:
```
p = N_p/N
```

2 Probability Mass Function (PMF)

[3]:
```
PMF = hypergeom.pmf(range(0, n+1), N, N_p, n)
print("PMF values:", PMF)
```

```
PMF values: [0.16666667 0.5        0.3        0.03333333]
```

[9]:
```
plt.figure(figsize=(12, 5))
x_values = range(0, n+1)
plt.stem(x_values, PMF, basefmt=" ")
plt.xlabel('x')
plt.ylabel('PMF')
plt.title(f'Hypergeometric PDF (N={N}, n = {n}, p={p})')
plt.grid(True)
plt.show()
```

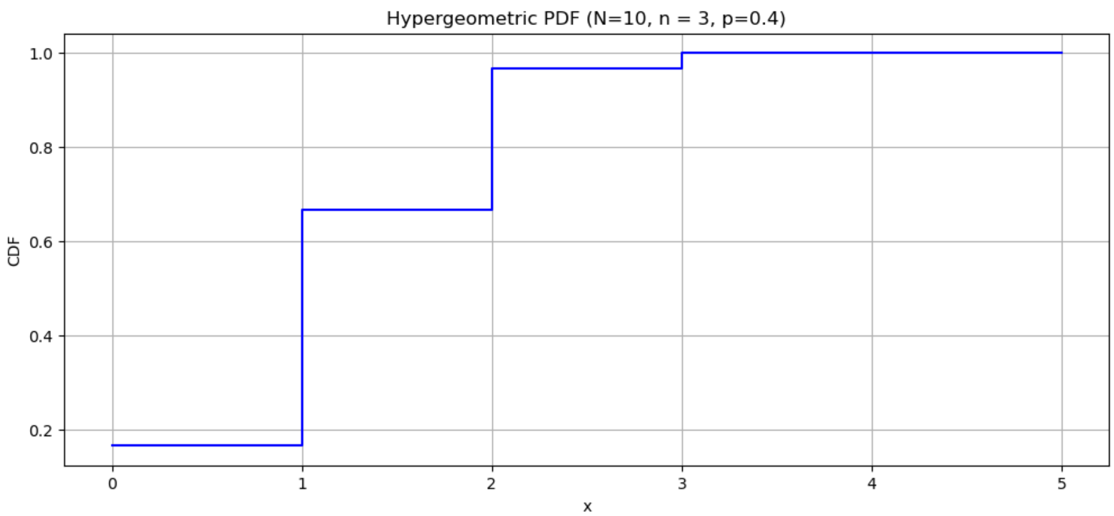


3 Cumulative Distribution Function (CDF)

[7]:
```
CDF = hypergeom.cdf(range(0, n+3), N, N_p, n)
print("CDF values:", CDF)
```

```
CDF values: [0.16666667 0.66666667 0.96666667 1.         1.         1.        ]
```

[10]:
```
plt.figure(figsize=(12, 5))
x_values = range(0, n+3)
plt.step(x_values, CDF, where='post',  color='blue')
plt.xlabel('x')
plt.ylabel('CDF')
plt.title(f'Hypergeometric PDF (N={N}, n = {n}, p={p})')
plt.grid(True)
plt.show()
```



4 Expected value or mean (μ)

[11]:
```
mean = hypergeom.mean(N, N_p, n)
display(Math(r"\mu = " + str(mean)))
```

$\mu = 1.2$

5 Variance (σ^2)

[12]:
```
variance = hypergeom.var(N, N_p, n)
display(Math(r"\sigma^2 = " + str(variance)))
```

$\sigma^2 = 0.56$

6 Skewness(γ_1)

[13]:
```
skewness = hypergeom.stats(N, N_p, n, moments='s')
display(Math(r"\gamma_1 = " + str(skewness)))
```

$\gamma_1 = 0.1336306209562122$

7 Excess Kurtosis (γ_2)

[14]:
```
excess_kurtosis = hypergeom.stats(N, N_p, n, moments='k')
display(Math(r"\gamma_2 = " + str(excess_kurtosis)))
```

$\gamma_2 = -0.3877551020408163$

8 Median M_e

[15]:
```
median = hypergeom.median(N, N_p, n)
display(Math(r"M_e = " + str(median)))
```

$M_e = 1.0$

9 Mode M_o

[16]:
```
mode = np.argmax(PMF)
display(Math(r"$M_o$ = " + str(mode)))
```

$M_o = 1$

2.5 Geometric Distribution

The geometric distribution is related to the binomial distribution in that the underlying probability experiment is the same, i.e., **independent trials with two possible outcomes**. However, the random variable defined in the geometric case highlights a different aspect of the experiment, namely the number of trials needed to obtain the first "success".

2.5.1 Definition

Suppose that a sequence of independent Bernoulli trials is performed, with $p = P(\text{"success"})$ for each trial. Define the random variable X to give the number of trial at which the first success occurs. Then X has a geometric distribution with parameter p. The probability mass function (PMF) of X is given by

$$\begin{aligned} p(x) &= P[X = x] \\ &= P[1^{\text{st}} \text{ Success on the } x^{\text{th}} \text{ Trial}] \\ &= P[1^{\text{st}}(x-1) \text{ trials are failures \& } x^{\text{th}} \text{ trial is success}] \\ &= (1-p)^{x-1}p; \quad x = 1, 2, 3, \ldots \end{aligned}$$

In other words, X is said to follow the geometric distribution (also known as Pascal distribution) if

$$X \sim \mathcal{G}(p) \Leftrightarrow P\left[X = x\right] = (1-p)^{x-1}\, p; \quad x \in \mathbb{N}^*$$

The graphical representation of the PMF of a geomtric(p = 0.3) distribution is given in Figure 2.5.

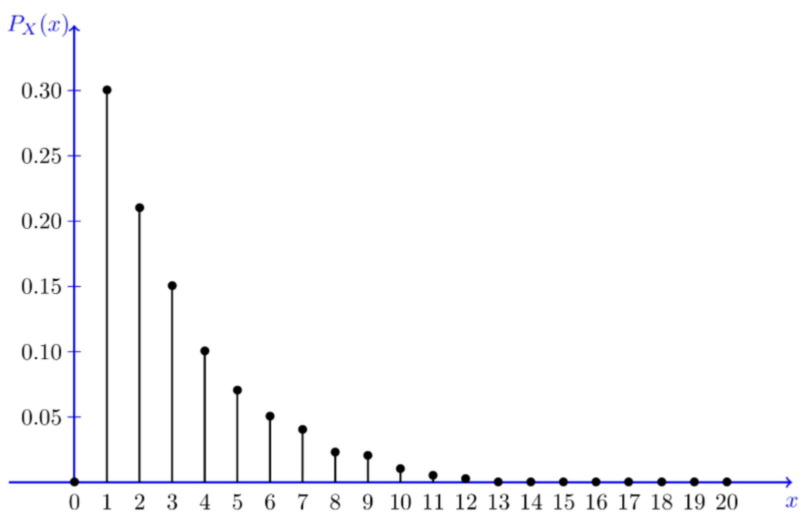


Figure 2.5: The PMF of $X \sim \mathcal{G}[0.3]$

2.5.2 Examples

Each of the following is an example of a random variable with the geometric distribution.

1. Toss a fair coin until the first heads occurs. In this case, a "success" is getting a heads ("failure" is getting tails) and so the parameter $p = p(H) = 0.5$.

2. Toss a biased coin until the first heads occurs (getting a face has double the chances of getting a head). In this case, a "success" is getting a heads ("failure" is getting tails) and so the parameter $p = p(H) = \frac{1}{3}$.
3. Roll a pair of fair dice until getting the first double 1's. In this case, a "success" is getting double 1's, and a "failure" is simply not getting double 1's (so anything else). The parameter is $p = p(1,1) = \frac{1}{36}$.
4. Historically the heir of the throne is usually a male. Suppose the King will have children until he has had a boy. What is the probability that this King will have at least one girl?

- Verify that the PMF for a geometric distribution is a valid PMF. (Hint: It is called geometric distribution for a reason!).

2.5.3 Moments

The probabilistic characteristics of the Geometric distribution are summarized in the following table.

Table 2.3: Probabilistic Characteristics of The Geometric Distribution

Parameter	**Geometric Distribution** $\mathcal{G}(p)$
$\boldsymbol{R_X}$	$\{1, 2, 3, \ldots\}$
Parameter	$0 < p < 1; q = 1 - p$
PMF	$p_X(x) = (1-p)^{x-1}\, p; \quad \text{if} \quad x \in R_X$
MGF	$M_X(t) = \dfrac{p\, e^t}{1 - q\, e^t}; t < -ln(q)$
Expected Value	$\mu = \dfrac{1}{p}$
Variance	$\sigma^2 = \dfrac{q}{p^2}$
Third Central Moment	$\mu_3 = \dfrac{q\,(q+1)}{p^3}$
Fourth Central Moment	$\mu_4 = \dfrac{q\,(9q + p^2)}{p^4}$
Skewness	$\gamma_1 = \dfrac{q+1}{\sqrt{q}}$
Excess Kurtosis	$\gamma_2 = 6 + \dfrac{p^2}{q}$

2.5.4 Remarks

1. To generalize the geometric distribution by considering any number of successes (r), we use the negative binomial distribution.

   Example 2.5.1
   Keep tossing a fair coin until you get 3 heads. What is the probability of getting 3 heads supposing that it took you 20 trails?

In this case, the parameter p is still given by $p = P(H) = 0.5$, but now we also have the parameter $r = 3$, the number of desired "successes", i.e., heads.
Let X be a R.V. that represents the number of necessary trials until getting 3 heads.

$$P[X = 20] = C_{19}^{2}\, 0.5^3\, 0.5^{17} = 0.003$$

The PMF of the Negative Binomial distribution is given as

$$X \sim \mathcal{NB}(r, p) \Leftrightarrow P\left[X = x\right] = C_{x-1}^{r-1}\, p^r (1-p)^{x-r};\ r \in \mathbb{N};\ r \geq 2;\ x = r, r+1, r+2, \ldots$$

2. In general, note that a geometric distribution can be thought of a negative binomial distribution with parameter $r = 1$.

$$\mathcal{NB}(r = 1,\, p) = \mathcal{G}(p)$$

3. Note that for both the **Geometric** and **Negative Binomial** distributions **the number of possible values the random variable can take is infinite**. These are still discrete distributions though, since we can "list" the values. In other words, the possible values are countable. This is in contrast to the **Bernoulli, Binomial, and Hypergeometric** distributions, where the number of possible values is finite.

4. We again note the distinction between the **Binomial** distribution and the **Geometric** and **Negative Binomial** distributions. In the **Binomial** distribution, the number of trials is fixed, and we count the number of "successes". Whereas, in the **geometric** and **negative binomial** distributions, the number of "successes" is fixed, and we count the number of trials needed to obtain the desired number of "successes".

Example 2.5.2
In some economically disadvantaged countries, a male child is considered necessary to help with physical work and family finances. Suppose a couple will have children until they have had two boys. What is the probability that this couple will have at least one girl?

2.5.5 Exercises

Exercise 2.5.1
A game of chance has only two possible outcomes to win or lose for each time you play. The probability of losing the game is 65%. You play this game until your first loss.
What is the probability of playing the game exactly 6 times in a row?

Exercise 2.5.2
Suppose that you are looking for a student at your college who lives within five kilometers of you. You know that 55% of the 25,000 students do live within five kilometers of you. You randomly contact students from the college until one says he or she lives within five kilometers of you.
What is the probability that you need to contact four people?

Exercise 2.5.3
The lifetime risk of developing pancreatic cancer is about one in 78 (1.28%). Let X be the number of people you ask until one says he or she has pancreatic cancer.

1. What is the probability that you ask ten people before one says he or she has pancreatic cancer?

2. What is the probability that you must ask 20 people?

3. Find the (i) mean and (ii) standard deviation of X.

2.5.6 Python Implementation

Geometric

[1]:
```python
import numpy as np
from scipy.stats import geom
from IPython.display import display, Math
import matplotlib.pyplot as plt
```

1 Parameters

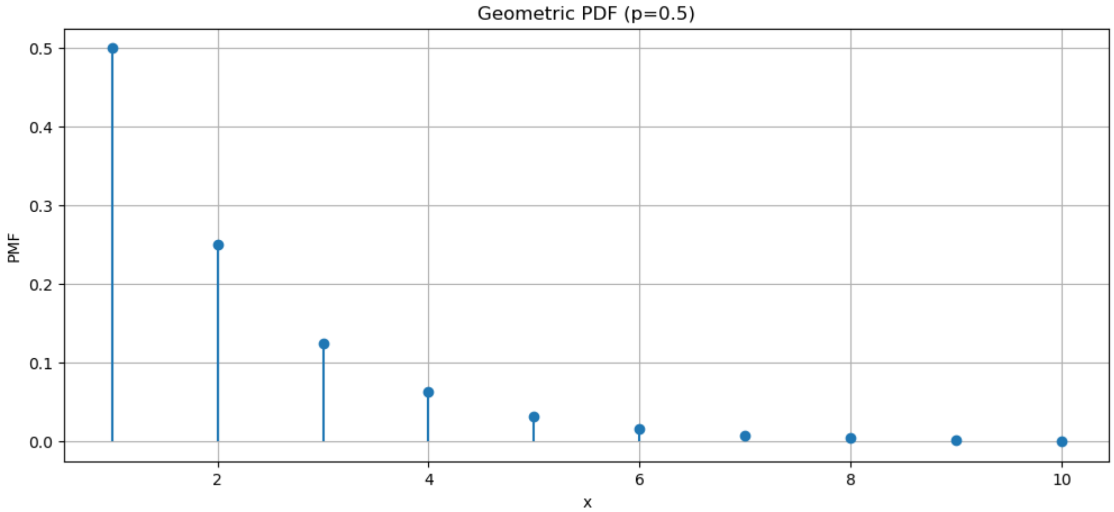
[2]:
```python
p = 0.5  # Probability of success
q = 1 - p  # Probability of failure
```

2 Probability Mass Function (PMF)

[3]:
```python
PMF = geom.pmf(range(1, 11), p)
print("PMF values:", PMF)
```

```
PMF values: [0.5        0.25       0.125      0.0625     0.03125    0.015625
 0.0078125  0.00390625 0.00195312 0.00097656]
```

[4]:
```python
plt.figure(figsize=(12, 5))
x_values = range(1, 11)
plt.stem(x_values, PMF, basefmt=" ")
plt.xlabel('x')
plt.ylabel('PMF')
plt.title(f'Geometric PDF (p={p})')
plt.grid(True)
plt.show()
```

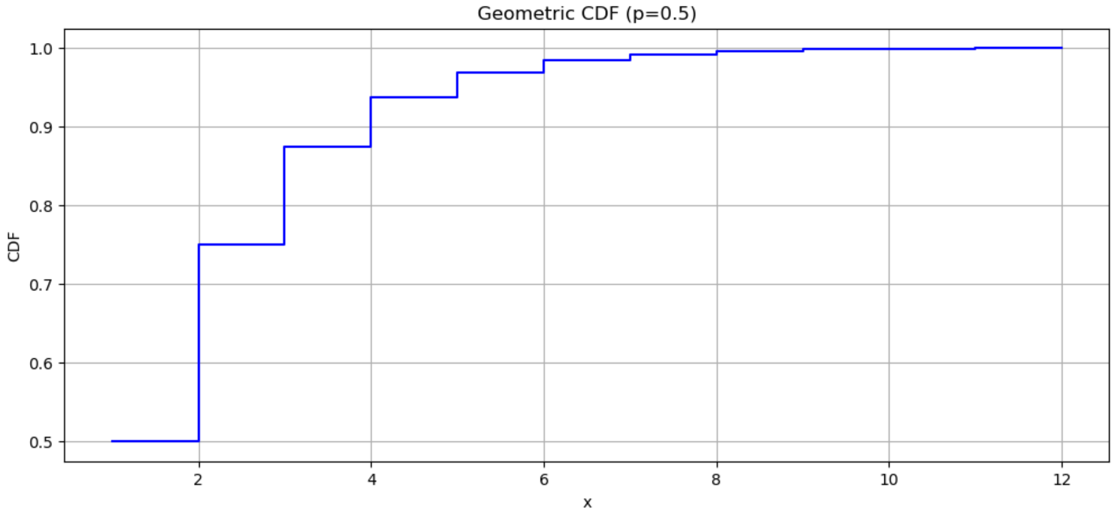


3 Cumulative Distribution Function (CDF)

[5]:
```
CDF = geom.cdf(range(1, 13), p)
print("CDF values:", CDF)
```

```
CDF values: [0.5        0.75       0.875      0.9375     0.96875    0.984375
 0.9921875  0.99609375 0.99804688 0.99902344 0.99951172 0.99975586]
```

[6]:
```
plt.figure(figsize=(12, 5))
x_values = range(1, 13)
plt.step(x_values, CDF, where='post',  color='blue')
plt.xlabel('x')
plt.ylabel('CDF')
plt.title(f'Geometric CDF (p={p})')
plt.grid(True)
plt.show()
```



4 Expected value or mean (μ)

[7]:
```
mean = geom.mean(p)
display(Math(r"\mu = " + str(mean)))
```

$\mu = 2.0$

5 Variance (σ^2)

[8]:
```
variance = geom.var(p)
display(Math(r"\sigma^2 = " + str(variance)))
```

$\sigma^2 = 2.0$

6 Skewness(γ_1)

[9]:
```
skewness = geom.stats(p, moments='s')
display(Math(r"\gamma_1 = " + str(skewness)))
```

$\gamma_1 = 2.1213203435596424$

7 Excess Kurtosis (γ_2)

[10]:
```
excess_kurtosis = geom.stats(p, moments='k')
display(Math(r"\gamma_2 = " + str(excess_kurtosis)))
```

$\gamma_2 = 6.5$

8 Median M_e

[11]:
```
median = geom.median(p)
display(Math(r"M_e = " + str(median)))
```

$M_e = 1.0$

9 Mode M_o

[13]:
```
mode_index = np.argmax(PMF)
mode = x_values[mode_index]
display(Math(r"\text{$M_o$} = " + str(mode)))
```

$M_o = 1$

2.6 Poisson Distribution

2.6.1 The Poisson Limit

The use of the Poisson distribution is appropriate when the probability of success is very small ($p \to 0$), the number of trials large ($n \to \infty$) and $n \times p$ remains constant. The Poisson's limit is used to approximate hard-to-calculate binomial probabilities where the values of n and p reflect the conditions of the limit.

Suppose

$$X \sim \mathcal{B}(n, p) \Leftrightarrow P\left[X = x\right] = C_n^x \, p^x q^{n-x}; \quad x \in \{0, 1, 2, \ldots, n\}$$

If ($n \to \infty$) and ($p \to 0$) in such way that $\lambda = n \times p$ remains constant, then

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ n\times p = \text{ constant}}} P\left[X = x\right] = \lim_{\substack{n\to\infty \\ p\to 0 \\ n\times p = \text{ constant}}} C_n^x \, p^x q^{n-x} = \frac{e^{-n\,p}\,(n\,p)^x}{x!} = \frac{e^{-\lambda}\,\lambda^x}{x!}$$

Proof

$$\lim_{n\to\infty} C_n^x \, p^x (1-p)^{n-x} = \lim_{n\to\infty} C_n^x \, (\frac{\lambda}{n})^x (1 - \frac{\lambda}{n})^{n-x}$$

$$\lim_{n\to\infty} \frac{n!}{x!\,(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{-x} \left(1 - \frac{\lambda}{n}\right)^n =$$

$$\frac{\lambda^x}{x!} \lim_{n\to\infty} \frac{n!}{(n-x)!} \frac{1}{(n-\lambda)^x} \left(1 - \frac{\lambda}{n}\right)^n$$

Given that

$$\lim_{n\to\infty} \left(1 + \frac{y}{n}\right)^n = e^y$$

Then

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

Since $\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$, we need $\lim_{n\to\infty} \frac{n!}{(n-x)!\,(n-\lambda)^x} = 1$

$$\lim_{n\to\infty} \frac{n!}{(n-x)!\,(n-\lambda)^x} = \lim_{n\to\infty} \frac{n\,(n-1)\,(n-2)\ldots\,(n-x+1)}{(n-\lambda)\,(n-\lambda)\,\ldots(n-\lambda)} \approx 1$$

The above theorem left unanswered question is how large n needs to be and how small p have to be before the Poisson distribution becomes a good approximation to the binomial distribution.
The following tables represent two cases where

1. $n = 5$, $p = \frac{1}{5}$, $\lambda = np = 1$. The probabilities are expressed in Table 2.4.

2. $n = 100$, $p = \frac{1}{100}$, $\lambda = np = 1$. The probabilities are expressed in Table 2.5.

Table 2.4: The PMF of Binomial($n = 5$, $p = 0.2$) and Poisson ($\lambda = 1$)

x	**Binomial** ($n = 5, p = 0.2$)	**Poisson** ($\lambda = 1$)
0	0.328	0.368
1	0.410	0.368
2	0.205	0.184
3	0.051	0.061
4	0.006	0.015
5	0.000	0.003
6+	0.000	0.001
Total	1.000	1.000

Table 2.5: The PMF of Binomial($n = 100$, $p = 0.01$) and Poisson ($\lambda = 1$)

x	**Binomial** ($n = 100, p = 0.01$)	**Poisson** ($\lambda = 1$)
0	0.366032	0.367879
1	0.369730	0.367879
2	0.184865	0.183940
3	0.060999	0.061313
4	0.014942	0.015328
5	0.002898	0.003066
6	0.000463	0.000511
7	0.000063	0.000073
8	0.000007	0.000009
9	0.000001	0.000001
10	0.000000	0.000000
Total	1.000000	0.999999

We see in Figure 2.6 that when $n = 5$ for some x values the agreement between the binomial distribution and Poisson's distribution is not very good. However, when $n = 100$, the agreement is remarkably good for all x.

As rule of thumb, **we approximate the binomial distribution to the Poisson distribution when**

$$n \geq 30,\ p \leq 0.1,\ np < 15$$

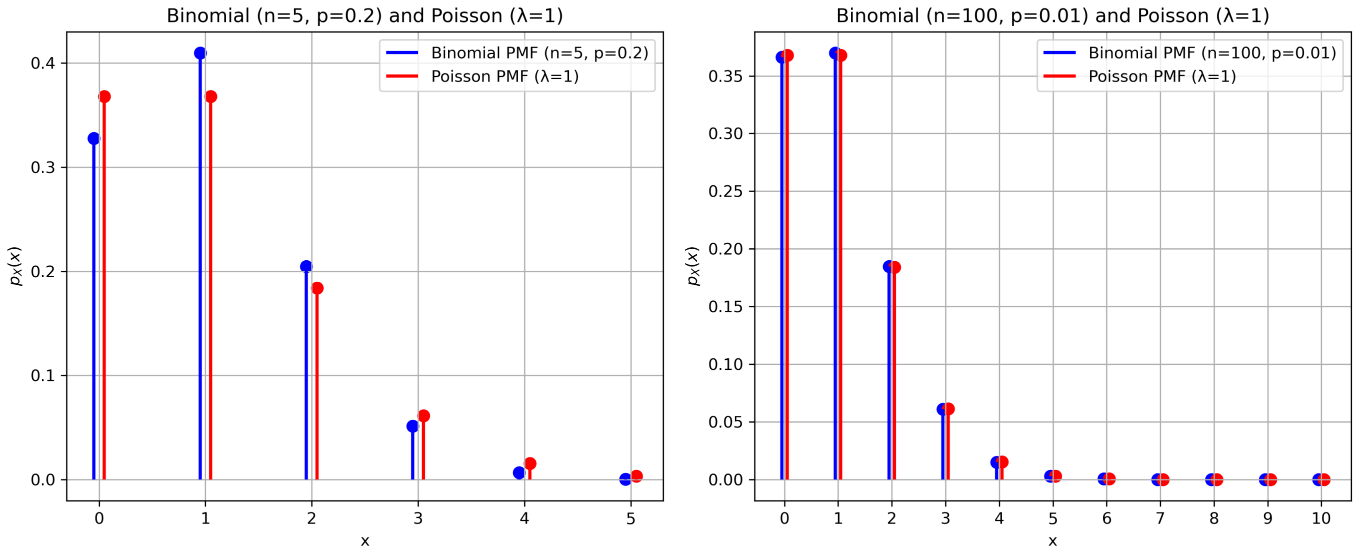


Figure 2.6: The PMFs of Binomial and Poisson distributions

2.6.2 Definition

A random variable X has a Poisson distribution, with parameter $\lambda > 0$, if its probability mass function (PMF) is given by

$$X \sim \mathcal{P}(\lambda) \Leftrightarrow P\left[X = x\right] = \frac{e^{-\lambda}\,\lambda^{x}}{x!}; \quad x \in \{0, 1, 2, \ldots, \} = \mathbb{N}$$

The main application of the Poisson distribution is to count the number of times some event occurs over a fixed interval of time or space. More specifically, if the random variable X denotes the number of times the event occurs during an interval of length T, and r denotes the average rate at which the event occurs per unit interval, then X has a Poisson distribution with parameter $\lambda = rT$. Consider the following examples:

1. The number of customers arriving at the local BNA's branch between 8 and 9 AM.
2. The number of calls made to the emergency police call number in Koléa on a Saturday.
3. The number of accidents at a particular intersection during the month of June.

Example 2.6.1
Suppose typos occur at an average rate of $r = 0.01$ per page in the Friday edition of the New York Times, which is 45 pages long. Let X denote the number of typos on the front page. Then X has a Poisson distribution with parameter $\lambda = r \times 1 = 0.01$.

$$X \sim \mathcal{P}(\lambda = 0.01) \Leftrightarrow P\left[X = x\right] = \frac{e^{-0.01}\,0.01^{x}}{x!}; \quad x \in \mathbb{N}$$

since we are considering an interval of length one page ($T = 1$). Thus, the probability that there is at least one typo on the front page is given by

$$P\left[X \geq 1\right] = 1 - P\left[X = 0\right] = 1 - \frac{e^{-0.01}\,0.01^{0}}{0!} = 0.00995$$

Now, if we let random variable Y denote the number of typos in the entire paper, then Y has a Poisson distribution with parameter $\lambda = r \times 45 = 0.45$.

$$Y \sim \mathcal{P}(\lambda = 0.45) \Leftrightarrow P\left[Y = y\right] = \frac{e^{-0.45}\, 0.45^{y}}{y!}; \quad y \in \mathbb{N}$$

Since we are considering an interval of $T = 45$ pages. The probability that there is at least one typo on the front page is given by

$$P\left[Y \geq 1\right] = 1 - P\left[Y = 0\right] = 1 - \frac{e^{-0.45}\, 0.45^{0}}{0!} = 0.3624$$

The Poisson distribution is similar to all previously considered families of discrete probability distributions in that it counts the number of times something happens. However, the Poisson distribution is different in that there is not an act that is being repeatedly performed. In other words, there are no set trials, but rather a set window of time or space to observe.

2.6.3 Moments

The probabilistic characteristics of the Poisson distribution are summarized in the following table.

Table 2.6: Probabilistic Characteristics of the Poisson Distribution

X	$\mathcal{P}(\lambda)$
R_X	$\mathbb{N}$
Parameters	$\lambda \in \mathbb{R}^{+}$
PMF	$\frac{e^{-\lambda}\, \lambda^{x}}{x!}; \quad \text{if} \quad x \in R_X$
MGF	$M_X(t) = \exp\left[\lambda\left(e^{t} - 1\right)\right]$
Expected Value	$\mu = \lambda$
Variance	$\sigma^2 = \lambda$
Third Central Moment	$\mu_3 = \lambda$
Fourth Central Moment	$\mu_4 = \lambda + 3\lambda^2$
Skewness	$\gamma_1 = \frac{1}{\sqrt{\lambda}}$
Kurtosis	$\gamma_2 = \frac{1}{\lambda}$

Example 2.6.2
Prove that $\mathcal{P}\left(\lambda\right)$ is a valid PMF.
Use the Maclaurin series expansion of the exponential function.

$$e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^{x}}{x!}$$

2.6.4 Exercises

Exercise 2.6.1
The number of emails that I get on a weekday can be modeled by a Poisson distribution with an average of 0.2 emails per minute.

1. What is the probability that I get no emails in an interval of length 5 minutes?
2. What is the probability that I get more than 3 emails in an interval of length 10 minutes?

Exercise 2.6.2
A bank expects to receive six bad checks per day, on average.

1. Determine the random variable of interest, its distribution, and its parameter.
2. What is the probability of the bank getting fewer than five bad checks on any given day?
3. Find the mean, standard deviation, and mode.

Exercise 2.6.3
Suppose 2% of the items made by a factory are defective.

1. Find the probability P that there are 3 defective items in a sample of 100 items.

2.6.5 Python Implementation

Poisson

[1]:
```
import numpy as np
from scipy.stats import poisson
import matplotlib.pyplot as plt
from IPython.display import display, Math
```

1 Parameters

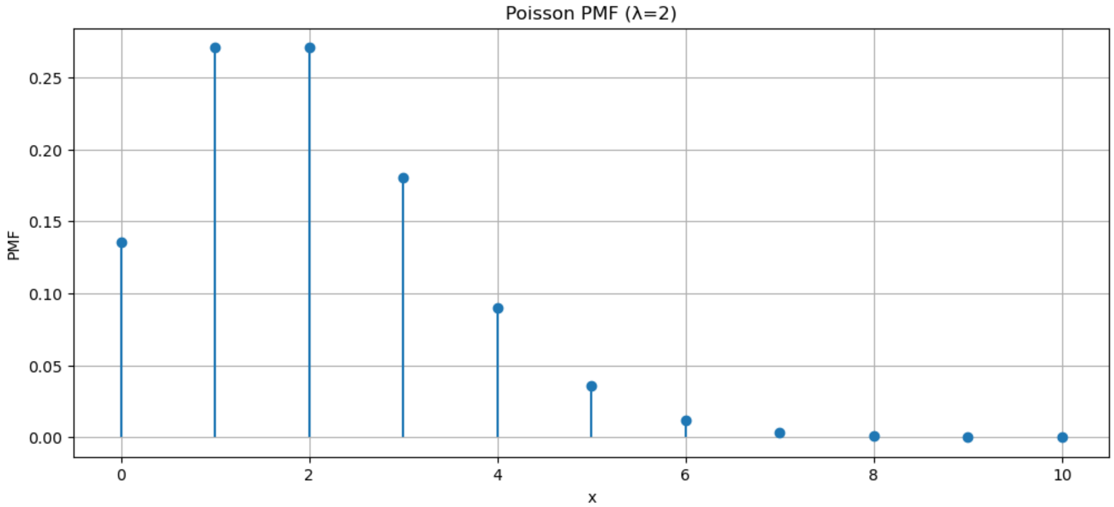
[2]:
```
# Parameters for the Poisson distribution
lambda_param = 2  # Rate parameter (λ)
```

2 Probability Mass Function (PMF)

[3]:
```
PMF = poisson.pmf(range(0, 11), lambda_param)
print("PMF values:", PMF)
```

```
PMF values: [1.35335283e-01 2.70670566e-01 2.70670566e-01 1.80447044e-01
 9.02235222e-02 3.60894089e-02 1.20298030e-02 3.43708656e-03
 8.59271640e-04 1.90949253e-04 3.81898506e-05]
```

[4]:
```
plt.figure(figsize=(12, 5))
x_values = range(0, 11)
plt.stem(x_values, PMF, basefmt=" ")
plt.xlabel('x')
plt.ylabel('PMF')
plt.title(f'Poisson PMF (λ={lambda_param})')
plt.grid(True)
plt.show()
```

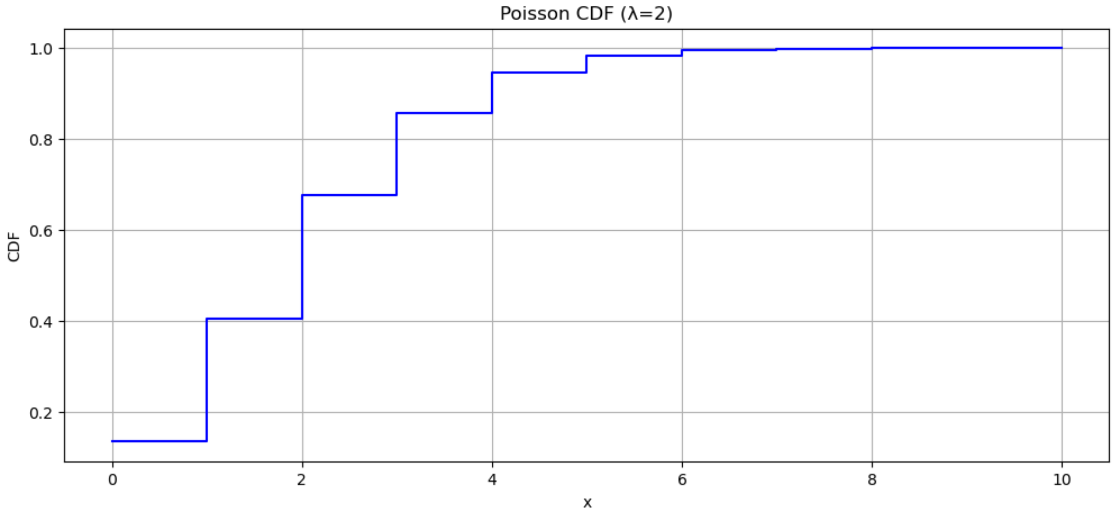


3 Cumulative Distribution Function (CDF)

[5]:
```
CDF = poisson.cdf(range(0, 11), lambda_param)
print("CDF values:", CDF)
```

```
CDF values: [0.13533528 0.40600585 0.67667642 0.85712346 0.94734698 0.98343639
 0.99546619 0.99890328 0.99976255 0.9999535  0.99999169]
```

[7]:
```
plt.figure(figsize=(12, 5))
x_values = range(0, 11)
plt.step(x_values, CDF, where='post', label=f'λ={lambda_param}', color='blue')
plt.xlabel('x')
plt.ylabel('CDF')
plt.title(f'Poisson CDF (λ={lambda_param})')
plt.grid(True)
plt.show()
```



4 Expected value or mean (μ)

[8]:
```
mean = poisson.mean(lambda_param)
display(Math(r"\mu = " + str(mean)))
```

$\mu = 2.0$

5 Variance (σ^2)

[9]:
```
variance = poisson.var(lambda_param)
display(Math(r"\sigma^2 = " + str(variance)))
```

$\sigma^2 = 2.0$

6 Skewness(γ_1)

[10]:
```
skewness = poisson.stats(lambda_param, moments='s')
display(Math(r"\gamma_1 = " + str(skewness)))
```

$\gamma_1 = 0.7071067811865476$

7 Excess Kurtosis (γ_2)

[11]:
```
excess_kurtosis = poisson.stats(lambda_param, moments='k')
display(Math(r"\gamma_2 = " + str(excess_kurtosis)))
```

$\gamma_2 = 0.5$

8 Median M_e

[12]:
```
median = poisson.median(lambda_param)
display(Math(r"M_e = " + str(median)))
```

$M_e = 2.0$

9 Mode M_o

[14]:
```
mode_index = np.argmax(PMF)
mode = x_values[mode_index]
display(Math(r"\text{$M_o$} = " + str(mode)))
```

$M_o = 1$

Chapter 3

Standard Continuous Probability distributions

3.1 Continuous Uniform Distribution

3.1.1 Definition

A random variable X has a uniform distribution on interval $[a, b]$, if it has PDF given by

$$X \sim \mathcal{U}[a,b] \Leftrightarrow f_X(x) = \begin{cases} \dfrac{1}{b-a}, & \text{if } x \in [a,b] \\ 0, & \text{otherwise} \end{cases}$$

The uniform distribution is also sometimes referred to as the Rectangular distribution, since the graph of its pdf looks like a rectangle. See the figure below.

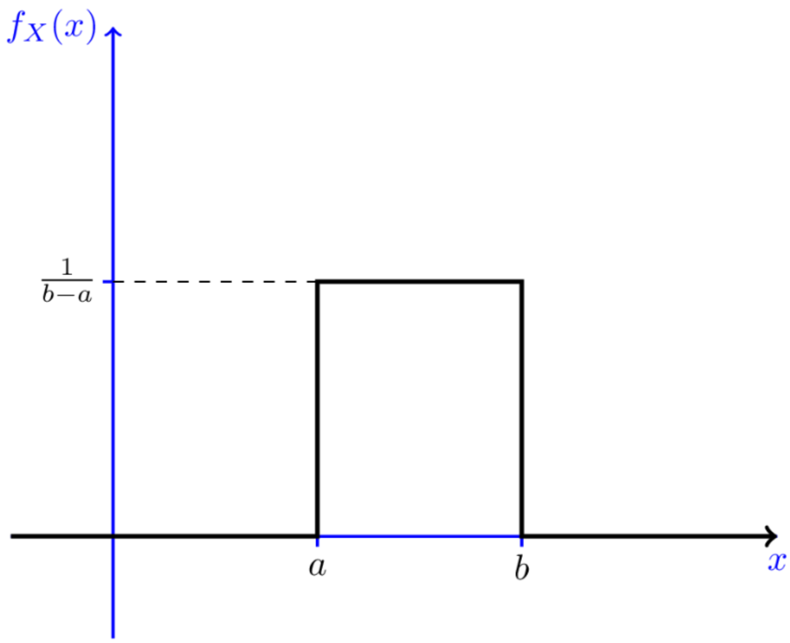


Figure 3.1: PDF for a continuous random variable uniformly distributed over $[a, b]$.

A typical application of the uniform distribution is to model randomly generated numbers. In other words, it provides the probability distribution for a random variable representing a randomly chosen number between numbers a and b.
The uniform distribution assigns equal probabilities to intervals of equal lengths, since it is a constant function, on the interval it is non-zero $[a, b]$. This is the continuous analogue to equally likely outcomes in the discrete setting.

3.1.2 CDF

The cumulative distribution function of a continuous uniform random variable is

$$F_X(x) = \begin{cases} 0, & \text{if } x < a \\ \dfrac{x-a}{b-a}, & \text{if } a \le x < b \\ 1, & \text{if } x \ge b \end{cases}$$

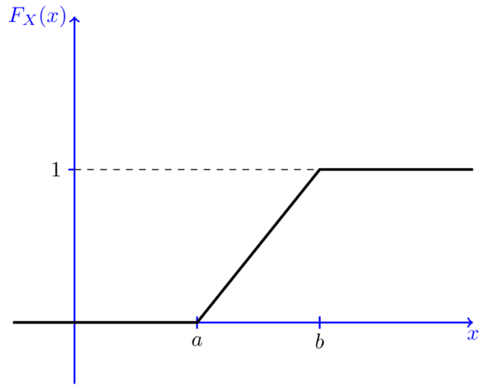


Figure 3.2: CDF for a continuous random variable uniformly distributed over $[a, b]$.

3.1.3 Moments

The probabilistic characteristics of the continuous uniform distribution are summarized in the following table.

Table 3.1: Probabilistic Characteristics of the Continuous Uniform Distribution.

X	$\mathcal{U}\,[a, b]$
Range of X	$R_X = [a, b]$
Parameters	$a, b \in \mathbb{R}; a < b$
PDF	$\frac{1}{b-a}; \quad \text{if} \quad x \in R_X$
CDF	$\frac{x-a}{b-a}; \quad \text{if} \quad x \in R_X$
Quantile Function	$Q(p) = a + p(b-a); \quad 0 < p < 1$
MGF	$M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}$
r^{th} **Raw Moment**	$\mu'_r = \frac{b^{r+1} - a^{r+1}}{(r+1)(b-a)}$
Expected Value	$\mu = \frac{a+b}{2}$
Variance	$\sigma^2 = \frac{(b-a)^2}{12}$
Third Central Moment	$\mu_3 = 0$
Fourth Central Moment	$\mu_4 = \frac{(b-a)^4}{80}$
Skewness	$\gamma_1 = 0$
Kurtosis	$\gamma_2 = \frac{-6}{5}$

Example 3.1.1
Given that X $\sim \mathcal{U}\,[0,1]$. Compute μ, M_e, M_o, σ, $P[X \leq 0.5\ X > 0.25]$

3.1.4 Exercises

Exercise 3.1.1
The arrival time between trains at a train stop is uniformly distributed between 0 and 15 minutes. A student does not check the schedule and has arrived at the train stop.

1. Compute the probability they wait more than 10 minutes.
2. Compute the probability of waiting between 2 and 8 minutes.

Exercise 3.1.2
A distribution is given as $X \sim U(0, 20)$.

1. What is $P(2 < x < 18)$?
2. Find the 90th percentile.

Exercise 3.1.3
The amount of time, in minutes, that a person must wait for a bus is uniformly distributed between zero and 15 minutes.

1. What is the probability that a person waits fewer than 12.5 minutes?
2. On average, how long must a person wait? Find the mean μ and the standard deviation σ.
3. Fifty percent of the time, the time a person must wait falls below what value?
4. Find the mode.

3.1.5 Python Implementation

Continuous_Uniform

[11]:
```python
import numpy as np
from scipy.stats import uniform
from IPython.display import display, Math
import matplotlib.pyplot as plt
```

1 Parameters

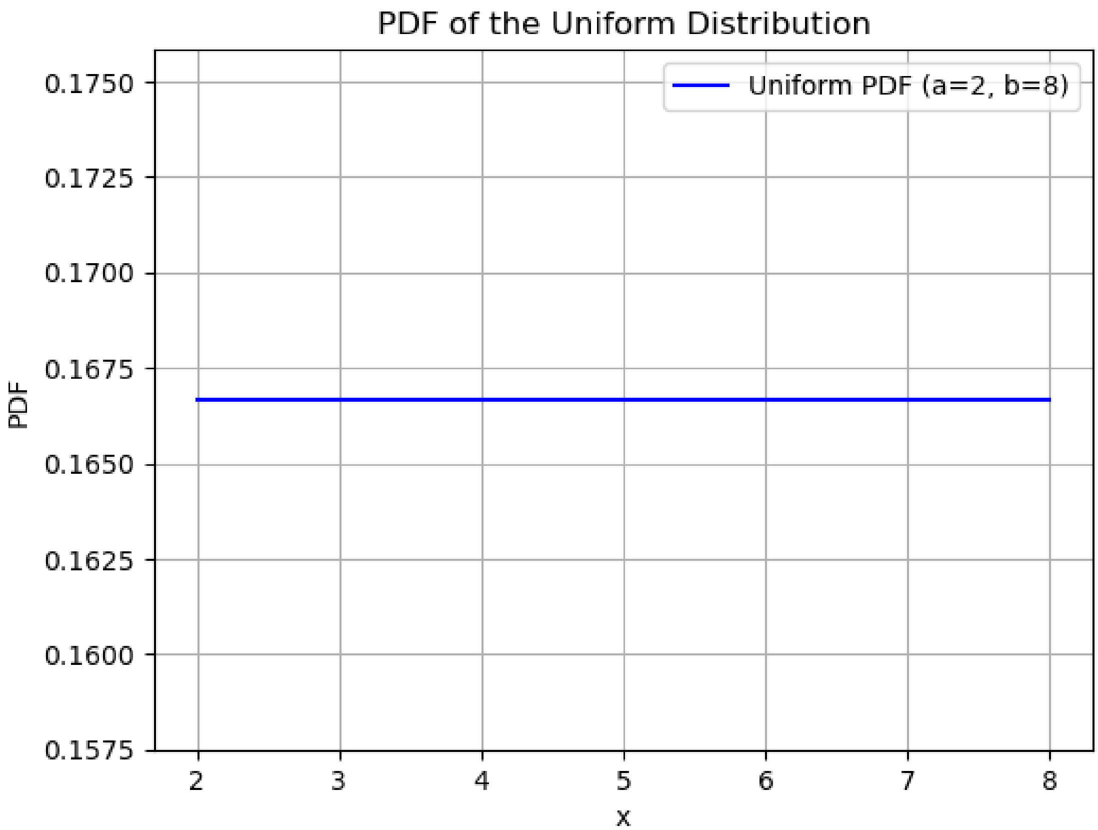
[12]:
```python
a = 2  # Lower bound
b = 8  # Upper bound
# Define the range of x-values
x_values = np.linspace(a, b, 100)
```

2 Probability Mass Function (PMF)

[13]:
```python
# Generate x values over the range of the uniform distribution
x_values = np.linspace(a, b, 1000)

# Calculate the PDF of the uniform distribution
pdf_values = uniform.pdf(x_values, loc=a, scale=b - a)

# Plot the PDF
plt.plot(x_values, pdf_values, label=f'Uniform PDF (a={a}, b={b})', color='blue')
plt.xlabel('x')
plt.ylabel('PDF')
plt.title('PDF of the Uniform Distribution')
plt.legend()
plt.grid(True)
plt.show()
```



[20]:
```python
x_values = np.linspace(a, b, 100)
PDF = uniform.pdf(x_values, loc=a, scale=b-a)
print("PMF values:", PDF)
```

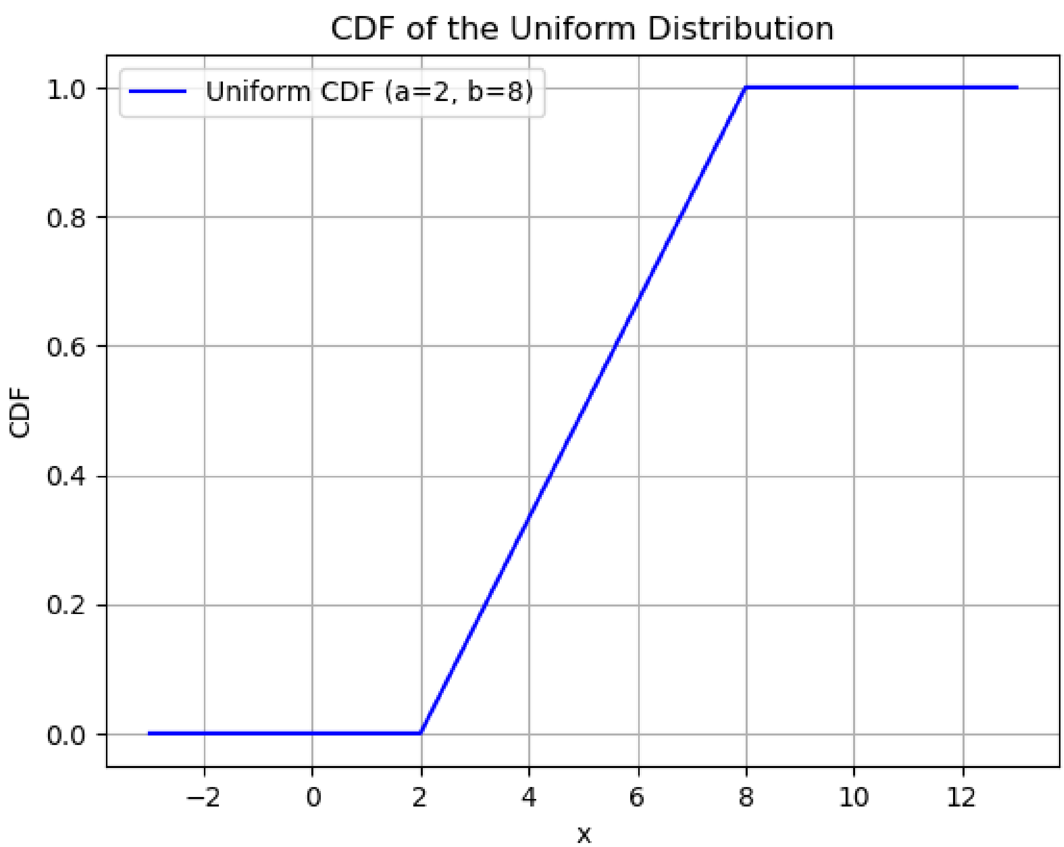
```
PMF values: [0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667 0.16666667
 0.16666667 0.16666667 0.16666667 0.16666667]
```

3 Cumulative Distribution Function (CDF)

[15]:
```
# Generate x values over the range of the uniform distribution
x_values = np.linspace(a-5, b+5, 1000)

# Calculate the PDF of the uniform distribution
cdf_values = uniform.cdf(x_values, loc=a, scale= b - a)

# Plot the PDF
plt.plot(x_values, cdf_values, label=f'Uniform CDF (a={a}, b={b})', color='blue')
plt.xlabel('x')
plt.ylabel('CDF')
plt.title('CDF of the Uniform Distribution')
plt.legend()
plt.grid(True)
plt.show()
```



[21]:
```
x_values = np.linspace(a, b, 100)
# CDF for specific values of x
CDF = uniform.cdf(x_values, loc=a, scale=b-a)
```

```python
print("CDF values:", CDF)
```

```
CDF values: [0.         0.01010101 0.02020202 0.03030303 0.04040404 0.05050505
 0.06060606 0.07070707 0.08080808 0.09090909 0.1010101  0.11111111
 0.12121212 0.13131313 0.14141414 0.15151515 0.16161616 0.17171717
 0.18181818 0.19191919 0.2020202  0.21212121 0.22222222 0.23232323
 0.24242424 0.25252525 0.26262626 0.27272727 0.28282828 0.29292929
 0.3030303  0.31313131 0.32323232 0.33333333 0.34343434 0.35353535
 0.36363636 0.37373737 0.38383838 0.39393939 0.4040404  0.41414141
 0.42424242 0.43434343 0.44444444 0.45454545 0.46464646 0.47474747
 0.48484848 0.49494949 0.50505051 0.51515152 0.52525253 0.53535354
 0.54545455 0.55555556 0.56565657 0.57575758 0.58585859 0.5959596
 0.60606061 0.61616162 0.62626263 0.63636364 0.64646465 0.65656566
 0.66666667 0.67676768 0.68686869 0.6969697  0.70707071 0.71717172
 0.72727273 0.73737374 0.74747475 0.75757576 0.76767677 0.77777778
 0.78787879 0.7979798  0.80808081 0.81818182 0.82828283 0.83838384
 0.84848485 0.85858586 0.86868687 0.87878788 0.88888889 0.8989899
 0.90909091 0.91919192 0.92929293 0.93939394 0.94949495 0.95959596
 0.96969697 0.97979798 0.98989899 1.        ]
```

4 Expected value or mean (μ)

[22]:
```python
mean = uniform.mean(loc=a, scale=b-a)
display(Math(r"\mu = " + str(mean)))
```

$\mu = 5.0$

5 Variance (σ^2)

[23]:
```python
variance = uniform.var(loc=a, scale=b-a)
display(Math(r"\sigma^2 = " + str(variance)))
```

$\sigma^2 = 3.0$

6 Skewness(γ_1)

[24]:
```python
# Skewness
skewness = uniform.stats(loc=a, scale=b-a, moments='s')
display(Math(r"\gamma_1 = " + str(skewness)))
```

$\gamma_1 = 0.0$

7 Excess Kurtosis (γ_2)

[25]:
```
# Kurtosis (Excess Kurtosis)
kurtosis = uniform.stats(loc=a, scale=b-a, moments='k')
display(Math(r"\gamma_2 = " + str(kurtosis)))
```

$\gamma_2 = -1.2$

8 Median M_e

[26]:
```
# Median
median = uniform.median(loc=a, scale=b-a)
display(Math(r"\text{Median} = " + str(median)))
```

Median $= 5.0$

9 Mode M_o

[29]:
```
# Mode (for continuous uniform, mode is not typically defined, as every point is␣
↪equally likely)
display(Math(r"\text{Mode: No defined mode for continuous uniform}"))
```

Mode: No defined mode for continuous uniform

3.2 Exponential Distribution

3.2.1 Definition

A random variable X has an exponential distribution with parameter $\lambda > 0$, if X has PDF given by

$$X \sim \mathcal{E}(\lambda) \Leftrightarrow f_X(x) = \begin{cases} \lambda\, e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

The parameter λ is referred as the rate parameter, it represents how quickly events occur.
A typical application of exponential distributions is to model waiting times or lifetimes. For example, each of the following gives an application of an exponential distribution.

- The time until the first accident occurs at a given intersection.
- length of interval between consecutive occurrences of Poisson distributed events.

To get some intuition for this interpretation of the exponential distribution, suppose you are waiting for an event to happen. For example, you are at a store and are waiting for the next customer. In each millisecond, the probability that a new customer enters the store is very small. You can imagine that, in each millisecond, a coin (with a very small P(H)) is tossed, and if it lands heads a new customers enters. If you toss a coin every millisecond, the time until a new customer arrives approximately follows an exponential distribution.

3.2.2 CDF

The cumulative distribution function of an exponential random variable is

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-\lambda x}, & \text{if } x \geq 0 \end{cases}$$

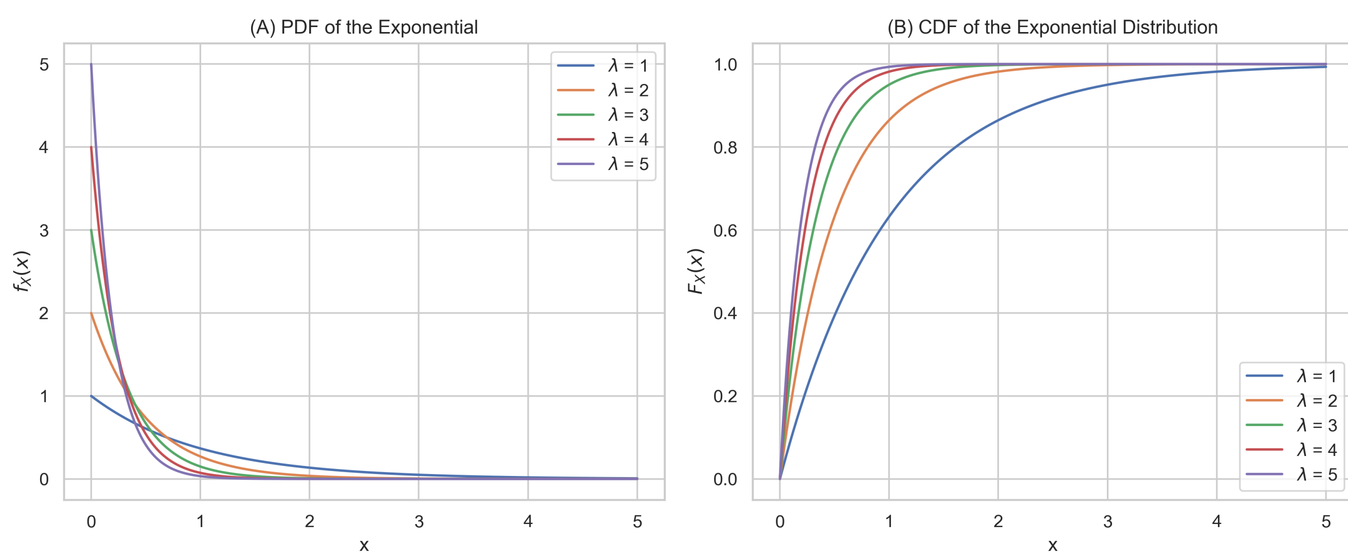


Figure 3.3: PDF and CDF for the exponential distribution.

3.2.3 Moments

The probabilistic characteristics of the exponential distribution are summarized in the following table.

X	$\mathcal{E}(\lambda)$
Range R_X	$[0, \infty[$
Parameters	$\lambda \in \mathbb{R}^+$
PDF	$\lambda\, e^{-\lambda x}, \quad \text{if } x \in R_X$
CDF	$1 - e^{-\lambda x}, \quad \text{if } x \in R_X$
Quantile Function	$Q(p) = \dfrac{-\ln(1-p)}{\lambda}, \quad 0 < p < 1$
MGF	$M_X(t) = \dfrac{\lambda}{\lambda - t}, \quad t < \lambda$
r^{th} **raw moment**	$\mu'_r = r!\, \lambda^{-r}$
Expected Value	$\mu = \dfrac{1}{\lambda}$
Variance	$\sigma^2 = \dfrac{1}{\lambda^2}$
Third Central Moment	$\mu_3 = \dfrac{2}{\lambda^3}$
Fourth Central Moment	$\mu_4 = \dfrac{9}{\lambda^4}$
Skewness	$\gamma_1 = 2$
Kurtosis	$\gamma_2 = 6$

Example 3.2.1
Laptops produced by company XYZ last, on average, for 5 years. The life span of each laptop follows an exponential distribution.

1. Calculate the rate parameter.
2. Write the PDF and graph it.
3. What is the probability that a laptop will last less than 3 years?
4. What is the probability that a laptop will last more than 10 years?
5. What is the probability that a laptop will last between 4 and 7 years?

3.2.4 Lack of Memory of the Exponential Distribution

Suppose that X follows the exponential distribution and represents the waiting time for an elevator, and suppose that you have already waited 3 minutes. Then the probability that you have to wait another 2 minutes is the

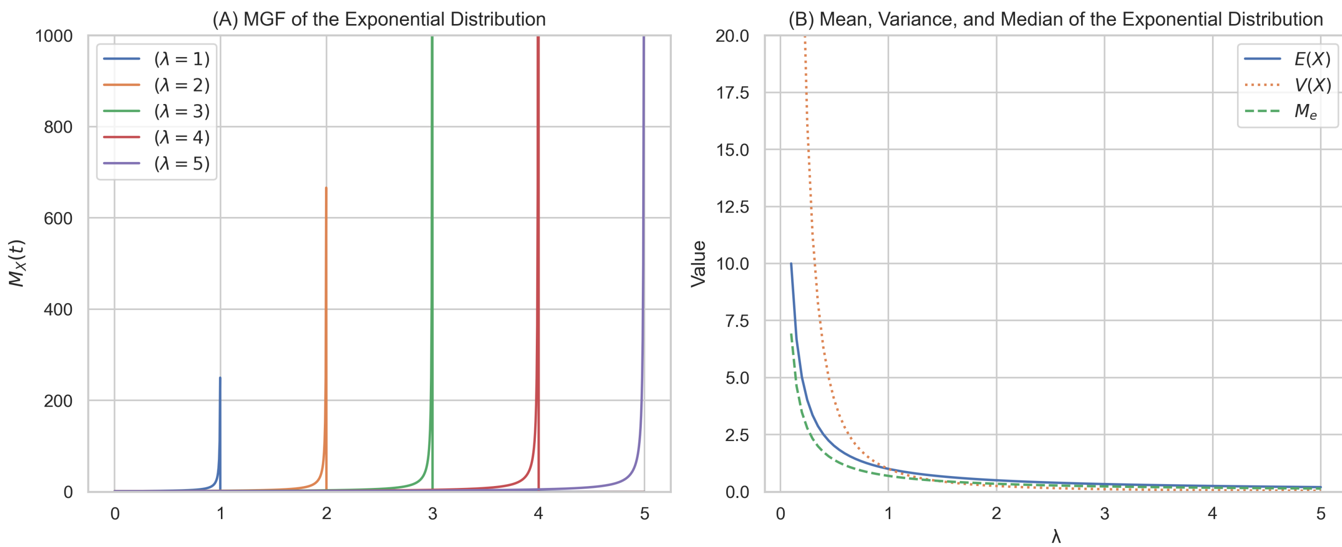


Figure 3.4: The MGF and other characteristics of the exponential distribution.

same as as the probability that you would have to wait two minutes if you just arrived. This is not true if the waiting time distribution is something other than an exponential.

$$P(X > 5\,/X > 2) = \frac{P(X > 5 \cap X > 2)}{P(X > 2)} = \frac{P(X > 5)}{P(X > 2)} = \frac{1 - F(5)}{1 - F(2)} = \frac{e^{-5\lambda}}{e^{-2\lambda}} = e^{-3\lambda} = P(X > 3)$$

Let $X \sim \mathcal{E}(\lambda)$, and let s and t be positive numbers. The lack of memory property is that

$$P(X > s + t\,/X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-(s+t)\lambda}}{e^{-s\lambda}} = e^{-t\lambda} = P(X > t)$$

3.2.5 Exercises

Exercise 3.2.1

The random variable X of the life lengths of batteries is associated with a probability density function of the form:

$$f(x) = \begin{cases} \frac{1}{2}e^{-x/2}, & \text{for } x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

1. Find the probability that the life of a particular battery of this type is less than 200 or greater than 400 hours.

2. Find the probability that a battery of this type lasts more than 300 hours, given that it already has been in use for more than 200 hours.

Exercise 3.2.2

The distribution function of the random variable X, the time (in months) from the diagnosis age until death for one population of patients with AIDS, is as follows:

$$F(x) = \begin{cases} 1 - e^{-0.03x^{1.2}}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

1. Find the probability that a randomly selected person from this population survives at least 12 months.
2. Find the probability density function of X.

Exercise 3.2.3
On average, a certain computer part lasts ten years. The length of time the computer part lasts is exponentially distributed.

1. What is the probability that a computer part lasts more than 7 years?
2. On the average, how long would five computer parts last if they are used one after another?
3. Eighty percent of computer parts last at most how long?
4. What is the probability that a computer part lasts between nine and 11 years?

3.2.6 Python Implementation

Exponential

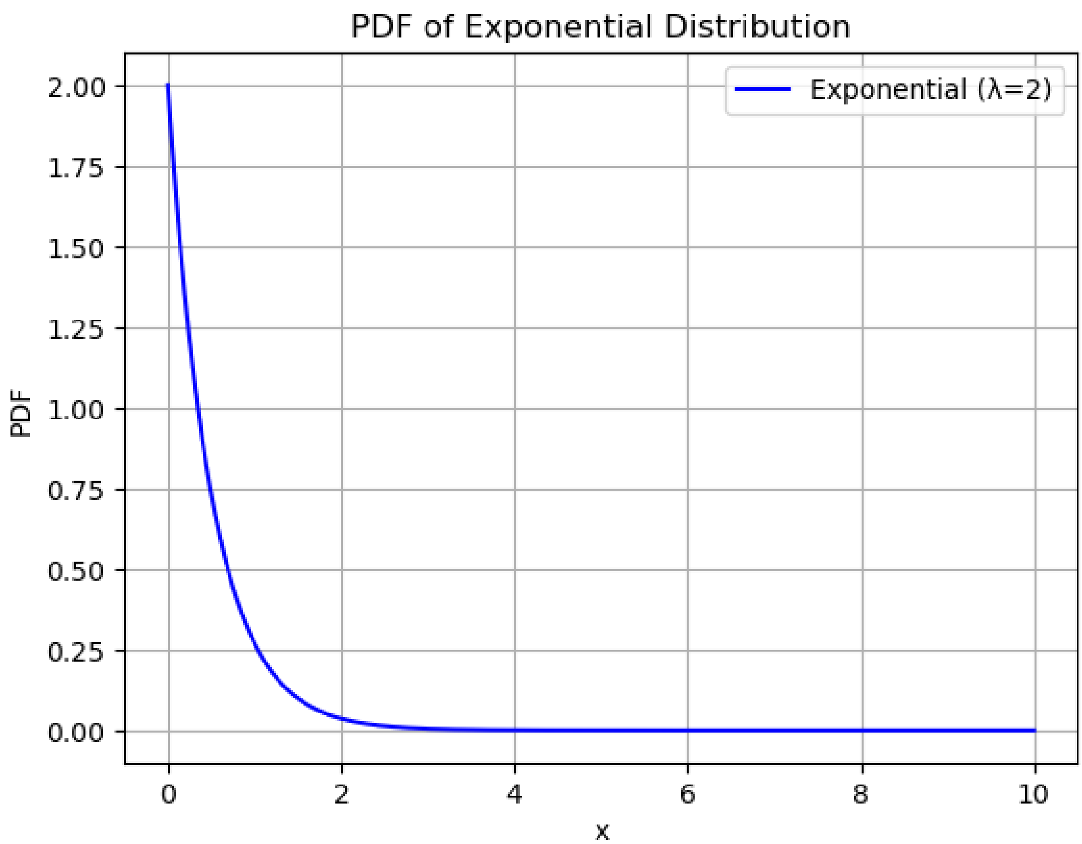
```
[1]: from IPython.display import display, Math
     import numpy as np
     import matplotlib.pyplot as plt
     from scipy.stats import expon
```

1 Parameters

```
[10]: lambda_param = 2  # Rate parameter (lambda)
```

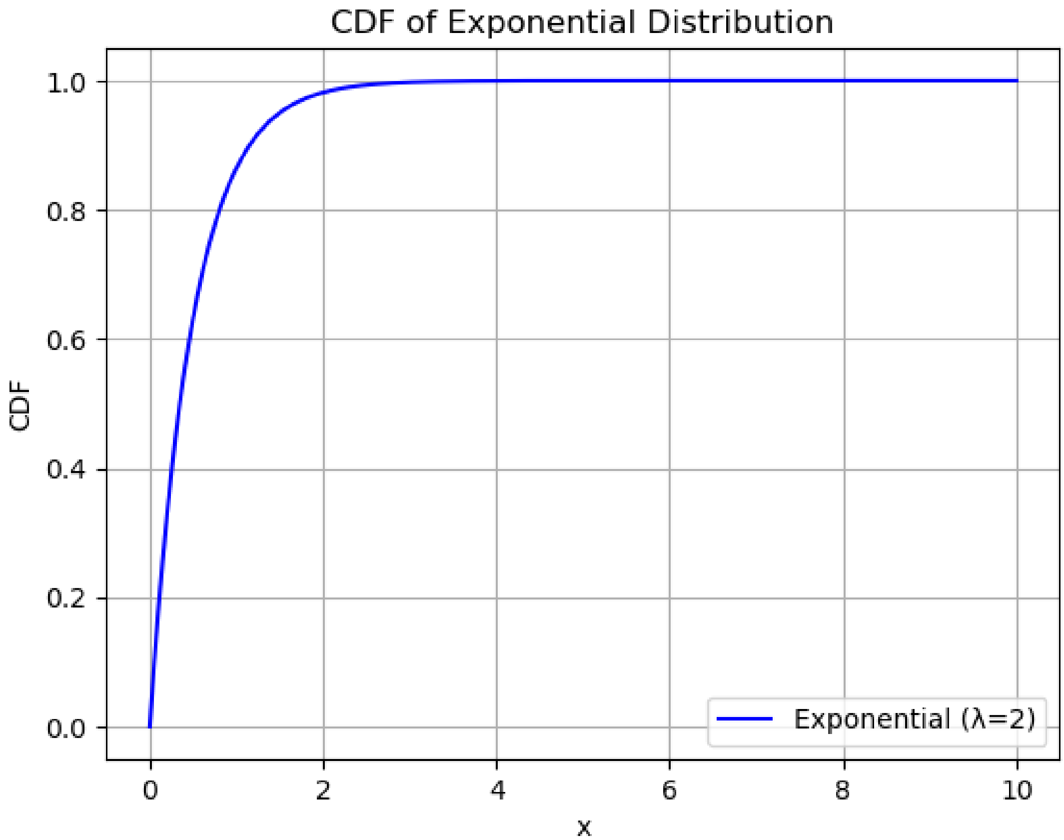
2 Probability Mass Function (PMF)

```
[11]: # Generate x values over the range of the uniform distribution
      x_values = np.linspace(0, 10, 1000)
      pdf_values = expon.pdf(x_values, scale=1 / lambda_param)
      plt.plot(x_values, pdf_values, label=f'Exponential (λ={lambda_param})',␣
       ↪color='blue')
      plt.xlabel('x')
      plt.ylabel('PDF')
      plt.title('PDF of Exponential Distribution')
      plt.legend()
      plt.grid(True)
```



3 Cumulative Distribution Function (CDF)

[12]:
```
cdf_values = expon.cdf(x_values, scale=1 / lambda_param)
plt.plot(x_values, cdf_values, label=f'Exponential (λ={lambda_param})',␣
 ↪color='blue')
plt.xlabel('x')
plt.ylabel('CDF')
plt.title('CDF of Exponential Distribution')
plt.legend()
plt.grid(True)
plt.show()
```



4 Expected value or mean (μ)

[17]:
```
mean = expon.mean(scale = 1/ lambda_param, loc = 0)
display(Math(r"\mu = " + str(mean)))
```

$\mu = 0.5$

5 Variance (σ^2)

[20]:
```
variance = expon.var(loc=0, scale=1/ lambda_param)
display(Math(r"\sigma^2 = " + str(variance)))
```

$\sigma^2 = 0.25$

6 Skewness(γ_1)

[22]:
```
# Skewness
skewness = expon.stats(loc=0 , scale= 1/ lambda_param, moments='s')
display(Math(r"\gamma_1 = " + str(skewness)))
```

$\gamma_1 = 2.0$

7 Excess Kurtosis (γ_2)

[23]:
```
# Kurtosis (Excess Kurtosis)
kurtosis = expon.stats(loc= 0, scale= 1/ lambda_param, moments='k')
display(Math(r"\gamma_2 = " + str(kurtosis)))
```

$\gamma_2 = 6.0$

8 Median M_e

[24]:
```
# Median
median = expon.median(loc= 0, scale= 1/ lambda_param)
display(Math(r"\text{M_e} = " + str(median)))
```

Median $= 0.34657359027997264$

9 Mode M_o

[26]:
```
mode = np.argmax(pdf_values)
display(Math(r"\text{$M_o$} = " + str(mode)))
```

$M_o = 0$

3.3 Gaussian Distribution

The normal distribution is in many ways the cornerstone of modern statistical theory. It was originally developed as an approximation to the binomial distribution when the number of trials is large and the Bernoulli probability p is not close to 0 or 1. The Normal distribution is a good approximation to the distribution of grades, heights or weights of a large group of individuals, lifetimes of various manufactured items.

3.3.1 Definition

A random variable X is said to have a normal distribution with parameters μ and σ if it has the density

$$X \sim \mathcal{N}(\mu, \sigma) \Leftrightarrow f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} ; -\infty < x < \infty$$

where the mean $\mu \in \mathbb{R}$ and the standard deviation $\sigma > 0$.

Clearly $f_X(x) > 0$; that $\int_{\mathbb{R}_X} f(x) = 1$ is proved as follows.

$$\int_{\mathbb{R}_X} f(x)\, dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Let $z = \dfrac{x-\mu}{\sigma}$, then $x = \mu + z\sigma$. This indicates that $dx = \sigma\, dz$

$$\Rightarrow \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \sigma\, dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz$$

Using **Gauss integration** $\left(\int_{-\infty}^{\infty} e^{-w^2} dw = \sqrt{\pi}\right)$.

$$\Rightarrow \frac{1}{\sqrt{2\pi}} \sqrt{2\pi} = 1$$

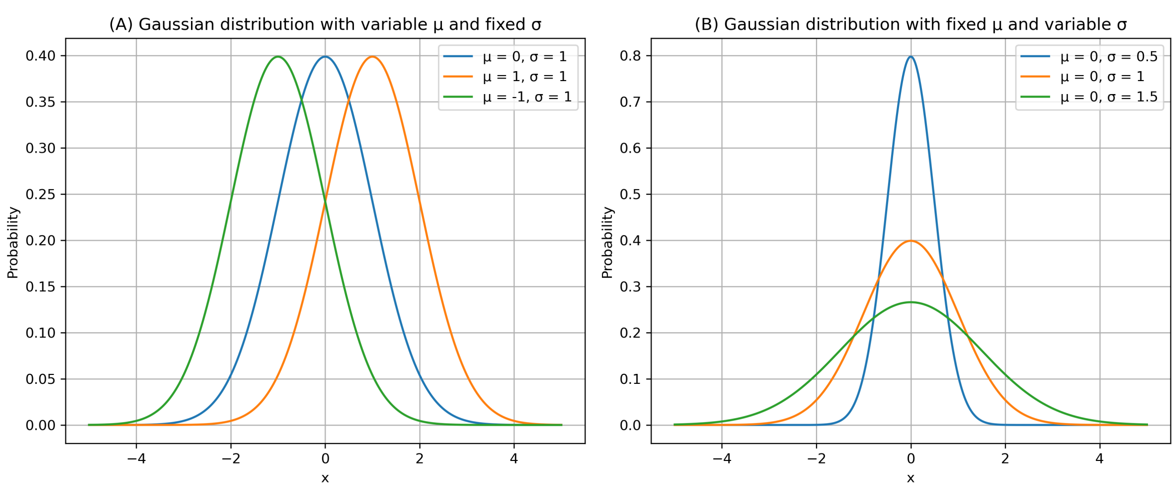
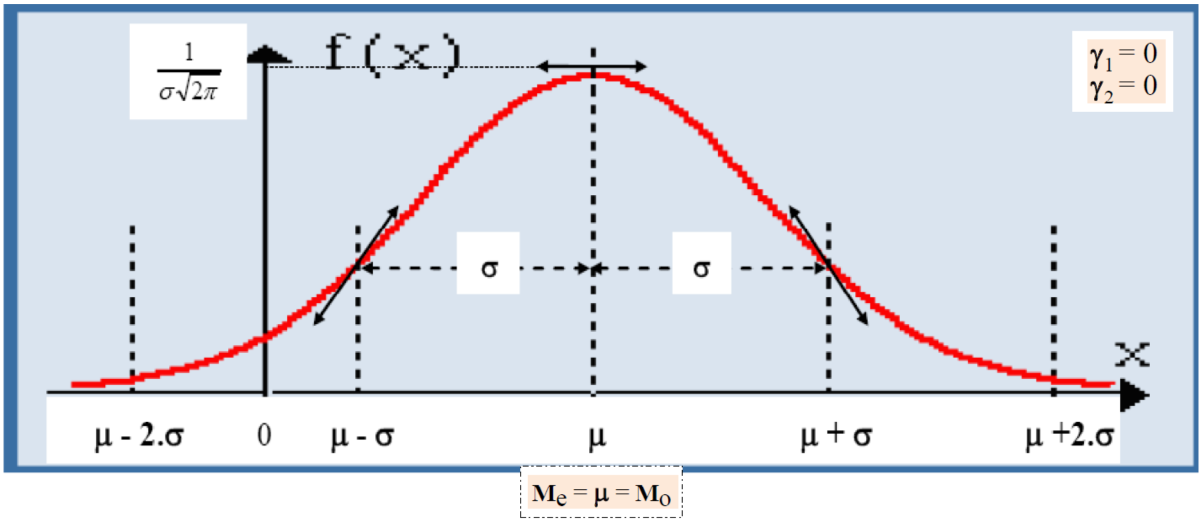
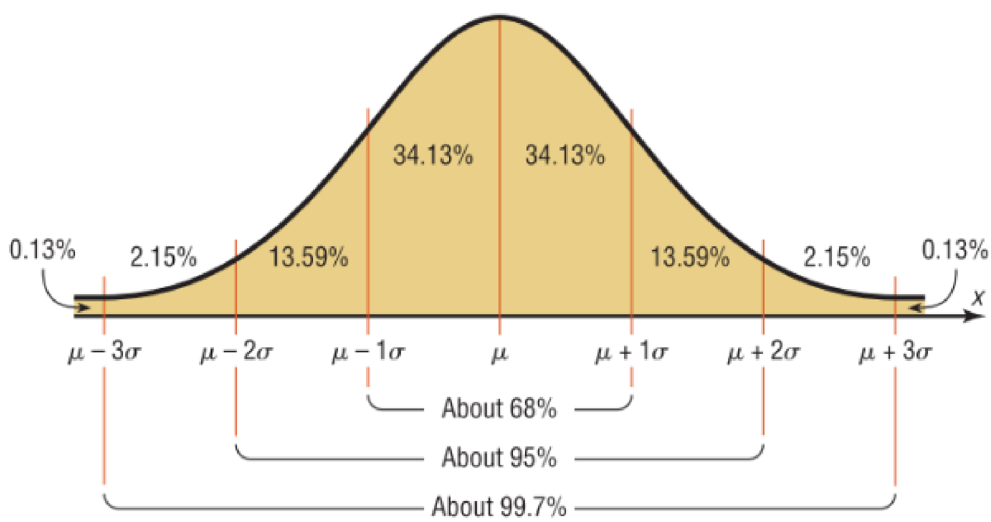


Figure 3.5: The effect of μ and σ on the Gaussian PDF.



(a) The probabilistic characteristics of $\mathcal{N}(\mu, \sigma)$.



(b) Important percentages.

Figure 3.6: The Gaussian PDF.

3.3.2 Moments

The Gaussian distribution is distinguishable by having the following characteristics:

- It is unimodal, having a central peak $\left(\text{Max} f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\right)$. If this were men's heights it would illustrate the fact that most men are clustered around the average height, with a few very tall and a few very short people.
- It is symmetric at $x = \mu$, $(f(\mu - x) = f(\mu + x))$, the left and right halves being mirror images of each other.
- It is bell-shaped.
- It extends continuously over all the values of X from minus infinity to plus infinity, though the value of $f(x)$ becomes extremely small as these values are approached.
- If $X \sim \mathbb{N}(\mu, \sigma)$, then $a\,X + b$ also follows a normal distribution with parameters $a\,\mu + b$ and $|a|\,\sigma$.

The probabilistic characteristics of the Gaussian distribution are summarized in the following table.

X	$\mathcal{N}(\mu, \sigma)$
Range of X	$R_X =]-\infty, \infty[$
Parameters	$\mu \in \mathbb{R}; \sigma \in \mathbb{R}^+$
PDF	$\dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$; if $x \in R_X$
MGF	$M_X(t) = e^{\mu\, t + \frac{\sigma^2}{2} t^2}$
Expected Value	$E(X) = \mu$
Variance	$V(X) = \sigma^2$
Third Central Moment	$\mu_3 = 0$
Fourth Central Moment	$\mu_4 = 3\,\sigma^4$
Skewness	$\gamma_1 = 0$
Excess Kurtosis	$\gamma_2 = 0$

3.3.3 Standard Normal Distribution

If $X \sim \mathcal{N}(\mu, \sigma)$, then $Z = \dfrac{X - \mu}{\sigma}$ follows the standard normal distribution, i.e., the normal distribution with parameters $\mu = 0$ and $\sigma = 1$. In other words, $Z \sim \mathcal{N}(0, 1)$.

This transformation, subtracting the mean and dividing by the standard deviation, is referred to as standardizing X, since the resulting random variable will always have the standard normal distribution with mean 0 and standard deviation 1. In this way, standardizing a normal random variable has the effect of removing the units. Before the prevalence of calculators and computer software capable of calculating normal probabilities, people would apply the standardizing transformation to the normal random variable and use a table of probabilities for the standard normal distribution.

Remarks

- The PDF of Z is $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}; z \in \mathbb{R}$
- $f(z) = f(-z)$ meaning that this distribution is symmetric around 0.
- The CDF of Z is not available in closed form.

$$P[Z \leq z_0] = \phi(z_0) = \int_{-\infty}^{z_0} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \sigma\, dz$$

- Because Z is symmetric around 0, $\phi(0) = 0.5$ and $\phi(-z) = 1 - \phi(z)$.

3.3.4 Reading the Standard Normal Distribution Table

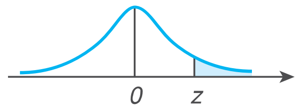
Example 3.3.1
Let X denotes the height of 2^{nd}-year male students. Assuming that X follows the normal distribution with mean 174 cm and standard deviation of 9.6 cm. Find the following:

1. The probability that a student chosen at random has a height above 180 cm.
2. The probability that a student chosen at random has a height below 180 cm.
3. The probability that a student chosen at random has a height equals 180 cm.

To answer the first question we use the table below. the following figure shows how to read it.

z	**0.00**	**0.01**	**0.02**	**0.03**	...	**0.09**
0.0	.5000	.4960	.4920	.4880	...	.4641
0.1	.4602	.4562	.4522	.4483	...	.4247
⋮	⋮	⋮	⋮	⋮	...	⋮
0.5	.3085	.3050	.3015	.2981	...	.2776
0.6	.2743	.2709	.2676	.2643	...	.2451
0.7	.2420	.2389	.2358	.2327	...	.2148

Table A2 The standard Normal distribution



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

$$P[X > 180] = P\left[\frac{X-\mu}{\sigma} > \frac{180-\mu}{\sigma}\right] = P\left[\frac{X-174}{9.6} > \frac{180-174}{9.6}\right] = P\left[Z > 0.63\right] = 0.2643$$

Or we can use the second table

$$P[X > 180] = P\left[Z > 0.63\right] = 1 - P\left[Z \leq 0.63\right] = 1 - \phi(z = 0.63) = 1 - 0.7357 = 0.2643$$

This indicates that 26.43% of 2^{nd}-year male students are over 180 cm.

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

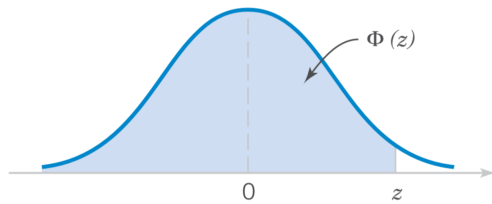


Table II Cumulative Standard Normal Distribution (*continued*)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

Example 3.3.2
Will the answers remain the same if the measurements were in inches?

Given that 1 inch = 2.54 cm, $X \sim \mathcal{N}(68.5, 3.78)$

$$\begin{aligned} P[X > 180\,\text{cm}] &= P[X > 70.87"] \\ &= P\left[\frac{X-\mu}{\sigma} > \frac{70.87-\mu}{\sigma}\right] \\ &= P\left[\frac{X-68.5}{3.78} > \frac{70.87-68.5}{3.78}\right] \\ &= P[Z > 0.63] \\ &= 0.2643 \end{aligned}$$

Example 3.3.3
Packets of cereal have a nominal weight of 750 grams, but there is some variation around this as the machines filling the packets are imperfect. Let us assume that the weights follow a Normal distribution. Suppose that the standard deviation around the mean of 750 is 5 grams.

1. What proportion of packets weigh more than 760 grams?
2. What proportion of packets weigh between 750 and 760 grams?

Example 3.3.4
Z is the standard Gaussian distribution.

1. Find $P[Z \leq 0.04]$, $P[Z > 0.33]$, $P[Z \leq -0.04]$, and $P[Z \leq -0.33]$
2. $P[-1 \leq Z \leq 1]$
3. Find the values of a_1 and a_2 where $P[Z \leq a_1] = 0.8413$ and $P[Z > a_2] = 0.975$
   Let $X \sim \mathcal{N}(10, \sigma)$.
4. If $\sigma = 2$, Compute $P[6 < X \leq 14.5]$ then find Q_3.
5. Determine σ, so $P[X \leq 11.2] = 0.67$.

Example 3.3.5
the logarithm of income is approximately Normally distributed. Suppose the log (to the base 10) of income has the distribution $X \sim \mathcal{N}(4.18, 2.56)$. Calculate the inter-quartile range for X and then take anti-logs to find the inter-quartile range of income.

Example 3.3.6
Solve exercise 22.

3.3.5 Exercises

Exercise 3.3.1
Packets of cereal have a nominal weight of 750 grams, but there is some variation around this as the machines filling the packets are imperfect. Let us assume that the weights follow a Normal distribution. Suppose that the standard deviation around the mean of 750 is 5 grams.

1. What proportion of packets weigh more than 760 grams?
2. What proportion of packets weigh between 750 and 760 grams?

Exercise 3.3.2
Z is the standard Gaussian distribution.

1. Find $P[Z \leq 0.04]$, $P[Z > 0.33]$, $P[Z \leq -0.04]$, and $P[Z \leq -0.33]$
2. $P[-1 \leq Z \leq 1]$
3. Find the values of a_1 and a_2 where $P[Z \leq a_1] = 0.8413$ and $P[Z > a_2] = 0.975$
   Let $X \sim \mathcal{N}(10, \sigma)$.
4. If $\sigma = 2$, Compute $P[6 < X \leq 14.5]$ then find Q_3.
5. Determine σ, so $P[X \leq 11.2] = 0.67$.

Exercise 3.3.3
the logarithm of income is approximately Normally distributed. Suppose the log (to the base 10) of income has the distribution $X \sim \mathcal{N}(4.18, 2.56)$. Calculate the inter-quartile range for X and then take anti-logs to find the inter-quartile range of income.

Exercise 3.3.4
Suppose X is $N(\mu, \sigma^2)$. For $a = 1, 2, 3$, find $P(|X - \mu| < a\sigma)$.

3.3.6 Python Implementation

Gaussian

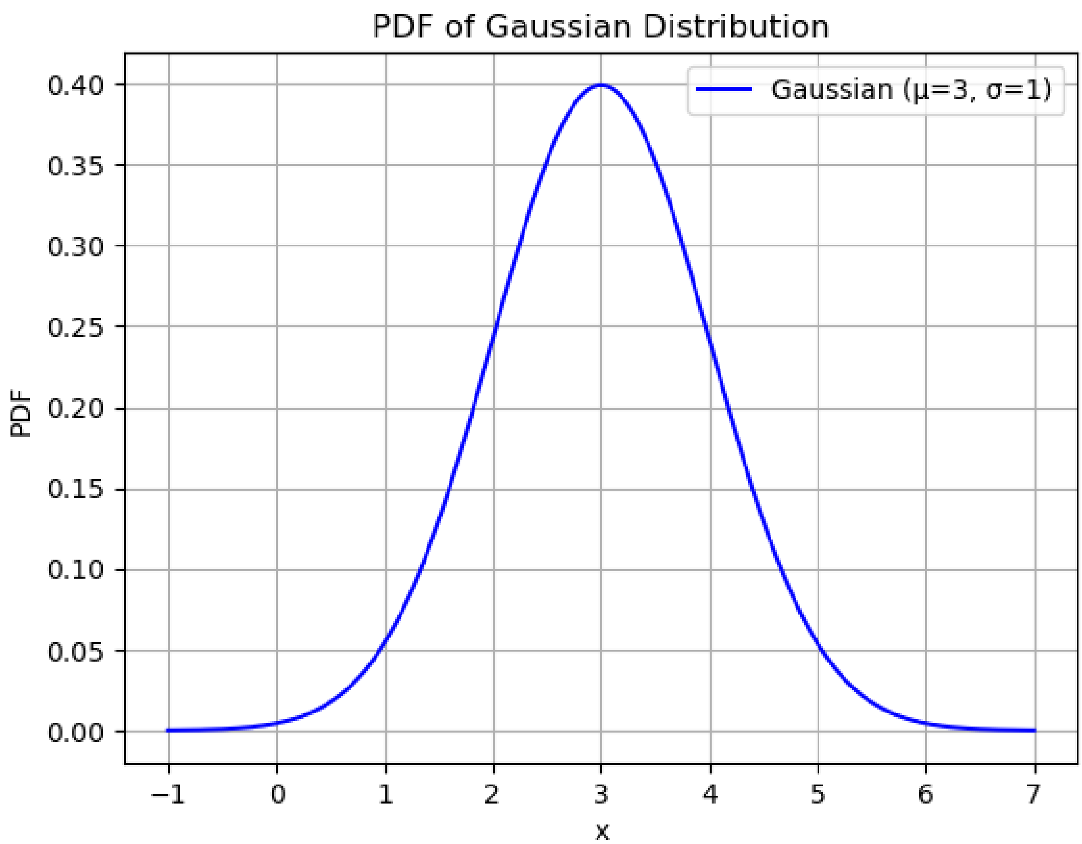
[3]:
```python
from IPython.display import display, Math
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm
```

1 Parameters

[2]:
```python
# Define the parameters for the Gaussian distribution
mu = 3      # Mean
sigma = 1   # Standard deviation
```

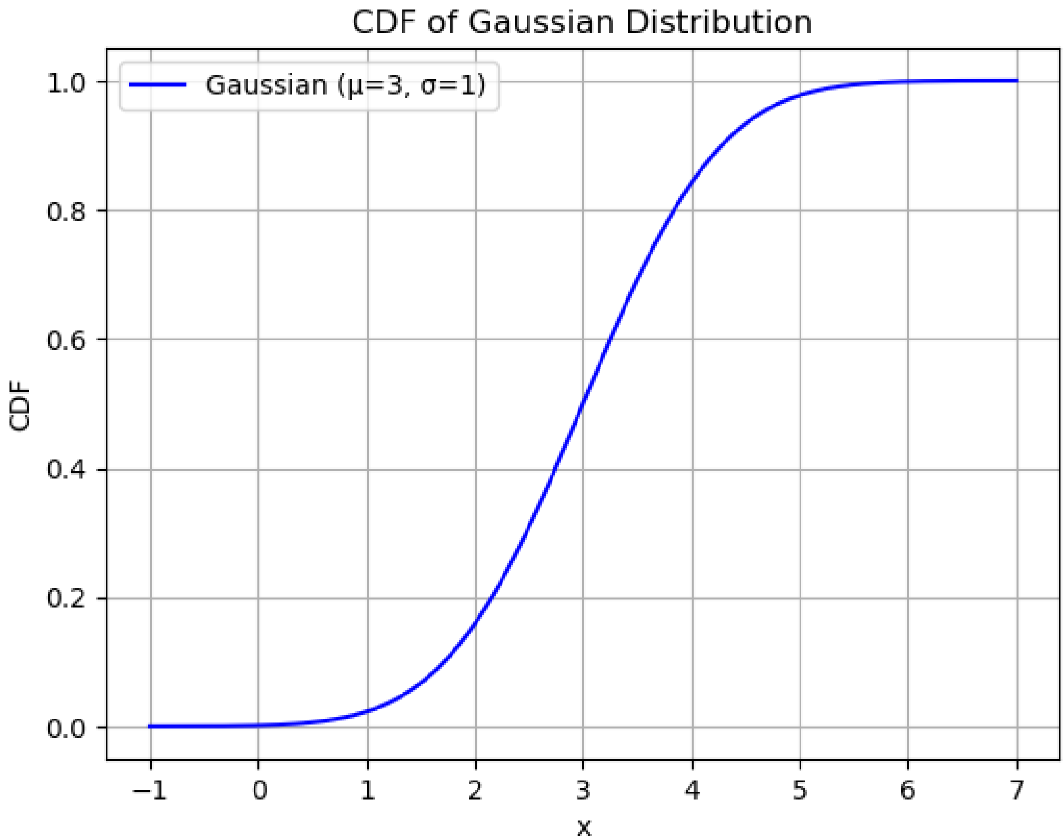
2 Probability Mass Function (PMF)

[8]:
```python
# Generate x values over the range of the uniform distribution
x_values = np.linspace(mu - 4 * sigma, mu + 4 * sigma, 1000)
pdf_values = norm.pdf(x_values, loc=mu, scale=sigma)
plt.plot(x_values, pdf_values, label=f'Gaussian (μ={mu}, σ={sigma})',␣
 ↪color='blue')
plt.xlabel('x')
plt.ylabel('PDF')
plt.title('PDF of Gaussian Distribution')
plt.legend()
plt.grid(True)
```



3 Cumulative Distribution Function (CDF)

[9]:
```
cdf_values = norm.cdf(x_values, loc=mu, scale=sigma)
plt.plot(x_values, cdf_values, label=f'Gaussian (μ={mu}, σ={sigma})',␣
 ↪color='blue')
plt.xlabel('x')
plt.ylabel('CDF')
plt.title('CDF of Gaussian Distribution')
plt.legend()
plt.grid(True)
plt.show()
```



4 Expected value or mean (μ)

[10]:
```
mean = norm.mean(loc=mu, scale=sigma)
display(Math(r"\mu = " + str(mean)))
```

$\mu = 3.0$

5 Variance (σ^2)

[11]:
```
variance = norm.var(loc=mu, scale=sigma)
display(Math(r"\sigma^2 = " + str(variance)))
```

$\sigma^2 = 1.0$

6 Skewness(γ_1)

[12]:
```
# Skewness
skewness = norm.stats(loc=mu, scale=sigma, moments='s')
display(Math(r"\gamma_1 = " + str(skewness)))
```

$\gamma_1 = 0.0$

7 Excess Kurtosis (γ_2)

[13]:
```
# Kurtosis (Excess Kurtosis)
kurtosis = norm.stats(loc=mu, scale=sigma, moments='k')
display(Math(r"\gamma_2 = " + str(kurtosis)))
```

$\gamma_2 = 0.0$

8 Median M_e

[15]:
```
# Median
median = norm.median(loc=mu, scale=sigma)
display(Math(r"\text{$M_e$} = " + str(median)))
```

$M_e = 3.0$

9 Mode M_o

[18]:
```
norm.pdf(x_values, loc=mu, scale=sigma)
mode_index = np.argmax(pdf_values)
mode = x_values[mode_index]
display(Math(r"\text{$M_o$} = " + str(mode)))
```

$M_o = 2.995995995995996$

3.4 Gamma Distribution

3.4.1 Definition

A non-negative random variable X has a gamma distribution with a positive shape parameter ($\beta > 0$) and a positive rate parameter ($\lambda > 0$), if X has the following PDF

$$X \sim \gamma(\beta, \lambda) \Leftrightarrow f_X(x) = \begin{cases} \dfrac{\lambda^\beta}{\Gamma(\beta)} \, x^{\beta-1} \, e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Where $\Gamma(.)$ is the gamma function.

3.4.2 Usage

- In queueing theory, the customers arrive at a bank in a given day according to a Poisson process. The service times for each customer is usually modeled using gamma distribution.
- In insurance, the gamma distribution is sometimes deployed to model claims severity (the total amount of claims divided by their number for an insurer).
- The distribution of certain incomes could be modeled satisfactorily by the gamma distribution.
- In the domain of quality and reliability engineering, the gamma distribution finds extensive application for fitting product lifetimes.

3.4.3 Gamma function

The gamma function is the continuous extension of the factorial function. For a positive number a, it is given as

$$\Gamma(a) = \int_0^\infty y^{a-1} \, e^{-y} \, dy$$

It has the following properties

1. $\Gamma(0) = 1$.
2. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.
3. The recursive property of the gamma function $\Gamma(a+1) = a\,\Gamma(a)$ for $a > 0$.
4. $\Gamma(n+1) = n!$ for n a positive integer.

The mathematical intuition behind the gamma distribution is as follows

$$\begin{aligned} \Gamma(\beta) &= \int_0^\infty y^{\beta-1} \, e^{-y} \, dy \\ \frac{\Gamma(\beta)}{\Gamma(\beta)} &= \int_0^\infty \frac{1}{\Gamma(\beta)} y^{\beta-1} \, e^{-y} \, dy \\ 1 &= \int_0^\infty \frac{1}{\Gamma(\beta)} y^{\beta-1} \, e^{-y} \, dy \end{aligned}$$

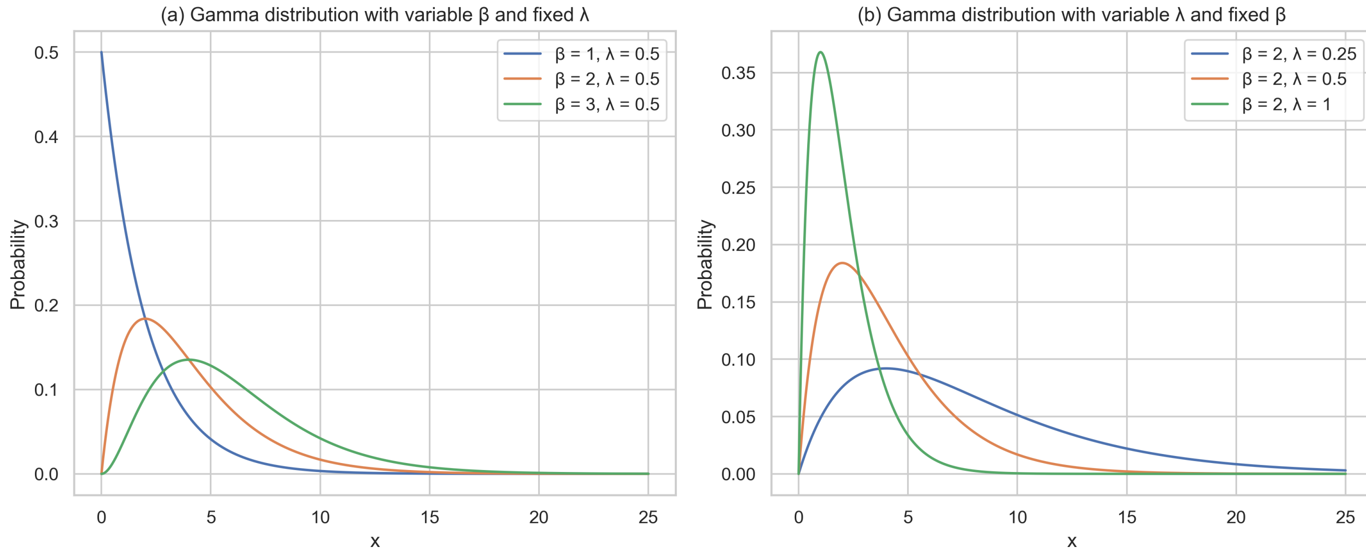


Figure 3.7: The effect of the β and λ parameters on the PDF of the Gamma distribution

By scaling the gamma function we obtain the standard gamma distribution. The quantity $f_Y(y) = \dfrac{1}{\Gamma(\beta)} y^{\beta-1}\, e^{-y}\, dy$ correspond to $Y \sim \gamma\,(\beta, 1)$. The gamma distribution is related to the exponential distribution $\mathcal{E}\,(\lambda)$. Let $X = \dfrac{Y}{\lambda}$. This is how we obtained the PDF of the R.V. X.

$$
\begin{aligned}
F_X(x) &= P\,[X \leq x] \\
&= P\left[\frac{Y}{\lambda} \leq x\right] \\
&= P\,[Y \leq \lambda x] \\
&= F_Y(\lambda x)
\end{aligned}
$$

To find the PDF of X, we differentiate $F_Y(\lambda x)$

$$
\begin{aligned}
f_X(x) &= \lambda\, f_Y(\lambda x) \\
&= \lambda\, \frac{1}{\Gamma(\beta)} (\lambda x)^{\beta-1}\, e^{-\lambda x}\, dy \\
&= \frac{\lambda^{\beta}}{\Gamma(\beta)}\, x^{\beta-1}\, e^{-\lambda x}
\end{aligned}
$$

3.4.4 Moments

The probabilistic characteristics of the gamma distribution are given in the following table.

Table 3.2: Probabilistic Characteristics of the Gamma Distribution.

X	$\boldsymbol{\gamma}\left(\boldsymbol{\beta},\, \boldsymbol{\lambda}\right)$
$\boldsymbol{R_X}$	$[0, \infty[$
Parameters	$\beta, \lambda \in \mathbb{R}^+$
PDF	$f_X\left(x\right) = \dfrac{\lambda^{\beta}}{\Gamma(\beta)}\, x^{\beta-1}\, e^{-\lambda x}$, if $x \in R_X$
MGF	$M_X(t) = \left(1 - \dfrac{t}{\lambda}\right)^{-\beta}$; $t < \lambda$
rth raw moment	$\mu'_r = \dfrac{\Gamma(r+\beta)}{\Gamma(\beta)\lambda^r}$
Expected Value	$E(X) = \dfrac{\beta}{\lambda}$
Mode	$M_o = \begin{cases} \dfrac{\beta - 1}{\lambda}, & \text{if } \beta \geq 1 \\ 0, & \text{otherwise} \end{cases}$
Variance	$V(X) = \dfrac{\beta}{\lambda^2}$
Skewness	$\gamma_1 = \dfrac{2}{\sqrt{\beta}}$
Kurtosis	$\gamma_2 = \dfrac{6}{\beta}$

Example 3.4.1
Solve exercise 20 of the 2^{nd} series.

3.4.5 Notes

- The gamma distribution $\gamma(\beta, \lambda)$ reduces to the **exponential** distribution when $\beta = 1$.
- The gamma distribution is called **Erlang** distribution when the shape parameter is a positive integer $n = \beta$; $n \in \mathbb{N}^*$.
- The gamma distribution $\gamma(\beta, \lambda)$ becomes the **Chi-square** distribution $\chi^2_{(n)}$ when $\beta = \dfrac{n}{2}$ and $\lambda = \dfrac{1}{2}$.
- The sum of n 'iid' (independent and identically distributed) $\mathcal{E}\left(\lambda\right)$ is $\gamma(n, \lambda)$.
- Another parametrization to the gamma distribution that you might encounter in textbooks is $\gamma(\beta, \alpha)$ via using the **Scale** parameter α rather than the rate parameter λ where $\alpha = \dfrac{1}{\lambda}$.

$$f_X\left(x\right) = \begin{cases} \dfrac{1}{\alpha^{\beta}\Gamma(\beta)}\, x^{\beta-1}\, e^{-\dfrac{x}{\alpha}}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

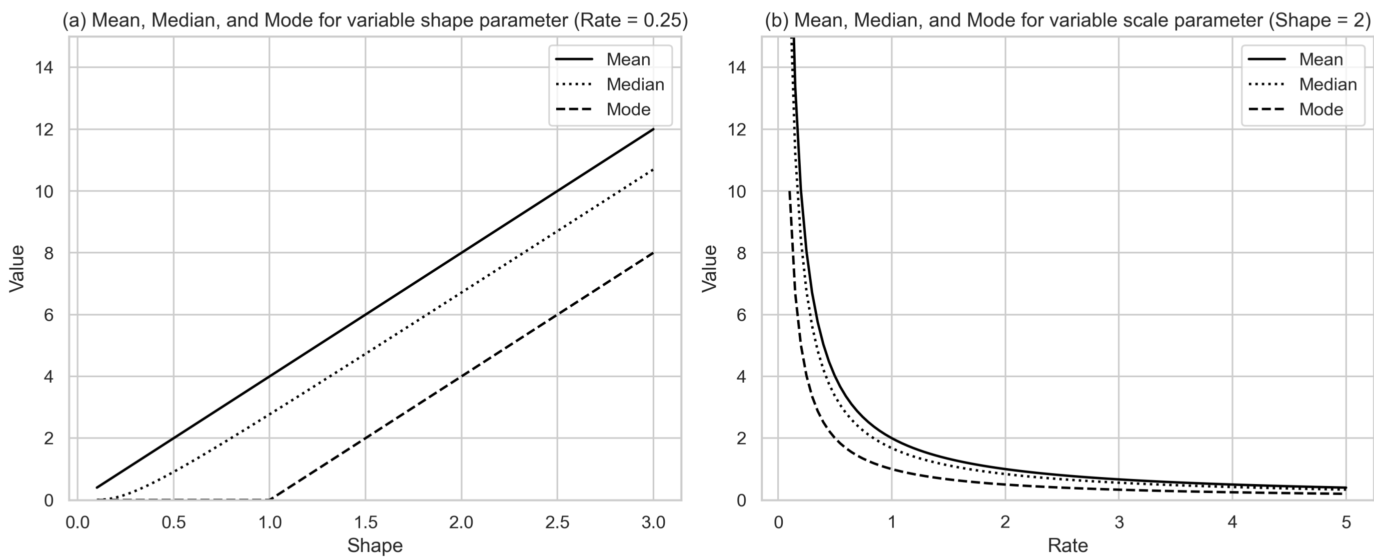


Figure 3.8: The Mean,Median, and Mode of $\gamma(\beta, \lambda)$

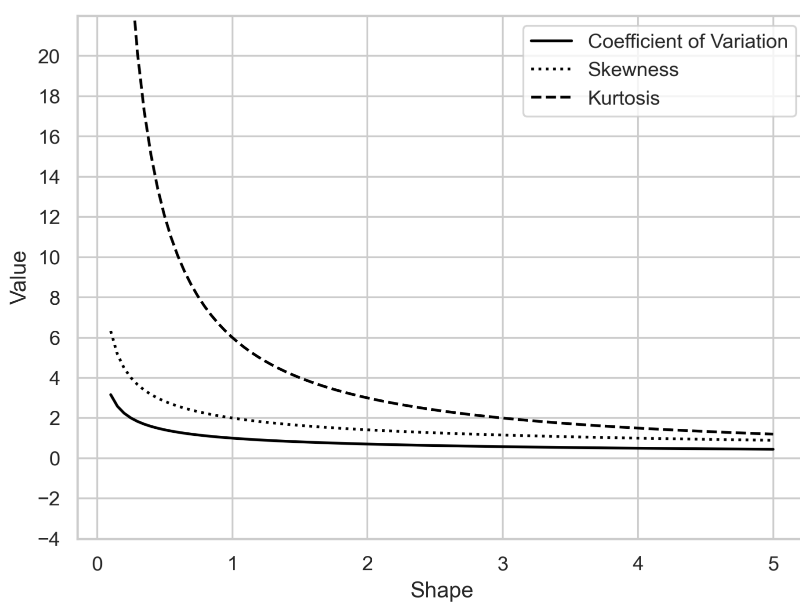


Figure 3.9: CV, Skewness, and Kurtosis of $\gamma(\beta, \lambda)$

3.4.6 Gamma and Exponential Distributions Connection via the Poisson Process

A Poisson process is any process that satisfies the following conditions:

- The number of events in disjoint time intervals is independent.
- The probability of more than one event in subinterval (an infinitesimally small time interval) is negligible.
- The probability that an event occurs during a given subinterval is constant over the entire interval from 0 to T.

Empirical Example

Let's say we are interested in modeling the arrivals of customers at the Gulf Bank branch located at the city center on Thursdays. Specifically, we would like to study the characteristics of the number of arrivals $N(t)$ and the inter-arrivals time X_j.

It is found that:

- The number of arrivals is poisson process, $N(t) \sim \mathcal{P}\left(\lambda t\right)$.
- The time until the first arrival, $X_1 \sim \mathcal{E}\left(\lambda\right)$.
- the inter-arrivals times are iid, $X_j \sim \mathcal{E}\left(\lambda\right)$.
- The time until the n-th arrival, $S_n \sim \gamma(n, \lambda)$.

The proof will be presented during the class session.
In other words, the **exponential** distribution is the continuous analogous of the **geometric** distribution, and the **gamma (Erlang)** distribution is the continuous counterpart of the **negative binomial** distribution.

Example 3.4.2
Solve Exercise 11 of the second series.

$$Y_t \sim \mathcal{P}\left(\lambda t\right) \Leftrightarrow P\left(y\right) = \begin{cases} e^{-\lambda t}\,\dfrac{(\lambda t)^y}{y!}, & \text{if } y \in \mathbb{N} \\ 0, & \text{otherwise} \end{cases}$$

$$P\left[X > x\right] = P\left[Y_t = 0\right] = e^{-\lambda t}\,\frac{(\lambda t)^0}{0!} = e^{-\lambda t}$$

This indicates that $X \sim \mathcal{E}\left(\lambda\right)$.
$E(X) = \dfrac{1}{\lambda}$
$E(Y_t) = \lambda t$

3.4.7 Gamma and Gaussian Distributions Connection

Let $Y = \sum_{j=1}^{n} X_j^2$ where $X_1, X_2, \ldots, X_n$ are independent standard normal random variables $X_j \sim \mathcal{N}(0,1); j = 1, 2, \ldots, n$. Then Y has a gamma distribution with shape $\beta = \dfrac{n}{2}$ and rate $\lambda = \dfrac{1}{2}$, which is $\chi^2_{(n)}$.

3.4.8 The CDF of the Gamma Distribution

Before addresing the CDF of $\gamma(\beta, \lambda)$, it is important to identify two functions that stem from the gamma function.

$$\begin{aligned}\Gamma(a) &= \int_0^\infty y^{a-1}\, e^{-y}\, dy \\ &= \int_0^x y^{a-1}\, e^{-y}\, dy + \int_x^\infty y^{a-1}\, e^{-y}\, dy \\ &= \gamma(a,\, x) + \Gamma(a,\, x)\end{aligned}$$

$\gamma(a,\, x)$ is called the lower incomplete gamma function, where $\Gamma(a,\, x)$ is known as the upper incomplete gamma function.

Let X be a R.V. that follows the gamma distribution with shape parameter β and rate parameter λ.

$$\begin{aligned}F_X(x_0) &= P\left[X \leq x_0\right] \\ &= \int_0^{x_0} \frac{\lambda^\beta}{\Gamma(\beta)} x^{\beta-1}\, e^{-\lambda x}\, dx \\ &= \frac{\lambda^\beta}{\Gamma(\beta)} \int_0^{x_0} x^{\beta-1}\, e^{-\lambda x}\, dx\end{aligned}$$

By taking $z = \lambda x$, The CDF becomes,

$$\begin{aligned}F_X(x_0) &= \frac{1}{\Gamma(\beta)} \int_0^{\lambda\, x_0} z^{\beta-1}\, e^{-z}\, dz \\ &= \frac{1}{\Gamma(\beta)}\, \gamma(\beta,\, \lambda x_0)\end{aligned}$$

In other words

$$F_X(x) = \frac{1}{\Gamma(\beta)}\, \gamma(\beta,\, \lambda x)$$

3.4.9 Exercises

Exercise 3.4.1
A certain electronic system has a life length of X_1, which follows an exponential distribution with a mean of 450 hours. The system is supported by an identical backup system that has a life length of X_2. The backup system takes over immediately when the primary system fails. If the systems operate independently, find the probability distribution and expected value for the total life length of the primary and backup systems.

Exercise 3.4.2
Suppose that Y has a gamma distribution with mean 40 and standard deviation 20. Find the shape parameter β and the rate parameter λ.

Exercise 3.4.3
Calls to a telephone system follow a Poisson distribution with a mean of five calls per minute.

1. What is the name applied to the distribution and parameter values of the time until the tenth call?

2. What is the mean time until the tenth call?

(c) What is the mean time between the ninth and tenth calls?

3. What is the probability that exactly four calls occur within one minute?

4. If 10 separate one-minute intervals are chosen, what is the probability that all intervals contain more than two calls?

3.4.10 Python Implementation

Gamma

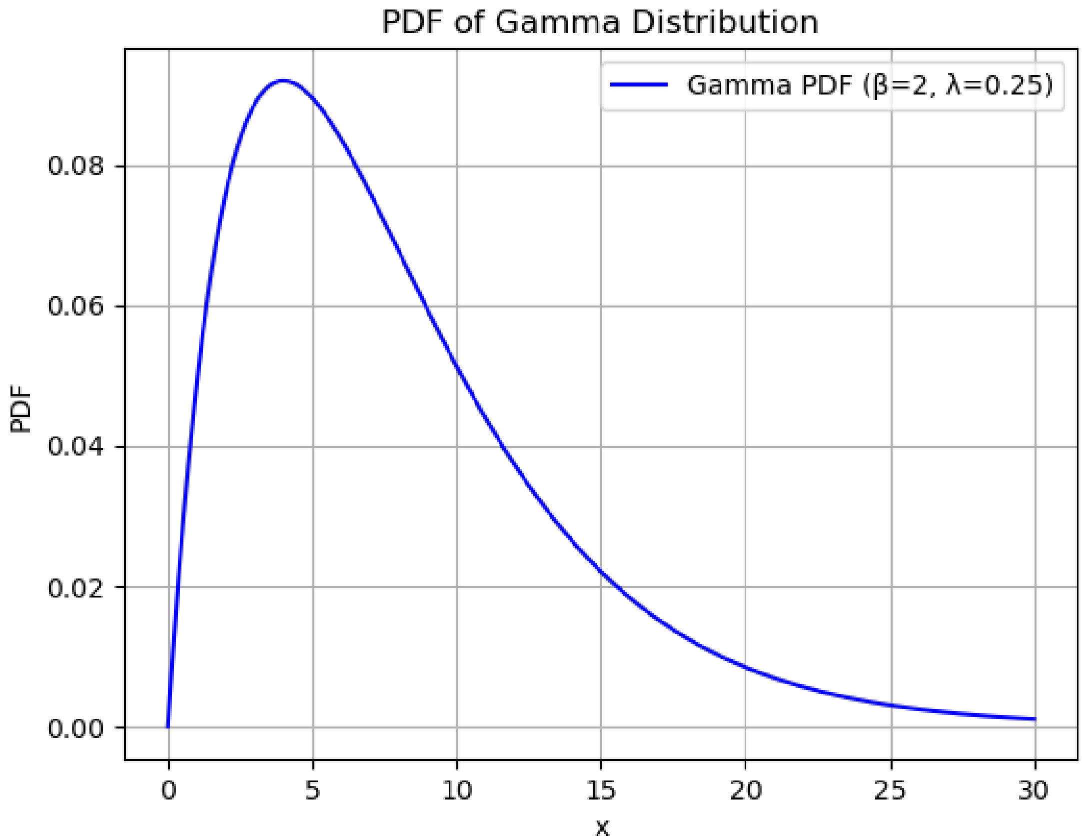
```
[1]: from IPython.display import display, Math
     import numpy as np
     import matplotlib.pyplot as plt
     from scipy.stats import gamma
```

1 Parameters

```
[6]: # Define the parameters for the Gaussian distribution
     beta_param = 2    # Shape parameter
     lambda_param = 0.25    # rate parameter
```

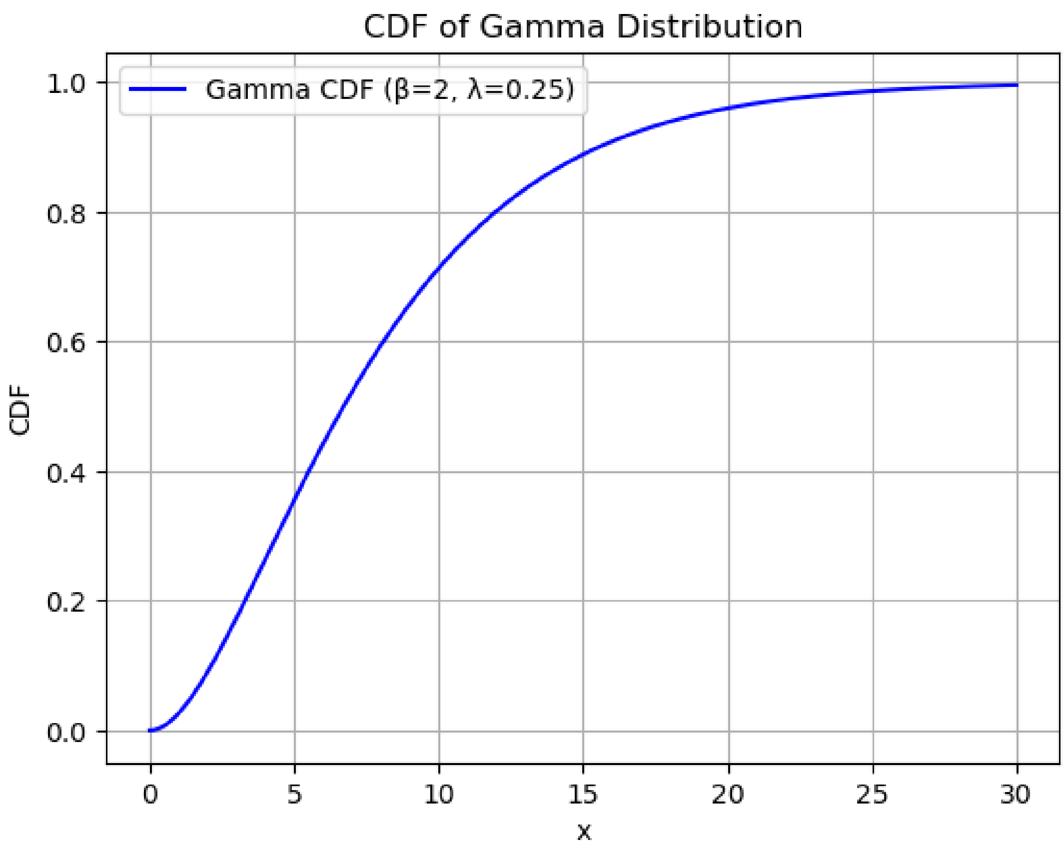
2 Probability Mass Function (PMF)

```
[25]: # Generate x values over the range of the uniform distribution
      x_values = np.linspace(0, 30, 1000)
      pdf_values = gamma.pdf(x_values, a= beta_param, scale= 1/lambda_param)
      plt.plot(x_values, pdf_values, label=f'Gamma PDF (β={beta_param},␣
       ↪λ={lambda_param})', color='blue')
      plt.xlabel('x')
      plt.ylabel('PDF')
      plt.title('PDF of Gamma Distribution')
      plt.legend()
      plt.grid(True)
```



3 Cumulative Distribution Function (CDF)

[24]:
```
cdf_values = gamma.cdf(x_values, a= beta_param, scale= 1/lambda_param)
plt.plot(x_values, cdf_values, label=f'Gamma CDF (β={beta_param},␣
↪λ={lambda_param})', color='blue')
plt.xlabel('x')
plt.ylabel('CDF')
plt.title('CDF of Gamma Distribution')
plt.legend()
plt.grid(True)
```



4 Expected value or mean (μ)

[10]:
```
mean = gamma.mean( a= beta_param, scale= 1/lambda_param)
display(Math(r"\mu = " + str(mean)))
```

$\mu = 8.0$

5 Variance (σ^2)

[12]:
```
variance = gamma.var( a= beta_param, scale= 1/lambda_param)
display(Math(r"\sigma^2 = " + str(variance)))
```

$\sigma^2 = 32.0$

6 Skewness(γ_1)

[14]:
```
# Skewness
skewness = gamma.stats( a= beta_param, scale= 1/lambda_param, moments = 's')
display(Math(r"\gamma_1 = " + str(skewness)))
```

$\gamma_1 = 1.414213562373095$

7 Excess Kurtosis (γ_2)

[15]:
```
# Kurtosis (Excess Kurtosis)
kurtosis = gamma.stats(a=beta_param, scale= 1/ lambda_param, moments='k')
display(Math(r"\gamma_2 = " + str(kurtosis)))
```

$\gamma_2 = 3.0$

8 Median M_e

[16]:
```
# Median
median = gamma.median(a=beta_param, scale=1/ lambda_param)
display(Math(r"\text{$M_e$} = " + str(median)))
```

$M_e = 6.713387960066645$

9 Mode M_o

[27]:
```
pdf_values = gamma.pdf(x_values, a= beta_param, scale= 1/lambda_param)
mode_index = np.argmax(pdf_values)
mode = x_values[mode_index]
display(Math(r"\text{$M_o$} = " + str(mode)))
```

$M_o = 3.993993993993994$

Bibliography

[1] Ross, S. M., & Ross, S. M. (1976). *A first course in probability* (Vol. 2). Macmillan New York.

[2] Montgomery, D. C. (2002). *Applied statistics and probability for engineers*. John Wiley & Sons, Inc.

[3] Larsen, R. J., & Marx, M. L. (2005). *An introduction to mathematical statistics*. Prentice Hall Hoboken, NJ.

[4] Spiegel, M. R. (2016). *Schaum's Outlines Probability and Statistics*.

[5] Tuckwell, H. C. (2018). *Elementary applications of probability theory*. Chapman and Hall/CRC.