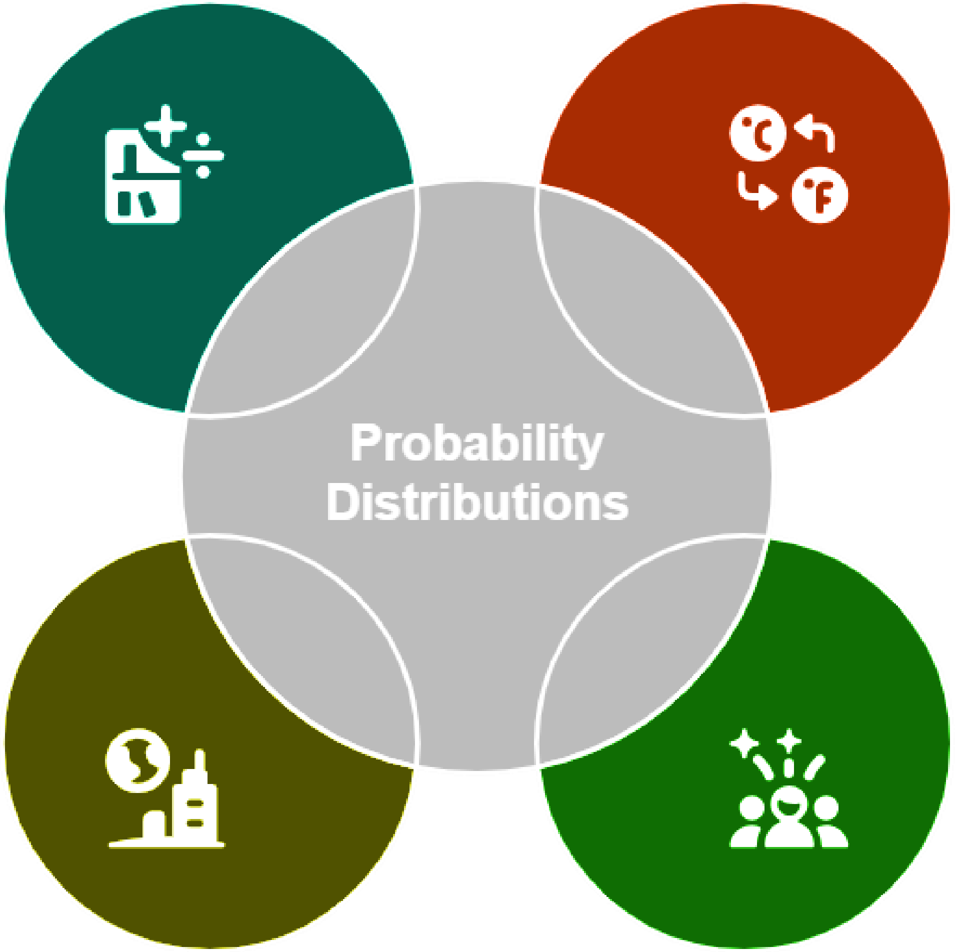
Epidemiology

Probability and Probability Distributions

Assis . Prof. Dr. Labeed Al - Saad

The objectives



Key Characteristics

Recognizing essential traits and formulas for each distribution.

Basic Probability

Understanding fundamental probability concepts and their real-life significance.

Real-World Applications

Identifying practical examples for each probability distribution.

Distribution Types

Detailed exploration of binomial, Poisson, normal, and exponential distributions.

Introduction to Probability

Probability is the study of how likely an event is to happen. It is a number between 0 and 1, where:

0 means the event will never happen.
1 means the event will always happen.

For example, the probability of flipping a fair coin and getting heads is 0.5 (or 50%).

Probability helps us make predictions and decisions in uncertain situations, like predicting the weather, winning a game, or even planning for traffic.

Probability Distributions

Probability Distributions: describes how probabilities are spread across the possible outcomes of a random event. Today, we'll focus on four important types:

Binomial Distribution

A distribution representing the number of successes in a fixed number of trials.

Poisson Distribution

A distribution that expresses the probability of a given number of events occurring in a fixed interval.

Normal Distribution

A continuous probability distribution characterized by its bell-shaped curve.

Exponential Distribution

A distribution that models the time between events in a Poisson process.

Binomial Distribution

1. Binomial Distribution

The binomial distribution is used when there are exactly two possible outcomes for an experiment: success or failure.

Key Characteristics:

- Fixed number of trials (n).
- Each trial is independent (the outcome of one doesn't affect the others).
- Probability of success (p) is the same for each trial.

Example:

Flipping a coin 10 times and counting the number of heads. Here:

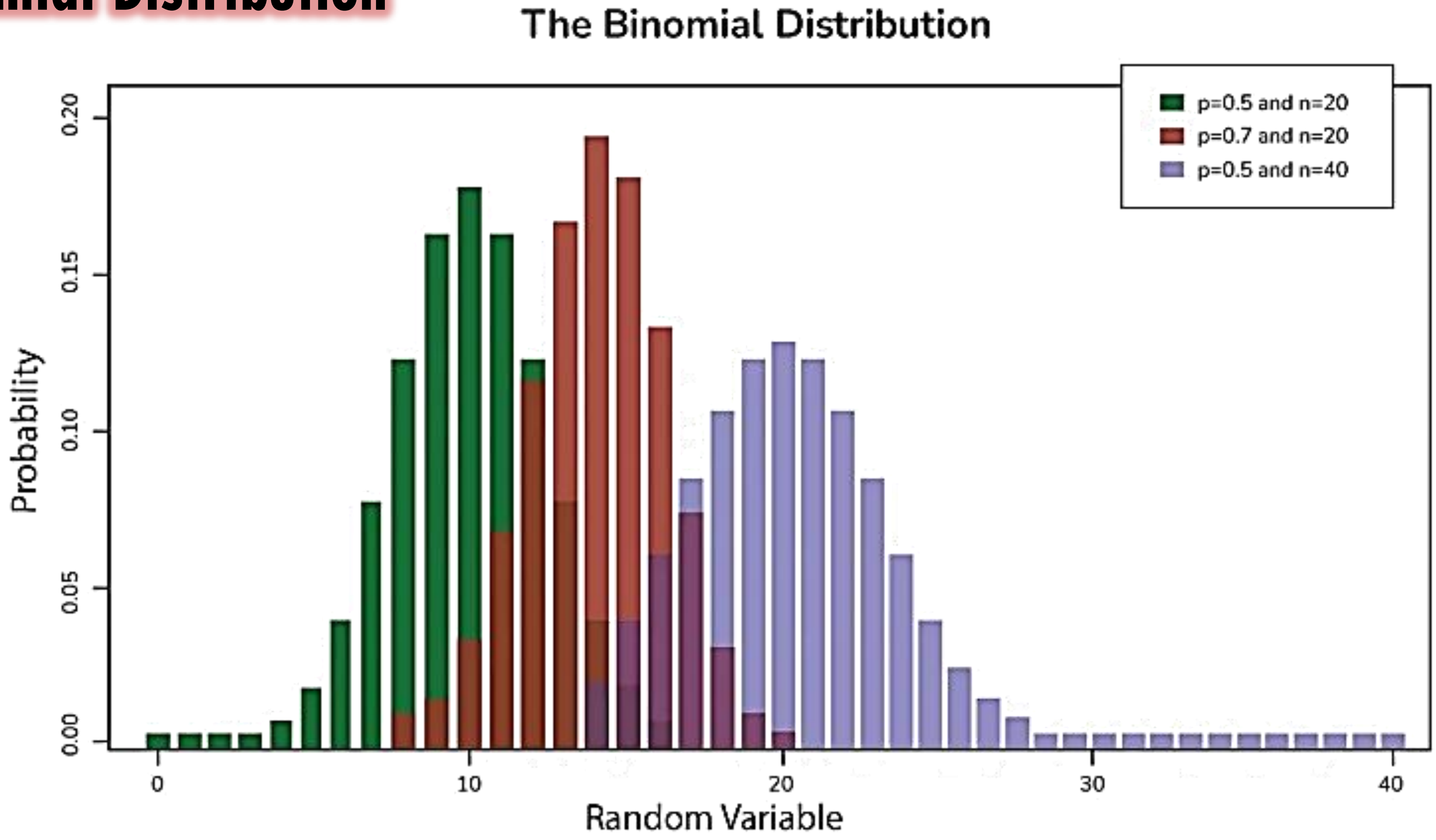
n = 10 (number of trials).

p = 0.5 (probability of getting heads).

Formula:

The probability of getting exactly k successes in n trials is: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

Binomial Distribution



https://www.technologynetworks.com/informatics/articles/the-binomial-test-366022

Binomial Distribution

Problem:

A doctor knows that 10% of patients experience side effects from a certain medication. If 5 patients are randomly selected, what is the probability that exactly 2 of them will experience side effects?

Solution:

The binomial distribution formula is: $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$

Where:

n=5 (number of trials),

k=2 (number of successes),

p=0.10 (probability of success).

Plugging in the values:

$$P(X=2)=\binom{5}{2}(0.10)^2(0.90)^3$$

$$P(X=2)=10\cdot(0.01)\cdot(0.729)$$

$$P(X=2)=0.0729$$

Poisson Distribution

2. Poisson Distribution

The Poisson distribution is used to count how many times an event happens in a fixed interval of time or space.

Key Characteristics:

- Events happen independently.
- The average rate (λ) of occurrence is constant.
- The probability of more than one event happening at the same time is very low.

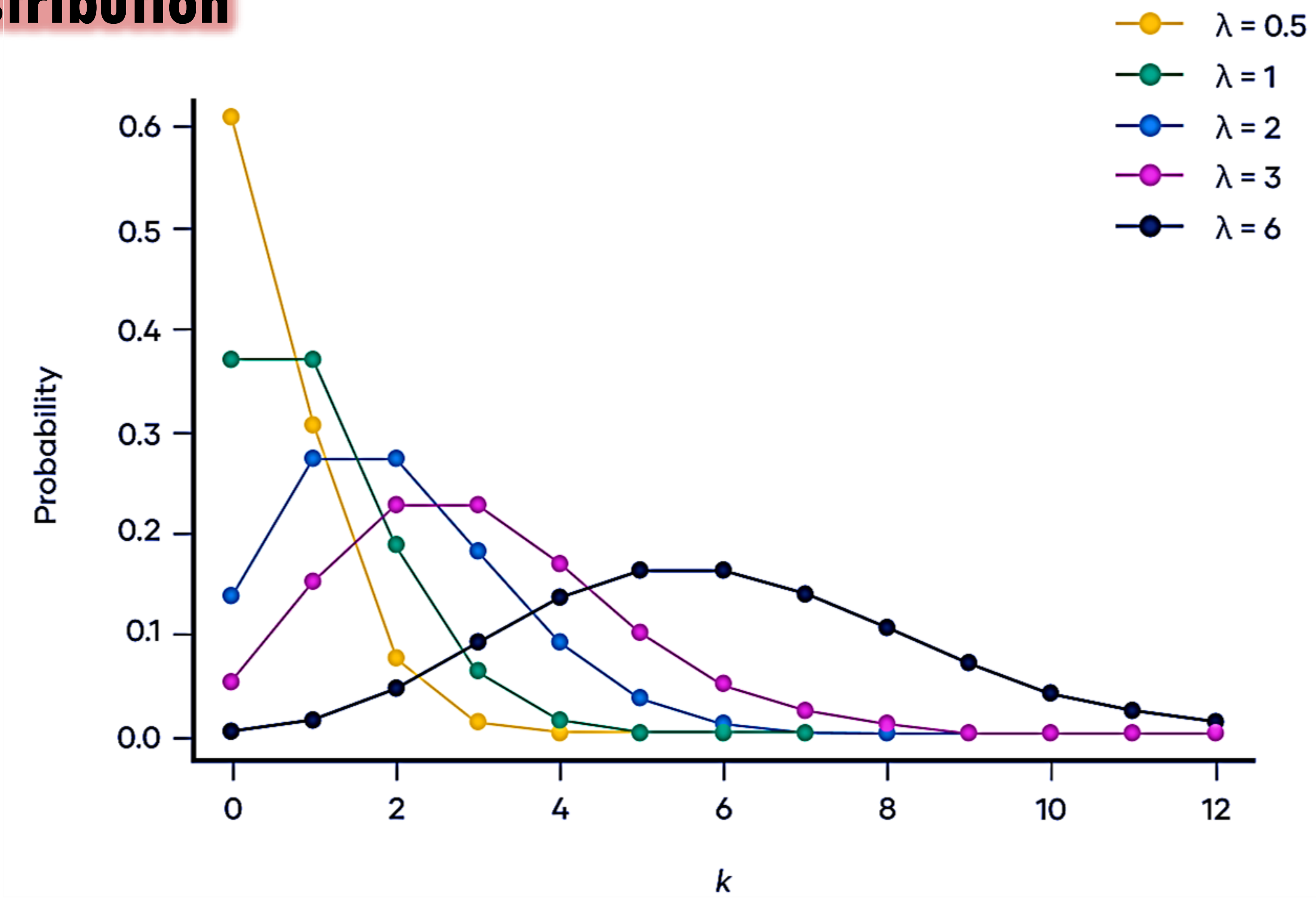
Example:

Counting the number of cars passing through a toll booth in an hour. If, on average, 5 cars pass every hour, $\lambda = 5$.

Formula:

The probability of k events happening in an interval is: $$P(X = k) = \frac{e^{-\lambda}\lambda^{k}}{k!}$$

Poisson Distribution



https://www.scribbr.nl/wp-content/uploads/2022/08/Poisson-distribution-graph.webp

Poisson Distribution

Problem:

A hospital emergency room receives an average of 4 patients per hour. What is the probability that exactly 3 patients will arrive in the next hour?

Solution:

The Poisson distribution formula is:

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

Where:

λ=4 (average rate),

k=3 (number of events).

Plugging in the values:

$$P(X = 3) = \frac{e^{-4} \cdot 4^3}{3!}$$

$$P(X = 3) = \frac{0.0183 \cdot 64}{6}$$

$$P(X = 3) = \frac{1.1712}{6}$$

$$P(X = 3) = 0.1952$$

Exponential Distribution

4. Exponential Distribution

The exponential distribution is used to model the time between events in a Poisson process.

Key Characteristics:

- Events happen continuously and independently at a constant average rate.
- Often used for waiting times or lifetimes.

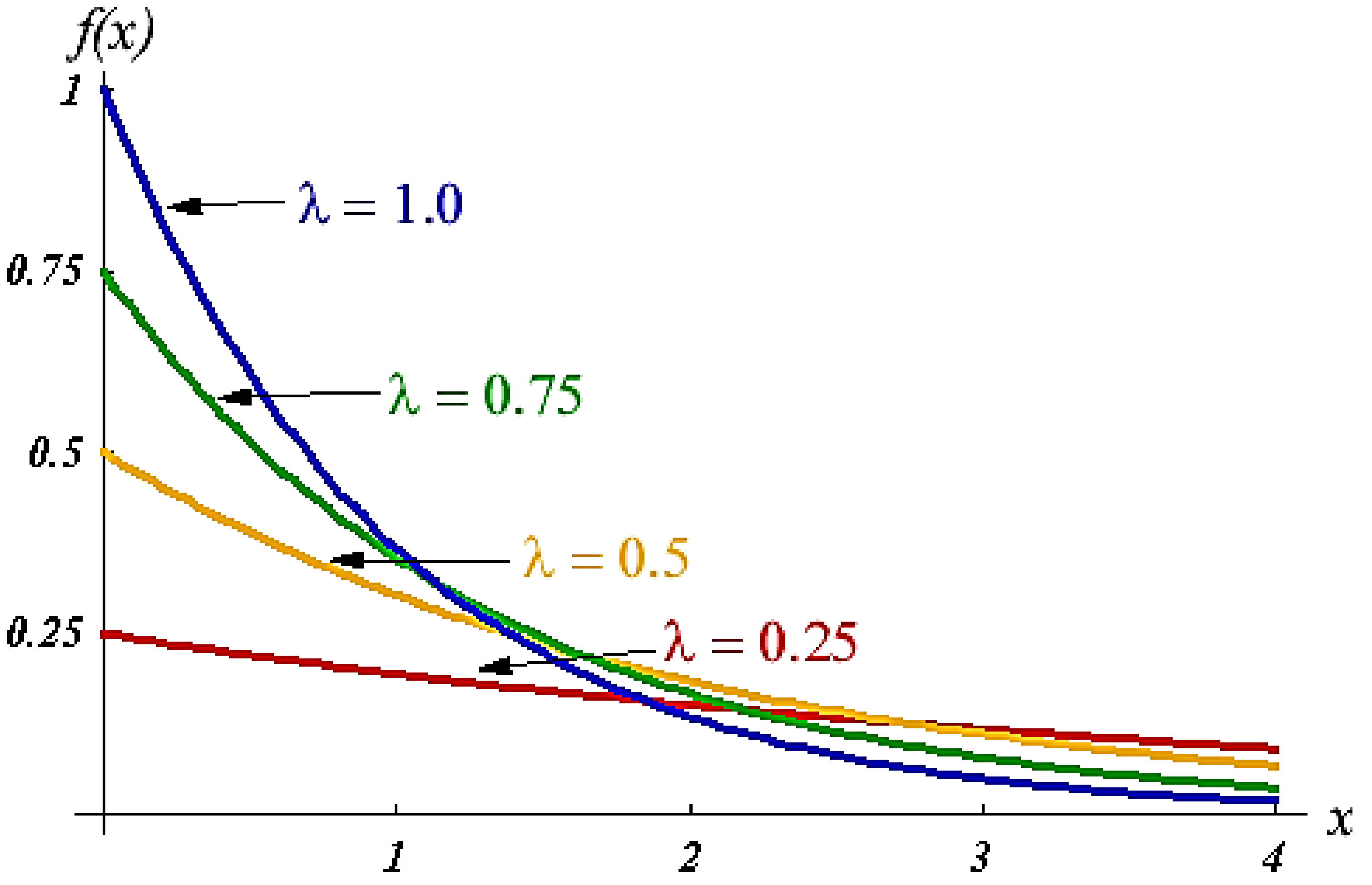
Example:

The time between arrivals of buses at a bus stop. If buses arrive, on average, every 10 minutes, the time between arrivals follows an exponential distribution.

Formula:

The exponential distribution formula is: $P(X \leq x) = 1 - e^{-\lambda x}$

Exponential Distribution



https://www.efunda.com/math/distributions/images/ExpDistPlot.gif

Exponential Distribution

Problem:

The average time between patient arrivals at a clinic is 10 minutes. What is the probability that the next patient will arrive within 5 minutes?

Solution:

The exponential distribution formula is: $P(X \leq x) = 1 - e^{-\lambda x}$

Where:

$\lambda = \frac{1}{10} = 0.1$ (rate parameter, in patients per minute),

$x = 5$ (time in minutes).

Plugging in the values:

$$P(X \leq 5) = 1 - e^{-0.1 \cdot 5}$$

$$P(X \leq 5) = 1 - e^{-0.5}$$

$$P(X \leq 5) = 1 - 0.6065$$

$$P(X \leq 5) = 0.3935$$

The probability is 39.35%.

Normal Distribution

3. Normal Distribution

The normal distribution, also called the "bell curve," is one of the most important distributions in statistics.

Key Characteristics:

- Symmetrical around the mean (average).
- Most of the data is clustered around the mean.
- Described by the mean (μ) and standard deviation (σ).

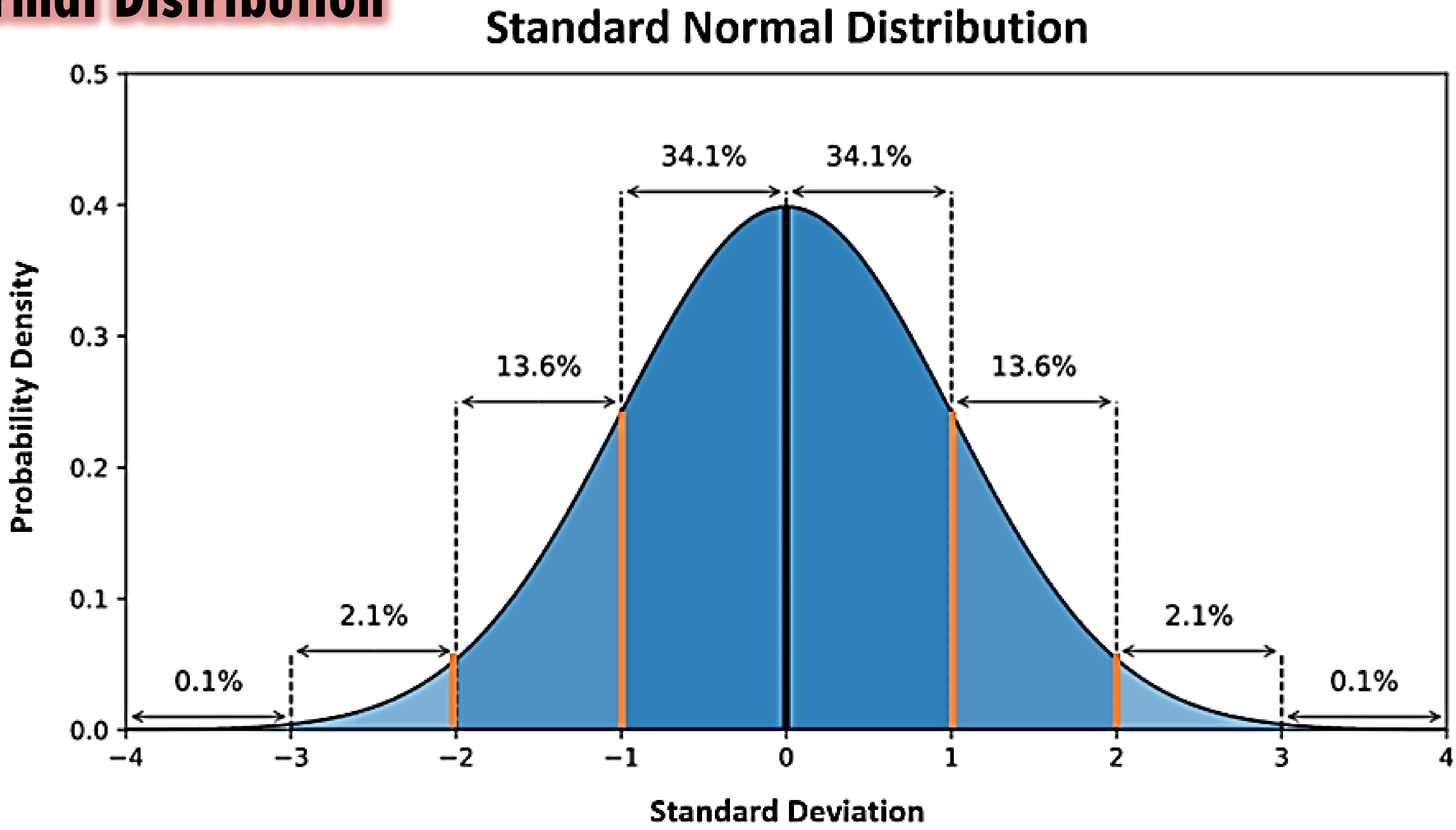
Example:

Heights of students in a class. Most students are of average height, with fewer being very tall or very short.

Formula: The probability density function is:

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

Normal Distribution



https://rocklandpeakperformance.com/wp-content/uploads/2022/07/Normal-Distribution.png

Normal Distribution

Problem:

The average height of adult males in a population is 70 inches, with a standard deviation of 3 inches. What is the probability that a randomly selected male is taller than 75 inches?

Solution:

First, calculate the z-score: $$z = \frac{X - \mu}{\sigma}$$

Where:

X = 75, μ = 70, σ = 3 ➔ z= 1.67

Using a z-table, the probability corresponding to z=1.67 is 0.9525. This represents the probability of being less than 75 inches.

To find the probability of being taller than 75 inches:

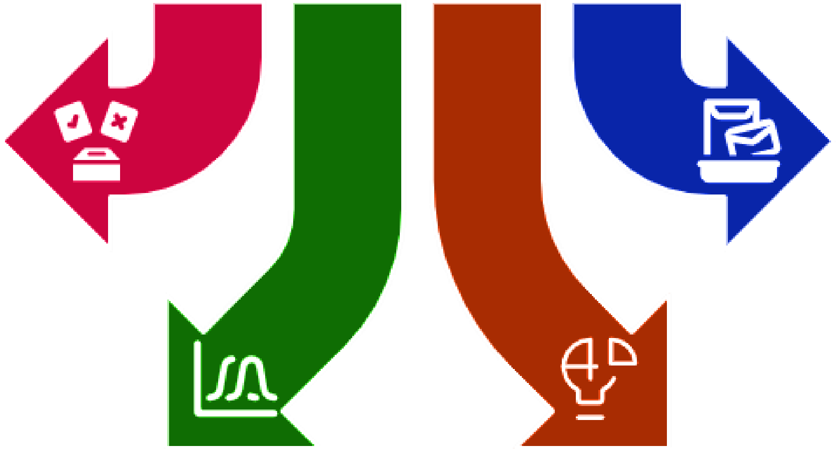
P(X>75)=1−0.9525=0.0475

➔ The probability is 4.75%

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9031	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	$\boxed{0.9525}$	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857

Real-World Applications

Which probability distribution should be used for analysis?



Binomial Distribution

Use when predicting the number of successes in a fixed number of trials, such as defective items in a batch.

Poisson Distribution

Use for counting occurrences in a fixed interval, like daily email receipts.

Normal Distribution

Use for analyzing data that clusters around a mean, such as test scores or heights.

Exponential Distribution

Use for modeling time until an event occurs, like lightbulb lifespan or earthquake intervals.

Probability Distribution vs. Statistical Tests

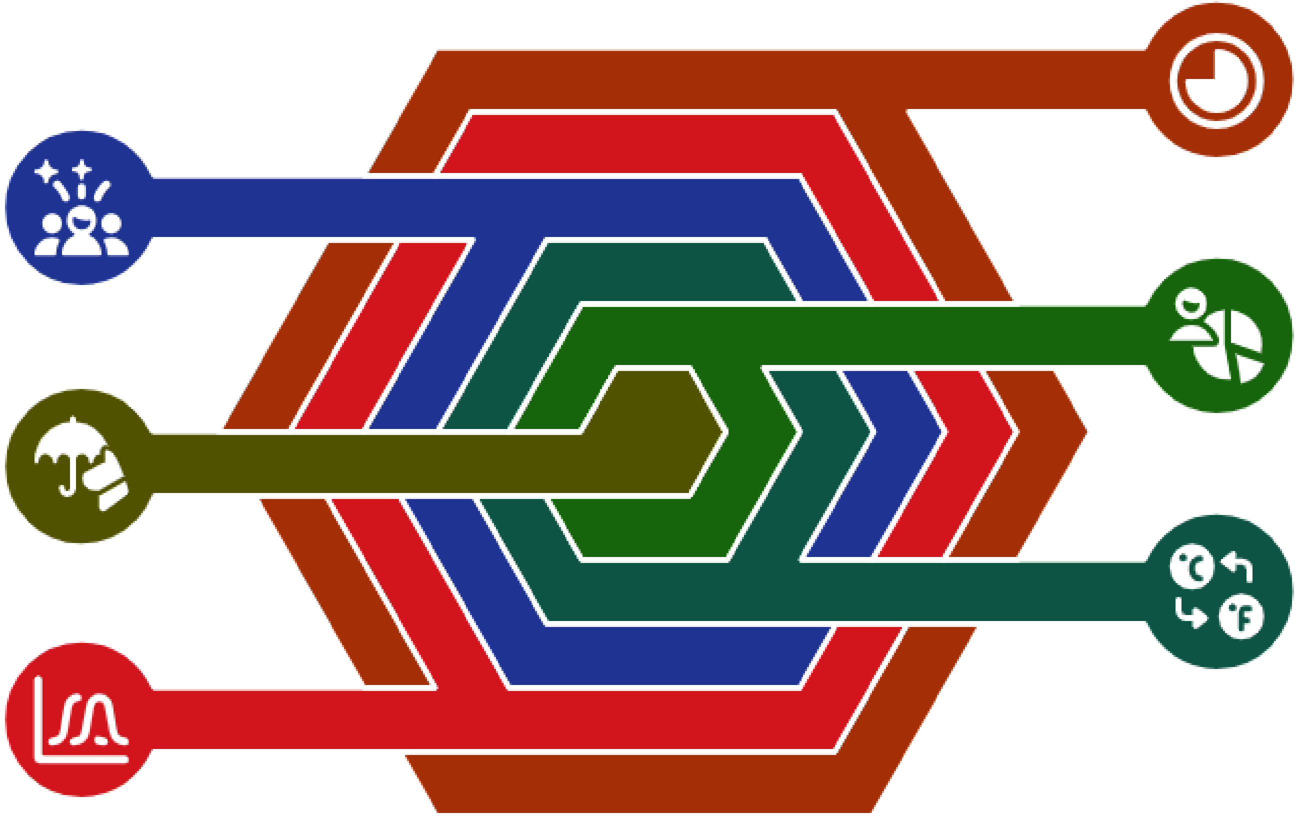
Probability Distribution	Statistical Test	Description	Medical Example
Normal Distribution	Z-test	Tests the mean of a normally distributed population with known variance.	Testing if the average blood pressure of patients on a new drug differs from the known population average.
	T-test	Tests the mean of a normally distributed population with unknown variance.	Comparing the mean cholesterol levels of two groups of patients (e.g., drug vs. placebo).
	ANOVA (Analysis of Variance)	Compares means of three or more groups with normally distributed data.	Comparing the effectiveness of three different doses of a medication on reducing blood sugar levels.
	Pearson's Correlation	Measures the linear relationship between two normally distributed variables.	Examining the relationship between age and blood pressure in a group of patients.
	Linear Regression	Models the relationship between a dependent and independent variable (normal).	Predicting a patient's BMI based on their calorie intake and exercise habits.
Binomial Distribution	Binomial Test	Tests the probability of success in a binomial experiment.	Testing if the proportion of patients recovering from a disease after treatment is greater than 50%.
	Chi-Square Goodness-of-Fit Test	Tests if observed frequencies match expected frequencies in categorical data.	Checking if the distribution of blood types in a hospital matches the general population.
	Chi-Square Test of Independence	Tests if two categorical variables are independent.	Testing if smoking status (smoker/non-smoker) is independent of lung cancer diagnosis.
Poisson Distribution	Poisson Test	Tests if data follows a Poisson distribution (e.g., counts of events).	Testing if the number of hospital admissions per day follows a Poisson distribution.
	Chi-Square Goodness-of-Fit Test	Tests if observed count data matches a Poisson distribution.	Checking if the number of heart attacks per month in a clinic matches a Poisson distribution.
Exponential Distribution	Exponential Test	Tests if data follows an exponential distribution (e.g., waiting times).	Testing if the time between patient arrivals at an emergency room follows an exponential distribution.
	Kolmogorov-Smirnov Test	Tests if a sample follows a specific distribution (e.g., exponential).	Testing if the survival time of patients after a specific treatment follows an exponential distribution.
Non-Parametric (No Assumed Distribution)	Mann-Whitney U Test	Compares two independent groups without assuming normality.	Comparing the pain scores of two groups of patients (e.g., drug vs. placebo) when the data is not normally distributed.
	Wilcoxon Signed-Rank Test	Compares two related samples without assuming normality.	Comparing the pain levels of patients before and after a physical therapy session.
	Kruskal-Wallis Test	Compares three or more independent groups without assuming normality.	Comparing the effectiveness of three different painkillers on reducing pain scores.
	Spearman's Rank Correlation	Measures the strength and direction of association between two ranked variables.	Examining the relationship between the severity of a disease and a patient's quality of life score.

Key Notes:

- Normal Distribution Tests: Used for continuous data that is normally distributed (e.g., blood pressure, cholesterol levels).
- Binomial Distribution Tests: Used for categorical data with two outcomes (e.g., recovery vs. no recovery, smoker vs. non-smoker).
- Poisson Distribution Tests: Used for count data (e.g., number of hospital admissions, heart attacks).
- Exponential Distribution Tests: Used for time-based data (e.g., waiting times, survival times).
- Non-Parametric Tests: Used when data does not meet the assumptions of normality or other parametric tests (e.g., pain scores, ranked data).

Summary

Understanding Probability Distributions



Poisson Distribution

Counts events in fixed interval

Probability

Foundation of understanding uncertainty

Normal Distribution

Bell curve, symmetric data

Exponential Distribution

Time between events

Probability Distributions

Describe spread of probabilities

Binomial Distribution

Two outcomes, fixed trials

Comprehensive Questions

1. What are the key characteristics of the binomial distribution? Give an example of a real-world scenario where the binomial distribution would be applicable.
2. Describe the Poisson distribution and its assumptions. How is it different from the binomial distribution?
3. What does the normal distribution represent, and why is it often referred to as the "bell curve"? Provide an example of a dataset that might follow a normal distribution.
4. Explain the concept of the exponential distribution. How is it related to the Poisson distribution?
5. In a clinical trial, the probability of a patient experiencing a side effect from a new drug is 0.2. If 20 patients are selected at random, what is the probability that at least 5 of them will experience side effects? Use the binomial distribution to solve this problem.
6. A hospital emergency room receives an average of 10 patients per hour. What is the probability that more than 12 patients will arrive in the next hour? Use the Poisson distribution to solve this problem.
7. Why is the normal distribution considered the most important distribution in statistics? Discuss its properties and applications.
8. How are the Poisson and exponential distributions related? Provide an example of a real-world scenario where both distributions might be used together.
9. What are the limitations of the binomial distribution? Under what conditions might the Poisson distribution be a better choice?
10. Discuss the concept of the "68-95-99.7 rule" for the normal distribution. How can this rule be applied in practical data analysis?
11. In what situations would you use a non-parametric test instead of a parametric test based on a specific probability distribution? Provide examples.


THANK YOU!