IT302 PROBABILITY AND STATISTICS
Lec-1

Dr. Anand Kumar M
Department of Information Technology
National Institute of Technology-Karnataka (NITK)
Surathkal, Mangalore.

Outline

- Statistics - Data Collection
- Population Sample
- Descriptive and Inferential Statistics
- Types of Variables
- Describing Dataset - Descriptive
  - Freq Table – Relative Freq Table – Data Grouping – Ogives – Stem and Leaf
- Summarizing Dataset
  - Mean – Median-Mode
  - Variance and SD
  - Percentiles and Box

Statistics - Data Collection

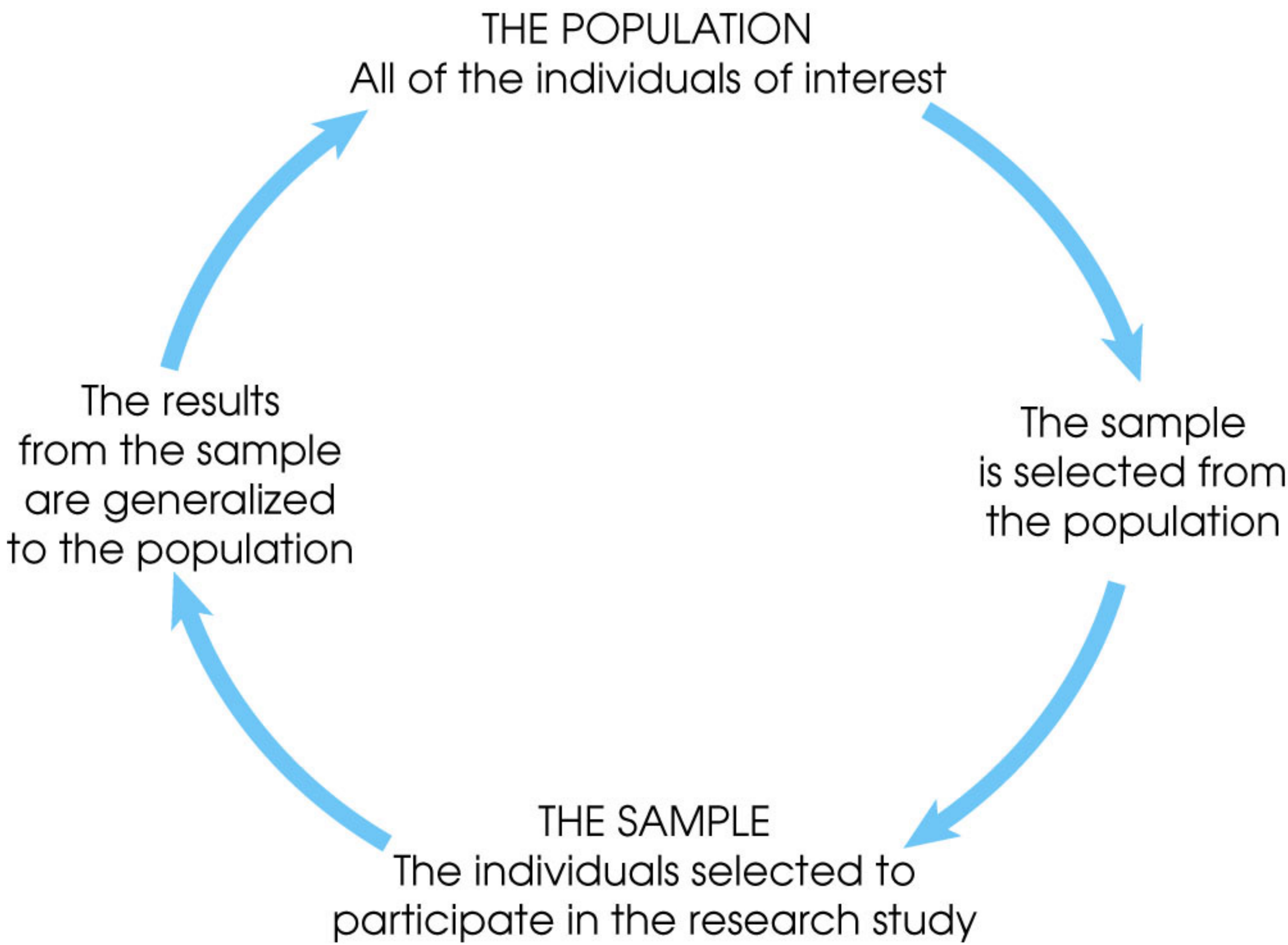
- Data is important to learn about something.
- “Statistics is the art of *Learning from Data* “
- *Describe-Analyze-Summarize* the data.
- Data is Not Available
  - Stat Theory used to design appropriate Experiment to generate data
  - *Example* – Class – Online/Offline
  - Not to be biased (Random)

Population

- The entire group of elements is called the *population*.
- For example, a researcher may be interested in the relation between class size (variable 1) and academic performance (variable 2) for the population of third-grade children.

Sample

- Usually populations are so large that a researcher **cannot examine the entire group**.
- Therefore, a **sample** is selected to represent the population in a research study.
- The goal is to use the *results obtained from the sample to help answer questions about the population*.


THE POPULATION
All of the individuals of interest
The sample is selected from the population
THE SAMPLE
The individuals selected to participate in the research study
The results from the sample are generalized to the population

Descriptive and Inferential Statistics

- End of Experiment data need to be described and summarized, i.e DS.
- Method concerned about drawing conclusions are IS.
- Ex: With 10 toss – 7 FAIR OR NOT
- With 100 toss- 92 FAIR OR NOT
- Assumptions required to draw conclusion – the assumptions are called Probability Model

Descriptive statistics

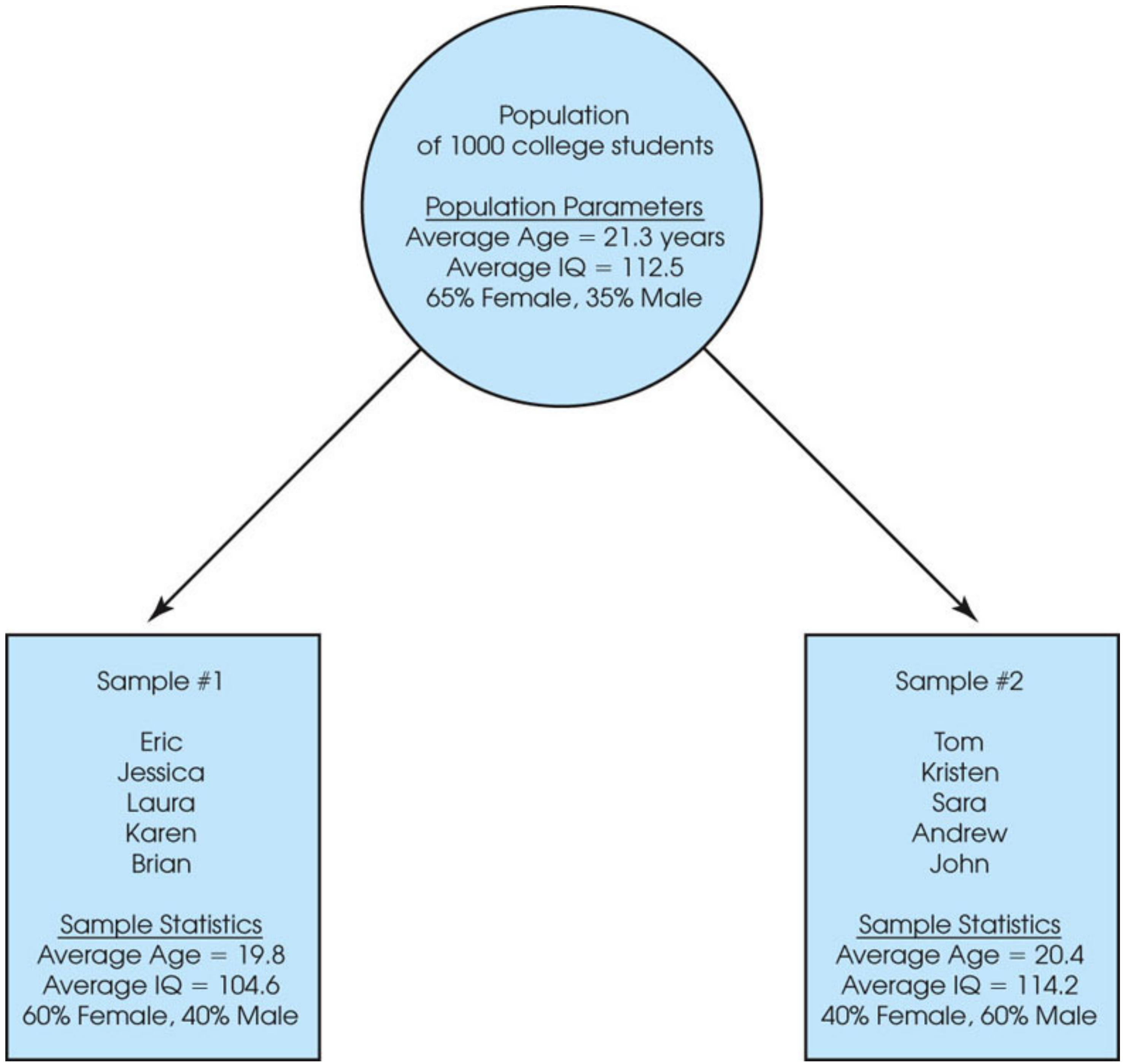
- **Descriptive statistics** are methods for organizing and summarizing data.
- For example, tables or graphs are used to organize data, and descriptive values such as the average score are used to summarize data.
- A descriptive value for a population is called a **parameter** and a descriptive value for a sample is called a **statistic**

Inferential Statistics

- **Inferential statistics** are methods for using sample data to make general conclusions (inferences) about populations.
- Because a sample is typically only a part of the whole population, sample data provide only limited information about the population. As a result, sample statistics are generally imperfect representatives of the corresponding population parameters.

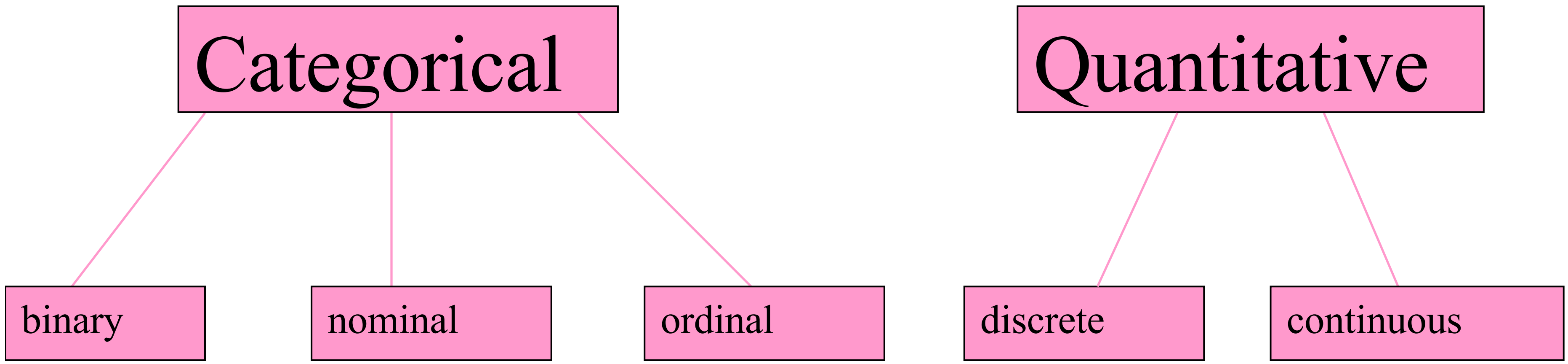
Sampling Error

- The discrepancy between a sample statistic and its population parameter is called **sampling error**.
- Defining and measuring sampling error is a large part of inferential statistics.



Population
of 1000 college students
Population Parameters
Average Age = 21.3 years
Average IQ = 112.5
65% Female, 35% Male
Sample #1
Eric
Jessica
Laura
Karen
Brian
Sample Statistics
Average Age = 19.8
Average IQ = 104.6
60% Female, 40% Male
Sample #2
Tom
Kristen
Sara
Andrew
John
Sample Statistics
Average Age = 20.4
Average IQ = 114.2
40% Female, 60% Male

Types of Variables: Overview



2 categories +

more categories +

order matters +

numerical +

uninterrupted

Categorical Variables

- Also known as “qualitative.”

- Categories.

    - treatment groups
    - exposure groups
    - disease status

Categorical Variables

- Dichotomous (binary) – two levels
  - Dead/alive
  - Treatment/placebo
  - Disease/no disease
  - Exposed/Unexposed
  - Heads/Tails
  - Pulmonary Embolism (yes/no)
  - Male/female

Categorical Variables

- Nominal variables – Named categories
  Order doesn’t matter!
  - The blood type of a patient (O, A, B, AB)
  - Marital status
  - Occupation

Categorical Variables

- Ordinal variable – Ordered categories. Order matters!
    - Staging in breast cancer as I, II, III, or IV
    - Birth order—1st, 2nd, 3rd, etc.
    - Letter grades (A, B, C, D, F)
    - Ratings on a scale from 1-5
    - Ratings on: always; usually; many times; once in a while; almost never; never
    - Age in categories (10-20, 20-30, etc.)
    - Shock index categories (Kline et al.)

Quantitative Variables

- Numerical variables; may be arithmetically manipulated.
  - Counts
  - Time
  - Age
  - Height

Quantitative Variables

- <u>Discrete Numbers</u> – a limited set of distinct values, such as whole numbers.
  - Number of new AIDS cases in CA in a year (counts)
  - Years of school completed
  - The number of children in the family (cannot have a half a child!)
  - The number of deaths in a defined time period (cannot have a partial death!)
  - Roll of a die

Quantitative Variables

- <u>Continuous Variables</u> - Can take on any number within a defined range.
    - Time-to-event (survival time)
    - Age
    - Blood pressure
    - Serum insulin
    - Speed of a car
    - Income
    - Shock index (Kline et al.)

Looking at Data

- ✓ How are the data distributed?
  - Where is the center?
  - What is the range?
  - What’s the shape of the distribution (e.g., Gaussian, binomial, exponential, skewed)?
- ✓ Are there “outliers”?
- ✓ Are there data points that don’t make sense?

The first rule of statistics:
USE COMMON SENSE!

90% of the information is contained in the graph.

Frequency Plots (univariate)

<u>Categorical variables</u>

– Bar Chart

<u>Continuous variables</u>

– Box Plot

– Histogram

Describing Dataset – Descriptive

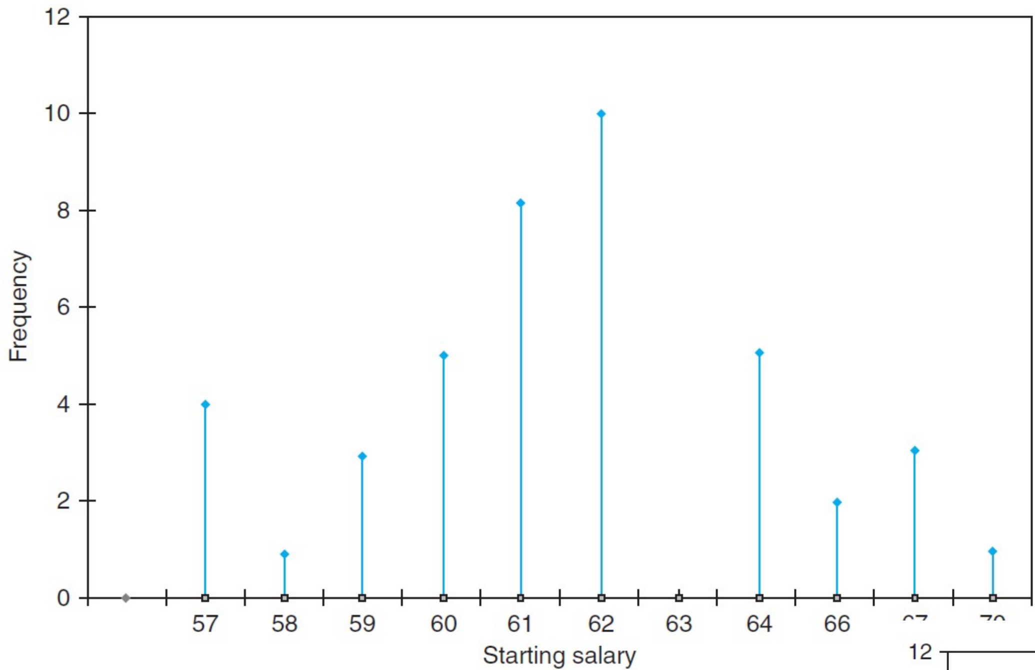
– Freq Table
– Relative Freq Table
– Data Grouping
– Ogives
– Stem and Leaf

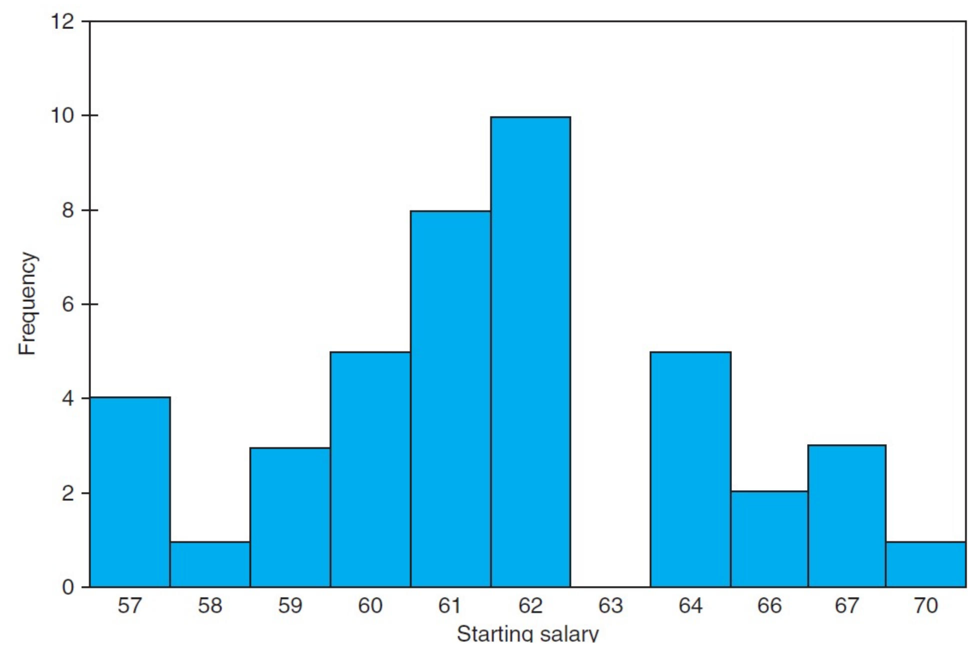
Freq Table AND Graphs

TABLE 2.1 *Starting Yearly Salaries*

Starting Salary	Frequency
57	4
58	1
59	3
60	5
61	8
62	10
63	0
64	5
66	2
67	3
70	1

Relative Frequency Graphs





Polygon

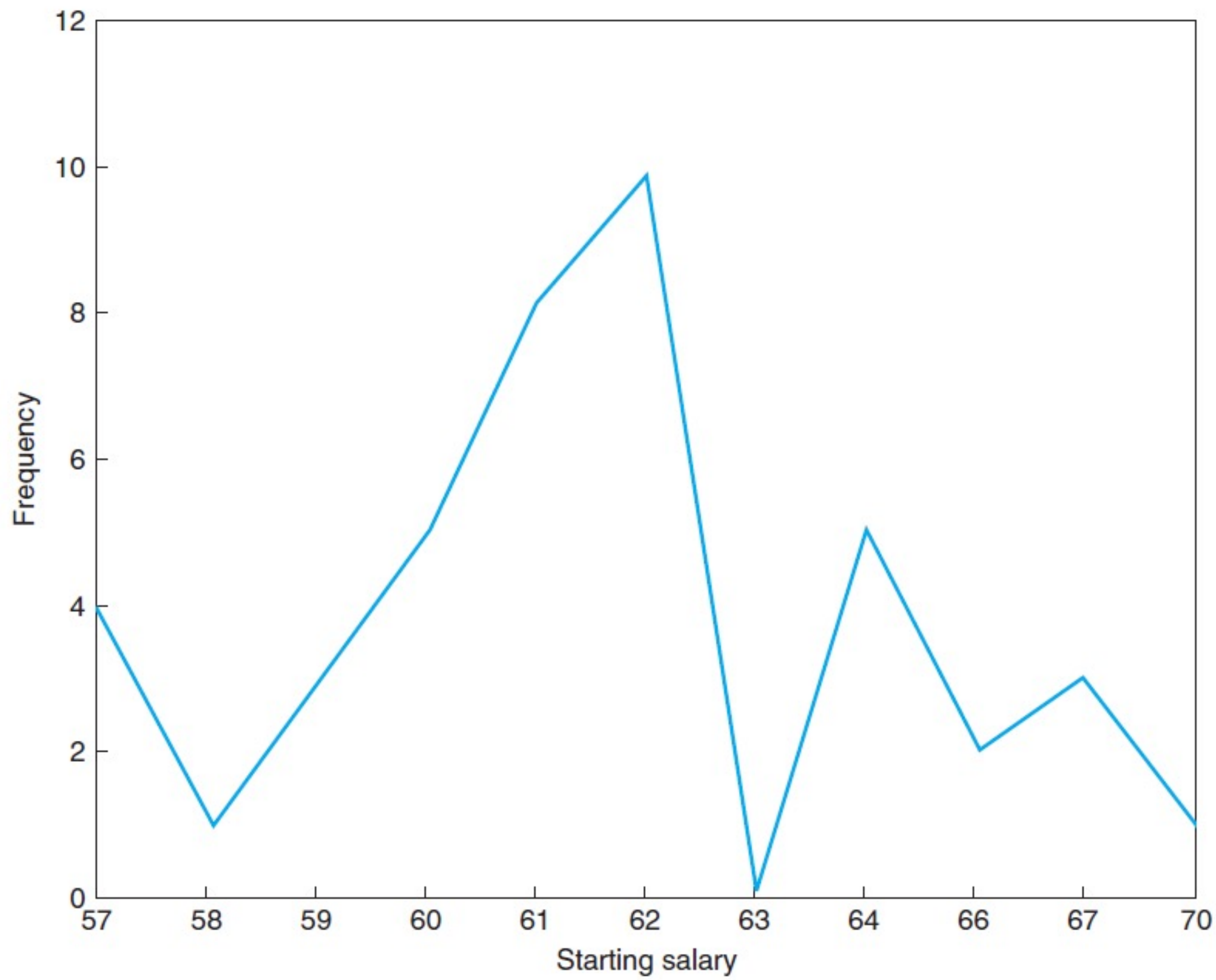
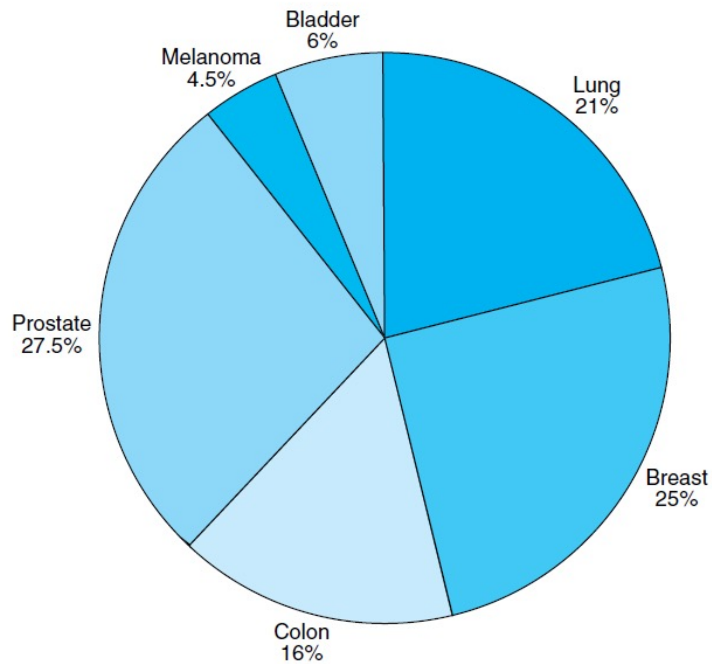


FIGURE 2.3 *Frequency polygon for starting salary data.*

TABLE 2.2

Starting Salary	Frequency
47	4/42 = .0952
48	1/42 = .0238
49	3/42
50	5/42
51	8/42
52	10/42
53	0
54	5/42
56	2/42
57	3/42
60	1/42



Type of Cancer	Number of New Cases	Relative Frequency
Lung	42	.21
Breast	50	.25
Colon	32	.16
Prostate	55	.275
Melanoma	9	.045
Bladder	12	.06

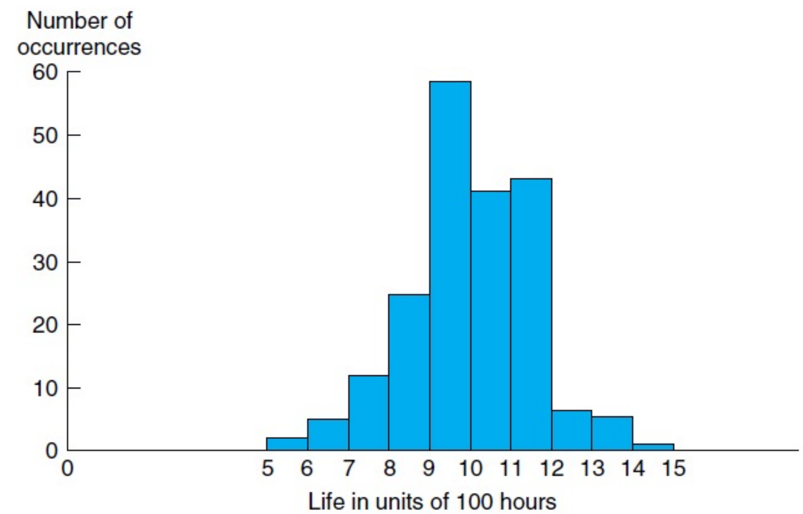
Data Grouping -Histograms

TABLE 2.3 *Life in Hours of 200 Incandescent Lamps*

Item Lifetimes									
1,067	919	1,196	785	1,126	936	918	1,156	920	948
855	1,092	1,162	1,170	929	950	905	972	1,035	1,045
1,157	1,195	1,195	1,340	1,122	938	970	1,237	956	1,102
1,022	978	832	1,009	1,157	1,151	1,009	765	958	902
923	1,333	811	1,217	1,085	896	958	1,311	1,037	702
521	933	928	1,153	946	858	1,071	1,069	830	1,063
930	807	954	1,063	1,002	909	1,077	1,021	1,062	1,157
999	932	1,035	944	1,049	940	1,122	1,115	833	1,320
901	1,324	818	1,250	1,203	1,078	890	1,303	1,011	1,102
996	780	900	1,106	704	621	854	1,178	1,138	951
1,187	1,067	1,118	1,037	958	760	1,101	949	992	966
824	653	980	935	878	934	910	1,058	730	980
844	814	1,103	1,000	788	1,143	935	1,069	1,170	1,067
1,037	1,151	863	990	1,035	1,112	931	970	932	904
1,026	1,147	883	867	990	1,258	1,192	922	1,150	1,091
1,039	1,083	1,040	1,289	699	1,083	880	1,029	658	912
1,023	984	856	924	801	1,122	1,292	1,116	880	1,173
1,134	932	938	1,078	1,180	1,106	1,184	954	824	529
998	996	1,133	765	775	1,105	1,081	1,171	705	1,425
610	916	1,001	895	709	860	1,110	1,149	972	1,002

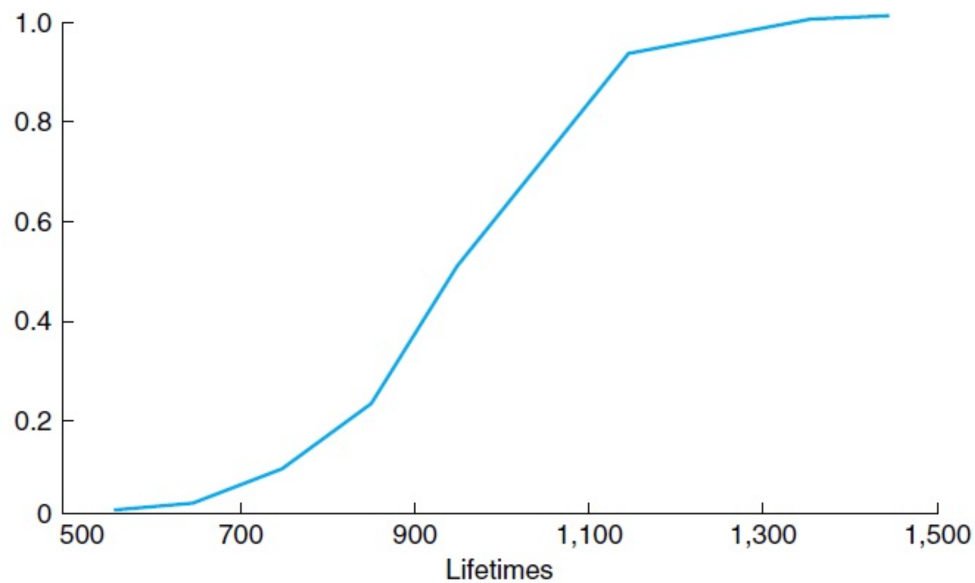
TABLE 2.4 *A Class Frequency Table*

Class Interval	Frequency (Number of Data Values in the Interval)
500–600	2
600–700	5
700–800	12
800–900	25
900–1000	58
1000–1100	41
1100–1200	43
1200–1300	7
1300–1400	6
1400–1500	1



- The number of class intervals chosen should be a trade-off
- between (1) choosing too few classes at a cost of losing too much information about the
- actual data values in a class and (2) choosing too many classes, which will result in the
- frequencies of each class being too small for a pattern to be discernible

Ogives



Stem and Leaf

7	0.0
6	9.0
5	1.0, 1.3, 2.0, 5.5, 7.1, 7.4, 7.6, 8.5, 9.3
4	0.0, 1.0, 2.4, 3.6, 3.7, 4.8, 5.0, 5.2, 6.0, 6.7, 8.1, 9.0, 9.2
3	3.1, 4.1, 5.3, 5.8, 6.2, 9.0, 9.5, 9.5
2	9.0, 9.8

Summarizing the datasets

Sample Mean, Sample Median, and Sample Mode

- Statistics that are used for describing the center of a set of data values.
- Suppose that we have a data set consisting of the *n numerical values x1, x2, . . . , xn.*
- *The sample mean is the arithmetic average of these values.*

$$\bar{x} = \sum_{i=1}^{n} x_i / n$$

Mean: example

<u>Some data</u>:

Age of participants: 17 19 21 22 23 23 23 38

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{17 + 19 + 21 + 22 + 23 + 23 + 23 + 38}{8} = 23.25$$

Mean

- $y_i = ax_i + b$

EXAMPLE 2.3a The winning scores in the U.S. Masters golf tournament in the years from 2004 to 2013 were as follows:

$$280,\ 278,\ 272,\ 276,\ 281,\ 279,\ 276,\ 281,\ 289,\ 280$$

Find the sample mean of these scores.

SOLUTION Rather than directly adding these values, it is easier to first subtract 280 from each one to obtain the new values $y_i = x_i - 280$:

$$0, -2, -8, -4, 1, -1, -4, 1, 9, 0$$

Because the arithmetic average of the transformed data set is

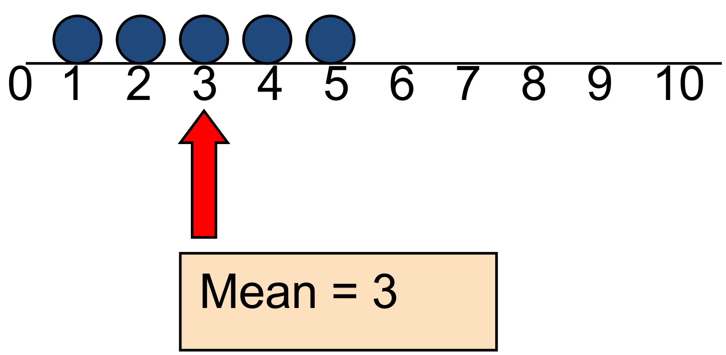
$$\bar{y} = -8/10$$

it follows that

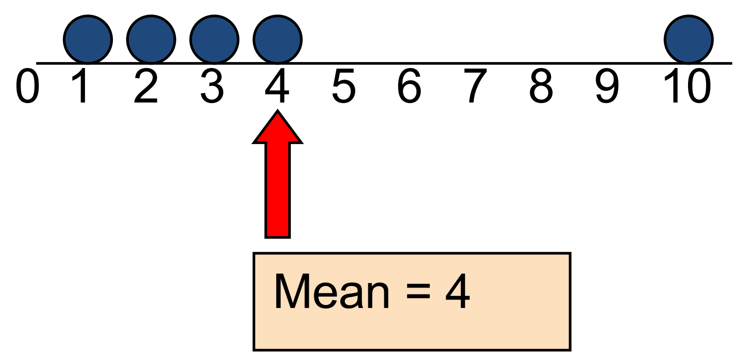
$$\bar{x} = \bar{y} + 280 = 279.2 \quad ■$$

Mean

- The mean is affected by extreme values (outliers)



$$\frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$



$$\frac{1+2+3+4+10}{5} = \frac{20}{5} = 4$$

Mean?

Age	Frequency
15	2
16	5
17	11
18	9
19	14
20	13

Central Tendency

- <u>Median</u> – the exact middle value

<u>Calculation:</u>

- If there are an odd number of observations, find the middle value
- If there are an even number of observations, find the middle two values and average them.

Median: example

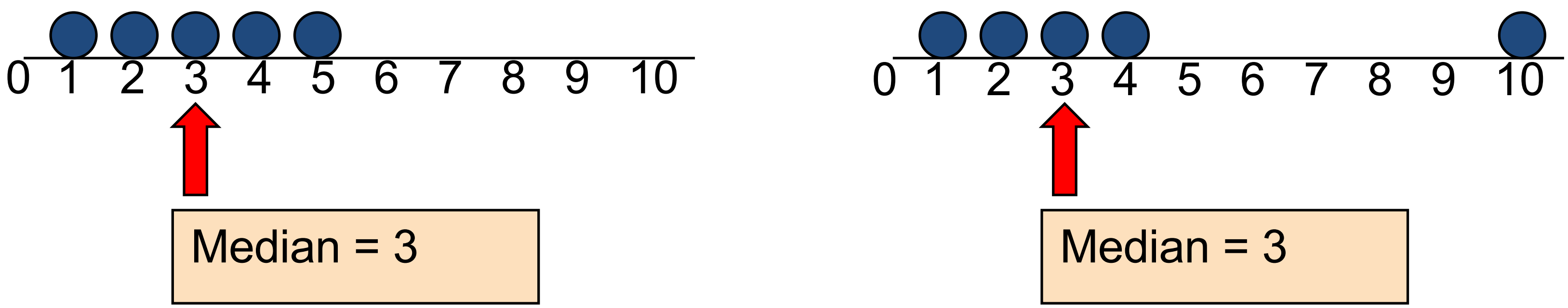
Some data:

Age of participants: 17 19 21 22 23 23 23 38

Median = (22+23)/2 = 22.5

Median

- The median is not affected by extreme values (outliers).



Central Tendency

- Mode – the value that occurs most frequently

Mode: example

Some data:
Age of participants: 17 19 21 22 23 23 23 38

Mode = 23 (occurs 3 times)

Value	Frequency
1	9
2	8
3	5
4	5
5	6
6	7

Practice

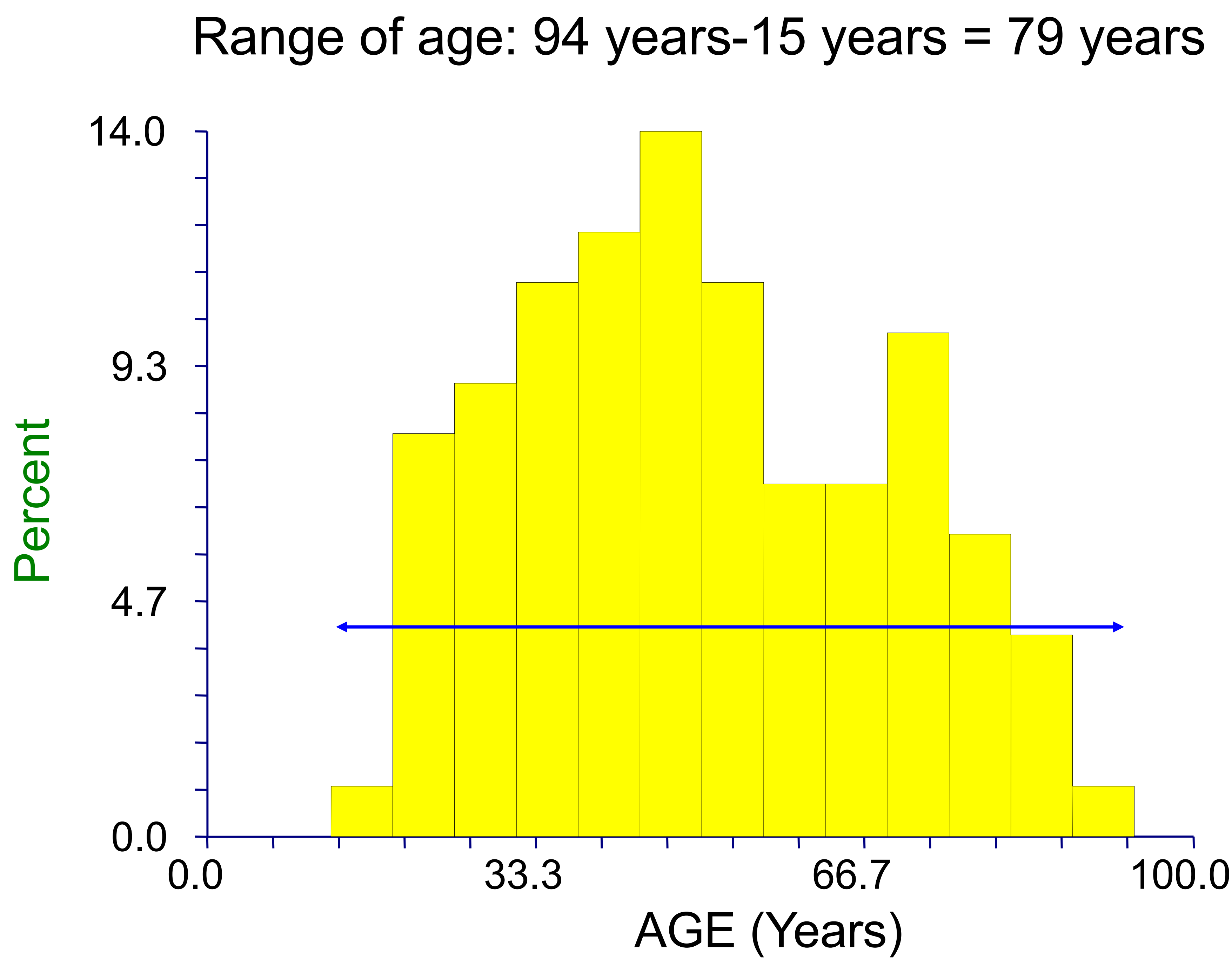
- Find the median of a series of all the even terms from 4 to 296.

Measures of Variation/Dispersion

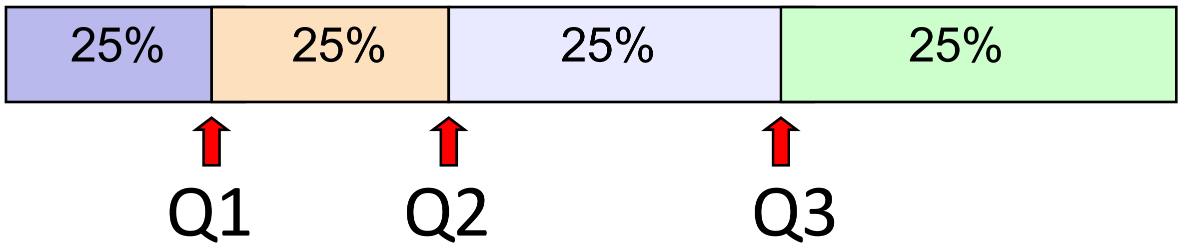
- Range
- Percentiles/quartiles
- Interquartile range
- Standard deviation/Variance

Range

- Difference between the largest and the smallest observations.


Range of age: 94 years-15 years = 79 years
14.0
9.3
4.7
0.0
Percent
0.0
33.3
66.7
100.0
AGE (Years)

Quartiles

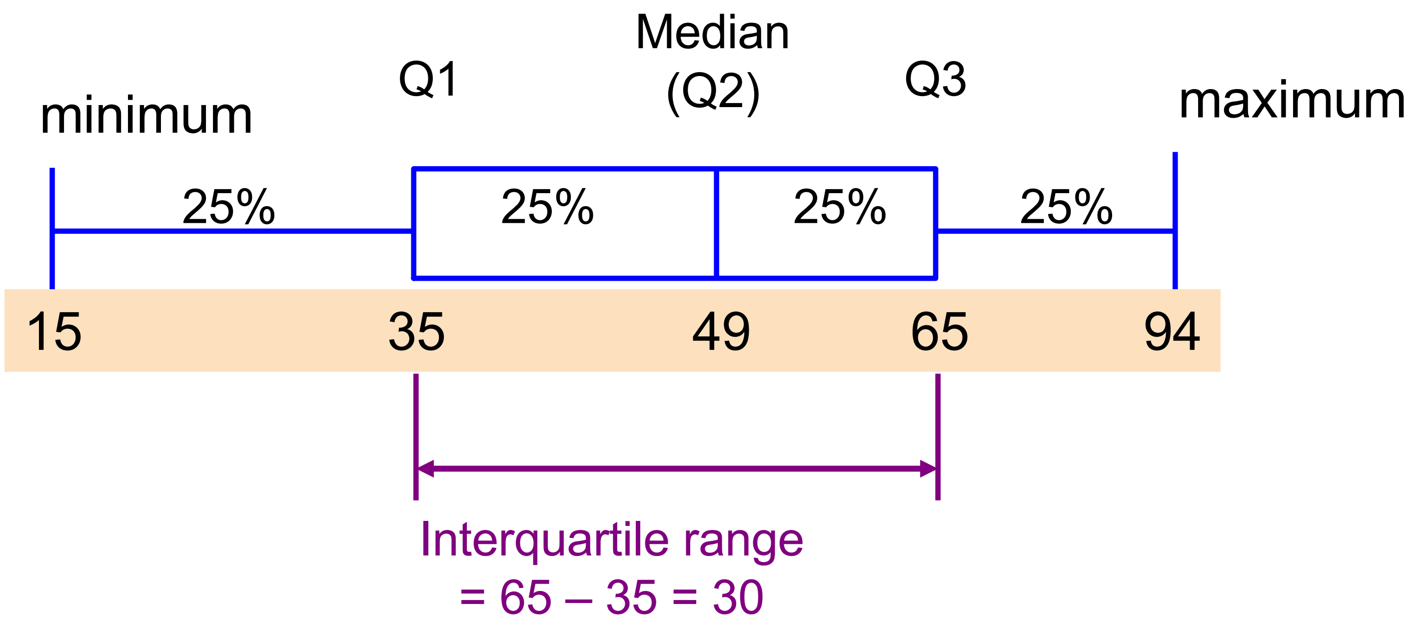


- The first quartile, Q_1, is the value for which 25% of the observations are smaller and 75% are larger
- Q_2 is the same as the median (50% are smaller, 50% are larger)
- Only 25% of the observations are greater than the third quartile

Interquartile Range

- Interquartile range = 3rd quartile – 1st quartile = $Q_3 - Q_1$

Interquartile Range: age



Variance

- Average (roughly) of squared deviations of values from the mean
- Describe the spread or variability of the data values

$$S^2 = \frac{\sum_{i}^{n}(x_i - \bar{X})^2}{n-1}$$

Why squared deviations?

- Adding deviations will yield a sum of 0.
- Absolute values are tricky!
- Squares eliminate the negatives.

- Result:
  - Increasing contribution to the variance as you go farther from the mean.

Practice

Find the sample variances of the data sets **A** and **B** given below.

$\mathbf{A}: 3, 4, 6, 7, 10 \qquad \mathbf{B}: -20, 5, 15, 24$

Standard Deviation

- Most commonly used measure of variation
- Shows variation about the mean
- Has the same units as the original data

$$S = \sqrt{\frac{\sum_{i}^{n} (x_i - \overline{X})^2}{n-1}}$$

Calculation Example: Sample Standard Deviation

Age data (n=8) : 17 19 21 22 23 23 23 38

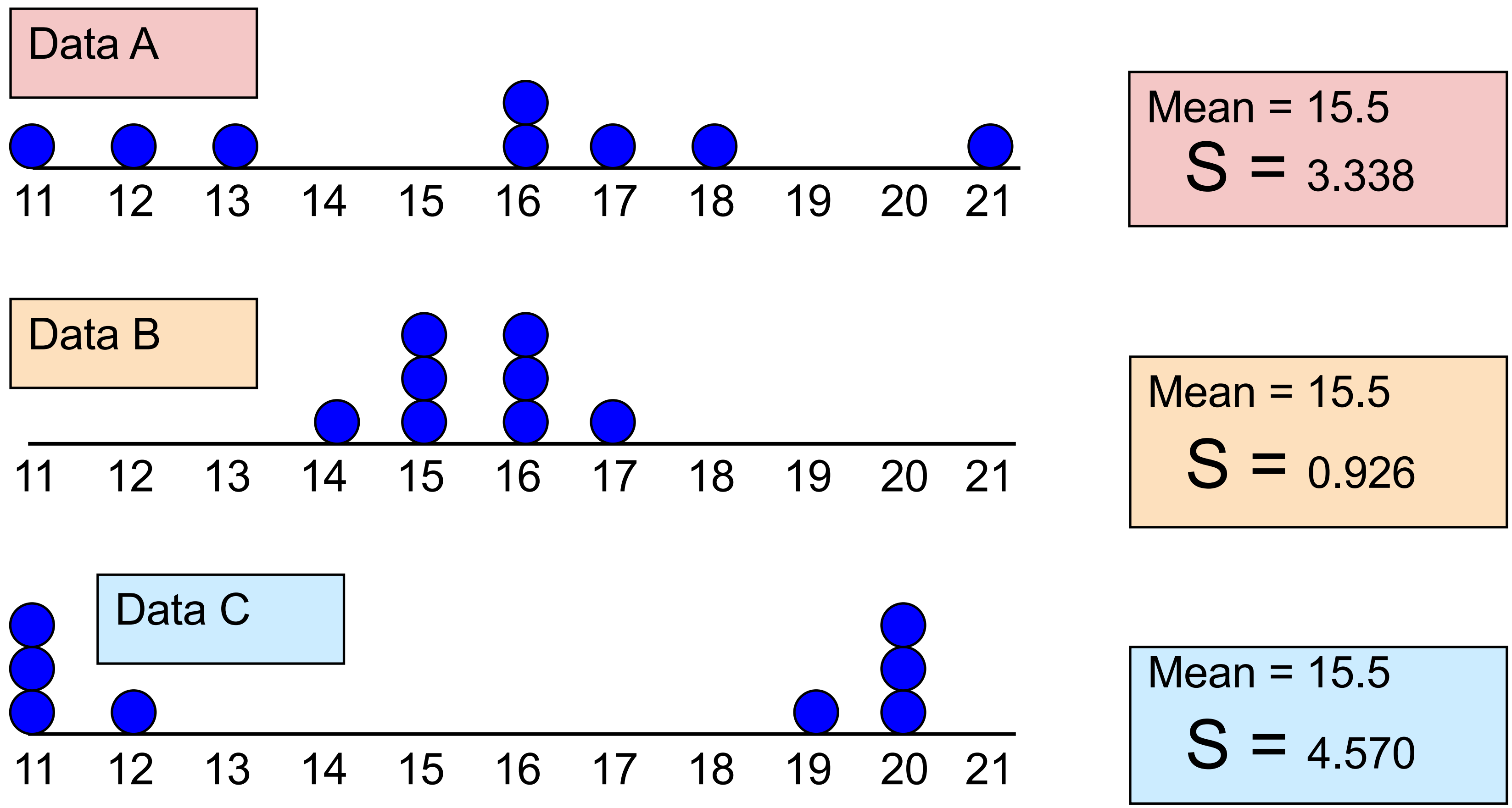
n = 8 Mean = $\overline{X}$ = 23.25

$$S = \sqrt{\frac{(17-23.25)^2 + (19-23.25)^2 + \cdots + (38-23.25)^2}{8-1}}$$

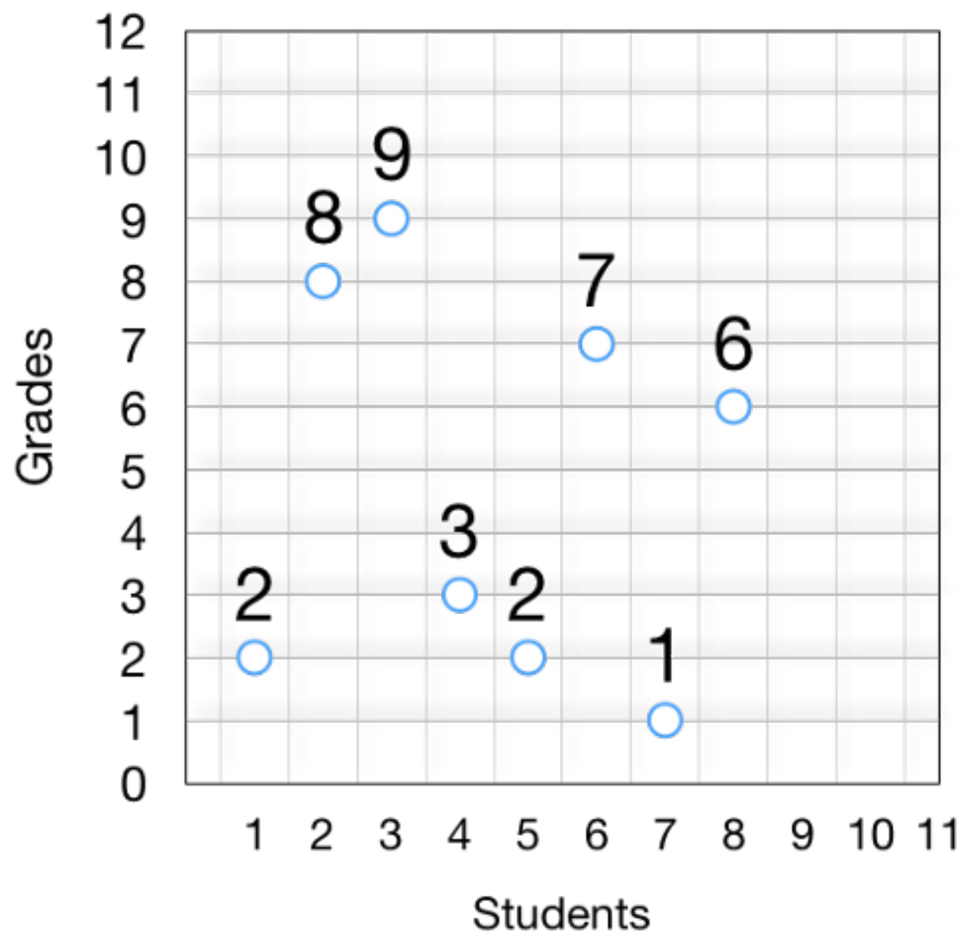
$$= \sqrt{\frac{280}{7}} = 6.3$$

standard deviation gets bigger when numbers are more spread out.

Comparing Standard Deviations

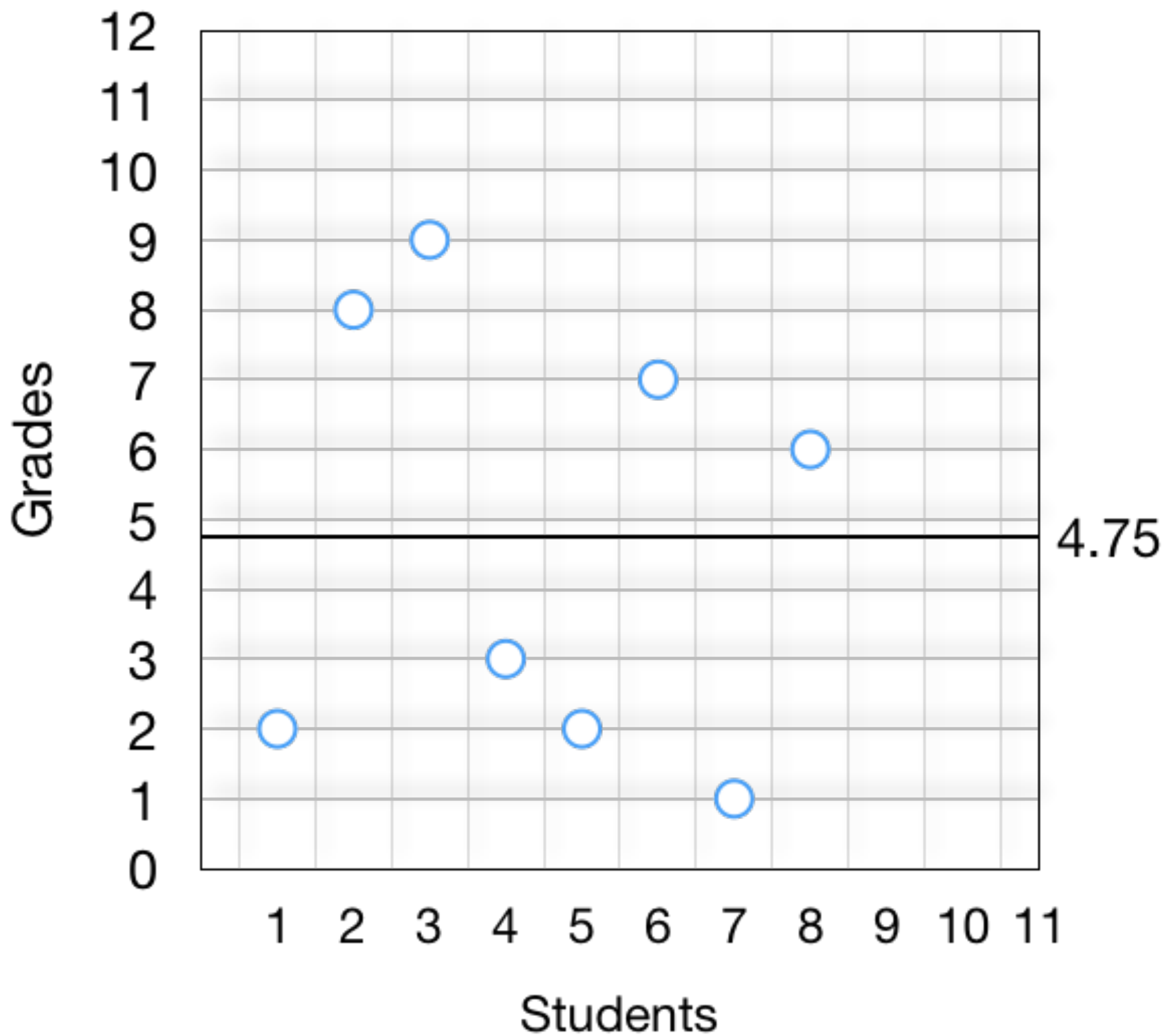


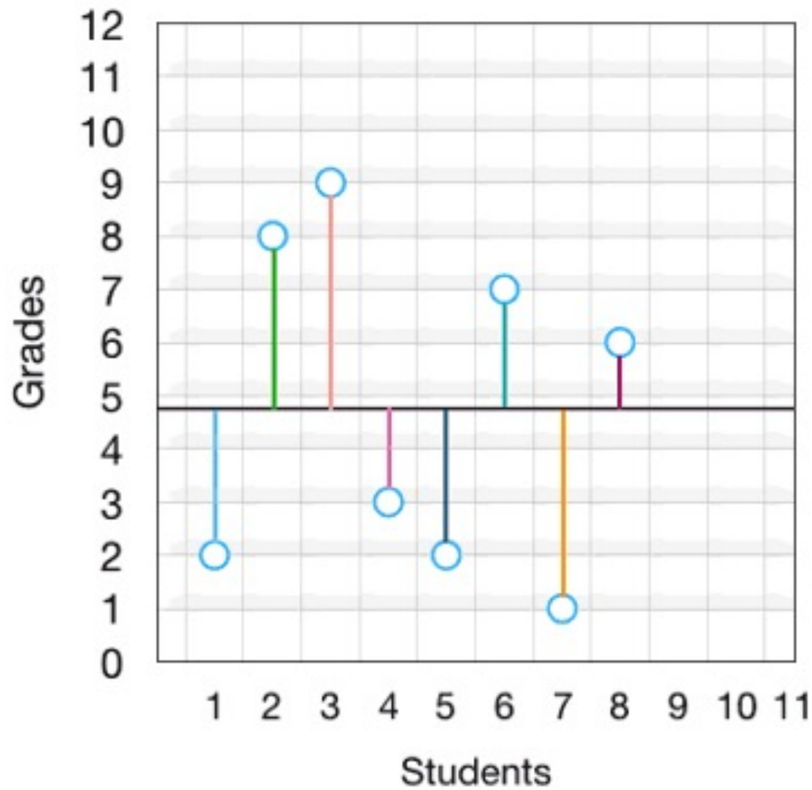
Visual Meaning



Student	Grade
1	2
2	8
3	9
4	3
5	2
6	7
7	1
8	6

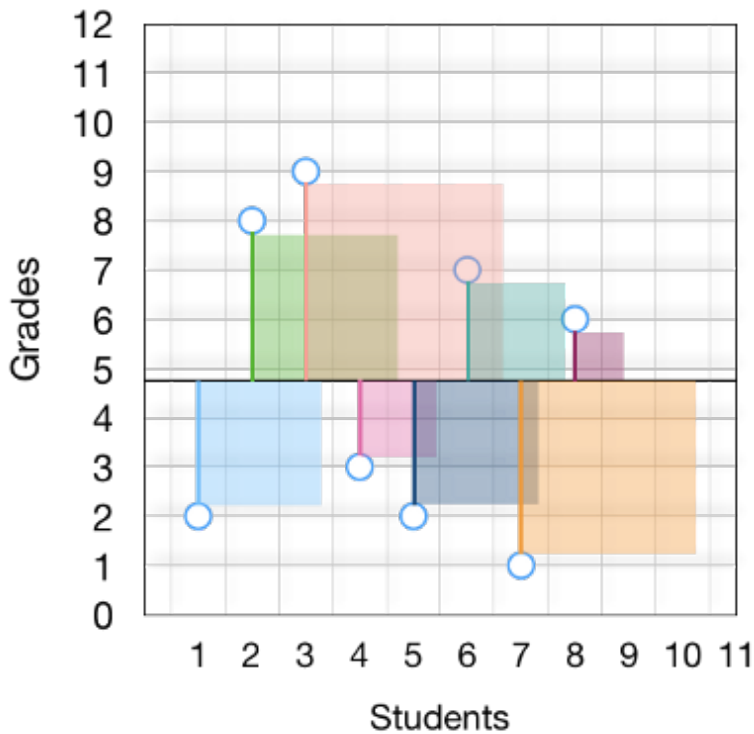
https://towardsdatascience.com/a-visual-interpretation-of-the-standard-deviation-30f4676c291c


Grades
12
11
10
9
8
7
6
5
4
3
2
1
0
4.75
1
2
3
4
5
6
7
8
9
10
11
Students



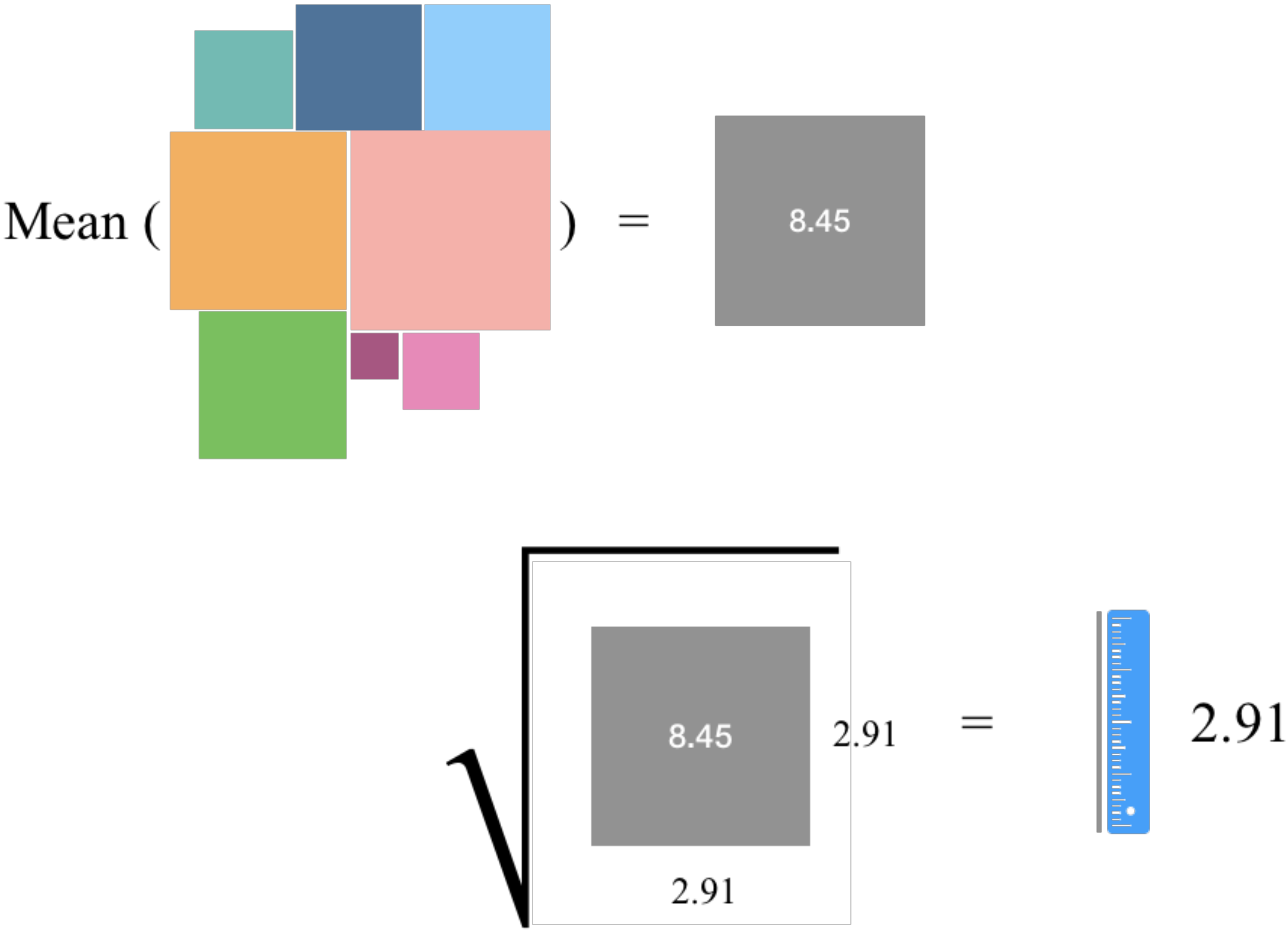
$$x - \bar{x} =$$

$$(2-4.75) + (8-4.75) + (9-4.75) + (3-4.75) + (2-4.75) + (7-4.75) + (1-4.75) + (6-4.75)$$



$$\sum (x_n - \bar{x})^2 =$$

7.5625 + **10.5625**
+ **18.0625** + **3.0625**
+ **7.5625** + **5.0625**
+ **14.0625** + **1.5625**
= **67.5**


Mean (
)
=
8.45
8.45
2.91
2.91
=
2.91

Bienaymé-Chebyshev Rule

- Regardless of how the data are distributed, a certain percentage of values must fall within K standard deviations from the mean:

Note use of μ (mu) to represent "mean".

Note use of σ (sigma) to represent "standard deviation."

At least	within
$(1 - 1/1^2) =$ 0% …………	k=1 $(\mu \pm 1\sigma)$
$(1 - 1/2^2) =$ 75% ………..	k=2 $(\mu \pm 2\sigma)$
$(1 - 1/3^2) =$ 89% …………	k=3 $(\mu \pm 3\sigma)$

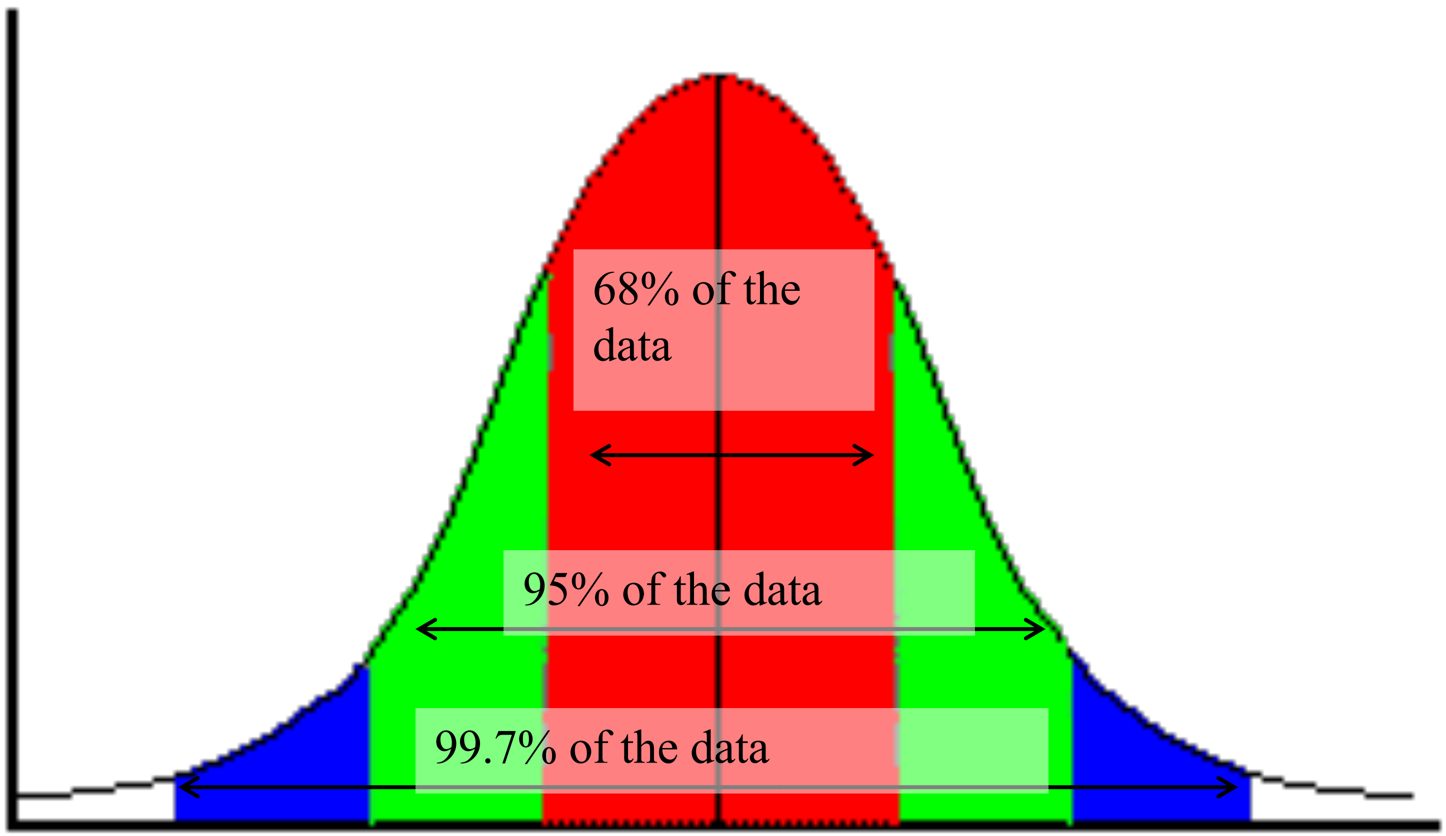
Symbol Clarification

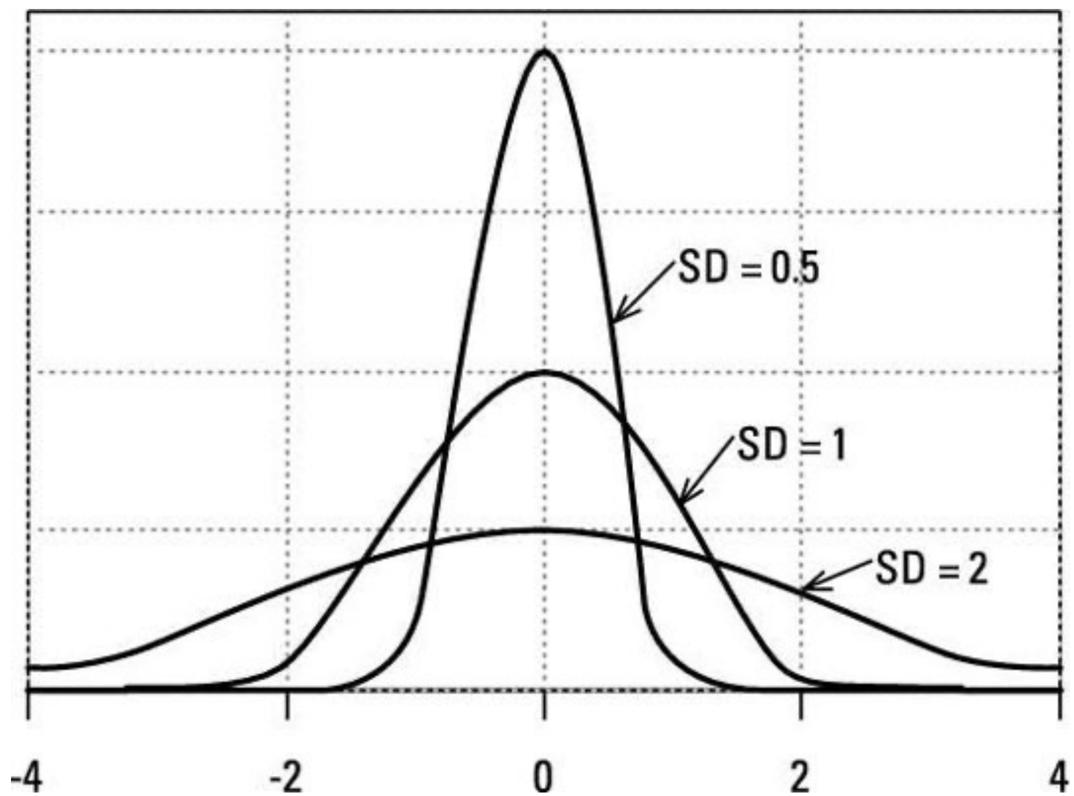
- S = Sample standard deviation (example of a "sample statistic")
- σ = Standard deviation of the entire population (example of a "population parameter") or from a theoretical probability distribution
- X = Sample mean
- μ = Population or theoretical mean

**The beauty of the normal curve:

No matter what μ and σ are, the area between μ-σ and μ+σ is about 68%; the area between μ-2σ and μ+2σ is about 95%; and the area between μ-3σ and μ+3σ is about 99.7%. Almost all values fall within 3 standard deviations.

68-95-99.7 Rule




SD = 0.5
SD = 1
SD = 2
-4
-2
0
2
4

LEC_2
Probability Theory

Dr. Anand Kumar M,
Assistant Professor
Department of Information Technology
National Institute of Technology- Karnataka (NITK)
m_anandkumar@nitk.edu

Outline

- Mathematical models-Deterministic and non Deterministic
- Sets
- Experiment-Sample Space –events
- Finite Sample space
- Equally Likely events
- Conditional Probability
- Bayes Theorem

Mathematical models

Deterministic Phenomena

- There exists a mathematical model that allows "***perfect***" prediction the phenomena's outcome. (Experiment is performed to predict th outcome)
- Many examples exist in Physics, Chemistry (the exact sciences).

Non-deterministic Phenomena

- **No** mathematical model exists that allows "***perfect***" prediction the phenomena's outcome.

Non-deterministic -probabilistic models

Random phenomena

- Unable to predict the outcomes, but in the long-run, the outcomes exhibit statistical regularity.

- Eg: We cannot predict the weather conditions accurately using deterministic – But the probabilistic model describes accurately

- Actual outcome is predicted from the conditions under which the experiments are carried out.
- The conditions of the experimentation determine the ***probabilistic behavior of the outcome. (specify the prob distribution)***

Random phenomena

– Unable to predict the outcomes, but in the long-run, the outcomes exhibit statistical regularity.

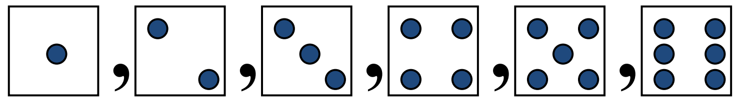
Examples

1. Tossing a coin – outcomes $S = \{\textbf{Head, Tail}\}$

Unable to predict on each toss whether is Head or Tail.

In the long run can predict that 50% of the time heads will occur and 50% of the time tails will occur

2. Rolling a die – outcomes

$S = \{$  $\}$

Unable to predict outcome but in the long run can one can determine that each outcome will occur 1/6 of the time.

Use symmetry. Each side is the same. One side should not occur more frequently than another side in the long run. If the die is not balanced this may not be true.

Sets -review

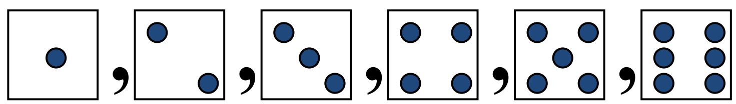
The sample Space, *S*

The **sample space**, *S*, for a random phenomena is the set of all possible outcomes.

Examples

1. Tossing a coin – outcomes $S = \{\textbf{Head, Tail}\}$

2. Rolling a die – outcomes

$S = \{$  $\}$

$$= \{1, 2, 3, 4, 5, 6\}$$

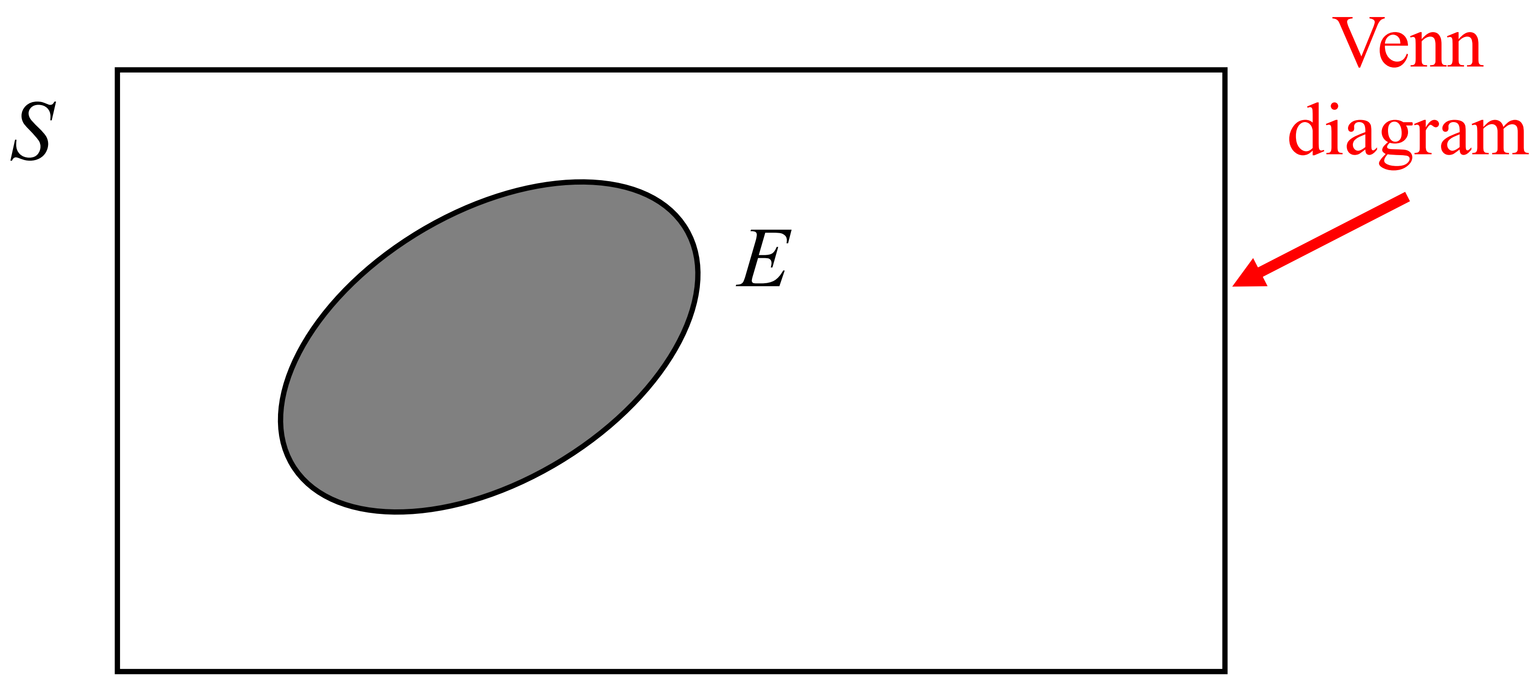
Events -E

- Events are sets:
- Can describe in words
- Can describe in notation
- Can describe with Venn diagrams

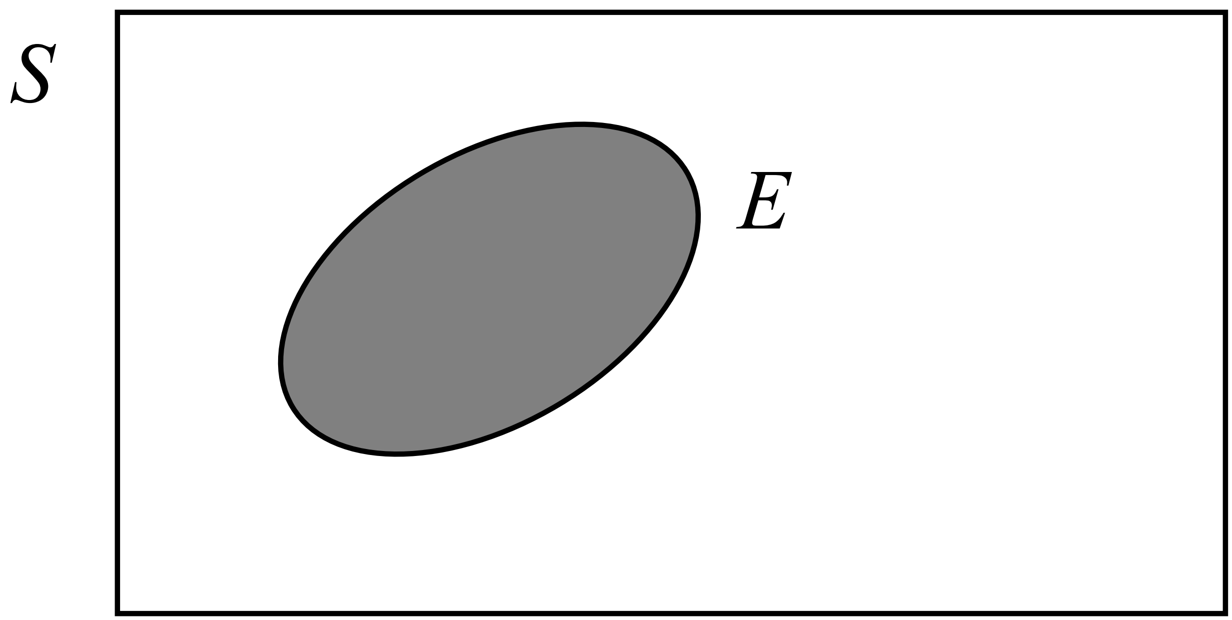
- Experiment: toss a coin 3 times.
- Event: You get 2 or more heads = { HHH, HHT, HTH, THH}

An Event , *E*

The **event**, *E*, is any subset of the **sample space**, *S*. i.e. any set of outcomes (not necessarily all outcomes) of the random phenomena



The **event**, *E*, is said to **have occurred** if after the outcome has been observed the outcome lies in *E*.



Examples

1. Rolling a die – outcomes

$S =$ {⚀, ⚁, ⚂, ⚃, ⚄, ⚅}

$= \{1, 2, 3, 4, 5, 6\}$

$E =$ the event that an even number is rolled

$= \{2, 4, 6\}$

$=$ {⚁, ⚃, ⚅}

Special Events

The Null Event, The empty event - ϕ

$\phi = \{\ \} =$ the event that contains no outcomes

The Entire Event, The Sample Space - S

$S =$ the event that contains all outcomes

The empty event, ϕ, never occurs.

The entire event, S, always occurs.

Experiment-SS-Event

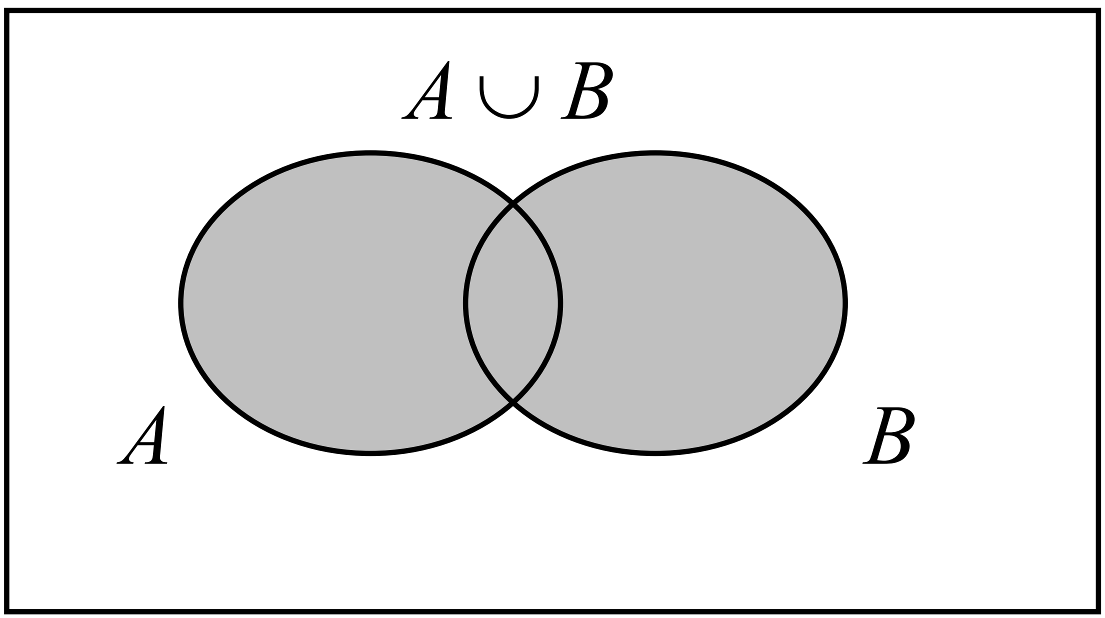
- Experiment: a repeatable procedure
- Sample space: set of all possible outcomes S (or Ω).
- Event: a subset of the sample space.
- Probability function, $P(\omega)$: gives the probability for each outcome $\omega \in S$
- Probability is between 0 and 1
- Total probability of all possible outcomes is 1.

Set operations on Events

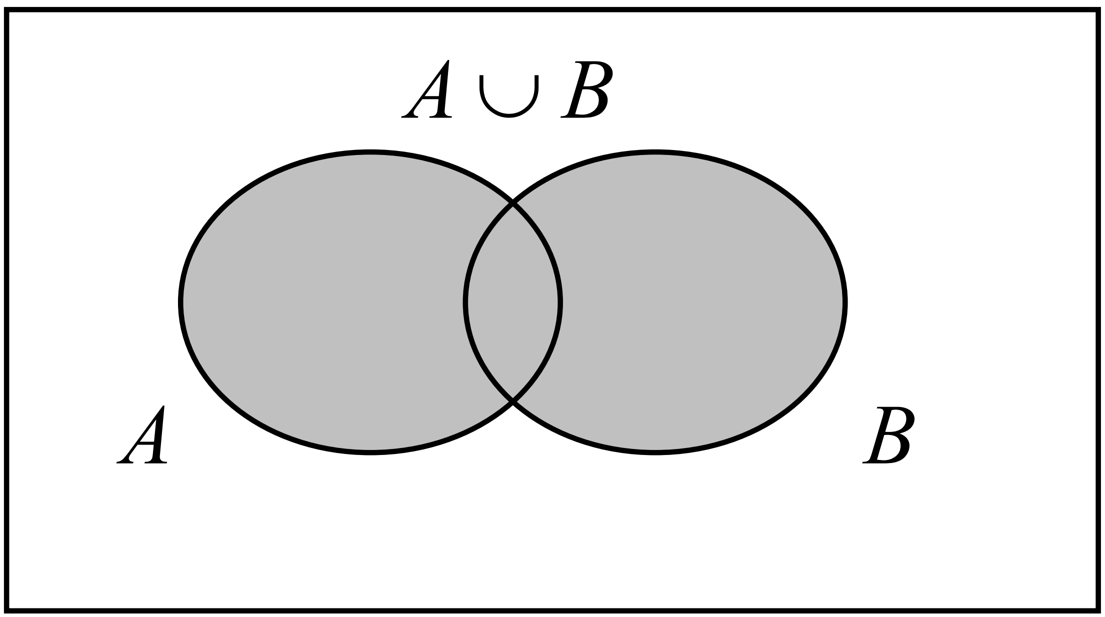
Union

Let A and B be two events, then the **union** of A and B is the event (denoted by $A \cup B$) defined by:

$$A \cup B = \{e \mid e \text{ belongs to } A \textbf{ or } e \text{ belongs to } B\}$$



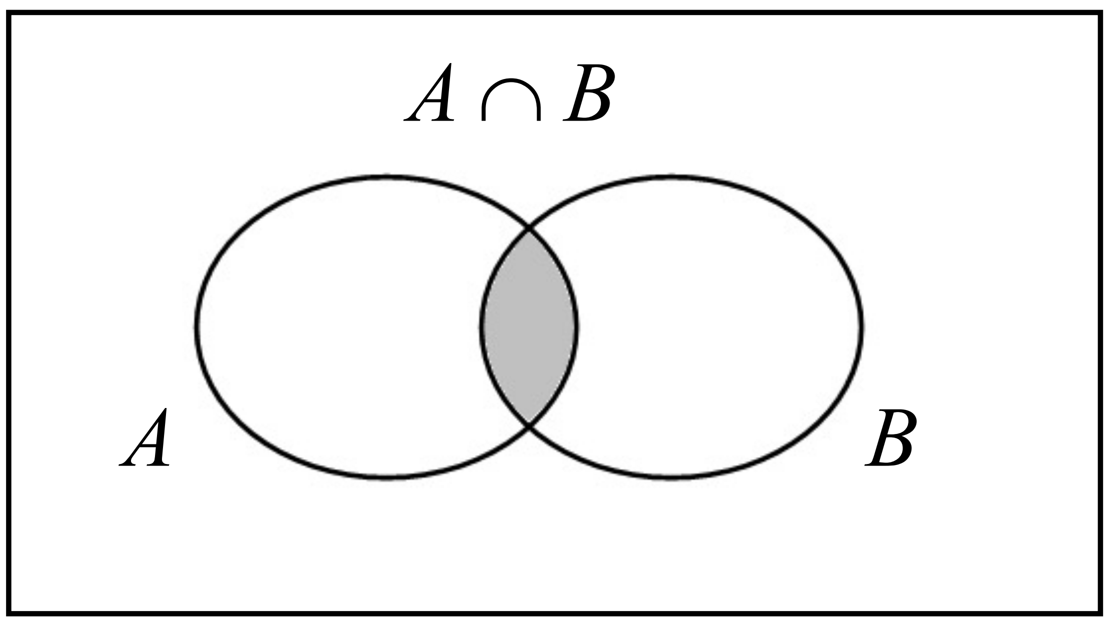
The event $A \cup B$ **occurs** if the event A **occurs or** the event and B **occurs** .



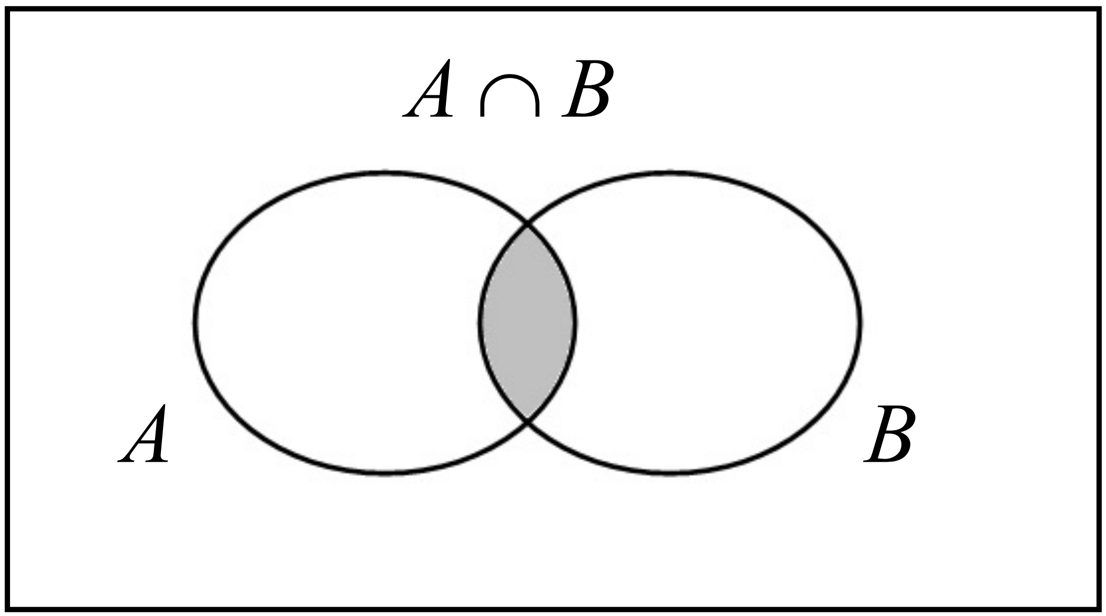
Intersection

Let A and B be two events, then the **intersection** of A and B is the event (denoted by $A \cap B$) defined by:

$$A \cap B = \{e|\ e \text{ belongs to } A \textbf{ and } e \text{ belongs to } B\}$$



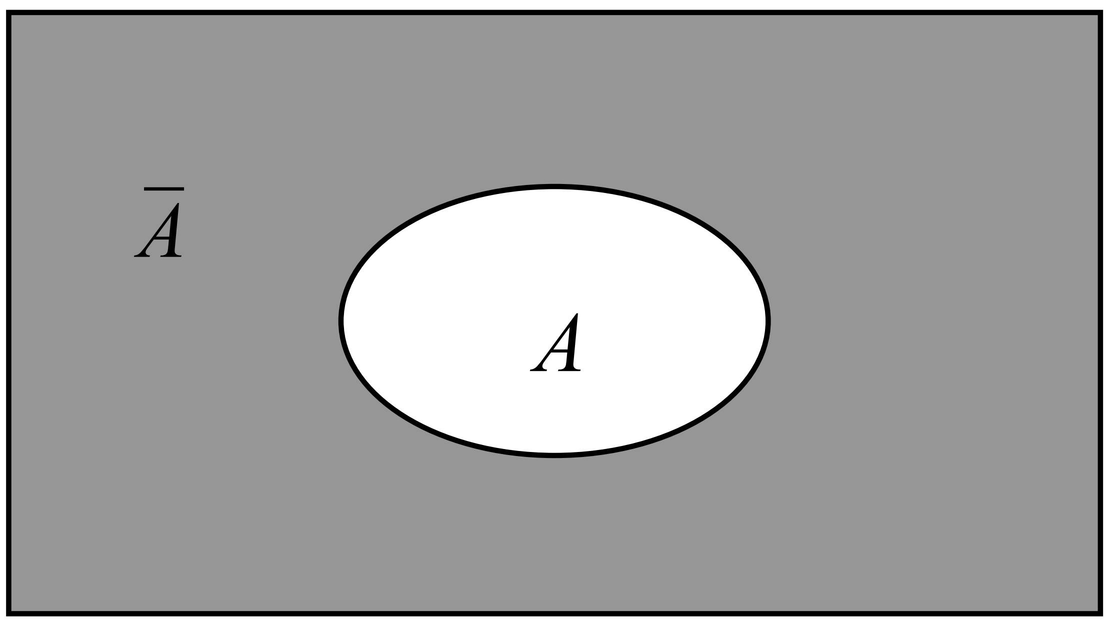
The event $A \cap B$ **occurs** if the event A **occurs and** the event and B **occurs** .



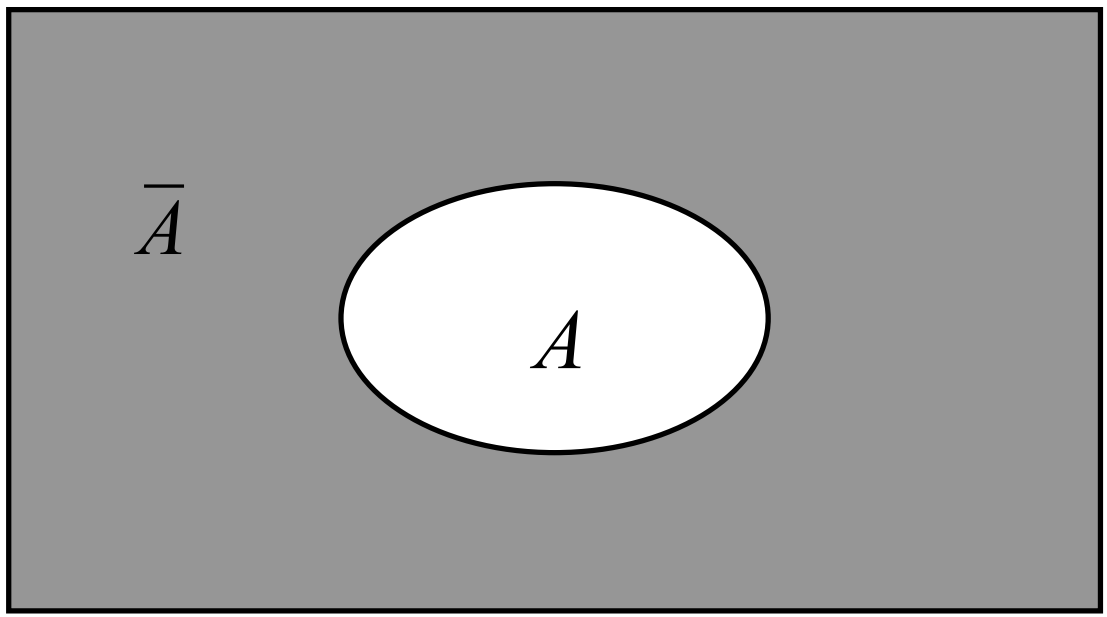
Complement

Let A be any event, then the **complement** of A (denoted by $\overline{A}$) defined by:

$$\overline{A} = \{e|\ e \textbf{ does not } \text{belongs to } A\}$$



The event $\bar{A}$ **occurs** if the event A **does not occur**



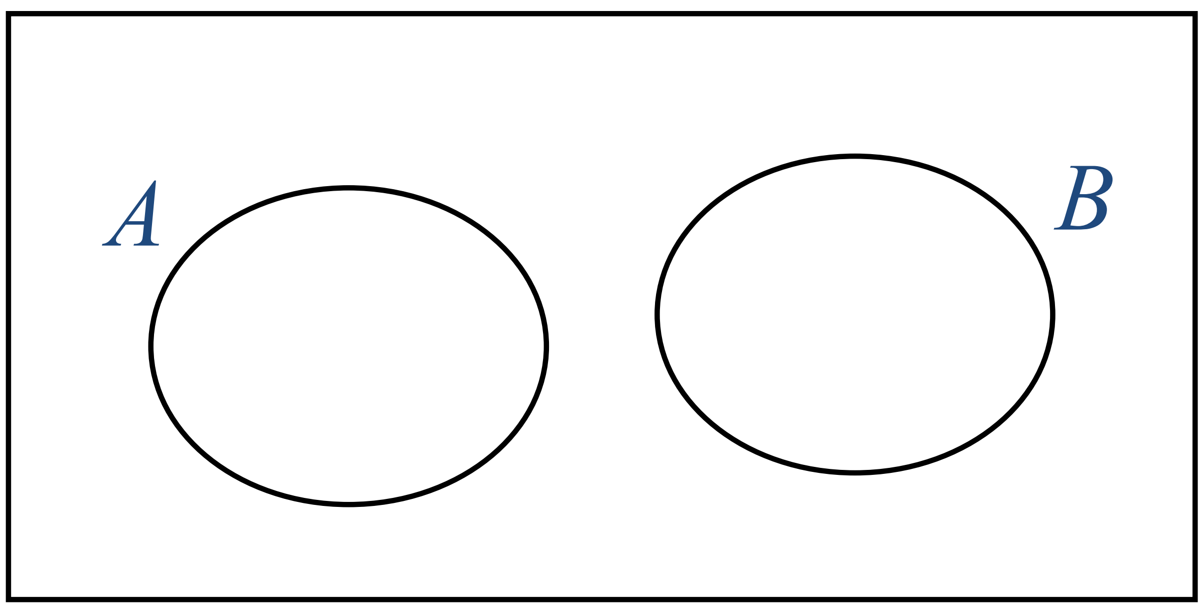
In problems you will recognize that you are working with:

1. **Union** if you see the word **or**,
2. **Intersection** if you see the word **and**,
3. **Complement** if you see the word **not**.

Definition: mutually exclusive

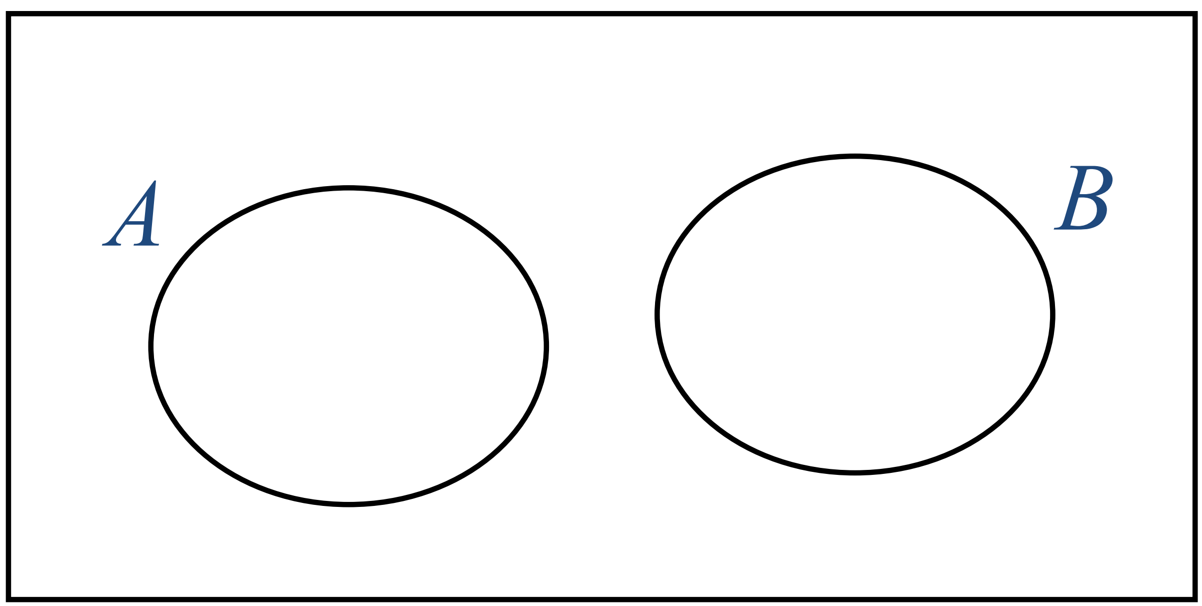
Two events A and B are called **mutually exclusive** if:

$$A \cap B = \phi$$



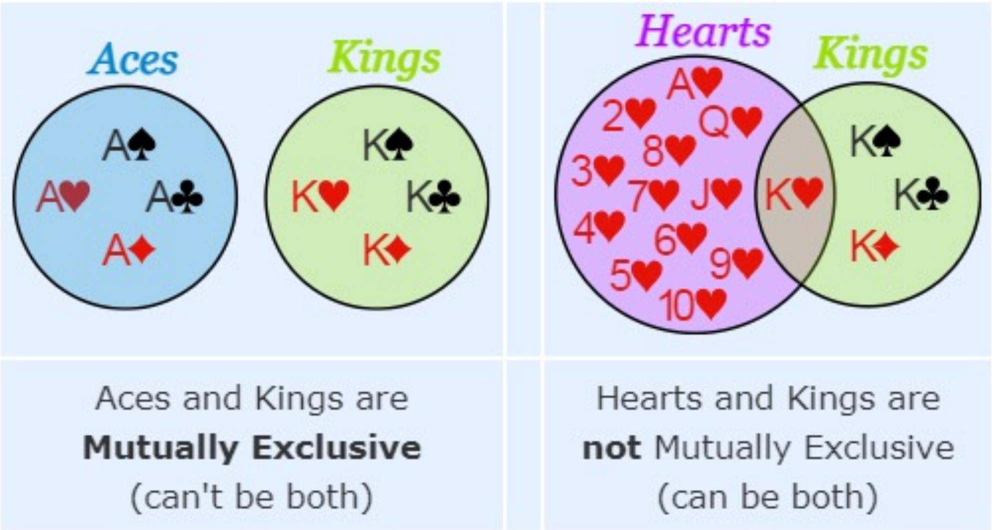
If two events A and B are are **mutually exclusive** then:

1. They have no outcomes in common.
   They can't occur at the same time. The outcome of the random experiment can not belong to both A and B.



Examples:

- Turning left and turning right are Mutually Exclusive (you can't do both at the same time)
- Tossing a coin: Heads and Tails are Mutually Exclusive
- Cards: Kings and Aces are Mutually Exclusive
- What is **not** Mutually Exclusive:
- Kings and Hearts, because we can have a King of Hearts!



Probability

Definition: probability of an Event *E.*

Suppose that the sample space $S = \{o_1, o_2, o_3, \ldots o_N\}$ has a finite number, *N*, of oucomes.

Also each of the outcomes is equally likely (because of symmetry).

Then for any event *E*

$$P[E] = \frac{n(E)}{n(S)} = \frac{n(E)}{N} = \frac{\text{no. of outcomes in } E}{\text{total no. of outcomes}}$$

$\textbf{Note}:$ the symbol $n(A) =$ no. of elements of A

Thus this definition of $P[E]$, i.e.

$$P[E] = \frac{n(E)}{n(S)} = \frac{n(E)}{N} = \frac{\text{no. of outcomes in } E}{\text{total no. of outcomes}}$$

Applies only to the special case when

1. The sample space has a finite no.of outcomes, and
2. Each outcome is equi-probable

   If this is not true a more general definition of probability is required.

Rules of Probability

Rule The additive rule

(Mutually exclusive events)

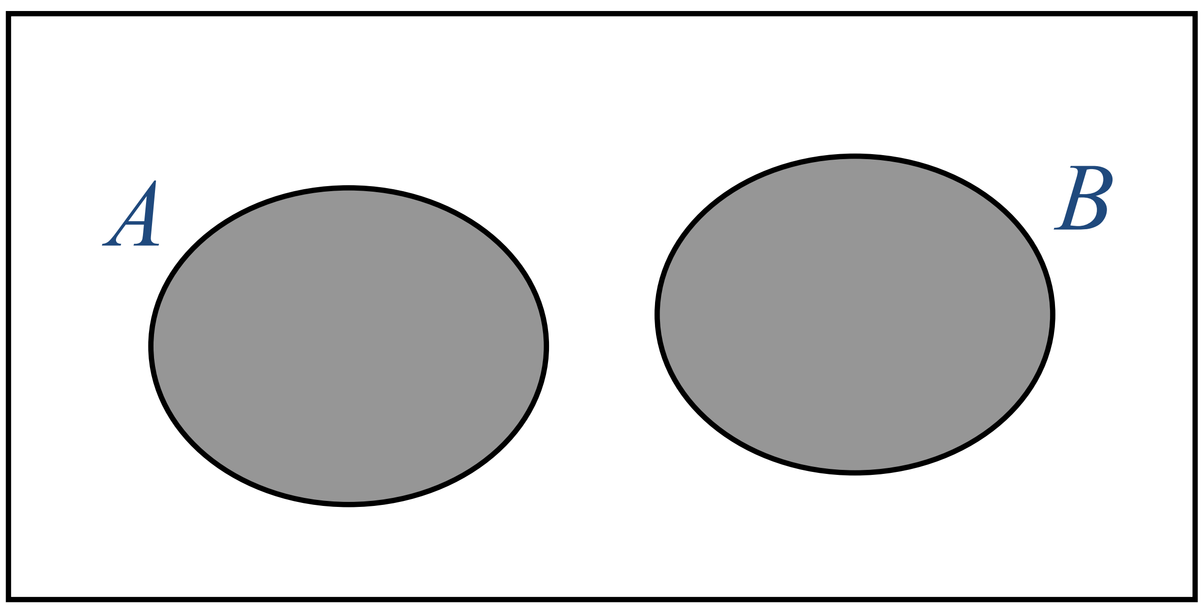
$$\boxed{\begin{array}{l} P[A \cup B] = P[A] + P[B] \\ \text{i.e.} \\ P[A \text{ or } B] = P[A] + P[B] \end{array}}$$

if $A \cap B = \phi$

(A and B mutually exclusive)

If two events A and B are are **mutually exclusive** then:

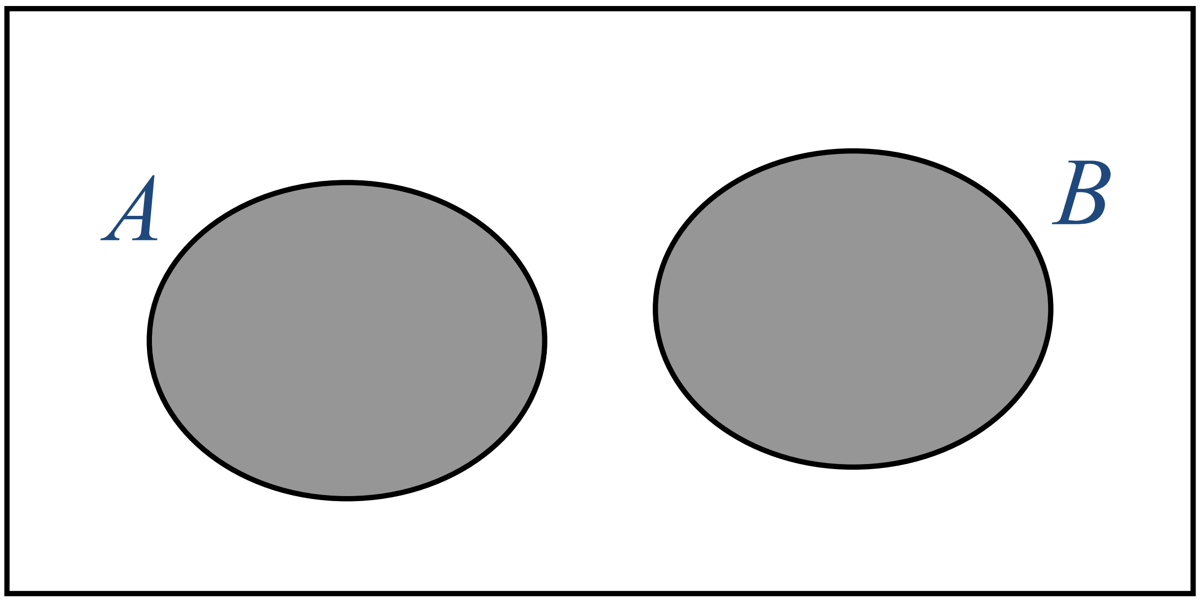
1. They have no outcomes in common.
   They can't occur at the same time. The outcome of the random experiment can not belong to both A and B.



$$P[A \cup B] = P[A] + P[B]$$

i.e.

$$P[A \text{ or } B] = P[A] + P[B]$$



Rule The additive rule

(In general)

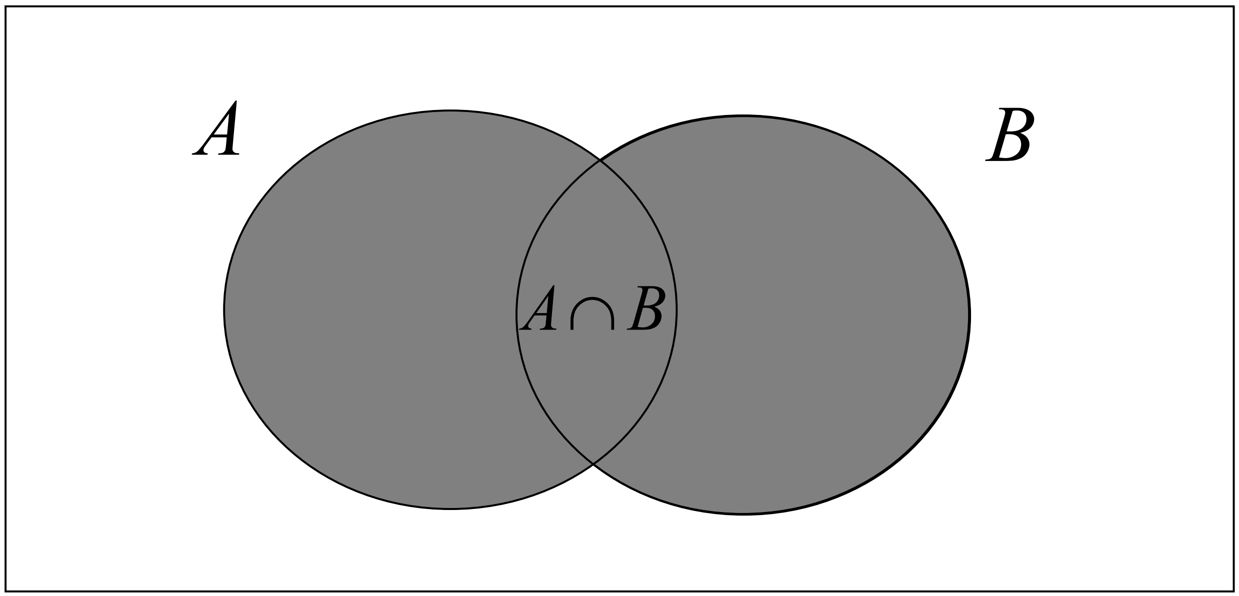
$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

or

$$P[A \text{ or } B] = P[A] + P[B] - P[A \text{ and } B]$$

Logic

$$A \cup B$$



When $P[A]$ is added to $P[B]$ the outcome in $A \cap B$ are counted twice

hence

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

Example:

Saskatoon and Moncton are two of the cities competing for the World university games. (There are also many others). The organizers are narrowing the competition to the **final 5 cities.**

There is a 20% chance that Saskatoon will be amongst the **final 5**. There is a 35% chance that Moncton will be amongst the **final 5** and an 8% chance that both Saskatoon and Moncton will be amongst the **final 5**.

What is the probability that Saskatoon or Moncton will be amongst the **final 5.**

Solution:

Let A = the event that Saskatoon is amongst the **final 5**.

Let B = the event that Moncton is amongst the **final 5**.

Given $P[A] = 0.20$, $P[B] = 0.35$, and $P[A \cap B] = 0.08$

What is $P[A \cup B]$?

Note: "and" $\equiv \cap$**,** "or" $\equiv \cup$.

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

$$= 0.20 + 0.35 - 0.08 = 0.47$$

Rule for complements

2. $$P\left[\bar{A}\right] = 1 - P\left[A\right]$$

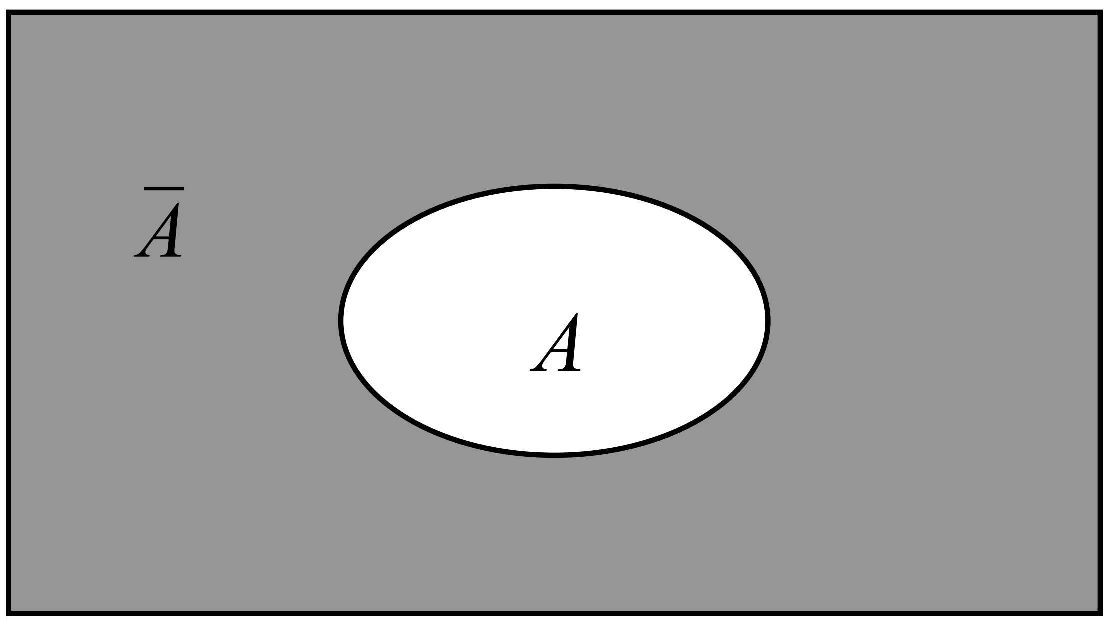
or

$$P\left[\text{not } A\right] = 1 - P\left[A\right]$$

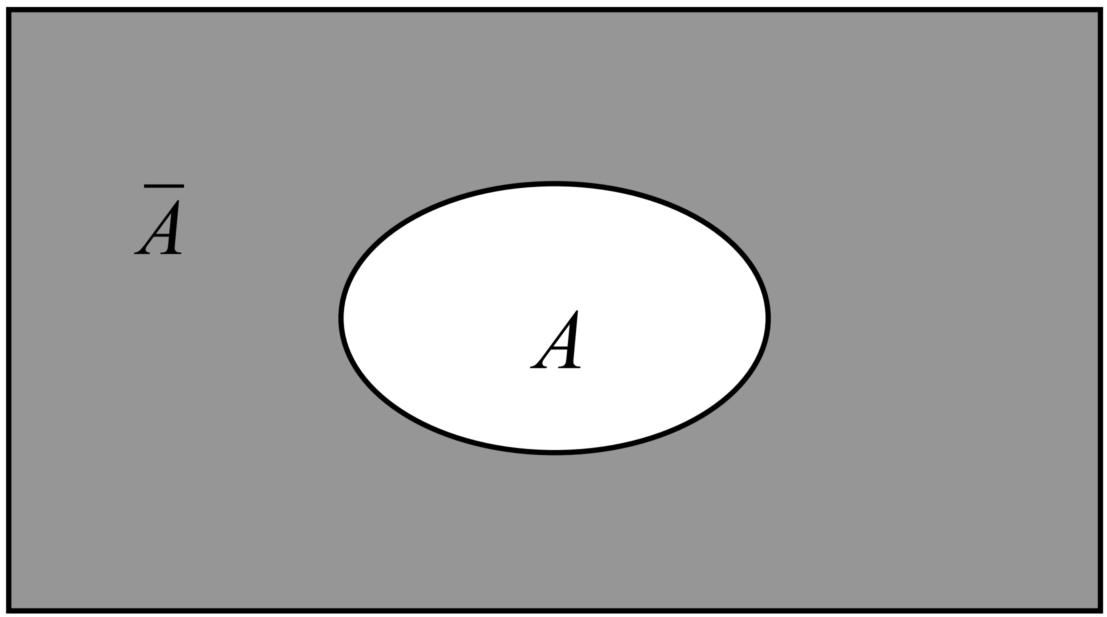
Complement

Let A be any event, then the **complement** of A (denoted by $\bar{A}$) defined by:

$$\bar{A} = \{e \mid e \textbf{ does not} \text{ belongs to } A\}$$



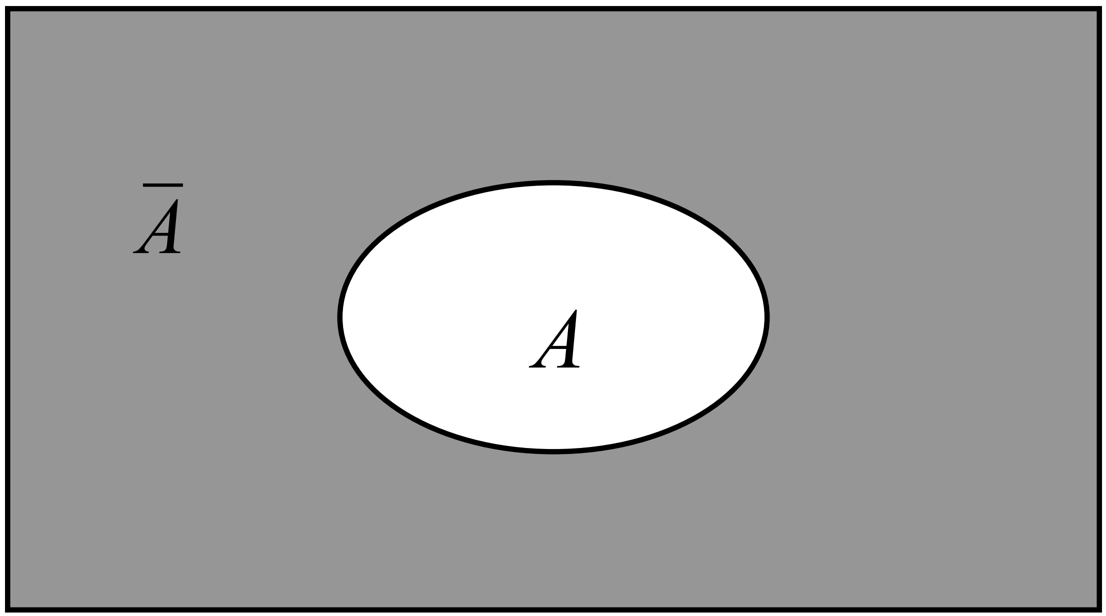
The event $\bar{A}$ **occurs** if the event A **does not occur**



Logic:

$\bar{A}$ and A are **mutually exclusive**.

and $S = A \cup \bar{A}$



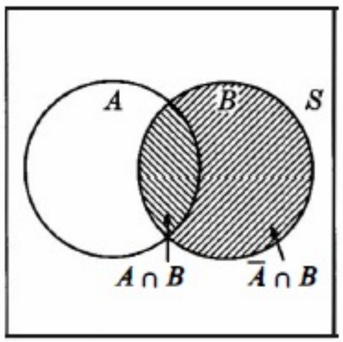
thus $1 = P[S] = P[A] + P[\bar{A}]$

and $P[\bar{A}] = 1 - P[A]$

Proof

If A and B are *any* two events, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$



Thus we write

$$A \cup B = A \cup (B \cap \bar{A}),$$
$$B = (A \cap B) \cup (B \cap \bar{A}).$$

Hence

$$P(A \cup B) = P(A) + P(B \cap \bar{A}),$$
$$P(B) = P(A \cap B) + P(B \cap \bar{A}).$$

Subtracting the second equation from the first yields

$$P(A \cup B) - P(B) = P(A) - P(A \cap B)$$

EXAMPLE

- An electronic device is tested and its total time of service, say t,
- is recorded. We shall assume the sample space to be {t I t >= O}. Let the three
- events A, B, and C be defined as follows:
- A = {t I t < 100}; B = {t I 50 <= t <= 200}; c = {tIt > 150}.
- AUB, A n B, B U C, B nC, A n C, A U C and A, C

EXAMPLE

- A u B = {t \ t <= 200}; A n B = {t I 50 <= t < 100};
- B u c = {t I t => 50}; B n c = {t I 150 < t :S 200}; A n c = null;
- A U C = {t I t < 100 or t > 150}; A = {t I t => 100}; C = {t \ t <=150}.

Conditional Probability

Conditional Probability

- Frequently before observing the outcome of a random experiment you are given information regarding the outcome
- How should this information be used in prediction of the outcome.
- Namely, how should probabilities be adjusted to take into account this information
- Usually the information is given in the following form: You are told that the outcome belongs to a given event. (i.e. you are told that a certain event has occurred)

Definition

Suppose that we are interested in computing the probability of event *A* and we have been told event *B* has occurred.

Then the conditional probability of *A* given *B* is defined to be:

$$P\left[A \middle| B\right] = \frac{P\left[A \cap B\right]}{P\left[B\right]} \quad \text{if} \quad P\left[B\right] \neq 0$$

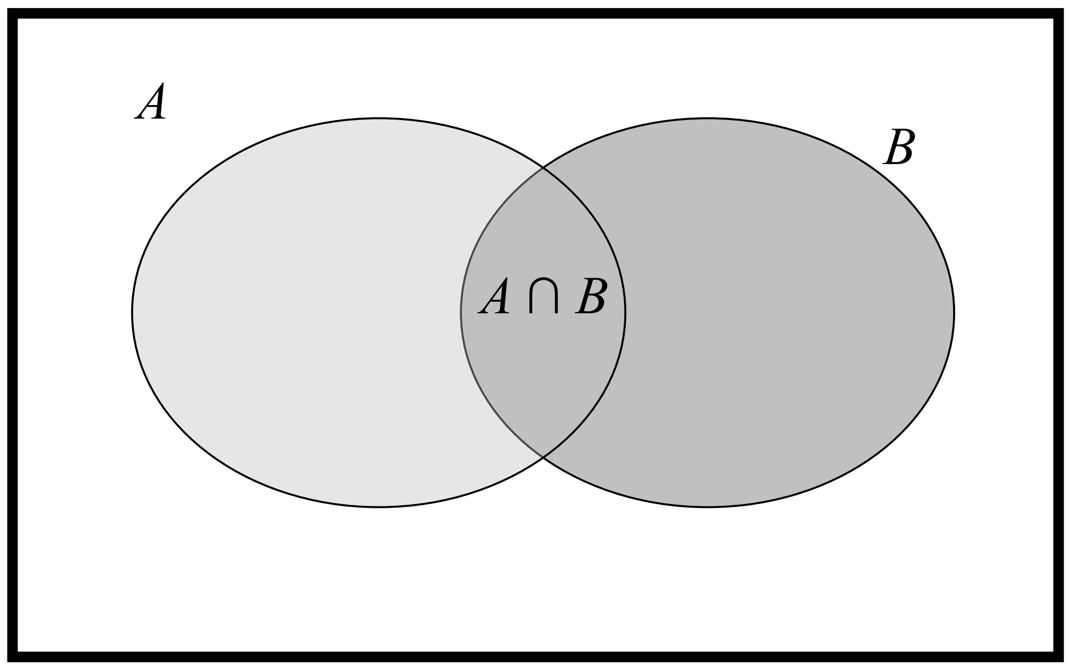
Rationale:

If we’re told that event *B* has occurred then the sample space is restricted to *B.*

The probability within *B* has to be normalized, This is achieved by dividing by *P*[*B*]

The event *A* can now only occur if the outcome is in of $A \cap B$. Hence the new probability of *A* is:

$$P\left[A\middle|B\right]=\frac{P\left[A\cap B\right]}{P\left[B\right]}$$



An Example

The academy awards is soon to be shown.

For a specific married couple the probability that the husband watches the show is 80%, the probability that his wife watches the show is 65%, while the probability that they both watch the show is 60%.

If the husband is watching the show, what is the probability that his wife is also watching the show

Solution:

The academy awards is soon to be shown.

Let B = the event that the husband watches the show

$P[B]$= 0.80

Let A = the event that his wife watches the show

$P[A]$= 0.65 and $P[A \cap B]$= 0.60

$$P\left[A|B\right] = \frac{P\left[A \cap B\right]}{P\left[B\right]} = \frac{0.60}{0.80} = 0.75$$

Independence

Definition

Two events A and B are called **independent** if

$$P[A \cap B] = P[A]P[B]$$

Note if $P[B] \neq 0$ and $P[A] \neq 0$ then

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[A]P[B]}{P[B]} = P[A]$$

and $$P[B|A] = \frac{P[A \cap B]}{P[A]} = \frac{P[A]P[B]}{P[A]} = P[B]$$

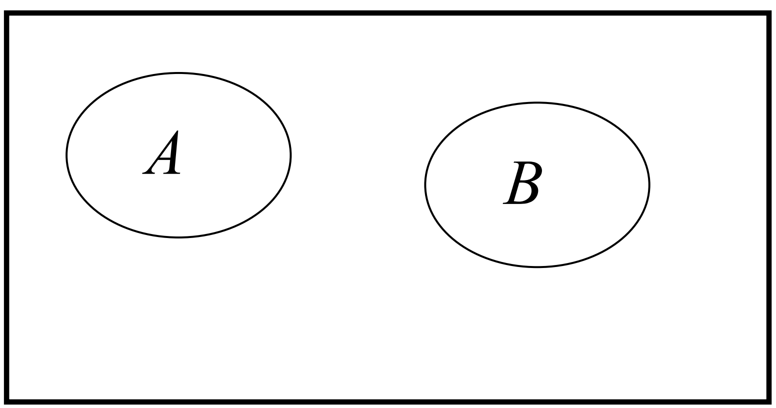
> Thus in the case of independence the conditional probability of an event is not affected by the knowledge of the other event

Difference between **independence** and **mutually exclusive**

mutually exclusive

Two mutually exclusive events are independent only in the special case where

$$P[A]=0 \text{ and } P[B]=0. \text{ (also } P[A \cap B]=0$$

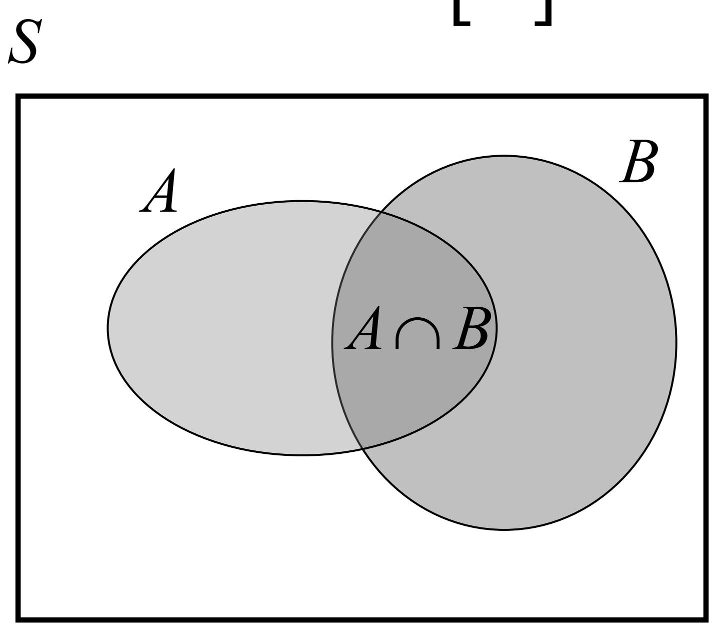


Mutually exclusive events are highly dependent otherwise. *A* and *B* **cannot** occur simultaneously. If one event occurs the other event does not occur.

Independent events

$$P[A \cap B] = P[A]P[B]$$

or $$\frac{P[A \cap B]}{P[B]} = P[A] = \frac{P[A]}{P[S]}$$



The ratio of the probability of the set A within B is the same as the ratio of the probability of the set A within the entire sample S.

The multiplicative rule of probability

$$P[A \cap B] = \begin{cases} P[A]P[B|A] & \text{if } P[A] \neq 0 \\ P[B]P[A|B] & \text{if } P[B] \neq 0 \end{cases}$$

and

$$P[A \cap B] = P[A]P[B]$$

if A and B are **independent**.

Probability

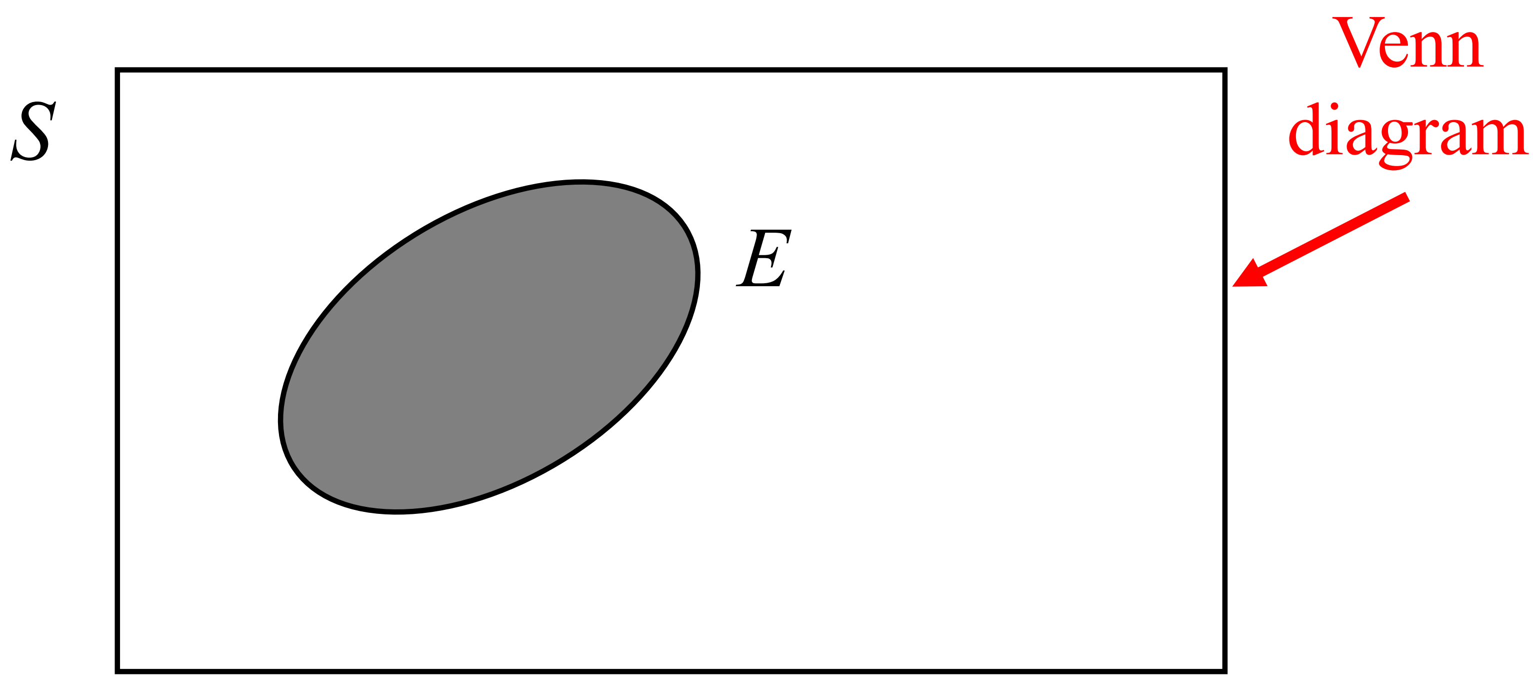
Models for **random** phenomena

The sample Space, *S*

The **sample space**, *S*, for a random phenomena is the set of all possible outcomes.

An Event , *E*

The **event**, *E*, is any subset of the **sample space**, *S*. i.e. any set of outcomes (not necessarily all outcomes) of the random phenomena



Definition: probability of an Event *E.*

Suppose that the sample space $S = \{o_1, o_2, o_3, \ldots o_N\}$ has a finite number, N, of oucomes.

Also each of the outcomes is equally likely (because of symmetry).

Then for any event E

$$P[E] = \frac{n(E)}{n(S)} = \frac{n(E)}{N} = \frac{\text{no. of outcomes in } E}{\text{total no. of outcomes}}$$

$\textbf{Note}:$ the symbol $n(A) =$ no. of elements of A

Thus this definition of $P[E]$, i.e.

$$P[E] = \frac{n(E)}{n(S)} = \frac{n(E)}{N} = \frac{\text{no. of outcomes in } E}{\text{total no. of outcomes}}$$

Applies only to the special case when

1. The sample space has a finite no.of outcomes, and
2. Each outcome is equi-probable

   If this is not true a more general definition of probability is required.

Summary of the Rules of Probability

The additive rule

$$\boxed{P[A \cup B] = P[A] + P[B] - P[A \cap B]}$$

and

$$P[A \cup B] = P[A] + P[B] \quad \text{if } A \cap B = \phi$$

The Rule for complements

for any event E

$$\boxed{P\left[\bar{E}\right] = 1 - P\left[E\right]}$$

Conditional probability

$$\boxed{P\left[A \middle| B\right] = \frac{P\left[A \cap B\right]}{P\left[B\right]}}$$

The multiplicative rule of probability

$$P[A \cap B] = \begin{cases} P[A]P[B|A] & \text{if } P[A] \neq 0 \\ P[B]P[A|B] & \text{if } P[B] \neq 0 \end{cases}$$

and

$$P[A \cap B] = P[A]P[B]$$

if A and B are **independent**.

This is the definition of independent

Warm-Up

- If P(A) = 0.3 and P(B) = 0.4 and if A and B are mutually exclusive events, find:

  a. $P(\overline{A})$

  b. $P(\overline{B})$

  c. $P(A\, or\, B)$

  d. $P(A\, and\, B)$

- a. 0.7
- b. 0.6
- c. 0.7
- d. 0

Multiplication Rule – Independent Events……

- When 2 events are **independent**, the probability of both occurring is

$$P(A\, and\, B) = P(A) \cdot P(B)$$

General Rule……

- “or” means to add
- “and” means to multiply (unless it is in a contingency table and you can actually see the intersection)

Example……

- If a coin is tossed twice, find the probability of getting 2 heads.

- Answer:

$$P(H\ and\ H) = P(H) \cdot P(H)$$

$$P(H\ and\ H) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Example……

- A coin is flipped and a die is rolled. Find the probability of getting a head on the coin ***and*** a 4 on the die.

- Answer:

$$P(H\ and\ 4) = P(H) \cdot P(4)$$

$$P(H\ and\ 4) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$$

Example……

- A card is drawn from a deck and replaced; then a 2^{nd} card is drawn. Find the probability of getting a queen ***and*** then an ace.

- Answer:

$$P(Q\, and\ A) = P(Q) \cdot P(A)$$

$$P(Q\, and\ A) = \frac{4}{52} \cdot \frac{4}{52} = \frac{1}{169}$$

Example……

- A box contains 3 red balls, 2 blue balls, and 5 white balls. A ball is selected and its color noted. Then it is replaced. A 2nd ball is selected and its color noted. Find the probability of

a. Selecting 2 blue balls

b. Selecting a blue ball and then a white ball

c. Selecting a red ball and then a blue ball

Answers……

a. Selecting 2 blue balls

$$P(B\ and\ B) = \frac{2}{10} \cdot \frac{2}{10} = \frac{1}{25}$$

b. Selecting a blue ball and then a white ball

$$P(B\ and\ W) = \frac{2}{10} \cdot \frac{5}{10} = \frac{1}{10}$$

c. Selecting a red ball and then a blue ball

$$P(R\ and\ B) = \frac{3}{10} \cdot \frac{2}{10} = \frac{3}{50}$$

Example……

- A poll found that 46% of Americans say they suffer from stress. If 3 people are selected at random, find the probability that **all three** will say they suffer from stress.

- Answer:

$$P(S \text{ and } S \text{ and } S) = P(S) \cdot P(S) \cdot P(S)$$

$$P(Stress) = (0.46)^3 = 0.097$$

Dependent Events……

- When the outcome or occurrence of the first event affects the outcome or occurrence of the second event in such a way that the probability is changed.

Examples of Dependent Events……

1. Draw a card from a deck. Do not replace it and draw another card.
2. Having high grades and getting a scholarship
3. Parking in a no parking zone and getting a ticket

Multiplication Rule – Dependent Events……

- When 2 events are dependent, the probability of both occurring is

$$\boxed{P(A\,and\,B) = P(A)\cdot P(B|A)}$$

- The slash reads:

  "The probability that B occurs given that A has already occurred."

Example……

- 53% of residents had homeowner's insurance. Of these, 27% also had car insurance. If a resident is selected at random, find the prob. That the resident has both homeowner's ***and*** car insurance.

- Answer:

$$P(H\ and\ C) = P(H) \cdot P(ClH)$$

$$P(H\ and\ C) = (.53)(.27) = .1431$$

Example……

- 3 cards are drawn from a deck and **<u>NOT</u>** replaced. Find the following probabilities.
  a. Getting 3 jacks
  b. Getting an ace, king, and queen
  c. Getting a club, spade, and heart
  d. Getting 3 clubs.

a. Getting 3 jacks……

$$\boxed{P(J\ and\ J\ and\ J) = \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} = \frac{1}{5525} = .000181}$$

b. Getting an ace, king, queen……

$$\boxed{P(A\ and\ K\ and\ Q) = \frac{4}{52}\cdot\frac{4}{51}\cdot\frac{4}{50} = \frac{8}{16575} = .000483}$$

c. Getting a club, spade, and heart……

$$\boxed{P(C\ and\ S\ and\ H) = \frac{13}{52}\cdot\frac{13}{51}\cdot\frac{13}{50} = \frac{169}{10200} = .017}$$

d. Getting 3 clubs……

$$\boxed{P(C\ and\ C\ and\ C) = \frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50} = \frac{11}{850}\ or\ .013}$$

Back to Conditional Probability - Remember……

$$\boxed{P(A\,and\,B) = P(A)\cdot P\left(B/A\right)}$$

- **Algebraically change this so that it is now in the form……**

“Given”

$$\boxed{P\left(B/A\right)=\frac{P(A\, and\; B)}{P(A)}}$$

$$\boxed{P\left(A/B\right)=\frac{P(A\, and\; B)}{P(B)}}$$

Example……

- In Rolling Acres Housing Plan, 42% of the houses have a deck **and** a garage; 60% have a deck. Find the probability that a home has a garage, ***given*** that it has a deck.

Answer……

$$P(Deck\ and\ Garage) = .42$$

$$P(Deck) = .60$$

$$Find\ P(G/D)$$

- Answer:

$$P(G/D) = \frac{P(G\ and\ D)}{P(D)}$$

$$\boxed{P(G/D) = \frac{.42}{.6} = .70}$$

Independent $\neq$ mutually exclusive

- Events A and ~A are mutually exclusive, but they are NOT independent.
- P(A&~A)= 0
- P(A)*P(~A) $\neq$ 0

Conceptually, once A has happened, ~A is impossible; thus, they are completely dependent.

Example……

- At an exclusive country club, 83% of the members play bridge; 75% of the members drink champagne ***given*** that he or she plays bridge. Find the probability that members drink champagne **and** play bridge.

Answer……

- Answer:

$$P(bridge) = .83$$

$$P(C/B) = \frac{P(C \; and \; B)}{P(B)}$$

$$P(champ/bridge) = .75$$

$$.75 = \frac{P(B \; and \; C)}{.83}$$

$$Find \; P(champ \; and \; bridge)$$

$$\boxed{P(C \; and \; B) = (.75)(.83) = .62}$$

Example

- Roll two dice and consider the following events A = ‘first die is 3’ B = ‘sum is 6’ C = ‘sum is 7’

A is independent of

(a) B and C (b) B alone (c) C alone (d) Neither B or C.

Examples

Toss a coin 4 times.

Let A = ‘at least three heads’ B = ‘first toss is tails’.

1. What is P(A|B)? (a) 1/16 (b) 1/8 (c) 1/4 (d) 1/5

2. What is P(B|A)? (a) 1/16 (b) 1/8 (c) 1/4 (d) 1/5

Counting techniques

Finite uniform probability space

Many examples fall into this category

1. Finite number of outcomes
2. All outcomes are equally likely
3. $$P[E] = \frac{n(E)}{n(S)} = \frac{n(E)}{N} = \frac{\text{no. of outcomes in } E}{\text{total no. of outcomes}}$$

$$\textbf{Note}: \quad n(A) = \text{ no. of elements of } A$$

To handle problems in case we have to be able to count. Count $n(E)$ and $n(S)$.

Techniques for counting

Rule 1

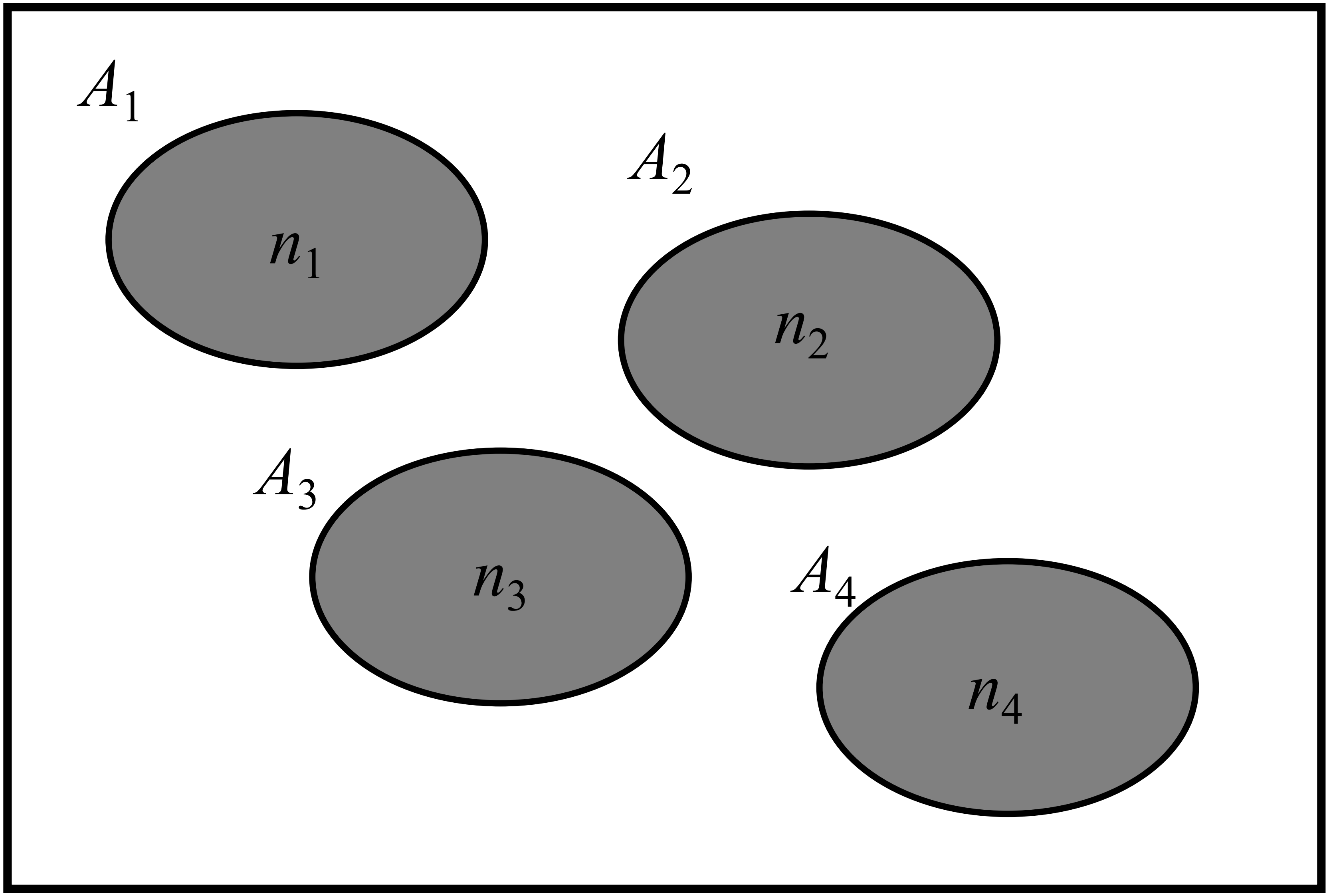
Suppose we carry out have a sets A_1, A_2, A_3, … and that any pair are mutually exclusive

(i.e. $A_1 \cap A_2 = \phi$) Let

$$n_i = n\,(A_i) = \text{the number of elements in } A_i.$$

$$\text{Let } A = A_1 \cup A_2 \cup A_3 \cup \ldots.$$

$$\text{Then } N = n(\,A\,) = \text{the number of elements in } A$$
$$= n_1 + n_2 + n_3 + \ldots$$



A_1
n_1
A_2
n_2
A_3
n_3
A_4
n_4

Rule 2

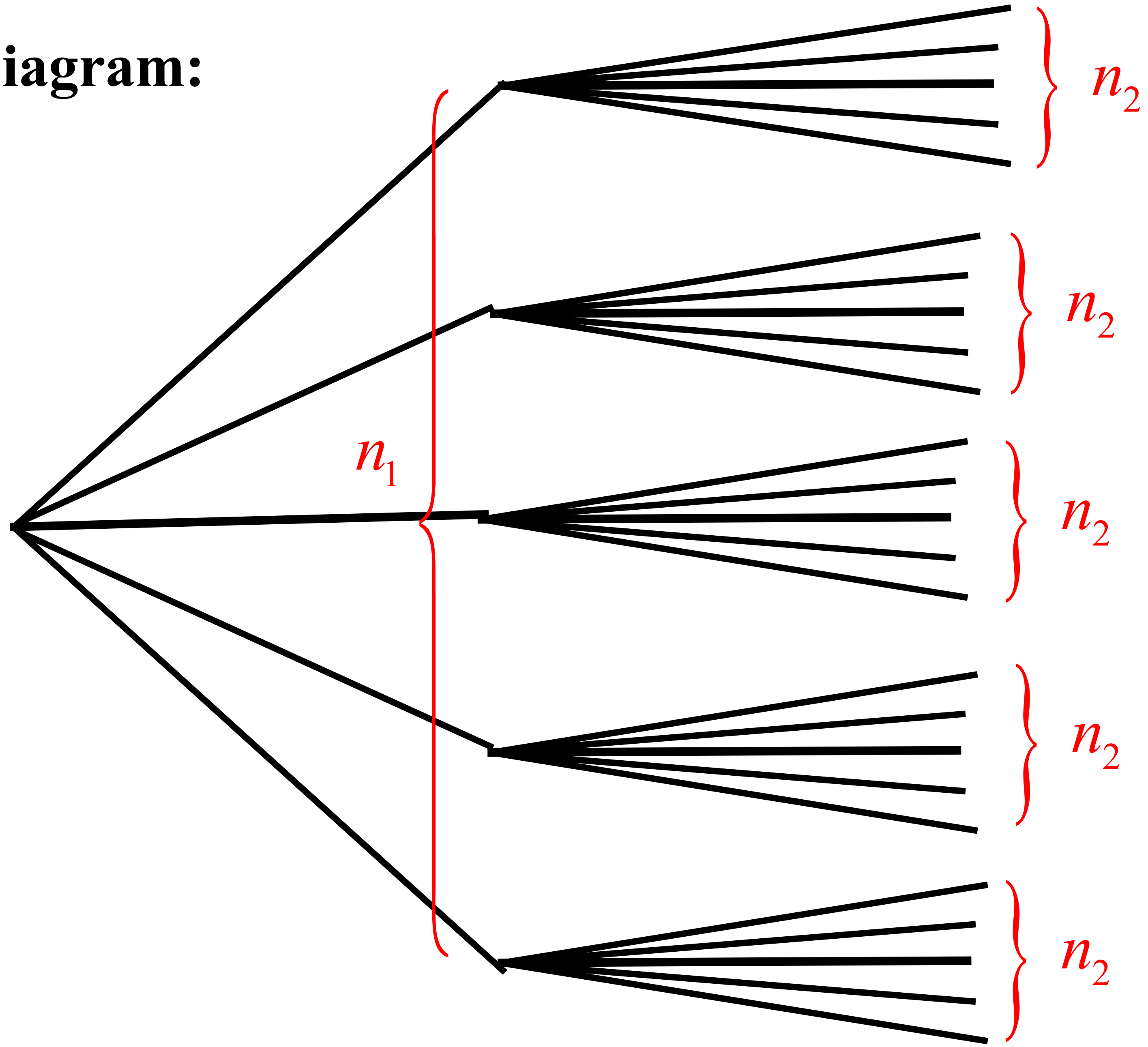
Suppose we carry out two operations in sequence

Let n_1 = the number of ways the first operation can be performed

n_2 = the number of ways the second operation can be performed once the first operation has been completed.

Then $\boxed{N = n_1\, n_2}$ = the number of ways the two operations can be performed in sequence.

Diagram:



Examples

1. We have a committee of 10 people. We choose from this committee, a chairman and a vice chairman. How may ways can this be done?

Solution:

Let n_1 = the number of ways the chairman can be chosen = 10.

Let n_2 = the number of ways the vice-chairman can be chosen once the chair has been chosen = 9.

Then $N = n_1 n_2 = (10)(9) = 90$

2. In **Black Jack** you are dealt 2 cards. What is the probability that you will be dealt a 21?

Solution:

The number of ways that two cards can be selected from a deck of 52 is $N = (52)(51) = 2652$.

A "21" can occur if the first card is an ace and the second card is a face card or a ten {10, J, Q, K} **or** the first card is a face card or a ten and the second card is an ace.

The number of such hands is $(4)(16) + (16)(4) = 128$

Thus the probability of a "21" = $128/2652 = 32/663$

The Multiplicative Rule of Counting

Suppose we carry out *k* operations in sequence

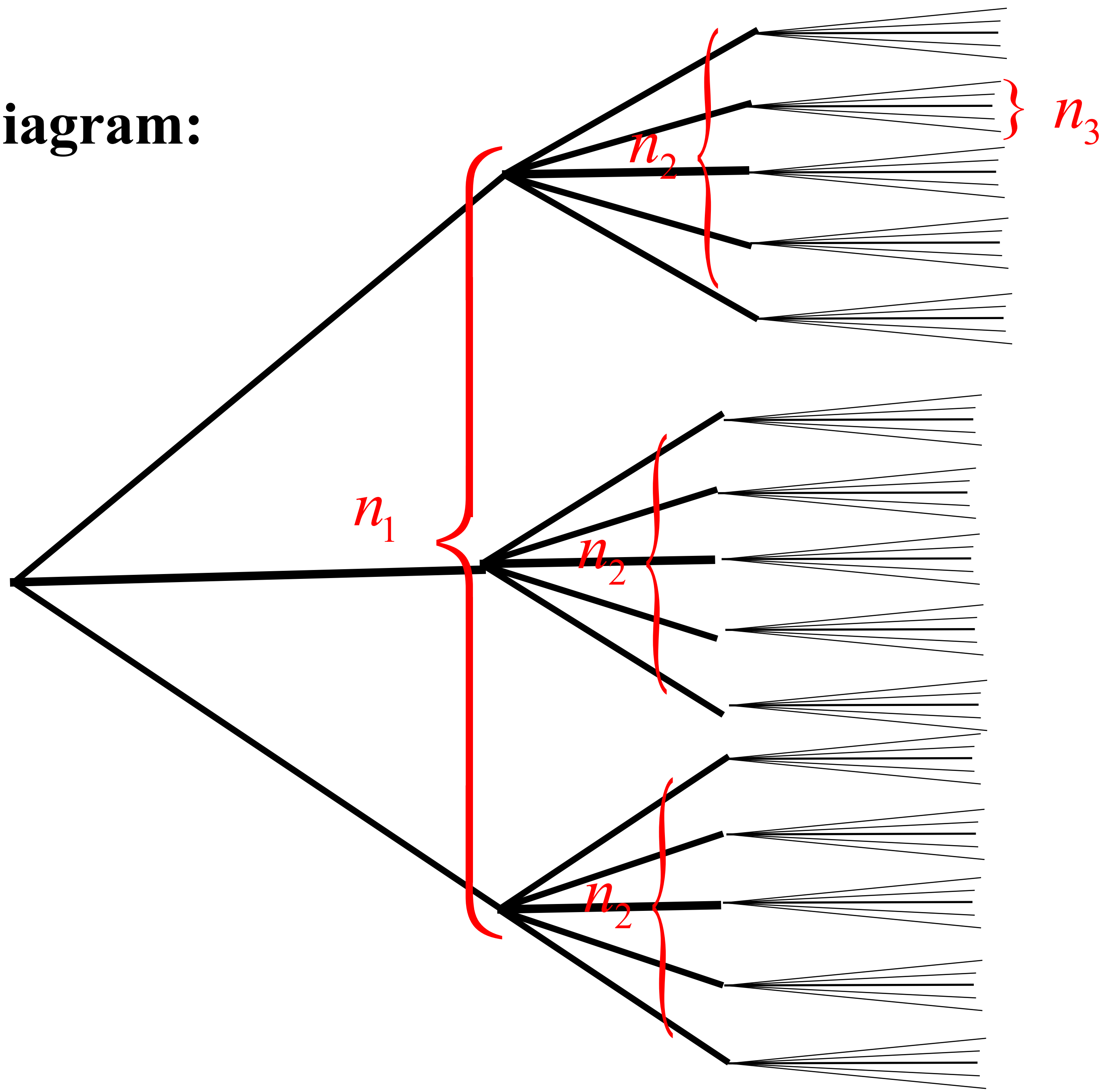
Let

n_1 = the number of ways the first operation can be performed

n_i = the number of ways the i^{th} operation can be performed once the first $(i - 1)$ operations have been completed. $i = 2, 3, \ldots , k$

Then $\boxed{N = n_1 n_2 \ldots n_k}$ = the number of ways the k operations can be performed in sequence.

Diagram:



Examples

1. **Permutations:** How many ways can you order *n* objects

Solution:

Ordering n objects is equivalent to performing n operations in sequence.

1. Choosing the first object in the sequence ($n_1 = n$)
2. Choosing the 2^{nd} object in the sequence ($n_2 = n - 1$).

…

k. Choosing the k^{th} object in the sequence ($n_k = n - k + 1$)

…

n. Choosing the n^{th} object in the sequence ($n_n = 1$)

The total number of ways this can be done is:

$$\boxed{N = n(n-1)\ldots(n-k+1)\ldots(3)(2)(1) = n!}$$

Example How many ways can you order the 4 objects {A, B, C, D}

Solution:

$N = 4! = 4(3)(2)(1) = 24$

Here are the orderings.

ABCD	*ABDC*	*ACBD*	*ACDB*	*ADBC*	*ADCB*
BACD	*BADC*	*BCAD*	*BCDA*	*BDAC*	*BDCA*
CABD	*CADB*	*CBAD*	*CBDA*	*CDAB*	*CDBA*
DABC	*DACB*	*DBAC*	*DBCA*	*DCAB*	*DCBA*

Examples - continued

2. **Permutations of size *k* (< *n*):** How many ways can you choose *k* objects from *n* objects in a specific order

Solution: This operation is equivalent to performing k operations in sequence.

1. Choosing the first object in the sequence ($n_1 = n$)
2. Choosing the 2^{nd} object in the sequence ($n_2 = n$ -1).

…

k. Choosing the k^{th} object in the sequence ($n_k = n - k + 1$)

The total number of ways this can be done is:

$$N = n(n-1)\ldots(n-k+1) = n!/\,(n-k)!$$

This number is denoted by the symbol

$$_nP_k = n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

Definition: $0! = 1$

This definition is consistent with

$$_{n}P_{k} = n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

for $k = n$

$$_{n}P_{n} = \frac{n!}{0!} = \frac{n!}{1} = n!$$

Example How many permutations of size 3 can be found in the group of 5 objects {*A, B, C, D, E*}

Solution: $${}_5P_3 = \frac{5!}{(5-3)!} = 5(4)(3) = 60$$

ABC	*ABD*	*ABE*	*ACD*	*ACE*	*ADE*	*BCD*	*BCE*	*BDE*	*CDE*
ACB	*ADB*	*AEB*	*ADC*	*AEC*	*AED*	*BDC*	*BEC*	*BED*	*CED*
BAC	*BAD*	*BAE*	*CAD*	*CAE*	*DAE*	*CBD*	*CBE*	*DBE*	*DCE*
BCA	*BDA*	*BEA*	*CDA*	*CEA*	*DEA*	*CDB*	*CEB*	*DEB*	*DEC*
CAB	*DAB*	*EAB*	*DAC*	*EAC*	*EAD*	*DBC*	*EBC*	*EBD*	*ECD*
CAB	*DBA*	*EBA*	*DCA*	*ECA*	*EDA*	*DCB*	*ECB*	*EDB*	*EDC*

Example We have a committee of n = 10 people and we want to choose a **chairperson**, **a vice-chairperson** and a **treasurer**

Solution: Essentually we want to select 3 persons from the committee of 10 in a specific order. (Permutations of size 3 from a group of 10).

$$ {}_{10}P_3 = \frac{10!}{(10-3)!} = \frac{10!}{7!} = 10(9)(8) = 720 $$

Example We have a committee of n = 10 people and we want to choose a **chairperson**, **a vice-chairperson** and a **treasurer**. Suppose that 6 of the members of the committee are male and 4 of the members are female. What is the probability that the three executives selected are all male?

Solution: Again we want to select 3 persons from the committee of 10 in a specific order. (Permutations of size 3 from a group of 10).The total number of ways that this can be done is:

$$_{10}P_3 = \frac{10!}{(10-3)!} = \frac{10!}{7!} = 10(9)(8) = 720$$

This is the size, $N = n(S)$, of the sample space S. Assume all outcomes in the sample space are equally likely.

Let E be the event that all three executives are male

$$n(E) = {}_6P_3 = \frac{6!}{(6-3)!} = \frac{6!}{3!} = 6(5)(4) = 120$$

Hence

$$P[E]=\frac{n(E)}{n(S)}=\frac{120}{720}=\frac{1}{6}$$

Thus if all candidates are equally likely to be selected to any position on the executive then the probability of selecting an all male executive is:

$$\frac{1}{6}$$

Examples - continued

3. **Combinations of size *k* (≤ *n*):** A combination of size *k* chosen from *n* objects is a subset of size *k* where the order of selection is irrelevant. How many ways can you choose a combination of size *k* objects from *n* objects (order of selection is irrelevant)

Here are the combinations of size 3 selected from the 5 objects $\{A, B, C, D, E\}$

$\{A,B,C\}$	$\{A,B,D\}$	$\{A,B,E\}$	$\{A,C,D\}$	$\{A,C,E\}$
$\{A,D,E\}$	$\{B,C,D\}$	$\{B,C,E\}$	$\{B,D,E\}$	$\{C,D,E\}$

Important Notes

1. In **combinations** ordering is **irrelevant**. Different orderings result in the same combination.
2. In **permutations** order is **relevant**. Different orderings result in the different permutations.

How many ways can you choose a combination of size *k* objects from *n* objects (order of selection is irrelevant)

Solution: Let n_1 denote the number of combinations of size k.
One can construct a permutation of size k by:

1. Choosing a combination of size k (n_1 = unknown)
2. Ordering the elements of the combination to form a permutation ($n_2 = k!$)

Thus $${}_nP_k = \frac{n!}{(n-k)!} = n_1 k!$$

and $$n_1 = \frac{{}_nP_k}{k!} = \frac{n!}{(n-k)!k!} = \text{ the \# of combinations of size } k.$$

The number:

$$n_1 = \frac{{}_nP_k}{k!} = \frac{n!}{(n-k)!k!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k(k-1)(k-2)\cdots(1)}$$

is denoted by the symbol

$${}_nC_k \quad \text{or} \quad \binom{n}{k} \qquad \text{read "}n\text{ choose }k\text{"}$$

It is the number of ways of choosing k objects from n objects (order of selection irrelevant).

${}_nC_k$ is also called a **binomial coefficient**.

It arises when we expand $(x + y)^n$ **(the binomial theorem)**

Summary of counting rules

Rule 1

$$n(A_1 \cup A_2 \cup A_3 \cup \ldots.) = n(A_1) + n(A_2) + n(A_3) + \ldots$$

if the sets A_1, A_2, A_3, … are pairwise mutually exclusive

(i.e. $A_i \cap A_j = \phi$)

Rule 2

$N = n_1 n_2$ = the number of ways that two operations can be performed in sequence if

n_1 = the number of ways the first operation can be performed

n_2 = the number of ways the second operation can be performed once the first operation has been completed.

Rule 3 $N = n_1 n_2 \ldots n_k$

= the number of ways the k operations can be performed in sequence if

n_1 = the number of ways the first operation can be performed

n_i = the number of ways the i^{th} operation can be performed once the first $(i - 1)$ operations have been completed. $i = 2, 3, \ldots, k$

Basic counting formulae

1. Orderings

$n! =$ the number of ways you can order n objects

2. Permutations

$${}_nP_k = \frac{n!}{(n-k)!} =$$ The number of ways that you can choose k objects from n in a specific order

3. Combinations

$$\binom{n}{k} = {}_nC_k = \frac{n!}{k!(n-k)!} =$$ The number of ways that you can choose k objects from n (order of selection irrelevant)

Applications to some counting problems

- The trick is to use the basic counting formulae together with the **Rules**
- We will illustrate this with examples
- Counting problems are not easy. The more practice better the techniques

Quick summary of probability

Bayes’ rule

0:00 / 20:13

Bayes' Rule: derivation

- <u>Definition:</u>

  Let A and B be two events with $P(B) \neq 0$. The conditional probability of A given B is:

$$P(A/B) = \frac{P(A \& B)}{P(B)} \qquad P(B) = 1$$

The idea: if we are given that the event B occurred, the relevant sample space is reduced to B {P(B)=1 because we know B is true} and conditional probability becomes a probability measure on B.

Bayes' Rule: derivation

$P(B/A)$

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

$P(A \cap B)$

can be re-arranged to:

$$P(A\&B) = P(\underline{A/B)P(}B)$$

and, since also:

$$P(B/A) = \frac{P(A\&B)}{P(A)} \quad \therefore P(A\&B) = P(B/A)P(A)$$

$$P(A/B)P(B) = P(A\&B) = P(B/A)P(A)$$

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

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

Bayes' Rule:

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

OR

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

From the "Law of Total Probability"

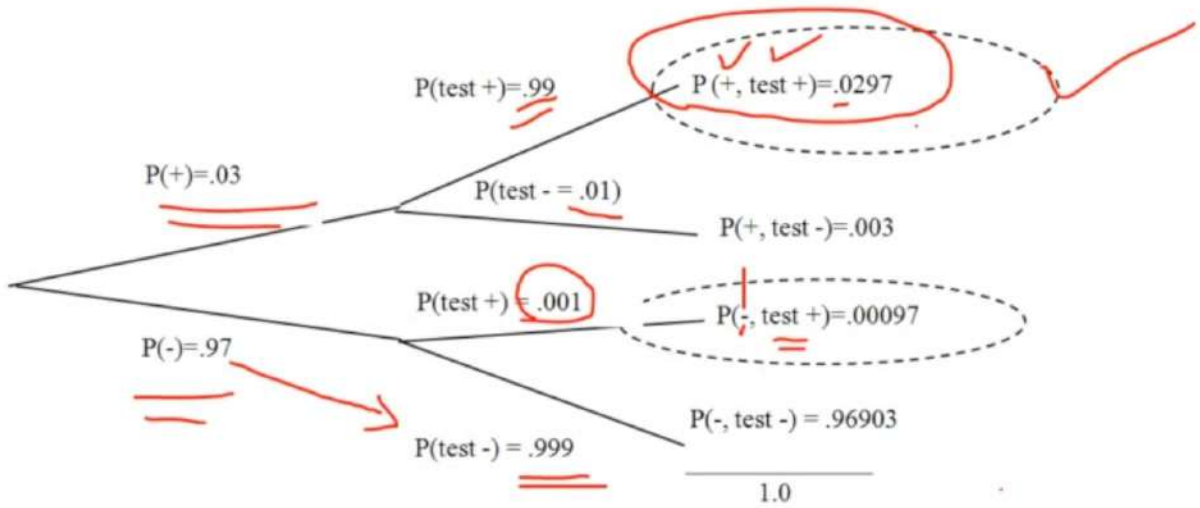
Bayes' Rule:

- Why do we care??
- Why is Bayes' Rule useful??
- It turns out that sometimes it is very useful to be able to "flip" conditional probabilities. That is, we may know the probability of A given B, but the probability of B given A may not be obvious. An example will help…

In-Class Exercise

- If Covid19 has a prevalence of 3% in Mangalore, and a particular Covid19 test has a false positive rate of .001 and a false negative rate of .01, what is the probability that a random person who tests positive is actually infected (also known as "positive predictive value")?

Answer: using probability tree



A positive test places one on either of the two "test +" branches.
But only the top branch also fulfills the event "true infection."
Therefore, the probability of being infected is the probability of being on the top branch given that you are on one of the two circled branches above.

$$P(+/test+) = \frac{P(test + \& true+)}{P(test+)} = \frac{.0297}{.0297 + .00097} = 96.8\%$$

Answer: using Bayes' rule

$$P(true+/test+) = \frac{P(test+/true+)P(true+)}{P(test+/true+)P(true+) + P(test+/true-)P(true-)} =$$

$$\frac{.99(.03)}{.99(.03) + .001(.97)} = 96.8\%$$

Practice problem

An insurance company believes that drivers can be divided into two classes—those that are of high risk and those that are of low risk. Their statistics show that a high-risk driver will have an accident at some time within a year with probability .4, but this probability is only .1 for low risk drivers.

a) Assuming that 20% of the drivers are high-risk, what is the probability that a new policy holder will have an accident within a year of purchasing a policy?

b) If a new policy holder has an accident within a year of purchasing a policy, what is the probability that he is a high-risk type driver?

19:36 / 20:13

Probability Distributions

Random Variable

- A random variable *x* takes on a defined set of values with different probabilities.
    - For example, if you roll a die, the outcome is random (not fixed) and there are 6 possible outcomes, each of which occur with probability one-sixth.
    - For example, if you poll people about their voting preferences, the percentage of the sample that responds "Yes on Proposition 100" is a also a random variable (the percentage will be slightly differently every time you poll).

- Roughly, <u>probability</u> is how frequently we expect different outcomes to occur if we repeat the experiment over and over ("frequentist" view)

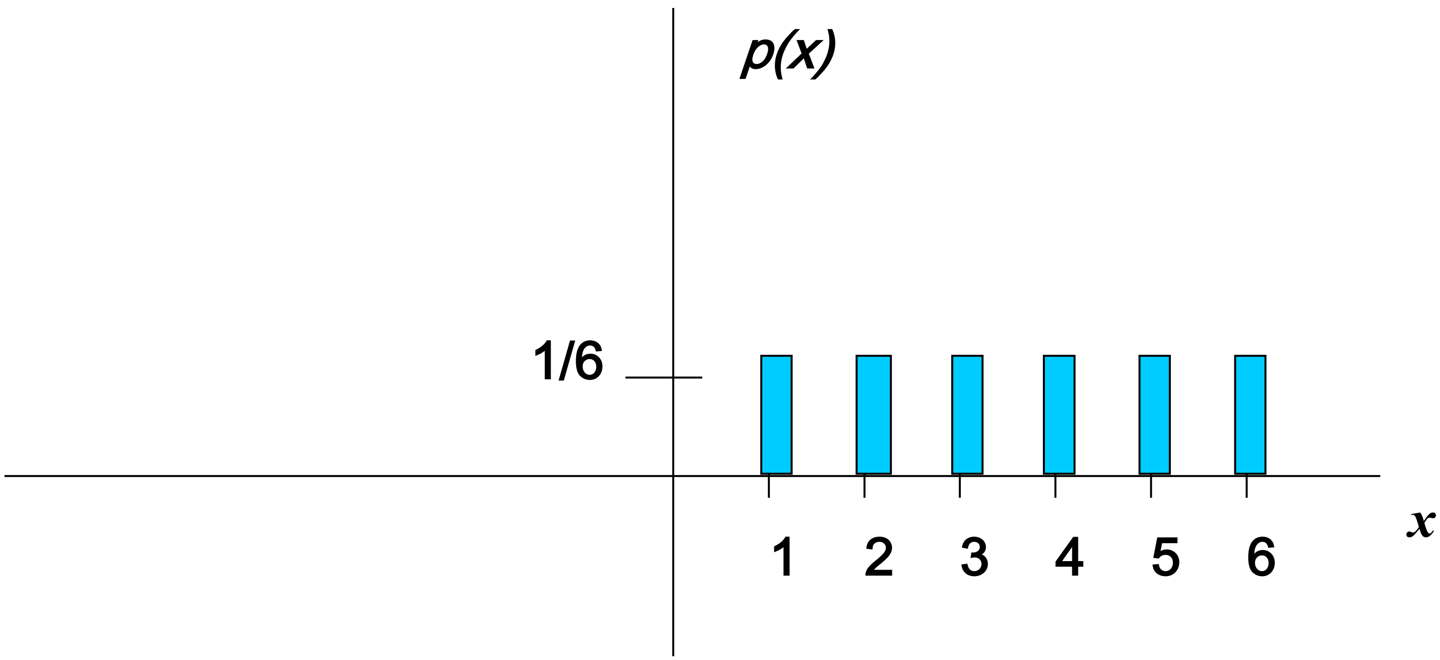
Random variables can be discrete or continuous

- **Discrete** random variables have a countable number of outcomes
  - Examples: Dead/alive, treatment/placebo, dice, counts, etc.
- **Continuous** random variables have an infinite continuum of possible values.
  - Examples: blood pressure, weight, the speed of a car, the real numbers from 1 to 6.

Probability functions

- A probability function maps the possible values of *x* against their respective probabilities of occurrence, *p(x)*
- *p(x)* is a number from 0 to 1.0.
- The area under a probability function is always 1.

Discrete example: roll of a die

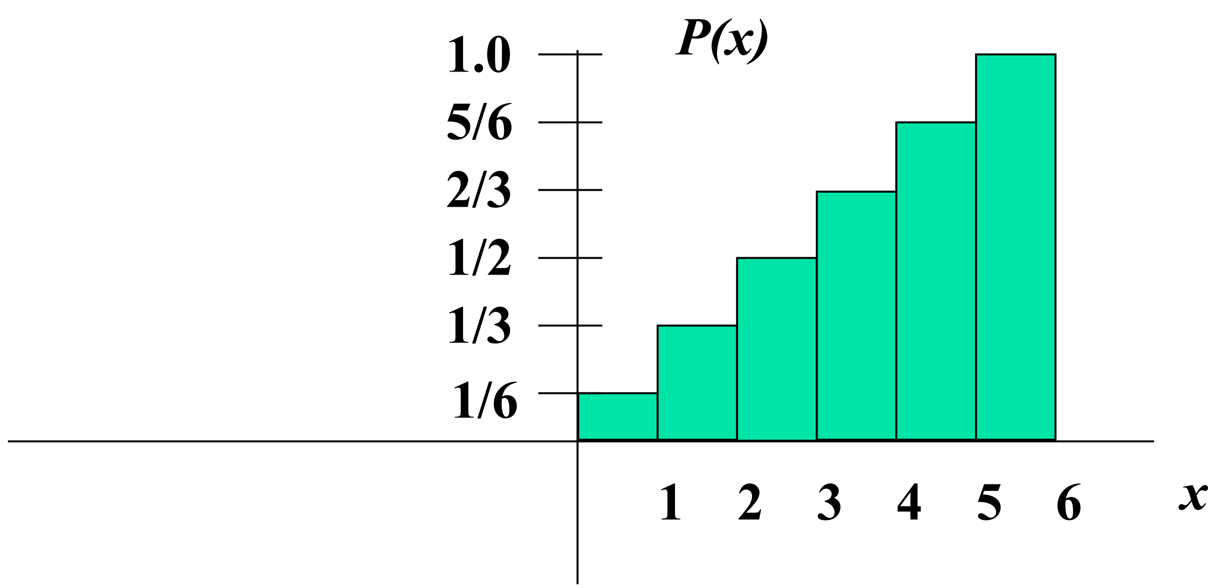


$$\sum_{\text{all } x} P(x) = 1$$

Probability mass function (pmf)

x	$p(x)$
1	$p(x=1)=1/6$
2	$p(x=2)=1/6$
3	$p(x=3)=1/6$
4	$p(x=4)=1/6$
5	$p(x=5)=1/6$
6	$p(x=6)=1/6$
	1.0

Cumulative distribution function (CDF)



Cumulative distribution function

x	$P(x \leq A)$
1	$P(x \leq 1) = 1/6$
2	$P(x \leq 2) = 2/6$
3	$P(x \leq 3) = 3/6$
4	$P(x \leq 4) = 4/6$
5	$P(x \leq 5) = 5/6$
6	$P(x \leq 6) = 6/6$

Examples

1. What's the probability that you roll a 3 or less?
$P(x \leq 3) = 1/2$

2. What's the probability that you roll a 5 or higher?
$P(x \geq 5) = 1 - P(x \leq 4) = 1 - 2/3 = 1/3$

Practice Problem

Which of the following are probability functions?

a. $f(x)=.25$ for x=9,10,11,12

b. $f(x)= (3-x)/2$ for x=1,2,3,4

c. $f(x)= (x^2+x+1)/25$ for x=0,1,2,3

Answer (a)

a. $f(x)=.25$ for x=9,10,11,12

x	*f(x)*
9	.25
10	.25
11	.25
12	.25
	1.0

Yes, probability function!

Answer (b)

b. $f(x) = (3-x)/2$ for x=1,2,3,4

x	*f(x)*
1	(3-1)/2=1.0
2	(3-2)/2=.5
3	(3-3)/2=0
4	(3-4)/2=-.5

Though this sums to 1, you can't have a negative probability; therefore, it's not a probability function.

Answer (c)

c. $f(x) = (x^2+x+1)/25$ for x=0,1,2,3

x	f(x)
0	1/25
1	3/25
2	7/25
3	13/25

24/25

Doesn't sum to 1. Thus, it's not a probability function.

Practice Problem:

- The number of ships to arrive at a harbor on any given day is a random variable represented by *x*. The probability distribution for *x* is:

x	10	11	12	13	14
$P(x)$	.4	.2	.2	.1	.1

Find the probability that on a given day:

a. exactly 14 ships arrive $p(x=14) = .1$

b. At least 12 ships arrive $p(x \geq 12) = (.2 + .1 + .1) = .4$

c. At most 11 ships arrive $p(x \leq 11) = (.4 + .2) = .6$

Practice Problem:

You are lecturing to a group of 1000 students. You ask them to each randomly pick an integer between 1 and 10. Assuming, their picks are truly random:

- What's your best guess for how many students picked the number 9?

  Since $p(x=9) = 1/10$, we'd expect about $1/10^{th}$ of the 1000 students to pick 9. 100 students.

- What percentage of the students would you expect picked a number less than or equal to 6?

  Since $p(x \leq 6) = 1/10 + 1/10 + 1/10 + 1/10 + 1/10 + 1/10 = .6$
  60%

Important discrete distributions in epidemiology…

- Binomial
  - Yes/no outcomes (dead/alive, treated/untreated, smoker/non-smoker, sick/well, etc.)
- Poisson
  - Counts (e.g., how many cases of disease in a given area)

Continuous case

- The probability function that accompanies a continuous random variable is a continuous mathematical function that integrates to 1.
- The probabilities associated with continuous functions are just areas under the curve (integrals!).
- Probabilities are given for a range of values, rather than a particular value (e.g., the probability of getting a math SAT score between 700 and 800 is 2%).

Continuous case

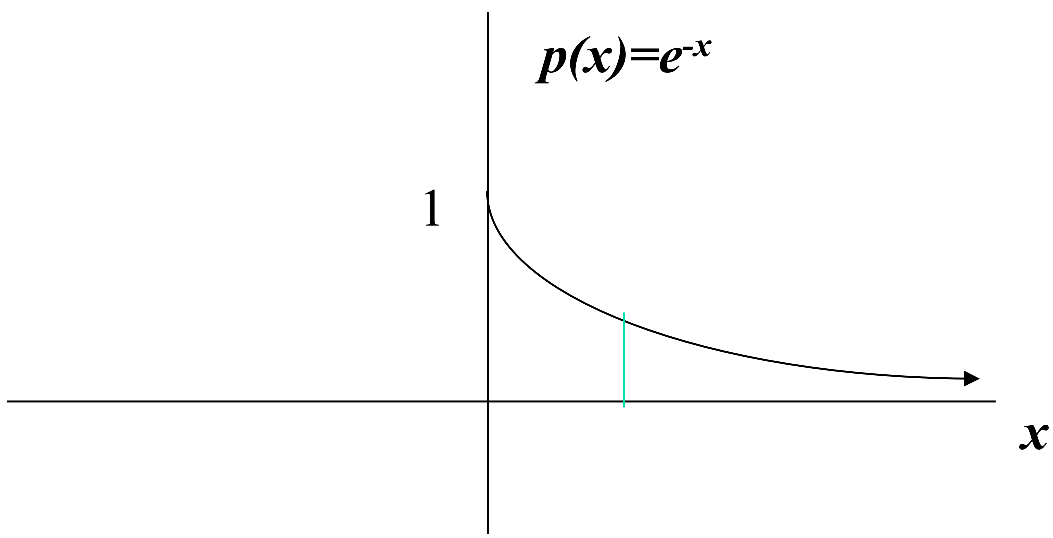
- For example, recall the negative exponential function (in probability, this is called an "exponential distribution"):

$$f(x) = e^{-x}$$

- This function integrates to 1:

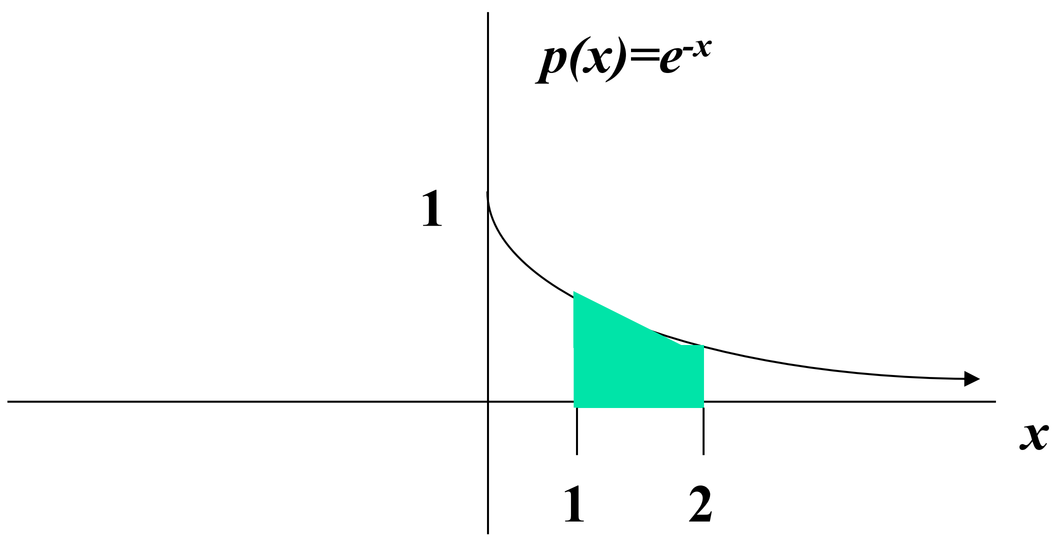
$$\int_{0}^{+\infty} e^{-x} = -e^{-x} \Big|_{0}^{+\infty} = 0 + 1 = 1$$

Continuous case: "probability density function" (pdf)



The probability that x is any exact particular value (such as 1.9976) is 0; we can only assign probabilities to possible ranges of x.

For example, the probability of *x* falling within 1 to 2:



$$P(1 \le x \le 2) = \int_1^2 e^{-x} = -e^{-x} \Big|_1^2 = -e^{-2} - -e^{-1} = -.135 + .368 = .23$$

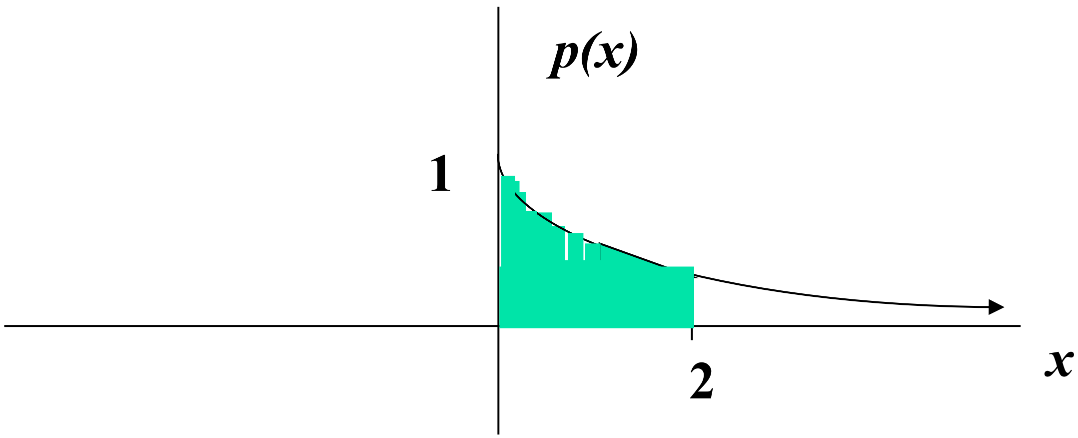
Cumulative distribution function

As in the discrete case, we can specify the "cumulative distribution function" (CDF):

The CDF here = P(x≤A)=

$$\int_0^A e^{-x} = -e^{-x} \Big|_0^A = -e^{-A} - -e^{0} = -e^{-A} + 1 = 1 - e^{-A}$$

Example

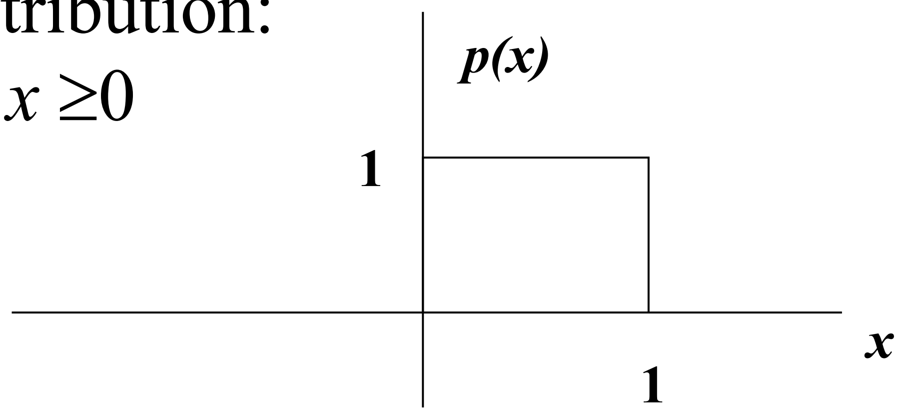


$$P(x \leq 2) = 1 - e^{-2} = 1 - .135 = .865$$

Example 2: Uniform distribution

The uniform distribution: all values are equally likely

The uniform distribution:
$f(x) = 1$, for $1 \geq x \geq 0$

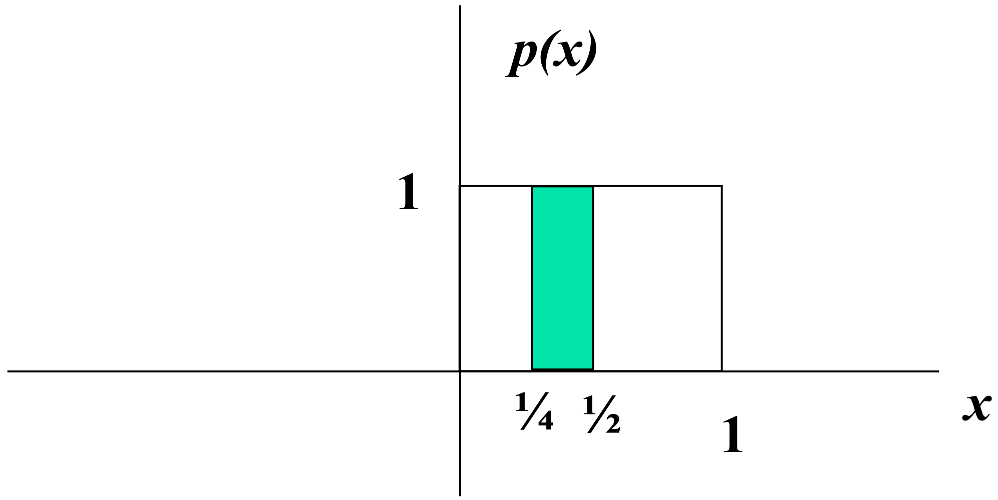


We can see it's a probability distribution because it integrates to 1 (the area under the curve is 1):

$$\int_0^1 1 = x \Big|_0^1 = 1 - 0 = 1$$

Example: Uniform distribution

What's the probability that x is between ¼ and ½?



P(½ ≥*x*≥ ¼)= ¼

Practice Problem

4. Suppose that survival drops off rapidly in the year following diagnosis of a certain type of advanced cancer. Suppose that the length of survival (or time-to-death) is a random variable that approximately follows an exponential distribution with parameter 2 (makes it a steeper drop off):

$$\text{probability function}: p(x=T) = 2e^{-2T}$$

$$[note: \int_0^{+\infty} 2e^{-2x} = -e^{-2x} \Big|_0^{+\infty} = 0 + 1 = 1]$$

What's the probability that a person who is diagnosed with this illness survives a year?

Answer

The probability of dying within 1 year can be calculated using the cumulative distribution function:

Cumulative distribution function is:

$$P(x \leq T) = -e^{-2x} \Big|_0^T = 1 - e^{-2(T)}$$

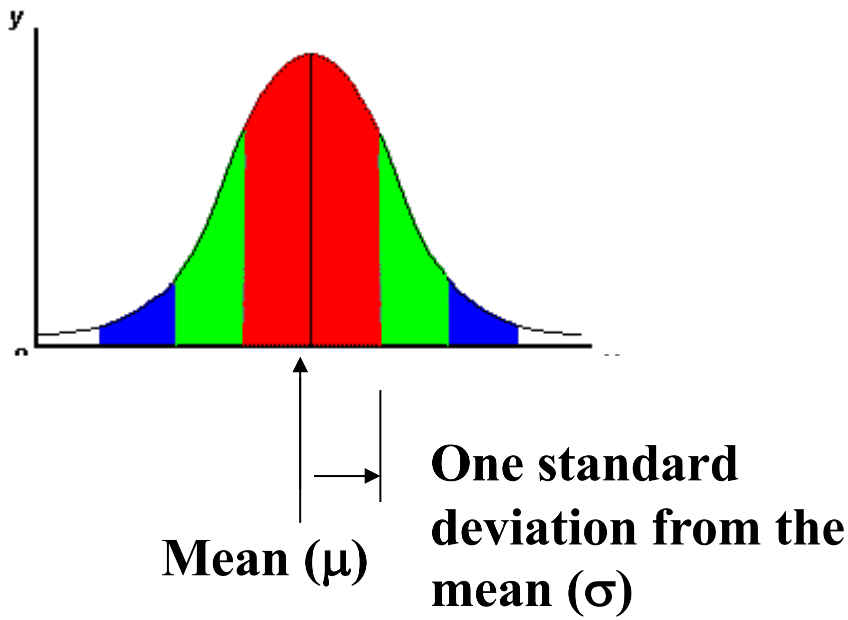
The chance of surviving past 1 year is: *P(x≥1) = 1 – P(x≤1)*

$$1 - (1 - e^{-2(1)}) = .135$$

Expected Value and Variance

- All probability distributions are characterized by an expected value and a variance (standard deviation squared).

For example, bell-curve (normal) distribution:



Expected value, or mean

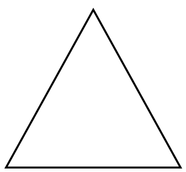
- If we understand the underlying probability function of a certain phenomenon, then we can make informed decisions based on how we expect *x* to behave on-average over the long-run…(so called "frequentist" theory of probability).

- Expected value is just the weighted average or mean (µ) of random variable *x*. Imagine placing the masses $p(x)$ at the points X on a beam; the balance point of the beam is the expected value of *x*.

Example: expected value

- Recall the following probability distribution of ship arrivals:

x	10	11	12	13	14
P(x)	.4	.2	.2	.1	.1



$$\sum_{i=1}^{5} x_i\, p(x) = 10(.4) + 11(.2) + 12(.2) + 13(.1) + 14(.1) = 11.3$$

Expected value, formally

Discrete case:

$$E(X) = \sum_{\text{all x}} x_i p(x_i)$$

Continuous case:

$$E(X) = \int_{\text{all x}} x_i p(x_i) dx$$

Empirical Mean is a special case of Expected Value…

Sample mean, for a sample of n subjects: =

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \sum_{i=1}^{n} x_i \left(\frac{1}{n}\right)$$

The probability (frequency) of each person in the sample is 1/n.

Expected value, formally

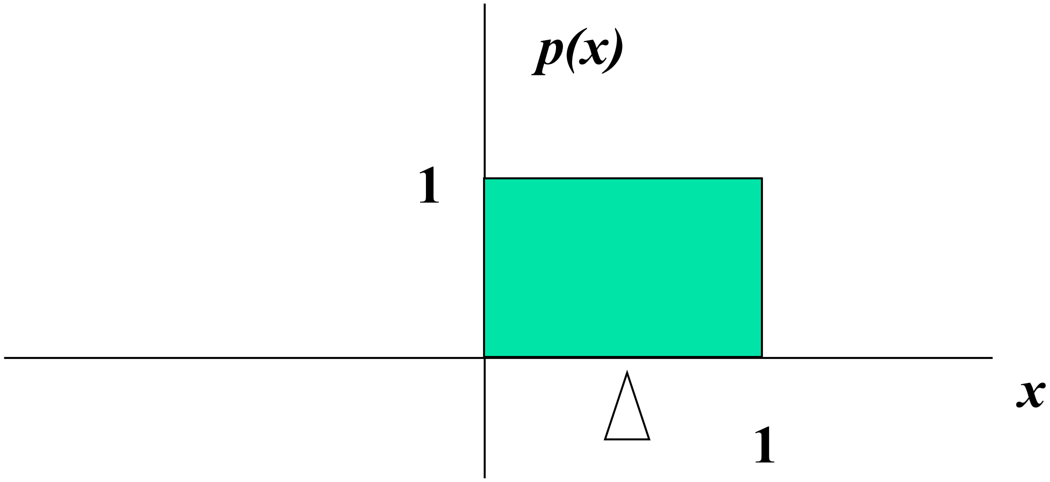
Discrete case:

$$E(X) = \sum_{\text{all x}} x_i p(x_i)$$

Continuous case:

$$E(X) = \int_{\text{all x}} x_i p(x_i) dx$$

Extension to continuous case: uniform distribution



$$E(X) = \int_0^1 x(1)dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2} - 0 = \frac{1}{2}$$

Symbol Interlude

- $E(X) = \mu$
  - these symbols are used interchangeably

Expected Value

- Expected value is an extremely useful concept for good decision-making!

Example: the lottery

- The Lottery (also known as a tax on people who are bad at math…)
- A certain lottery works by picking 6 numbers from 1 to 49. It costs $1.00 to play the lottery, and if you win, you win $2 million after taxes.

- *If you play the lottery once, what are your expected winnings or losses?*

Lottery

Calculate the probability of winning in 1 try:

$$\frac{1}{\binom{49}{6}} = \frac{1}{\frac{49!}{43!6!}} = \frac{1}{13{,}983{,}816} = 7.2 \text{ x } 10^{-8}$$

"49 choose 6"

Out of 49 numbers, this is the number of distinct combinations of 6.

The probability function (note, sums to 1.0):

x$	*p(x)*
-1	.999999928
+ 2 million	*7.2 x 10^-8*

Expected Value

The probability function

$x\$$	$p(x)$
-1	.999999928
+ 2 million	7.2×10^{-8}

Expected Value

$E(X) = P(\text{win})*\$2{,}000{,}000 + P(\text{lose})*-\$1.00$

$= 2.0 \times 10^{6} * 7.2 \times 10^{-8} + .999999928\,(-1) = .144 - .999999928 = -\$.86$

Negative expected value is never good!
You shouldn't play if you expect to lose money!

Expected Value

If you play the lottery every week for 10 years, what are your expected winnings or losses?

520 x (-.86) = -$447.20

Gambling (or how casinos can afford to give so many free drinks…)

A roulette wheel has the numbers 1 through 36, as well as 0 and 00. If you bet \$1 that an odd number comes up, you win or lose \$1 according to whether or not that event occurs. If random variable X denotes your net gain, X=1 with probability 18/38 and X= -1 with probability 20/38.

$E(X) = 1(18/38) – 1 (20/38) = -\$.053$

On average, the casino wins (and the player loses) 5 cents per game.

The casino rakes in even more if the stakes are higher:

$E(X) = 10(18/38) – 10 (20/38) = -\$.53$

If the cost is \$10 per game, the casino wins an average of 53 cents per game. If 10,000 games are played in a night, that's a cool \$5300.

**A few notes about Expected Value as a mathematical operator:

If c= a constant number (i.e., not a variable) and *X* and *Y* are any random variables…

- $E(c) = c$
- $E(cX)=cE(X)$
- $E(c + X)=c + E(X)$
- $E(X+ Y)= E(X) + E(Y)$

E(c) = c

- $E(c) = c$
  - Example: If you cash in soda cans in CA, you always get 5 cents per can.
  - Therefore, there's no randomness. You always expect to (and do) get 5 cents.

$E(cX)=cE(X)$

- $E(cX)=cE(X)$
  - Example: If the casino charges \$10 per game instead of \$1, then the casino expects to make 10 times as much on average from the game (See roulette example above!)

$E(c + X) = c + E(X)$

$E(c + X) = c + E(X)$

Example, if the casino throws in a free drink worth exactly $5.00 every time you play a game, you always expect to (and do) gain an extra $5.00 regardless of the outcome of the game.

$E(X+Y) = E(X) + E(Y)$

$E(X+Y) = E(X) + E(Y)$

Example: If you play the lottery twice, you expect to lose: -$.86 + -$.86.

NOTE: This works even if X and Y are dependent!! Does not require independence!! Proof left for later…

Practice Problem

If a disease is fairly rare and the antibody test is fairly expensive, in a resource-poor region, one strategy is to take half of the serum from each sample and pool it with *n* other halved samples, and test the pooled lot. If the pooled lot is negative, this saves *n-1* tests. If it's positive, then you go back and test each sample individually, requiring *n+1* tests total.

a. Suppose a particular disease has a prevalence of 10% in a third-world population and you have 500 blood samples to screen. If you pool 20 samples at a time (25 lots), how many tests do you expect to have to run (assuming the test is perfect!)?
b. What if you pool only 10 samples at a time?
c. 5 samples at a time?

Answer (a)

a. Suppose a particular disease has a prevalence of 10% in a third-world population and you have 500 blood samples to screen. If you pool 20 samples at a time (25 lots), how many tests do you expect to have to run (assuming the test is perfect!)?

Let X = a random variable that is the number of tests you have to run per lot:

E(X) = P(pooled lot is negative)(1) + P(pooled lot is positive) (21)

$E(X) = (.90)^{20}\ (1) + [1-.90^{20}]\ (21) = 12.2\%\ (1) + 87.8\%\ (21) = 18.56$

E(total number of tests) = 25*18.56 = 464

Answer (b)

b. What if you pool only 10 samples at a time?

$E(X) = (.90)^{10}\,(1) + [1-.90^{10}]\,(11) = 35\%\,(1) + 65\%\,(11) = 7.5$ average per lot

50 lots * 7.5 = 375

Answer (c)

c. 5 samples at a time?

E(X) = $(.90)^5$ (1) + [$1-.90^5$] (6) = 59% (1) + 41% (6) = 3.05 average per lot

100 lots * 3.05 = 305

Practice Problem

If X is a random integer between 1 and 10, what's the expected value of X?

Answer

If X is a random integer between 1 and 10, what's the expected value of X?

$$\mu = E(x) = \sum_{i=1}^{10} i(\frac{1}{10}) = \frac{1}{10}\sum_{i}^{10} i = (.1)\frac{10(10+1)}{2} = 55(.1) = 5.5$$

Expected value isn't everything though…

- Take the show "Deal or No Deal"
- Everyone know the rules?
- Let's say you are down to two cases left. $1 and $400,000. The banker offers you $200,000.
- So, Deal or No Deal?

Deal or No Deal…

- This could really be represented as a probability distribution and a non-random variable:

x$	*p(x)*
+1	.50
+$400,000	.50

x$	*p(x)*
+$200,000	1.0

Expected value doesn't help…

x$	*p(x)*
+1	.50
+$400,000	.50

$$\mu = E(X) = \sum_{\text{all } x} x_i p(x_i) = +1(.50) + 400{,}000(.50) = 200{,}000$$

x$	*p(x)*
+$200,000	1.0

$$\mu = E(X) = 200{,}000$$

How to decide?

Variance!

- **If you take the deal, the variance/standard deviation is 0.**
- **If you don't take the deal, what is average deviation from the mean?**
- **What's your gut guess?**

Variance/standard deviation

"The average (expected) squared distance (or deviation) from the mean"

$$\sigma^2 = Var(x) = E[(x-\mu)^2] = \sum_{\text{all } x} (x_i - \mu)^2 \, p(x_i)$$

***We square because squaring has better properties than absolute value. Take square root to get back linear average distance from the mean (="standard deviation").*

Variance, formally

Discrete case:

$$Var(X) = \sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 \, p(x_i)$$

Continuous case:

$$Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x_i - \mu)^2 \, p(x_i) dx$$

Similarity to empirical variance

The variance of a sample: s^2 =

$$\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{n-1} = \sum_{i=1}^{N}(x_i - \bar{x})^2 \left(\frac{1}{n-1}\right)$$

Division by n-1 reflects the fact that we have lost a "degree of freedom" (piece of information) because we had to estimate the sample mean before we could estimate the sample variance.

Symbol Interlude

- $Var(X) = \sigma^2$
  - these symbols are used interchangeably

Variance: Deal or No Deal

$$\sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 \, p(x_i)$$

$$\sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 \, p(x_i) =$$

$$= (1 - 200{,}000)^2 (.5) + (400{,}000 - 200{,}000)^2 (.5) = 200{,}000^2$$

$$\sigma = \sqrt{200{,}000^2} = 200{,}000$$

Now you examine your personal risk tolerance…

Practice Problem

A roulette wheel has the numbers 1 through 36, as well as 0 and 00. If you bet \$1.00 that an odd number comes up, you win or lose \$1.00 according to whether or not that event occurs. If X denotes your net gain, X=1 with probability 18/38 and X= -1 with probability 20/38.

- We already calculated the mean to be = -\$.053. What's the variance of X?

Answer

$$\sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)$$

$$= (+1 - -.053)^2 (18/38) + (-1 - -.053)^2 (20/38)$$

$$= (1.053)^2 (18/38) + (-1 + .053)^2 (20/38)$$

$$= (1.053)^2 (18/38) + (-.947)^2 (20/38)$$

$$= .997$$

$$\sigma = \sqrt{.997} = .99$$

Standard deviation is $.99. Interpretation: On average, you're either 1 dollar above or 1 dollar below the mean, which is just under zero. Makes sense!

Handy calculation formula!

Handy calculation formula (if you ever need to calculate by hand!):

$$Var(X) = \sum_{\text{all x}} (x_i - \mu)^2 p(x_i) = \sum_{\text{all x}} x_i^{\,2} p(x_i) - (\mu)^2$$

Intervening algebra!

$$= E(x^2) - [E(x)]^2$$

$Var(x) = E(x-\mu)^2 = E(x^2) - [E(x)]^2$ (your calculation formula!)

Proofs (optional!):

$E(x-\mu)^2 = E(x^2 - 2\mu x + \mu^2)$ remember "FOIL"?!

$= E(x^2) - E(2\mu x) + E(\mu^2)$ Use rules of expected value: $E(X+Y) = E(X) + E(Y)$

$= E(x^2) - 2\mu E(x) + \mu^2$ $E(c) = c$

$= E(x^2) - 2\mu\mu + \mu^2$ $E(x) = \mu$

$= E(x^2) - \mu^2$

$= E(x^2) - [E(x)]^2$

OR, equivalently:

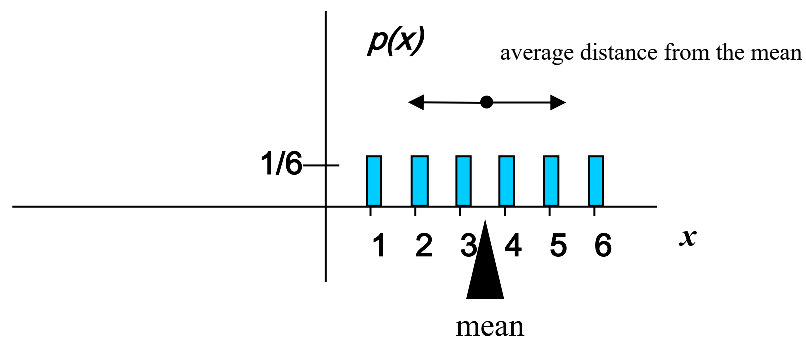
$E(x-\mu)^2 =$

$$\sum_{all x}[(x-\mu)^2]p(x) = \sum_{all x}[(x^2 - 2\mu x + \mu^2]p(x) = \sum_{all x} x^2 p(x) - 2\mu\sum xp(x) + \mu^2 \sum p(x) = E(x^2) - 2\mu E(x) + \mu^2(1) =$$

$$E(x^2) - 2\mu^2 + \mu^2(1) = E(x^2) - \mu^2$$

For example, what's the variance and standard deviation of the roll of a die?

x	*p(x)*
1	*p(x=1)*=1/6
2	*p(x=2)*=1/6
3	*p(x=3)*=1/6
4	*p(x=4)*=1/6
5	*p(x=5)*=1/6
6	*p(x=6)*=1/6
	1.0



$$E(x)=\sum_{\text{all } x} x_i\, p(x_i)=(1)(\frac{1}{6})+2(\frac{1}{6})+3(\frac{1}{6})+4(\frac{1}{6})+5(\frac{1}{6})+6(\frac{1}{6})=\frac{21}{6}=3.5$$

$$E(x^2)=\sum_{\text{all } x} x_i^{\,2}\, p(x_i)=(1)(\frac{1}{6})+4(\frac{1}{6})+9(\frac{1}{6})+16(\frac{1}{6})+25(\frac{1}{6})+36(\frac{1}{6})=15.17$$

$$\sigma_x^2=Var(x)=E(x^2)-[E(x)]^2=15.17-3.5^2=2.92$$

$$\sigma_x=\sqrt{2.92}=1.71$$

**A few notes about Variance as a mathematical operator:

If c= a constant number (i.e., not a variable) and X and Y are random variables, then

- Var(c) = 0
- Var $(c+X)$= Var(X)
- Var(cX)= c^2Var(X)
- Var($X+Y$)= Var(X) + Var(Y) ***ONLY IF X and Y are independent!!!!***
- {Var($X+Y$)= Var(X) + Var(Y)+2Cov(X,Y) ***IF X and Y are not independent}***

Var(c) = 0

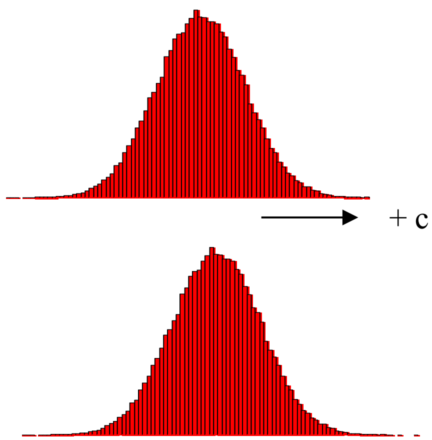
$\text{Var}(c) = 0$

Constants don't vary!

Var (c+X)= Var(X)

Var (c+X)= Var(X)

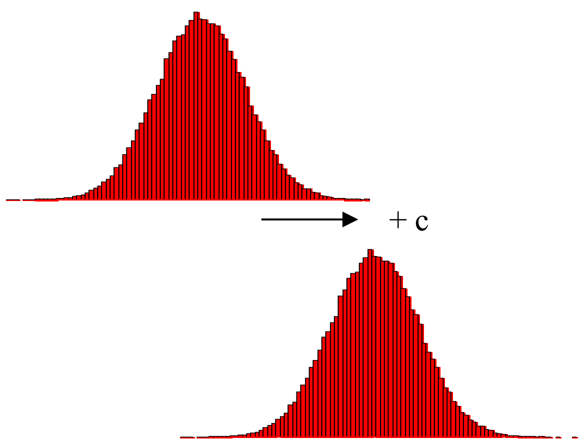
Adding a constant to every instance of a random variable doesn't change the variability. It just shifts the whole distribution by c. If everybody grew 5 inches suddenly, the variability in the population would still be the same.



Var (c+X)= Var(X)

Var (c+X)= Var(X)

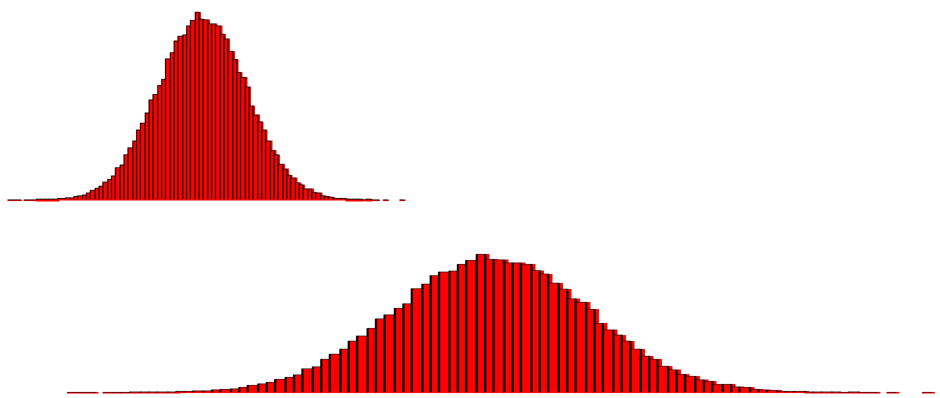
Adding a constant to every instance of a random variable doesn't change the variability. It just shifts the whole distribution by c. If everybody grew 5 inches suddenly, the variability in the population would still be the same.



$Var(cX) = c^2 Var(X)$

$Var(cX) = c^2 Var(X)$

Multiplying each instance of the random variable by c makes it c-times as wide of a distribution, which corresponds to c^2 as much variance (deviation squared). For example, if everyone suddenly became twice as tall, there'd be twice the deviation and 4 times the variance in heights in the population.



$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

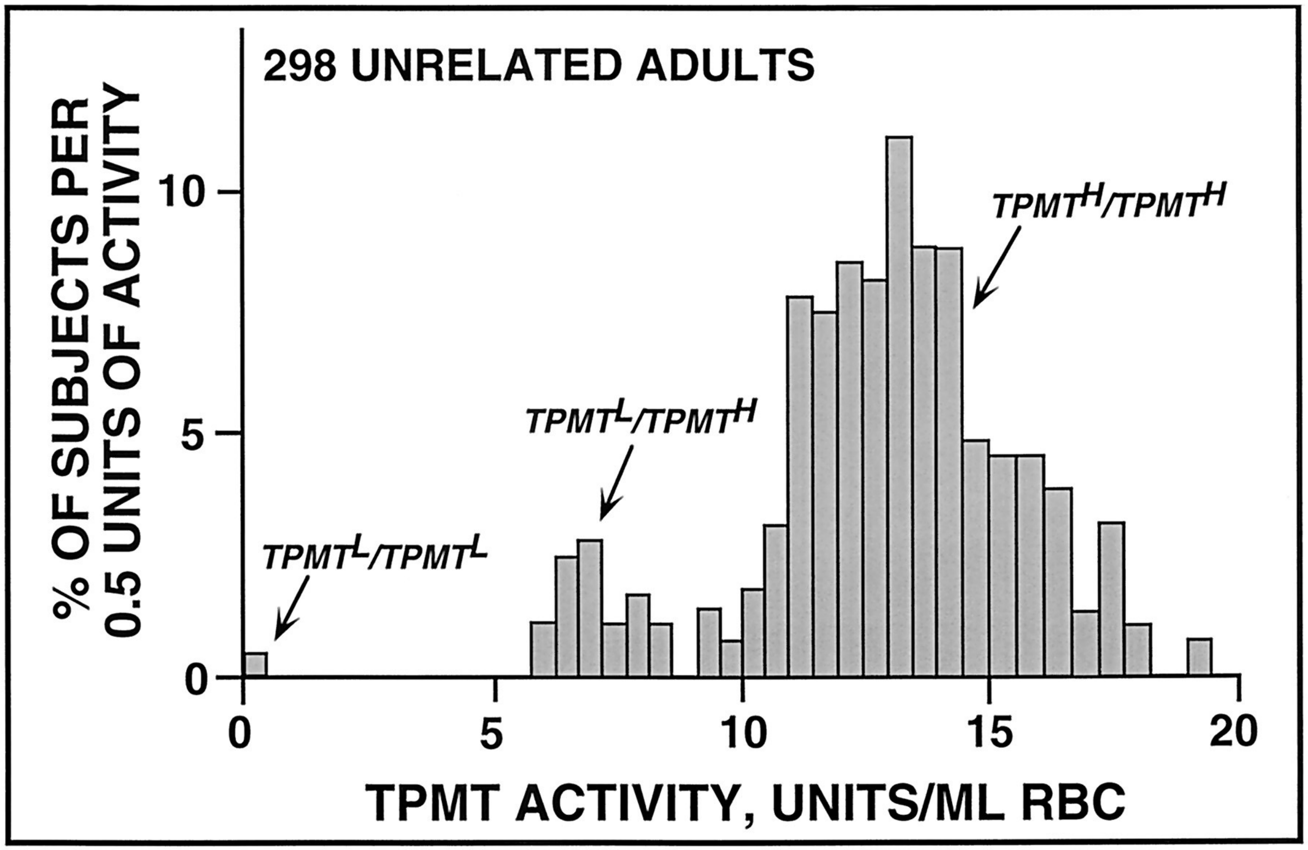
$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ ***ONLY IF X and Y are independent!!!!!!!!***

With two random variables, you have more opportunity for variation, unless they vary together (are dependent, or have covariance): $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$

Example of Var(X+ Y)= Var(X) + Var(Y): TPMT

- TPMT metabolizes the drugs 6-mercaptopurine, azathioprine, and 6-thioguanine (chemotherapy drugs)
- People with TPMT^{-}/ TPMT^{+} have reduced levels of activity (10% prevalence)
- People with TPMT^{-}/ TPMT^{-} have no TPMT activity (prevalence 0.3%).
- They cannot metabolize 6-mercaptopurine, azathioprine, and 6-thioguanine, and risk bone marrow toxicity if given these drugs.

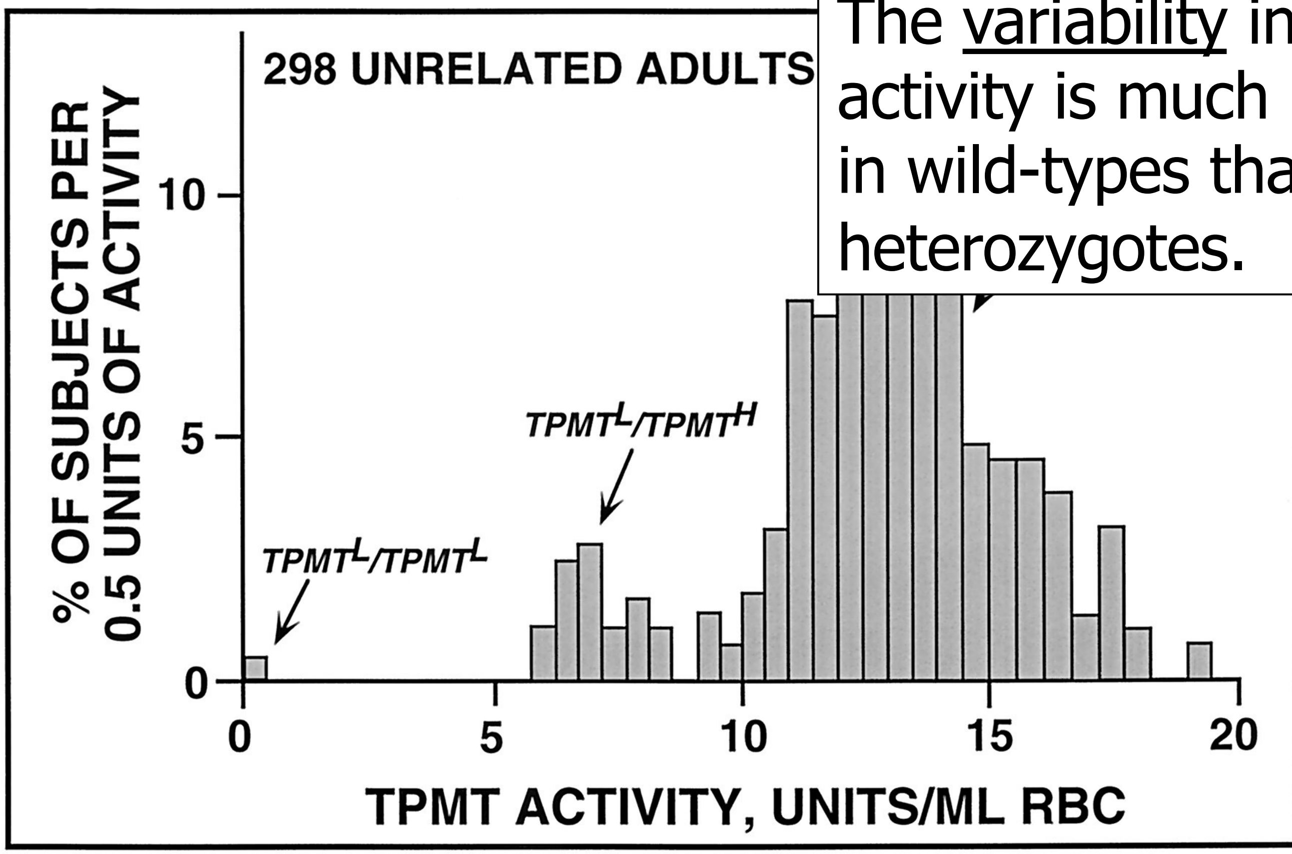
TPMT activity by genotype



Weinshilboum R. Drug Metab Dispos. 2001 Apr;29(4 Pt 2):601-5

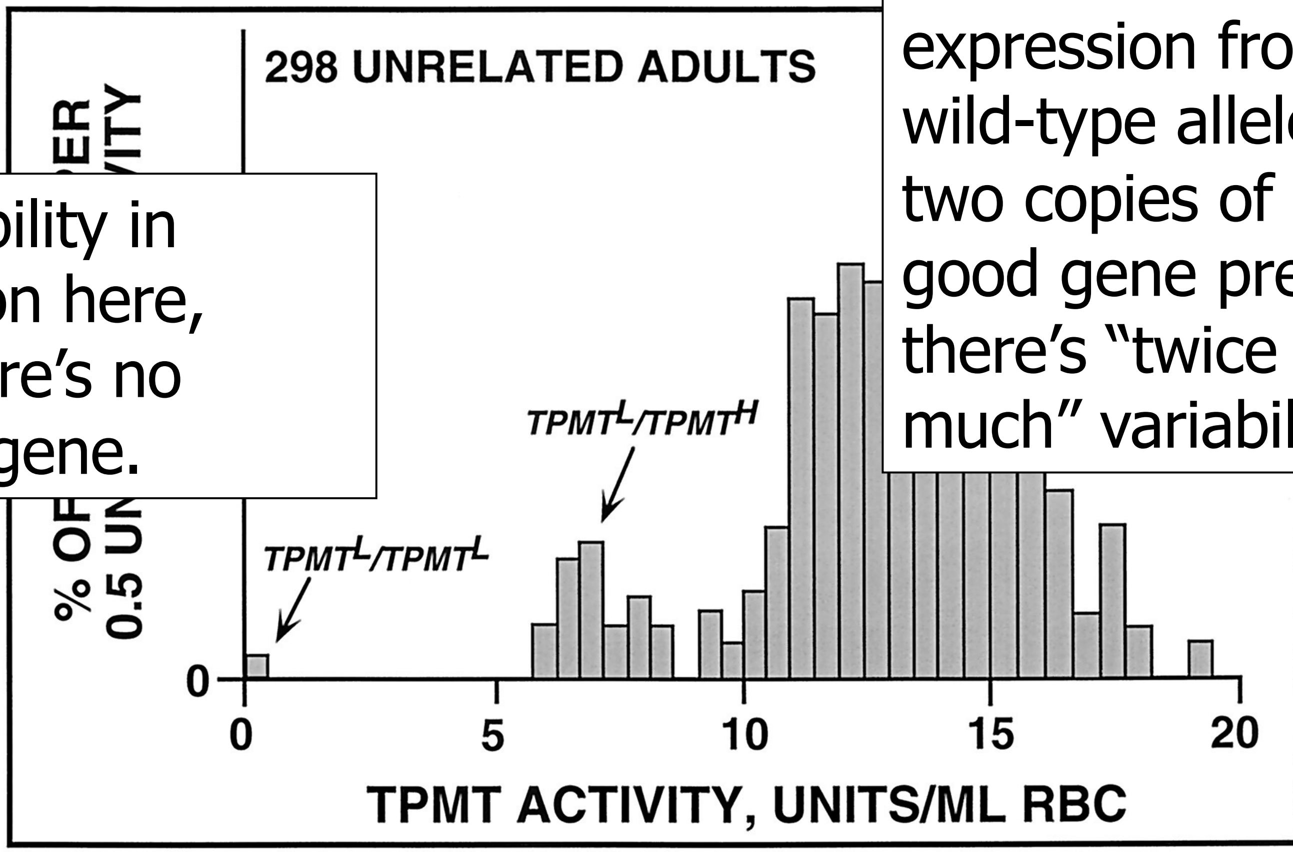
TPMT activity by genotype

The <u>variability</u> in TPMT activity is much higher in wild-types than heterozygotes.



Weinshilboum R. Drug Metab Dispos. 2001 Apr;29(4 Pt 2):601-5

TPMT activity by genotype



No variability in expression here, since there's no working gene.

There is variability in expression from each wild-type allele. With two copies of the good gene present, there's "twice as much" variability.

Weinshilboum R. Drug Metab Dispos. 2001 Apr;29(4 Pt 2):601-5

Practice Problem

Find the variance and standard deviation for the number of ships to arrive at the harbor (recall that the mean is 11.3).

x	10	11	12	13	14
$P(x)$	.4	.2	.2	.1	.1

Answer: variance and std dev

x^2	100	121	144	169	196
$P(x)$	.4	.2	.2	.1	.1

$$E(x^2) = \sum_{i=1}^{5} x_i^{\,2}\, p(x_i) = (100)(.4) + (121)(.2) + 144(.2) + 169(.1) + 196(.1) = 129.5$$

$$Var(x) = E(x^2) - [E(x)]^2 = 129.5 - 11.3^2 = 1.81$$

$$stddev(x) = \sqrt{1.81} = 1.35$$

Interpretation: On an average day, we expect 11.3 ships to arrive in the harbor, plus or minus 1.35. This gives you a feel for what would be considered a usual day!

Practice Problem

You toss a coin 100 times. What's the expected number of heads? What's the variance of the number of heads?

Answer: expected value

Intuitively, we'd probably all agree that we expect around 50 heads, right?

Another way to show this→

Think of tossing 1 coin. E(X=number of heads) = (1) P(heads) + (0)P(tails)

$\therefore$E(X=number of heads) = 1(.5) + 0 = .5

If we do this 100 times, we're looking for the sum of 100 tosses, where we assign 1 for a heads and 0 for a tails. (these are 100 "independent, identically distributed (i.i.d)" events)

$E(X_1 + X_2 + X_3 + X_4 + X_5 \ldots + X_{100}) = E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5) \ldots + E(X_{100}) =$

$100\, E(X_1) = 50$

Answer: variance

What's the variability, though? More tricky. But, again, we could do this for 1 coin and then use our rules of variance.

Think of tossing 1 coin.

$E(X^2 = \text{number of heads squared}) = 1^2\, P(\text{heads}) + 0^2\, P(\text{tails})$

$\therefore E(X^2) = 1(.5) + 0 = .5$

$Var(X) = .5 - .5^2 = .5 - .25 = .25$

Then, using our rule: $Var(X+Y) = Var(X) + Var(Y)$ (coin tosses are independent!)

$Var(X_1 + X_2 + X_3 + X_4 + X_5 \ldots.. + X_{100}) = Var(X_1) + Var(X_2) + Var(X_3) + Var(X_4) + Var(X_5) \ldots.. + Var(X_{100}) =$

$100\, Var(X_1) = 100\,(.25) = 25$

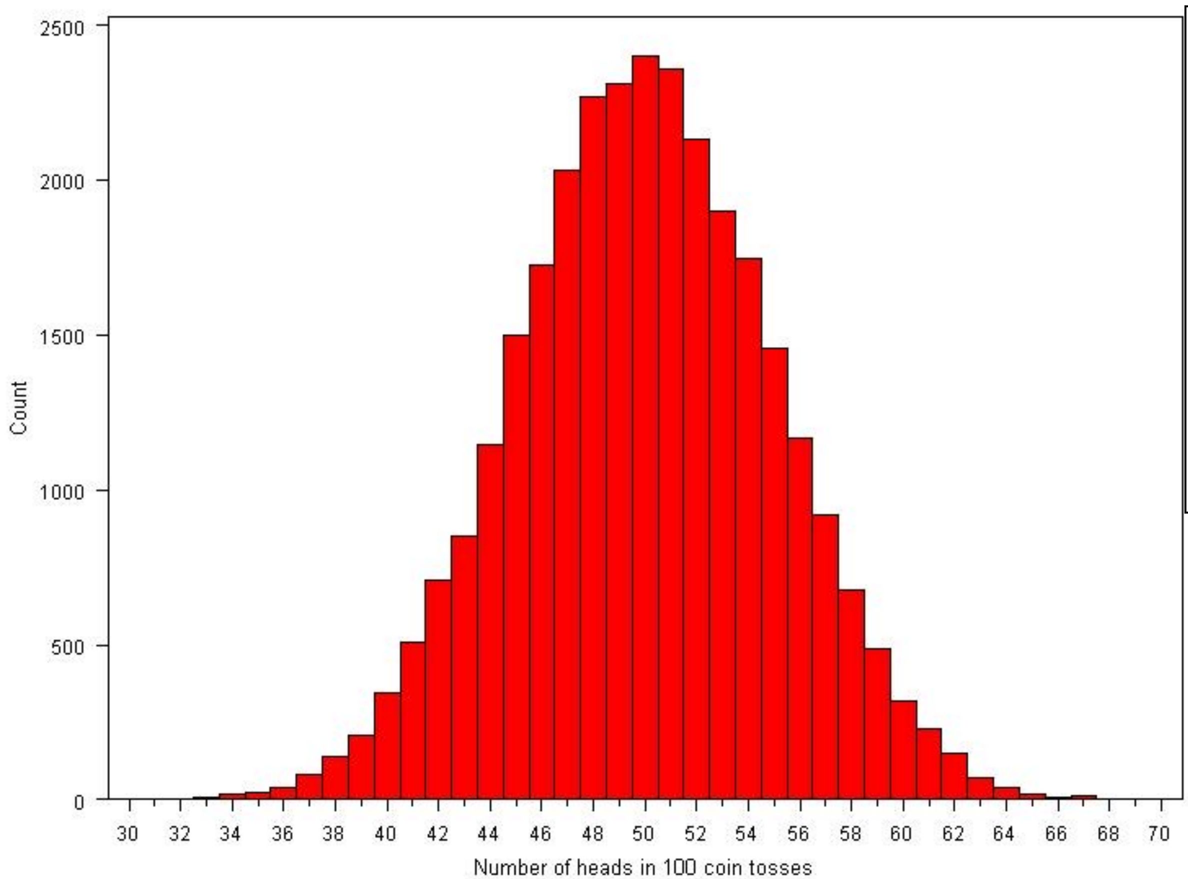
$SD(X) = 5$

Interpretation: When we toss a coin 100 times, we expect to get 50 heads plus or minus 5.

Or use computer simulation…

- Flip coins virtually!
  - Flip a virtual coin 100 times; count the number of heads.
  - Repeat this over and over again a large number of times (we’ll try 30,000 repeats!)
  - Plot the 30,000 results.

Coin tosses…



Mean = 50

Std. dev = 5

Follows a normal distribution

∴95% of the time, we get between 40 and 60 heads…

Covariance: joint probability

- The covariance measures the strength of the linear relationship between two variables
- The covariance: $E[(x-\mu_x)(y-\mu_y)]$

$$\boxed{\sigma_{xy} = \sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)\, P(x_i, y_i)}$$

The Sample Covariance

- The sample covariance:

$$\boxed{\operatorname{cov}(x, y) = \frac{\sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y})}{n-1}}$$

Interpreting Covariance

- **Covariance** between two random variables:

cov(X,Y) > 0 $\longrightarrow$ X and Y are positively correlated

cov(X,Y) < 0 $\longrightarrow$ X and Y are inversely correlated

cov(X,Y) = 0 $\longrightarrow$ X and Y are independent

Probability Distributions

Anand Kumar M

Probability distribution

- Discrete probability distribution
  - Discrete uniform probability distribution
  - Binomial distribution
  - Multinomial distribution
  - Hypergeometric distribution
  - Poisson distribution
- Continuous probability distribution
  - Continuous uniform probability distribution
  - Normal distribution
  - Standard normal distribution
  - Chi-squared distribution
  - Gamma distribution
  - Exponential distribution
  - Lognormal distribution
  - Weibull distribution

Binomial Distribution

- In many situations, an outcome has only two outcomes: success and failure.
  - Such outcome is called dichotomous outcome.
- An experiment when consists of repeated trials, each with dichotomous outcome is called Bernoulli process. Each trial in it is called a Bernoulli trial.

Example : Firing bullets to hit a target.

- Suppose, in a Bernoulli process, we define a random variable $X \equiv$ the number of successes in trials.
- Such a random variable obeys the binomial probability distribution, if the experiment satisfies the following conditions:

  1) The experiment consists of n trials.

  2) Each trial results in one of two mutually exclusive outcomes, one labelled a *"success"* and the other a *"failure"*.

  3) The probability of a success on a single trial is equal to $\boldsymbol{p}$. The value of p remains constant throughout the experiment.

  4) The trials are independent.

Binomial Probability Distribution

- A fixed number of observations (trials), n
  - e.g., 15 tosses of a coin; 20 patients; 1000 people surveyed
- A binary random variable
  - e.g., head or tail in each toss of a coin; defective or not defective light bulb
  - Generally called "success" and "failure"
  - Probability of success is p, probability of failure is 1 – p
- Constant probability for each observation
  - e.g., Probability of getting a tail is the same each time we toss the coin

Defining Binomial Distribution

Binomial distribution

The function for computing the probability for the binomial probability distribution is given by

$$f(x) = \frac{n!}{x!\,(n-x)!}\, p^x (1-p)^{n-x}$$

for x = 0, 1, 2, …., n

Here, $f(x) = P(X = x)$, where X denotes "the number of success" and $X = x$ denotes the number of success in x trials.

Binomial example

Take the example of 5 coin tosses. What's the probability that you flip exactly 3 heads in 5 coin tosses?

Binomial distribution

Solution:

One way to get exactly 3 heads: HHHTT

What's the probability of this <u>exact</u> arrangement?

P(heads)xP(heads) xP(heads)xP(tails)xP(tails) =$(1/2)^3 x (1/2)^2$

Another way to get exactly 3 heads: THHHT

Probability of this exact outcome = $(1/2)^1 x (1/2)^3 x (1/2)^1 = (1/2)^3 x (1/2)^2$

Binomial distribution

In fact, $(1/2)^3 \, x \, (1/2)^2$ is the probability of each unique outcome that has exactly 3 heads and 2 tails.

So, the overall probability of 3 heads and 2 tails is:

$(1/2)^3 \, x \, (1/2)^2 + (1/2)^3 \, x \, (1/2)^2 + (1/2)^3 \, x \, (1/2)^2 + \ldots..$ for as many unique arrangements as there are—but how many are there??

$\binom{5}{3}$ ways to arrange 3 heads in 5 trials

$_5C_3 = 5!/3!2! = 10$

Outcome	Probability
THHHT	$(1/2)^3 \, x \, (1/2)^2$
HHHTT	$(1/2)^3 \, x \, (1/2)^2$
TTHHH	$(1/2)^3 \, x \, (1/2)^2$
HTTHH	$(1/2)^3 \, x \, (1/2)^2$
HHTTH	$(1/2)^3 \, x \, (1/2)^2$
THTHH	$(1/2)^3 \, x \, (1/2)^2$
HTHTH	$(1/2)^3 \, x \, (1/2)^2$
HHTHT	$(1/2)^3 \, x \, (1/2)^2$
THHTH	$(1/2)^3 \, x \, (1/2)^2$
HTHHT	$(1/2)^3 \, x \, (1/2)^2$

10 arrangements $x \, (1/2)^3 \, x \, (1/2)^2$

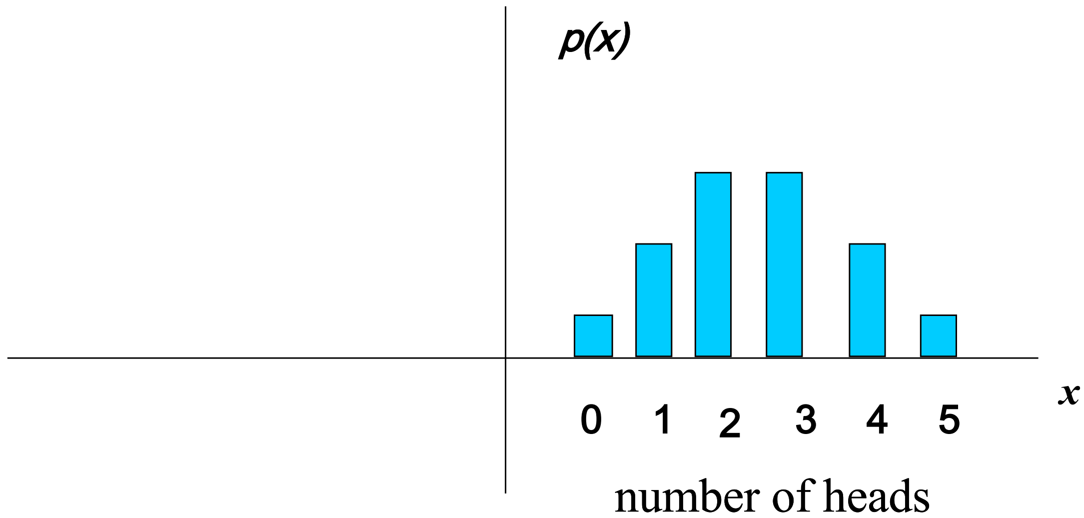
The probability of each unique outcome (note: they are all equal)

$\therefore$ **P(3 heads and 2 tails) =** $\boxed{\binom{5}{3}}$ ***x P(heads)³ x P(tails)² =***

10 x (½)⁵= 31.25%

Binomial distribution function:

X= the number of heads tossed in 5 coin tosses



***Example**: A component has a 20% chance of being a dud. If five are selected from a large batch, what is the probability that more than one is a dud?*

Answer:

Let X = number of duds in selection of 5

Bernoulli trial: dud or not dud, $X \sim B(5,0.2)$

P(More than one dud) $= P(X > 1) = 1 - P(X \le 1) = 1 - \mathrm{P}(\mathrm{X} = 0) - \mathrm{P}(\mathrm{X} = 1)$

$= 1 - C_0^5 0.2^0 (1 - 0.2)^5 - C_1^5 0.2^1 (1 - 0.2)^4$

$= 1 - 1 \times 1 \times 0.8^5 - 5 \times 0.2 \times 0.8^4$

$= 1 - 0.32768 - 0.4096 \approx 0.263.$

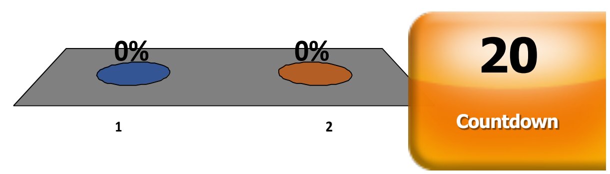
Binomial or not?

A mixed box of 10 screws contains 5 that are galvanized and 5 that are non-galvanized.

Three screws are picked at random without replacement. I want galvanized screws, so consider picking a galvanized screw to be a success.

Does the number of successes have a Binomial distribution?

1. Yes
2. No



Binomial or not?

A mixed box of 10 screws contains 5 that are galvanized and 5 that are non-galvanized.

Three screws are picked at random without replacement. I want galvanized screws, so consider picking a galvanized screw to be a success.

Does the number of successes have a Binomial distribution?

No, the picks are not independent

Independent events have $P(A|B) = P(A)$ but

$$P(\text{second galvanized}|\text{first galvanized}) \neq P(\text{second galvanized})$$

- If the first is galvanized, then only $\frac{4}{9}$ of the remaining screws are galvanized, which is $\neq \frac{1}{2}$

Note: If the box were much larger then consecutive picks would be *nearly* independent - Binomial then a good approximation.

e.g. if box of 1000 screws, with 500 galvanized

$$P(\text{second galvanized}|\text{first galvanized}) = \frac{499}{999} = 0.4995 \approx \frac{1}{2}$$

Mean and variance of a binomial distribution

If $X \sim B(n, p)$, $\mu = E(X) = np$ σ^2 = var(X) = $np(1-p)$

Derivation

Suppose first that we have a single Bernoulli trial. Assign the value 1 to success, and 0 to failure, the first occurring with probability p and the second having probability $1 - p$.

The expected value for one trial is

$$\mu_1 = \sum_k kP(X = k) = 1\times p + 0\times(1-p) = p$$

Since the n trials are independent, the total expected value is just the sum of the expected values for each trial, hence $\sum_{i=1}^{n} \mu_1 = np$.

The variance in a single trial is:

$$\sigma_1^2 = \langle X^2 \rangle - \langle X \rangle^2 = 1^2\times p + 0^2\times(1-p) - p^2 = p(1-p)$$

Hence the variance for the sum of n independent trials is, by the rule for summing the variances of independent variables, $\sigma^2 = \sum_{i=1}^{n} \sigma_1^2 = np(1-p)$.

Polling

In the French population about 20% of people prefer Le Pen to other candidates (inc. Hollande and Sarkozy).

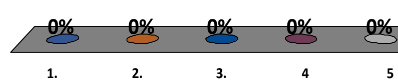
An opinion poll asks 1000 people if they will vote for Le Pen (*YES*) or not (*NO*). The expected number of Le Pen voters (*YES*s) in the poll is therefore $\mu = np = 200$

What is the standard deviation (approximately)?

Mean and variance of a binomial distribution

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

1. 6.3
2. 12.6
3. 25.3
4. 120
5. 160



30
Countdown

Binomial Distribution Summary

Discrete random variable $X \sim \mathrm{B}(n, p)$ if X is the number of successes in n independent Bernoulli trials each with probability p of success

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

Mean and variance

$$\mu = np$$

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

Binomial distribution, generally

Note the general pattern emerging → if you have only two possible outcomes (call them 1/0 or yes/no or success/failure) in n independent trials, then the probability of exactly X "successes"=

n = number of trials

$$\boxed{\binom{n}{X} p^X (1-p)^{n-X}}$$

$1\text{-}p$ = probability of failure

X = # successes out of n trials

p = probability of success

Q

<u>Polling</u>

In the French population about 20% of people prefer Le Pen to other candidates (inc. Hollande and Sarkozy).

An opinion poll asks 1000 people if they will vote for Le Pen (*YES*) or not (*NO*). The expected number of Le Pen voters (*YES*s) in the poll is therefore $\mu = np = 200$

What is the standard deviation (approximately)?

A

The number of *YES* votes has a distribution $X \sim B(1000{,}0.2)$

The variance is therefore $\sigma^2 = np(1-p) = 1000 \times 0.2 \times (1 - 0.2)$

$$= 1000 \times 0.2 \times 0.8 = 160$$

$$\Rightarrow \sigma = \sqrt{\sigma^2} = \sqrt{160} \approx 12.6$$

$\Rightarrow$ expect 200 ± 12.6 in poll to say Le Pen

i.e. the fractional error of $\frac{12.6}{1000} \approx 1.3\%$

Note: quoted errors in polls are not usually the standard deviation – see later

Poisson distribution

If events happen independently of each other, with average number of events in some fixed interval λ, then the distribution of the number of events k in that interval is Poisson.

A random variable X has the Poisson distribution with parameter $\lambda(>0)$ if

$$P(X=k)=\frac{e^{-\lambda}\lambda^{k}}{k!} \quad (k=0,1,2,\ldots)$$

If you are interested in the derivation, see the notes.

Examples of possible Poisson distributions

1) Number of messages arriving at a telecommunications system in a day

2) Number of flaws in a metre of fibre optic cable

3) Number of radio-active particles detected in a given time

4) Number of photons arriving at a CCD pixel in some exposure time
   (e.g. astronomy observations)

Sum of Poisson variables

If X is Poisson with average number λ_{X} and Y is Poisson with average number λ_Y

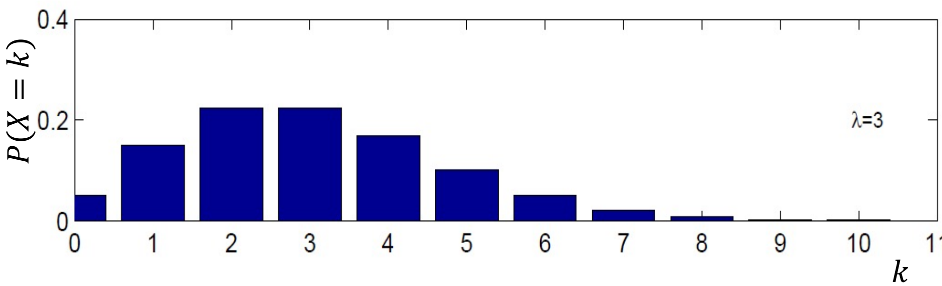
Then $X + Y$ is Poisson with average number $\lambda_X + \lambda_Y$

The probability of events per unit time does not have to be constant for the total number of events to be Poisson – can split up the total into a sum of the number of events in smaller intervals.

Example: On average lightning kills three people each year in the UK, $\lambda = 3$. What is the probability that only one person is killed this year?

Answer:

Assuming these are independent random events, the number of people killed in a given year therefore has a Poisson distribution:



Let the random variable X be the number of people killed in a year.

Poisson distribution $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$ with $\lambda = 3$

$$\Rightarrow P(X = 1) = \frac{e^{-3}3^1}{1!} \approx 0.15$$

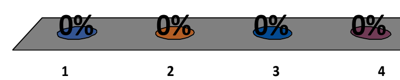
Poisson distribution

Suppose that trucks arrive at a receiving dock with an average arrival rate of 3 per hour. What is the probability exactly 5 trucks will arrive in a two-hour period?

Question from Derek Bruff

1. $\frac{e^{-3}3^5}{5!}$
2. $\frac{e^{-3}3^{2.5}}{2.5!}$
3. $\frac{e^{-5}5^6}{6!}$
4. ✔ $\frac{e^{-6}6^5}{5!}$

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



20

Countdown

Q

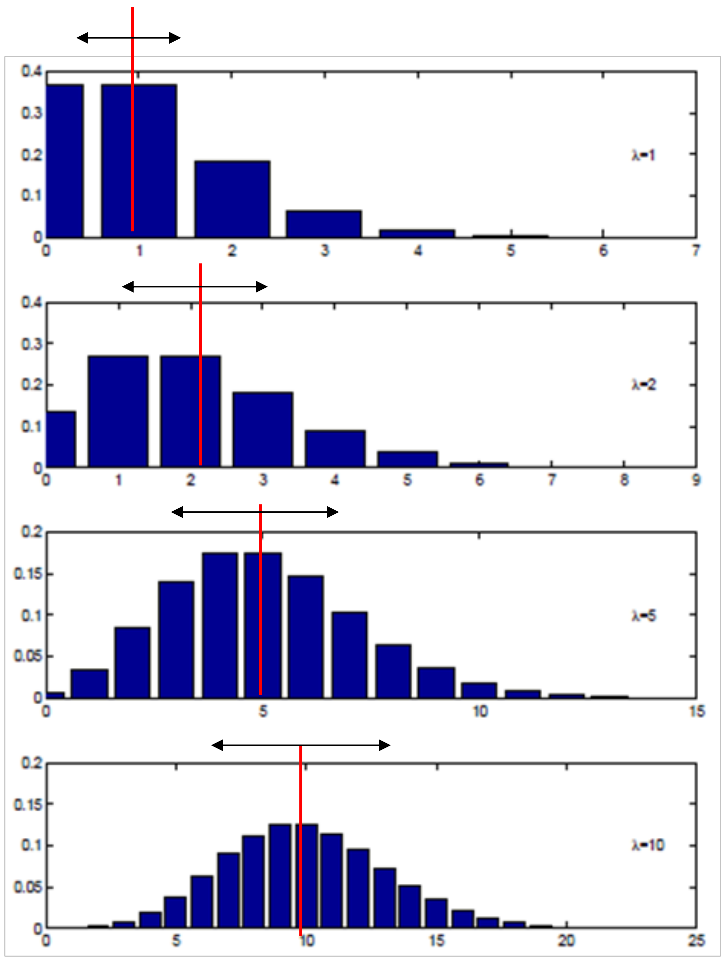
Poisson distribution

Suppose that trucks arrive at a receiving dock with an average arrival rate of 3 per hour. What is the probability exactly 5 trucks will arrive in a two-hour period?

A

In two hours mean number is $\lambda = 2 \times 3 = 6$.

$$P(X = k = 5) = \frac{e^{-\lambda}\lambda^{k}}{k!} = \frac{e^{-6}6^{5}}{5!}$$



Mean and variance

If X ~ Poisson with mean λ, then

$$\mu = E(X) = \lambda$$

$$\sigma^2 = \text{var}(X) = \lambda$$

Example: ***Telecommunications***

Messages arrive at a switching centre at random and at an average rate of 1.2 per second.

(a) Find the probability of 5 messages arriving in a 2-sec interval.

(b) For how long can the operation of the centre be interrupted, if the probability of losing one or more messages is to be no more than 0.05?

Answer:

Times of arrivals form a Poisson process, rate $\nu = 1.2/\text{sec}$.

(a) Let Y = number of messages arriving in a 2-sec interval.

Then Y ~ Poisson, mean number $\lambda = \nu t = 1.2 \times 2 = 2.4$

$$P(Y = k = 5) = \frac{e^{-\lambda}\lambda^k}{k!} = \frac{e^{-2.4}2.4^5}{5!} = 0.060$$

Question: (b) For how long can the operation of the centre be interrupted, if the probability of losing one or more messages is to be no more than 0.05?

Answer:

(b) Let the required time = *t* seconds. Average rate of arrival is 1.2/second.

Let k = number of messages in *t* seconds, so that

$k \sim$ Poisson, with $\lambda = 1.2 \times t = 1.2t$

Want P(At least one message) = $P(k \geq 1) = 1 - P(k = 0) \leq 0.05$

$$P(k = 0) = \frac{e^{-\lambda}\lambda^k}{k!} = \frac{e^{-1.2t}(1.2t)^0}{0!} = e^{-1.2t}$$

$$\Rightarrow 1 - e^{-1.2t} \leq 0.05$$

$$\Rightarrow \quad -e^{-1.2t} \leq 0.05 - 1$$

$$\Rightarrow e^{-1.2t} \geq 0.95$$

$$\Rightarrow -1.2t \geq \ln(0.95) = -0.05129$$

$$\boxed{\Rightarrow t \leq 0.043 \text{ seconds}}$$

1. **Binomial distribution**

The binomial probability distribution is characterized with p (the probability of success) and n (is the number of trials). Then

$$\mu = n.p$$

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

3. **Poisson Distribution**

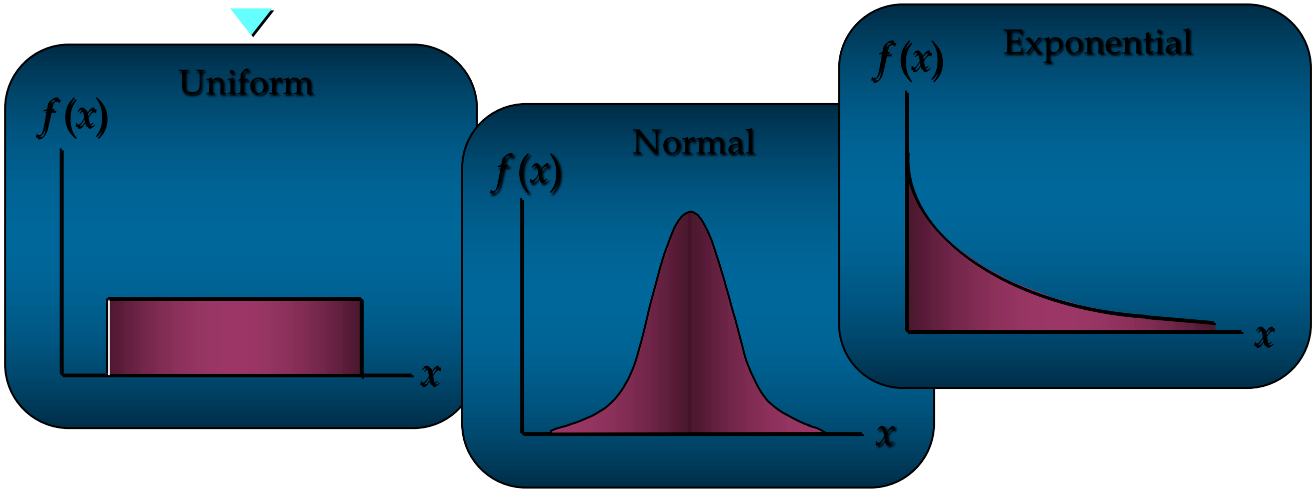
The Poisson distribution is characterized with λt where $\lambda = \textit{the mean of outcomes}$ and $t = \textit{time interval}.$

$$\mu = \lambda t$$

$$\sigma^2 = \lambda t$$

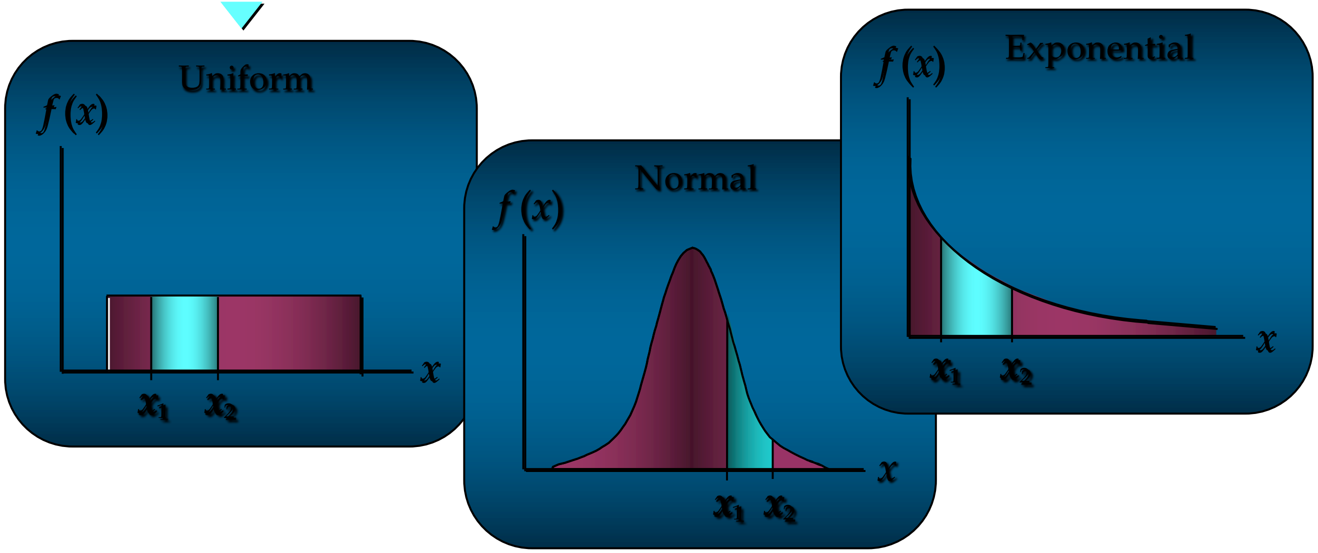
Uniform Probability Distribution

- Uniform Probability Distribution
- Normal Probability Distribution
- Exponential Probability Distribution



Continuous Probability Distributions

- The probability of the random variable assuming a value within some given interval from x_1 to x_2 is defined to be the area under the graph of the probability density function between x_1 and x_2.



Uniform Probability Distribution

- A random variable is uniformly distributed whenever the probability is proportional to the interval's length.
- The uniform probability density function is:

$$f(x) = 1/(b - a) \quad \text{for } a \leq x \leq b$$
$$= 0 \quad \text{elsewhere}$$

where: a = smallest value the variable can assume

b = largest value the variable can assume

Uniform Probability Distribution

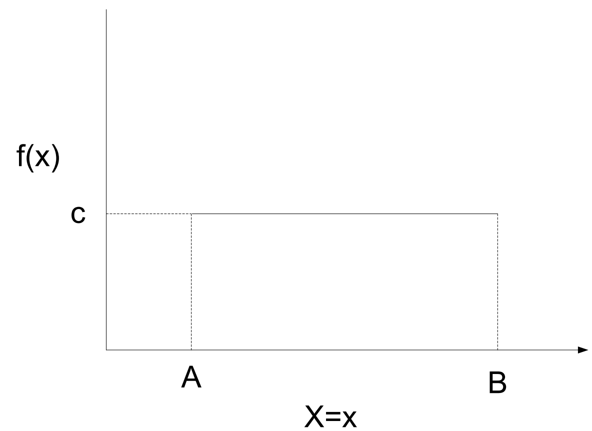
- Expected Value of x

$$E(x) = (a + b)/2$$

- Variance of x

$$Var(x) = (b - a)^2/12$$

Continuous Uniform Distribution



Note:

a) $\int_{\infty}^{-\infty} f(x)dx = \frac{1}{B-A} \times (B - A) = 1$

b) $P(c < x < d) = \frac{d-c}{B-A}$ where both c and d are in the interval (A,B)

c) $\mu = \frac{A+B}{2}$

d) $\sigma^2 = \frac{(B-A)^2}{12}$

Uniform Probability Distribution

- Example: Slater's Buffet

  - Slater customers are charged for the amount of salad they take. Sampling suggests that the amount of salad taken is uniformly distributed between 5 ounces and 15 ounces.

Uniform Probability Distribution

- Uniform Probability Density Function

$$f(x) = 1/10 \quad \text{for } 5 \leq x \leq 15$$
$$= 0 \quad \text{elsewhere}$$

where:

x = salad plate filling weight

Uniform Probability Distribution

- Expected Value of x

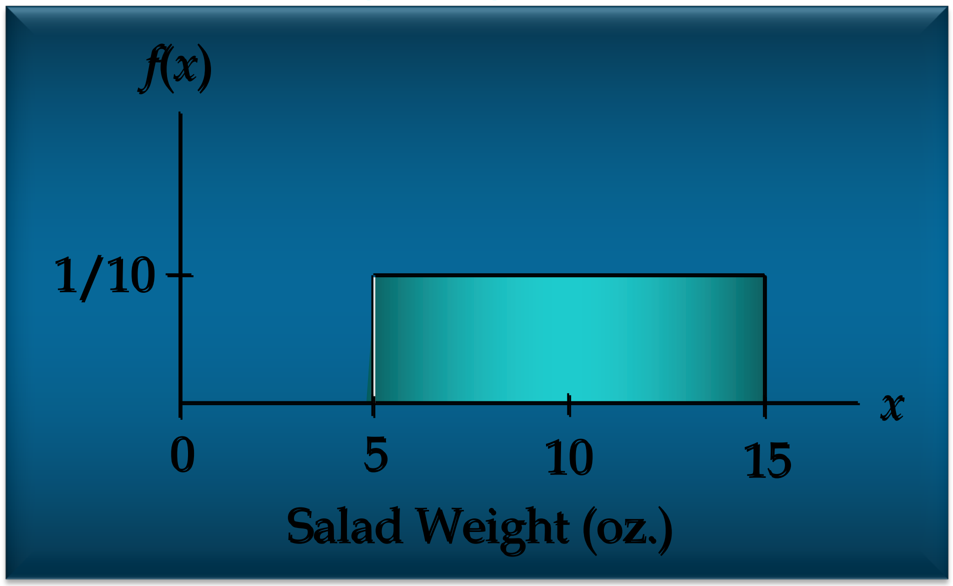
$$E(x) = (a + b)/2 = (5 + 15)/2 = 10$$

- Variance of x

$$Var(x) = (b - a)^2/12 = (15 - 5)^2/12 = 8.33$$

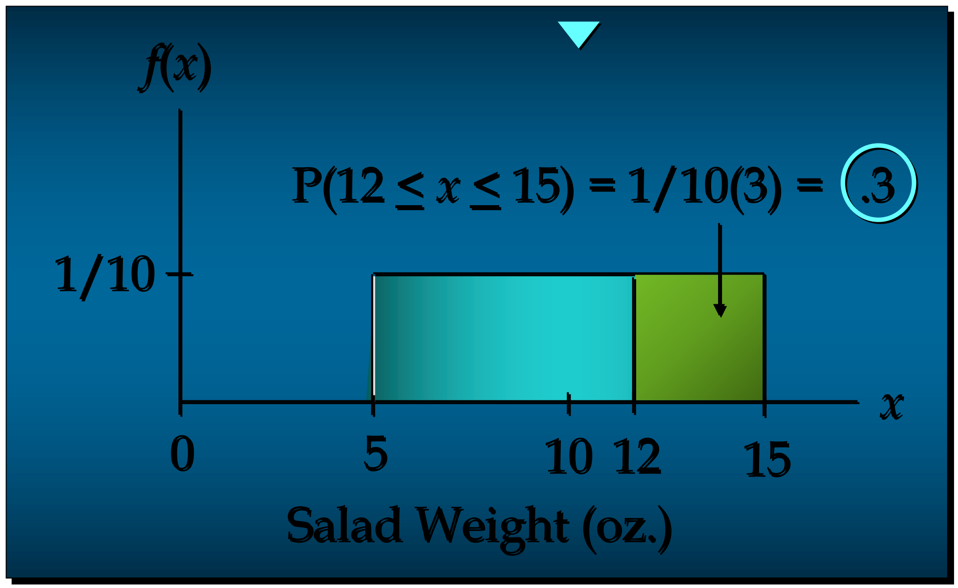
Uniform Probability Distribution

- Uniform Probability Distribution for Salad Plate Filling Weight

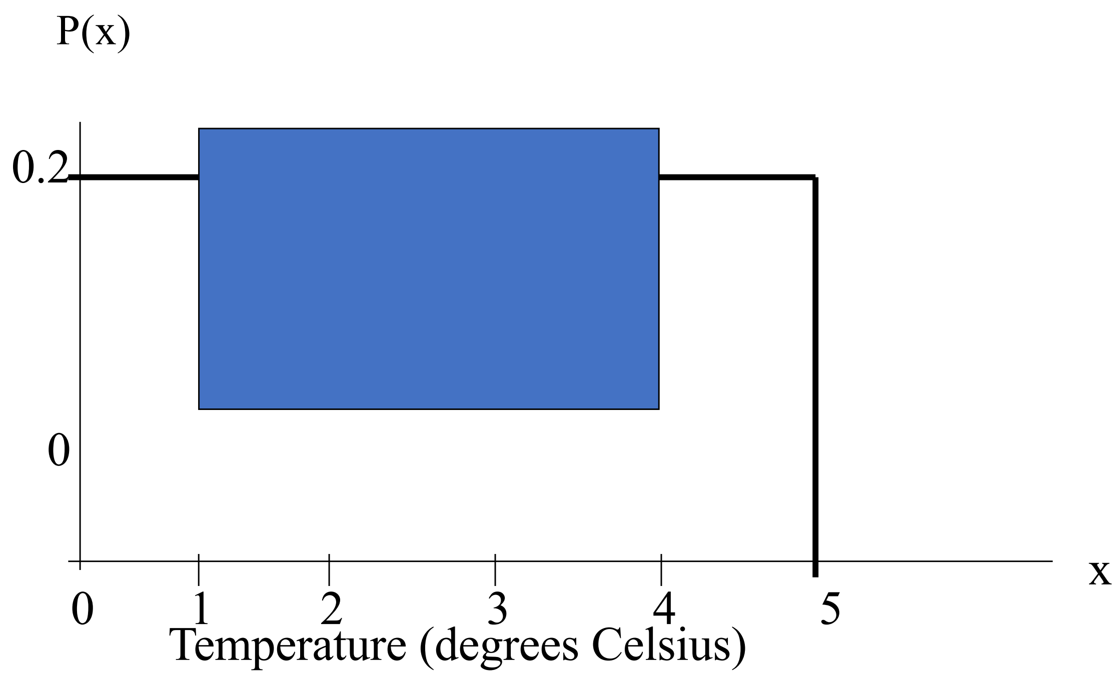


Uniform Probability Distribution

What is the probability that a customer will take between 12 and 15 ounces of salad?



What is the probability the temperature is between 1^0C and 4^0C?



We know that the total area of the rectangle is 1, and we can see that the part of the rectangle between 1 and 4 is 3/5 of the total, so $P(1 \leq x \leq 4) = 3/5*(1) = 0.6$.

Area as a Measure of Probability

- The area under the graph of $f(x)$ and probability are identical.
- This is valid for all continuous random variables.
- The probability that x takes on a value between some lower value x_1 and some higher value x_2 can be found by computing the area under the graph of $f(x)$ over the interval from x_1 to x_2.

Normal Probability Distribution

- The normal probability distribution is the most important distribution for describing a continuous random variable.
- It is widely used in statistical inference.
- It has been used in a wide variety of applications including:
  - Heights of people
  - Rainfall amounts
  - Test scores
  - Scientific measurements
- Abraham de Moivre, a French mathematician, published *The Doctrine of Chances* in 1733.
- He derived the normal distribution.

Normal Probability Distribution

- Normal Probability Density Function

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

where:

μ = mean

σ = standard deviation

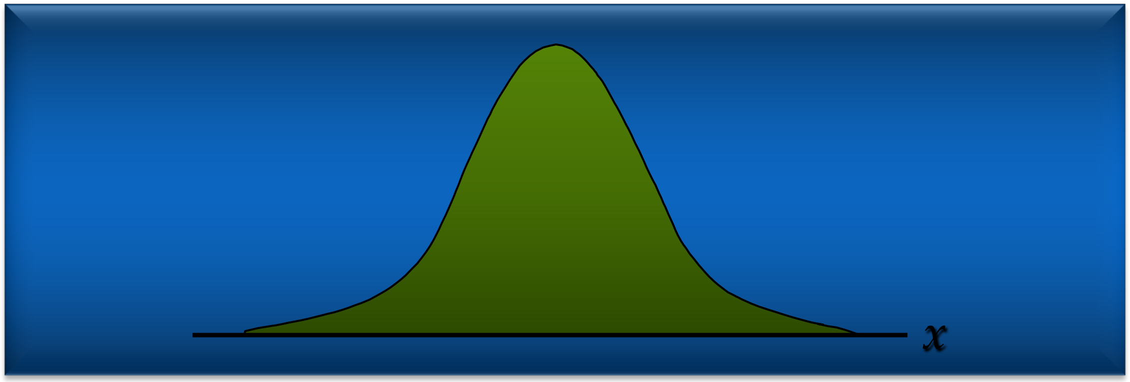
π = 3.14159

e = 2.71828

Normal Probability Distribution

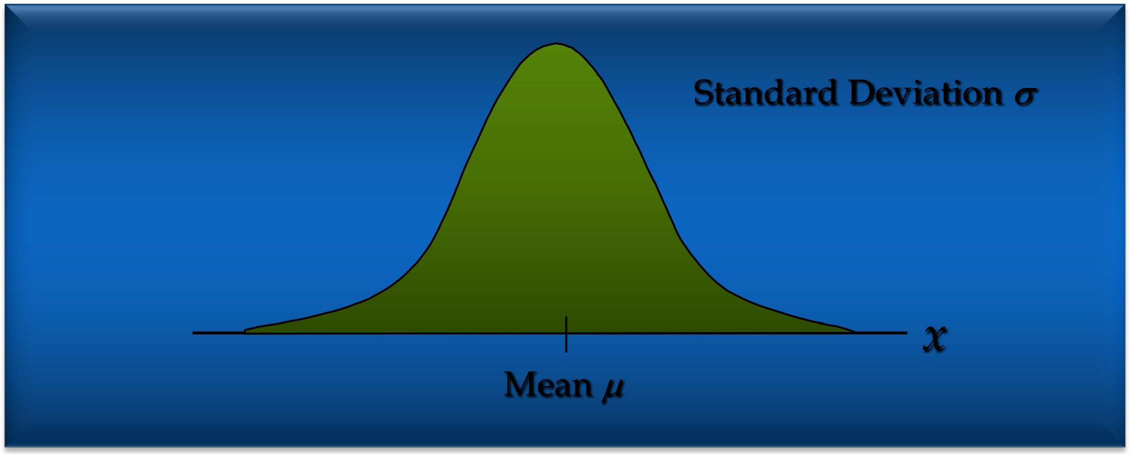
- Characteristics

The distribution is <u>symmetric</u>; its skewness measure is zero.



Normal Probability Distribution

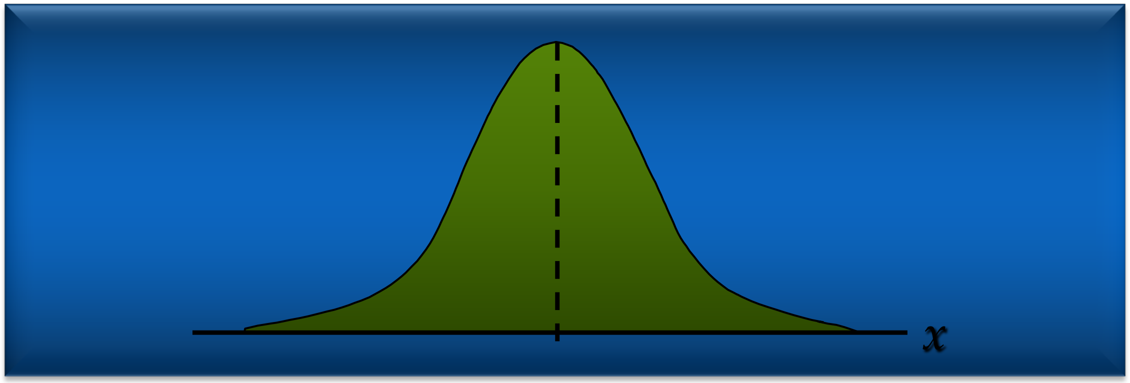
- Characteristics
  - The entire family of normal probability distributions is defined by its <u>mean</u> μ and its <u>standard deviation</u> σ.



Normal Probability Distribution

- Characteristics

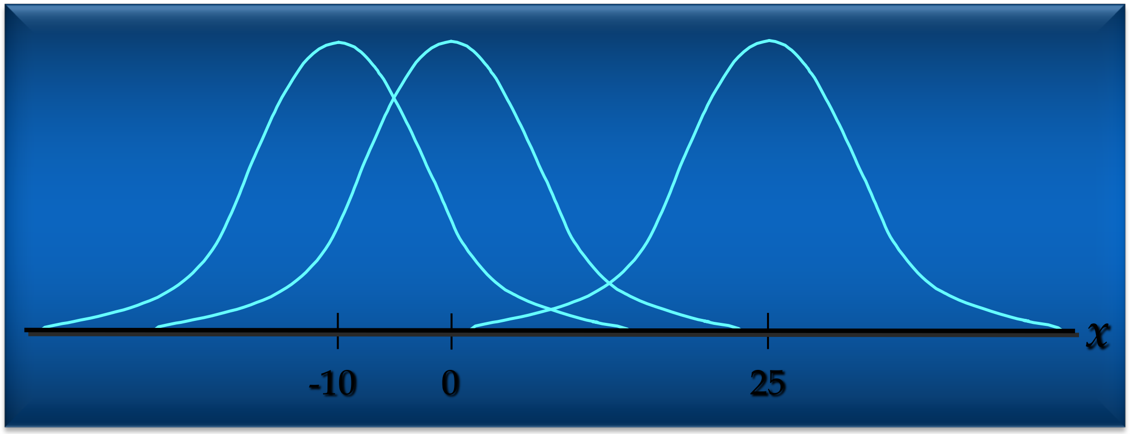
  - The <u>highest point</u> on the normal curve is at the <u>mean</u>, which is also the <u>median</u> and <u>mode</u>.



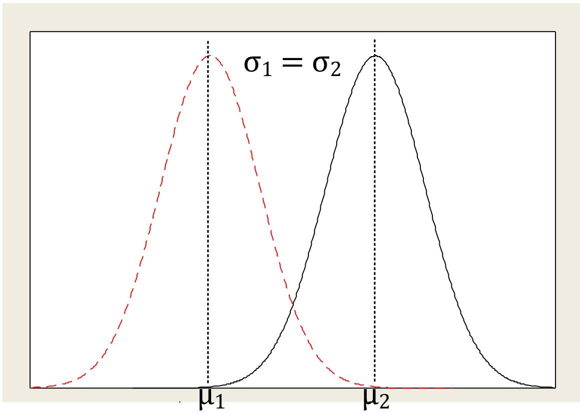
Normal Probability Distribution

- Characteristics

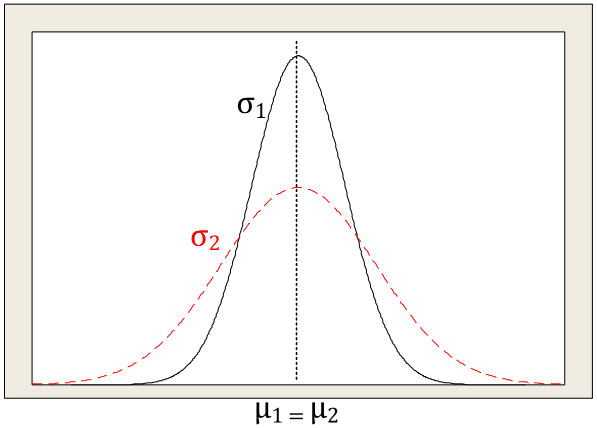
  - The mean can be any numerical value: negative, zero, or positive.



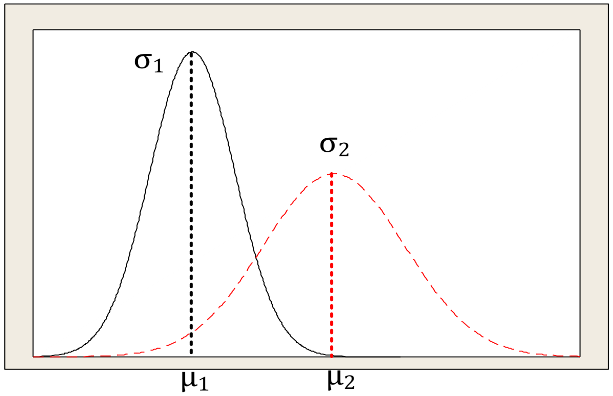
Normal Distribution



Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 = \sigma_2$



Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$

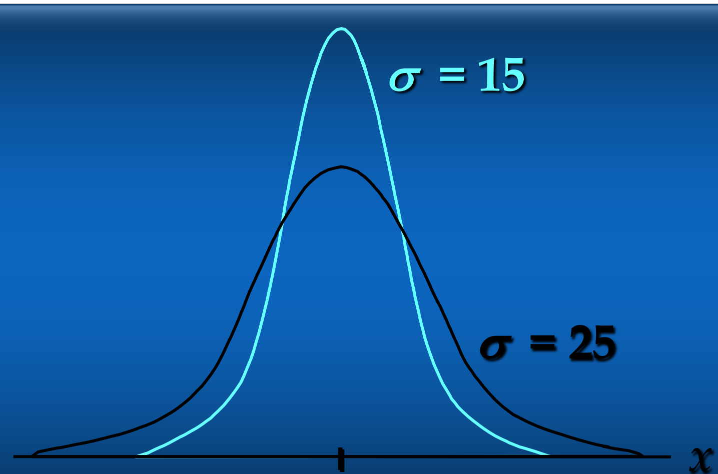


Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$

Normal Probability Distribution

- Characteristics

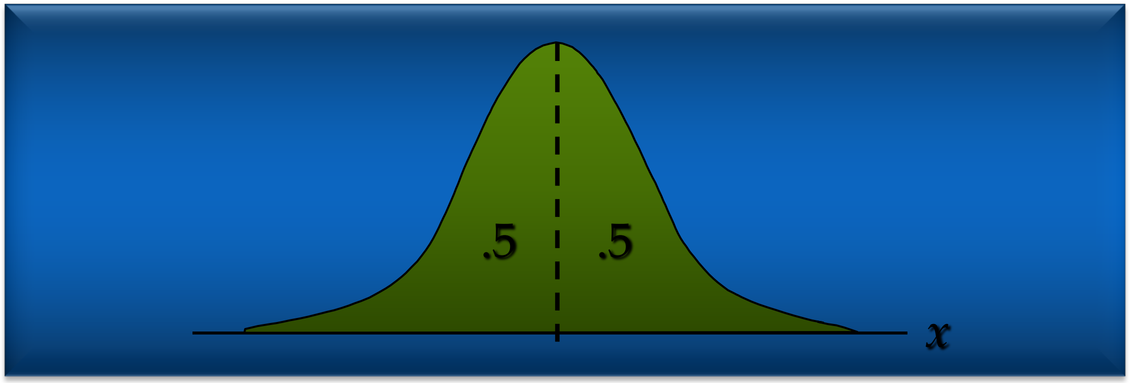
The standard deviation determines the width of the curve: larger values result in wider, flatter curves.



Normal Probability Distribution

- Characteristics

  - Probabilities for the normal random variable are given by <u>areas under the curve</u>. The total area under the curve is 1 (.5 to the left of the mean and .5 to the right).

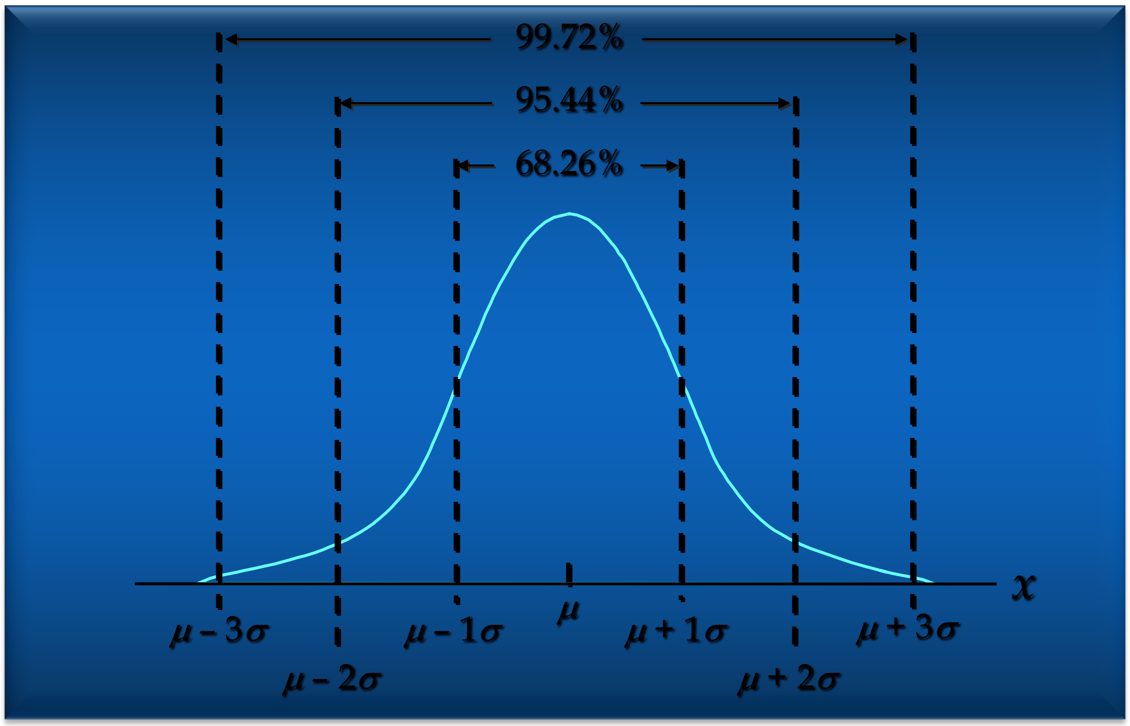


Normal Probability Distribution

- Characteristics (basis for the empirical rule)
  - 68.26% of values of a normal random variable are within +/- 1 standard deviation of its mean.
  - 95.44% of values of a normal random variable are within +/- 2 standard deviations of its mean.
  - 99.72% of values of a normal random variable are within +/- 3 standard deviations of its mean.

Normal Probability Distribution

- Characteristics (basis for the empirical rule)



Standard Normal Probability Distribution

- Characteristics

  - A random variable having a normal distribution with a mean of 0 and a standard deviation of 1 is said to have a standard normal probability distribution.

Standard Normal Distribution

The **Standard Normal Distribution** is a normal probability distribution that has a mean of 0 and a standard deviation of 1.

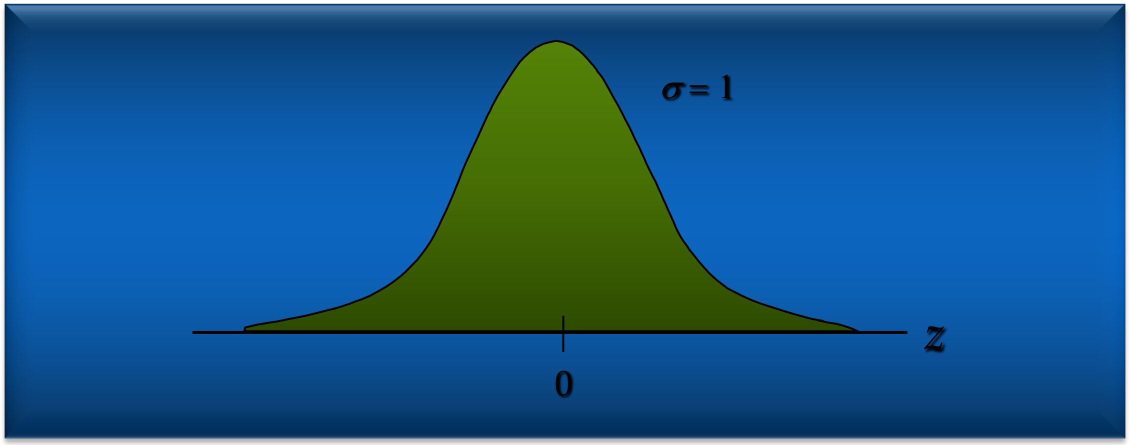
$$\mu = 0, \qquad \sigma = 1$$

In this way the formula giving the heights of the normal curve is simplified greatly.

Standard Normal Probability Distribution

- Characteristics

  - The letter z is used to designate the standard normal random variable.



Standard Normal Probability Distribution

- Converting to the Standard Normal Distribution

$$z = \frac{x - \mu}{\sigma}$$

We can think of z as a measure of the number of standard deviations x is from μ.

Standard Normal Distribution

- The normal distribution has computational complexity to calculate $P(x_1 < x < x_2)$ for any two (x_1, x_2) and given μ and σ
- To avoid this difficulty, the concept of z-transformation is followed.

$$\boxed{\mathrm{z} = \frac{x-\mu}{\sigma} \qquad \text{[Z-transformation]}}$$

- X: Normal distribution with mean μ and variance σ^2.
- Z: Standard normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$.
- Therefore, if *f(x)* assumes a value, then the corresponding value of $f(z)$ is given by

$$f(x{:}\,\mu,\sigma): \; P(x_1 < x < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2} dz$$

$$= f(z{:}\,0,\sigma)$$

Exercises

Example: Poisson distribution

Suppose that a rare disease has an incidence of 1 in 1000 person-years. Assuming that members of the population are affected independently, find the probability of k cases in a population of 10,000 (followed over 1 year) for k=0,1,2.

The expected value (mean) $=\lambda = .001*10{,}000 = 10$
10 new cases expected in this population per year→

$$P(X=0)=\frac{(10)^0 e^{-(10)}}{0!}=.0000454$$

$$P(X=1)=\frac{(10)^1 e^{-(10)}}{1!}=.000454$$

$$P(X=2)=\frac{(10)^2 e^{-(10)}}{2!}=.00227$$

Example

For example, if new cases of West Nile in New England are occurring at a rate of about 2 per month, then what's the probability that exactly 4 cases will occur in the next 3 months?

X ~ Poisson (λ=2/month)

$$\mathrm{P(X = 4\ in\ 3\ months)} = \frac{(2*3)^4 e^{-(2*3)}}{4!} = \frac{6^4 e^{-(6)}}{4!} = 13.4\%$$

Exactly 6 cases?

$$\mathrm{P(X = 6\ in\ 3\ months)} = \frac{(2*3)^6 e^{-(2*3)}}{6!} = \frac{6^6 e^{-(6)}}{6!} = 16\%$$

Ex Bus is uniformly late between 2 and 10 minutes. How long can you expect to wait? With what S.D.? If it's >7 mins late, you'll be late for work. What's the prob. of you being late?

Practice problems

1a. If calls to your cell phone are a Poisson process with a constant rate λ=2 calls per hour, what's the probability that, if you forget to turn your phone off in a 1.5 hour movie, your phone rings during that time?

1b. How many phone calls do you expect to get during the movie?

Answer

1a. If calls to your cell phone are a Poisson process with a constant rate λ=2 calls per hour, what's the probability that, if you forget to turn your phone off in a 1.5 hour movie, your phone rings during that time?

X ~ Poisson (λ=2 calls/hour)

P(X≥1)=1 – P(X=0)

$$P(X=0)=\frac{(2*1.5)^0 e^{-2(1.5)}}{0!}\frac{(3)^0 e^{-3}}{0!}=e^{-3}=.05$$

$\therefore$ P(X≥1)=1 – .05 = 95% chance

1b. How many phone calls do you expect to get during the movie?

E(X) = λt = 2(1.5) = 3

Examples

1) A study involving testing the effectiveness of a new drug, the number of cured patients among all the patients who use such a drug approximately follows a binomial distribution.

2) Operation of ticketing system in a busy public establishment (e.g., airport), the arrival of passengers can be simulated using Poisson distribution.

Example

If the random variable X follows a Poisson distribution with mean 3.4, find $P(X=6)$.

Solution

This can be written more quickly as: if $X \sim Po(3.4)$ find $P(X=6)$.

Now
$$P(X=6) = \frac{e^{-\lambda}\lambda^6}{6!}$$

$$= \frac{e^{-3.4}(3.4)^6}{6!} \quad (\text{ mean, } \lambda = 3.4)$$

$$= 0.071\,604\,409 = 0.072 \quad (\text{to 3 d.p.}).$$

Example

The number of industrial injuries per working week in a particular factory is known to follow a Poisson distribution with mean 0.5.

Find the probability that

(a) in a particular week there will be:

  (i) less than 2 accidents,

  (ii) more than 2 accidents;

(b) in a three week period there will be no accidents.

Solution

Let A be 'the number of accidents in one week', so $A \sim P_0(0.5)$.

(a) (i) $P(A<2) = P(A \le 1)$

$= 0.9098$ (from tables in Appendix 3 (p257), to 4 d.p.)

or, from the formula,

$$P(A<2) = P(A=0) + P(A=1)$$

$$= e^{-0.5} + \frac{e^{-0.5} \times 0.5}{1!}$$

$$= \frac{3}{2} e^{-0.5}$$

$$\approx 0.9098.$$

(ii) $P(A>2) = 1 - P(A \le 2)$

$= 1 - 0.9856$ (from tables)

$= 0.0144$ (to 4 d. p.)

or

$$1 - \left[P(A=0) + P(A=1) + P(A=2)\right]$$

$$= 1 - \left[e^{-0.5} + e^{-0.5}0.5 + \frac{e^{-0.5}(0.5)^2}{2!}\right]$$

$$= 1 - e^{-0.5}(1 + 0.5 + 0.125)$$

$$= 1 - 1.625\, e^{-0.5}$$

$$\approx 0.0144.$$

(b) $P\,(0 \text{ in } 3 \text{ weeks}) = \left(e^{-0.5}\right)^3 \approx 0.223.$

$$= \frac{3}{2} e^{-0.5}$$

$$\approx 0.9098.$$

(ii) $P(A > 2) = 1 - P(A \le 2)$

$$= 1 - 0.9856 \quad \text{(from tables)}$$

$$= 0.0144 \quad \text{(to 4 d. p.)}$$

or

$$1 - \left[P(A = 0) + P(A = 1) + P(A = 2)\right]$$

$$= 1 - \left[e^{-0.5} + e^{-0.5} 0.5 + \frac{e^{-0.5}(0.5)^2}{2!}\right]$$

$$= 1 - e^{-0.5}(1 + 0.5 + 0.125)$$

$$= 1 - 1.625\, e^{-0.5}$$

$$\approx 0.0144.$$

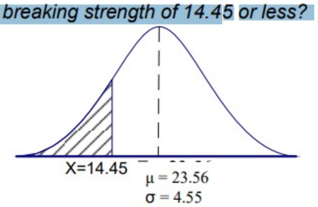
(b) $P\text{ (0 in 3 weeks)} = \left(e^{-0.5}\right)^3 \approx 0.223.$

Example

Wool fibre breaking strengths are normally distributed with mean μ = 23.56 Newtons and standard deviation, σ = 4.55.

What proportion of fibres would have a breaking strength of 14.45 or less?

- **Draw a diagram, label and shade area required:**



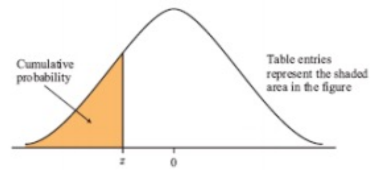
- **Convert** raw score (X) to standard score (Z) : $Z = \dfrac{14.45 - 23.56}{4.55} = -2.0$

  That is, the raw score of 14.45 is equivalent to a standard score of -2.0.
  It is negative because it is on the left hand side of the curve.

- **Use tables** to find probability and adjust this result to required probability:

$$
\begin{aligned}
&p(X < 14.45) = p(Z < -2.0) = 0.5 - p(0 < Z < 2) \\
&= 0.5 - 0.4772 \\
&= 0.0228
\end{aligned}
$$

Cumulative Probabilities for the Standard Normal Distribution



FIRST DIGIT OF z					SECOND DIGIT OF z						FIRST DIGIT OF z
z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**	z
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010	−3.0
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014	−2.9
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019	−2.8
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026	−2.7
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036	−2.6
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048	−2.5
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064	−2.4
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084	−2.3
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110	−2.2
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143	−2.1
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183	−2.0
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233	−1.9
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294	−1.8
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367	−1.7
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455	−1.6
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559	−1.5
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681	−1.4
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823	−1.3
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985	−1.2
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170	−1.1
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379	−1.0
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611	−0.9
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867	−0.8
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148	−0.7
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451	−0.6
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776	−0.5
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121	−0.4
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483	−0.3
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859	−0.2
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247	−0.1
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641	−0.0
z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**	z
FIRST DIGIT OF z					SECOND DIGIT OF z						FIRST DIGIT OF z

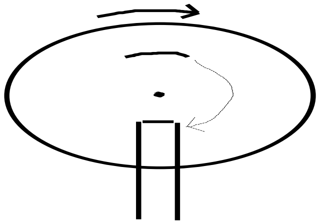
***Example*: Disk wait times**

In a hard disk drive, the disk rotates at 7200rpm. The wait time is defined as the time between the read/write head moving into position and the beginning of the required information appearing under the head.

(a) Find the distribution of the wait time.

(b) Find the mean and standard deviation of the wait time.

(c) Booting a computer requires that 2000 pieces of information are read from random positions. What is the total expected contribution of the wait time to the boot time, and rms deviation?



Answer: (a) Rotation rate of 7200rpm gives rotation time $= \frac{1}{7200}\, s =$ 8.33ms.

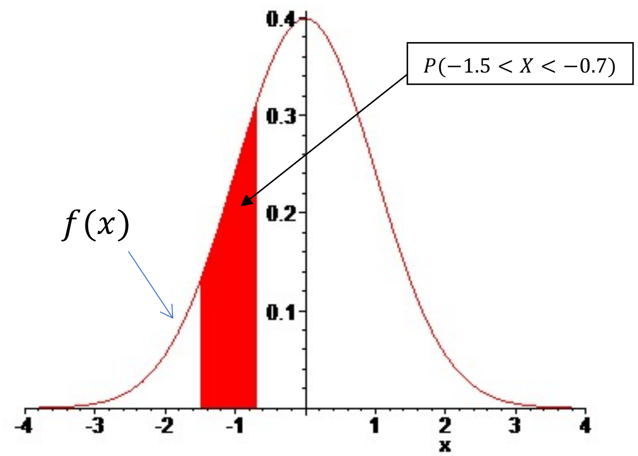
Wait time can be anything between 0 and 8.33ms and each time in this range is as likely as any other time.

Therefore, distribution of the wait time is uniform, U(0, 8.33ms)

Continuous Random Variables

A continuous random variable is a random variable which can take values measured on a continuous scale e.g. weights, strengths, times or lengths.

For any pre-determined value x, $P(X = x) = 0$, since if we measured X accurately enough, we are never going to hit the value x exactly. However the probability of some region of values near x can be non-zero.



Probability density function (pdf): $f(x)$

$$P(a \leq X \leq b) = \int_a^b f(x')dx'$$

Probability of X in the range a to b.

Normalization:

Since X has to have some value

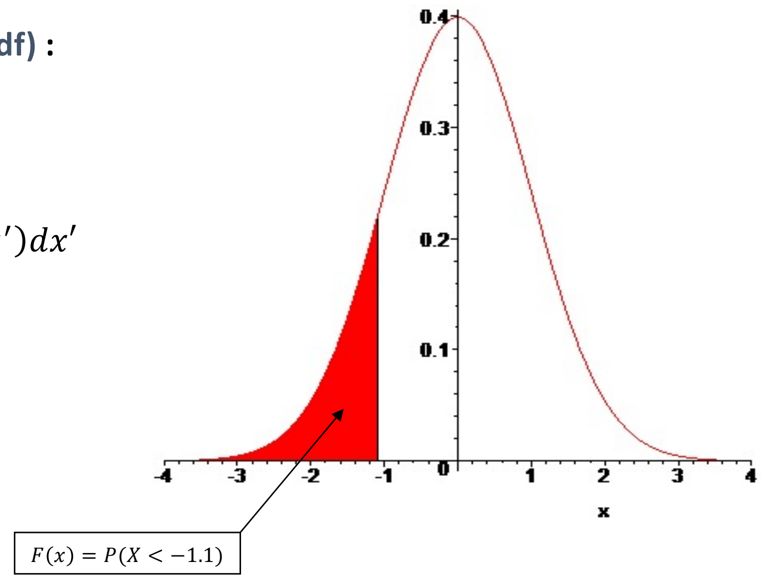
$$\int_{-\infty}^{\infty} f(x)dx = P(-\infty < X < \infty) = 1$$

And since $0 \leq P \leq 1$, for a pdf, $f(x) \geq 0$ for all x.

Cumulative distribution function (cdf) :

This is the probability of $X < x$.

$$F(x) \equiv P(X < x) = \int_{-\infty}^{x} f(x')dx'$$



Mean Expected value (mean) of X: $\mu = \int_{-\infty}^{\infty} x f(x) dx$

Variance Variance of X: $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$

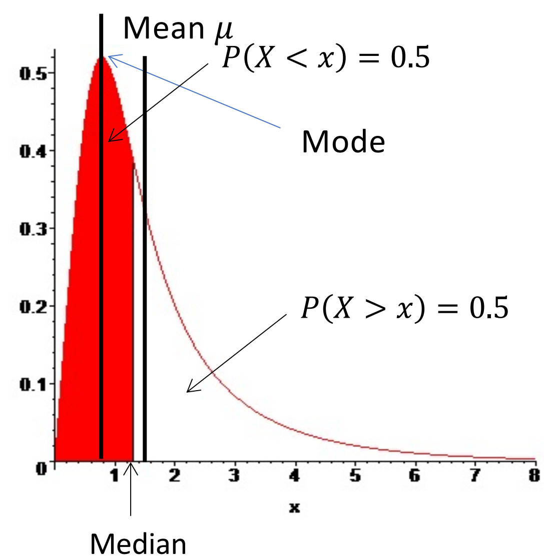
$$= \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$$

Note: the mean and variance may not be well defined for distributions with broad tails.

The *mode* is the value of x where $f(x)$ is maximum (which may not be unique).

The *median* is given by the value of *x* where

$$\int_{-\infty}^{x} f(x') dx' = \frac{1}{2}.$$



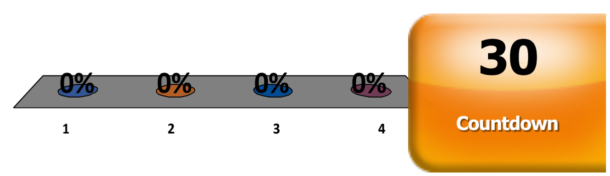
Question from Derek Bruff

<u>Probability density function</u>

Consider the continuous random variable X = the weight in pounds of a randomly selected new-born baby. Let f be the probability density function for X.
It is safe to assume that P(X < 0) = 0 and P(X < 20) = 1.

Which of the following is *not* a justifiable conclusion about f given this information?

1. No portion of the graph of f can lie below the x-axis.
2. f is non-zero for x in the range $0 \leq x < 20$
3. The area under the graph of f between x = 0 and x = 20 is 1.
4. The non-zero portion of the graph of f lies entirely between x = 0 and x = 20.



Q

<u>Probability density function</u>

Consider the continuous random variable X = the weight in pounds of a randomly selected new-born baby. Let f be the probability density function for X.
It is safe to assume that P(X < 0) = 0 and P(X < 20) = 1.

Which of the following is *not* a justifiable conclusion about f given this information?

A

1. No portion of the graph of f can lie below the x-axis.

- Correct, $f(x) \geq 0$ for all probabilities to be ≥ 0

2. f is non-zero for x in the range $0 \leq x < 20$

- Incorrect, $f(x)$ can be zero
e.g babies must weigh more than an embryo, so at least $f(x < \text{embryo weight}) = 0$

3. The area under the graph of f between x = 0 and x = 20 is 1.

- Correct. $\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{20} f(x)\,dx = 1$

4. The non-zero portion of the graph of f lies entirely between x = 0 and x = 20.

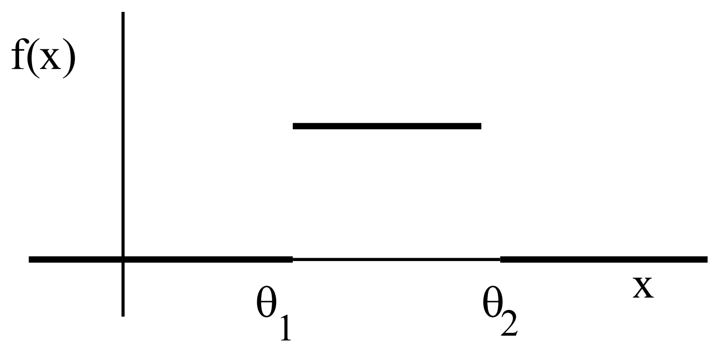
- Correct. $P(x < 0) = 0 \Rightarrow f(x < 0) = 0$ and $P(x < 20) = 1 \Rightarrow \int_{20}^{\infty} f(x)\,dx = 0$

Uniform distribution

The continuous random variable X has the Uniform distribution between θ_1 and θ_2, with $\theta_1 < \theta_2$ if

$$f(x) = \begin{cases} \dfrac{1}{\theta_2 - \theta_1} & \theta_1 \le x \le \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

$X \sim U(\theta_1, \theta_2)$, for short.



Occurrence of the Uniform distribution

1) Waiting times from random arrival time until a regular event (see later)

2) Simulation: programming languages often have a standard routine for simulating the U(0, 1) distribution. This can be used to simulate other probability distributions.

Actually not very common

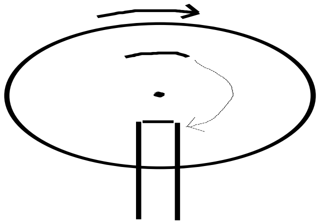
***Example*: Disk wait times**

In a hard disk drive, the disk rotates at 7200rpm. The wait time is defined as the time between the read/write head moving into position and the beginning of the required information appearing under the head.

(a) Find the distribution of the wait time.

(b) Find the mean and standard deviation of the wait time.

(c) Booting a computer requires that 2000 pieces of information are read from random positions. What is the total expected contribution of the wait time to the boot time, and rms deviation?



Answer: (a) Rotation rate of 7200rpm gives rotation time $= \frac{1}{7200}\,s =$ 8.33ms.

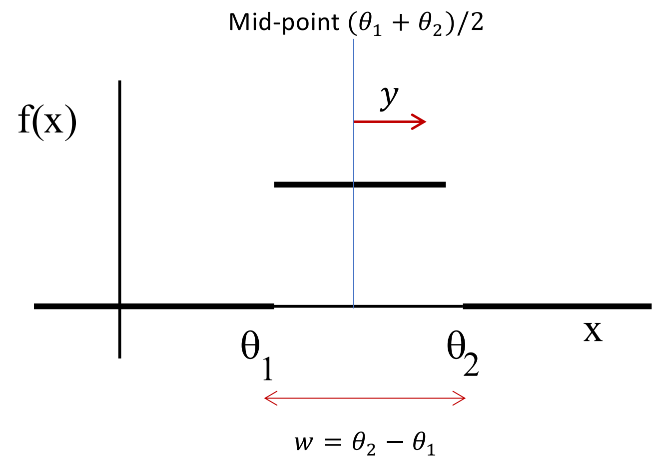
Wait time can be anything between 0 and 8.33ms and each time in this range is as likely as any other time.

Therefore, distribution of the wait time is uniform, U(0, 8.33ms)

Mean and variance: for $U(\theta_1, \theta_2)$

$$\mu = \frac{(\theta_1+\theta_2)}{2} \qquad \sigma^2 = \frac{(\theta_2-\theta_1)^2}{12}$$

Proof:



Let y be the distance from the mid-point,

$$y = x - (\theta_2 + \theta_1)/2$$

and the width be

$$w = \theta_2 - \theta_1.$$

Then since $x = \frac{\theta_1+\theta_2}{2} + y$, and means add

$$\mu = \langle x \rangle = \frac{\theta_2 + \theta_1}{2} + \langle y \rangle$$

$$= \frac{\theta_2 + \theta_1}{2} + \int_{-\frac{w}{2}}^{\frac{w}{2}} y f(y) dy \quad = \frac{\theta_2 + \theta_1}{2} + \int_{-\frac{w}{2}}^{\frac{w}{2}} y \frac{1}{w} dy \quad = \frac{\theta_1 + \theta_2}{2} \quad +0$$

Unsurprisingly the mean is the midpoint!

Variance:

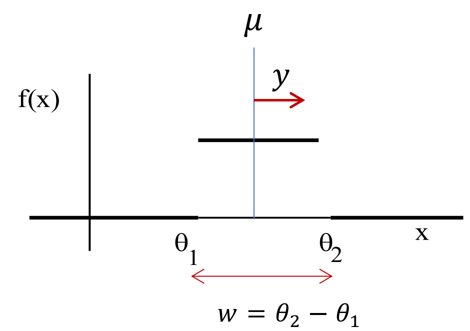
$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

$$= \int_{-\frac{w}{2}}^{\frac{w}{2}} y^2 \frac{1}{w}\; dy$$

$$= \frac{1}{w}\left[\frac{y^3}{3}\right]_{-\frac{w}{2}}^{\frac{w}{2}}$$

$$= \frac{1}{3w}\left(\frac{w^3}{8} + \frac{w^3}{8}\right) = \frac{w^2}{12}$$

$$= \frac{(\theta_2 - \theta_1)^2}{12}$$



Example: Disk wait times

In a hard disk drive, the disk rotates at 7200rpm. The wait time is defined as the time between the read/write head moving into position and the beginning of the required information appearing under the head.

(b) Find the mean and standard deviation of the wait time.

Answer: (b)

$$\mu = \frac{(\theta_1+\theta_2)}{2} = \frac{0+8.33}{2} \text{ ms} = 4.17 \text{ ms}$$

$$\sigma^2 = \frac{(\theta_2-\theta_1)^2}{12} = \frac{(8.33-0)^2}{12} \text{ ms} = 5.8 \text{ ms}^2$$

$$\Rightarrow \sigma = 2.4 \text{ ms}$$

http://en.wikipedia.org/wiki/Hard_disk_drive

Latency [edit]

Latency is the delay for the rotation of the disk to bring the required disk sector under the read-write mechanism. It depends on rotational speed of a disk, measured in revolutions per minute (RPM). Average rotational delay is shown in the table below, based on the empirical relation that the average latency in milliseconds for such a drive is one-half the rotational period:

Spindle [rpm]	Average latency [ms]
4200	7.14
5400	5.56
7200	4.17
10000	3
15000	2

Example*: *Disk wait times

In a hard disk drive, the disk rotates at 7200rpm. The wait time is defined as the time between the read/write head moving into position and the beginning of the required information appearing under the head.

(c) Booting a computer requires that 2000 pieces of information are read from random positions. What is the total expected contribution of the wait time to the boot time, and rms deviation?

Answer: (c)

$\mu = 4.2$ ms   For 2000 reads the mean total time is μ_{tot} =2000 × 4.2ms = 8.3s.

Note: rms = Root Mean Square = standard deviation

$\sigma = 2.4$ ms   So the variance is $\sigma_{tot}^2 = 2000 \times \sigma^2$ = 2000 ×5.8ms^2= 0.012s^2

$$\Rightarrow \sigma_{tot} = \sqrt{0.012\ s^2} = 0.11\text{s}$$

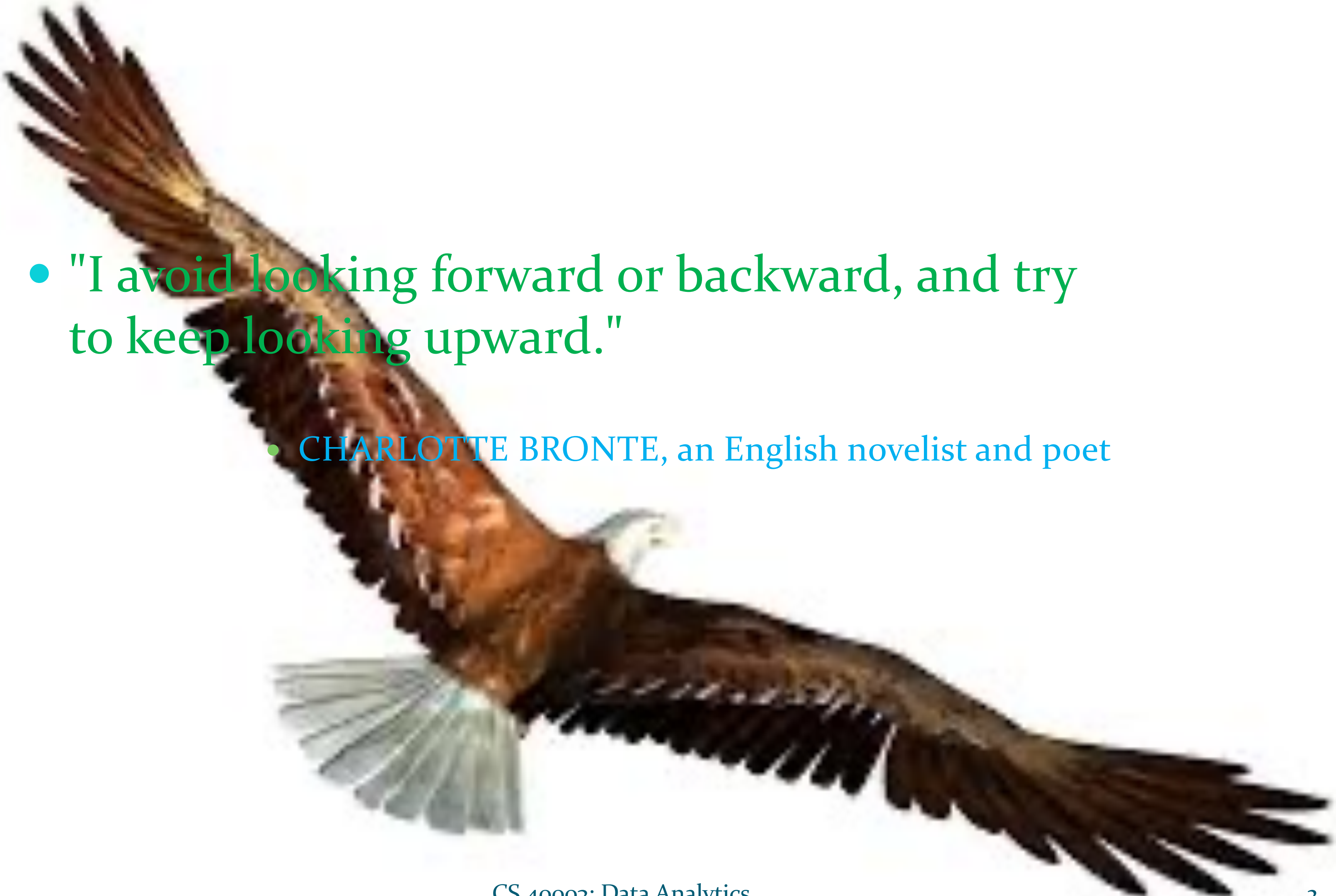
Data Analytics (CS40003)

Lecture #5

Probability Distributions

Dr. Debasis Samanta
Associate Professor
Department of Computer Science & Engineering

Quote of the day..



- "I avoid looking forward or backward, and try to keep looking upward."
    - CHARLOTTE BRONTE, an English novelist and poet

Today's discussion…

- Probability vs. Statistics
- Concept of random variable
- Probability distribution concept
- Discrete probability distribution
  - Discrete uniform probability distribution
  - Binomial distribution
  - Multinomial distribution
  - Hypergeometric distribution
  - Poisson distribution

Today's discussion

- Continuous probability distribution
  - Continuous uniform probability distribution
  - Normal distribution
  - Standard normal distribution
  - Chi-squared distribution
  - Gamma distribution
  - Exponential distribution
  - Lognormal distribution
  - Weibull distribution

Just a minute to mark your attendance

Probability and Statistics

Probability is the chance of an outcome in an experiment (also called event).

Event: Tossing a fair coin
Outcome: Head, Tail

Probability deals with **predicting** the likelihood of **future** events.	Statistics involves the **analysis of the frequency** of **past** events

Example: Consider there is a drawer containing 100 socks: 30 red, 20 blue and 50 black socks.

We can use probability to answer questions about the selection of a random sample of these socks.

- **PQ1.** What is the probability that we draw two blue socks or two red socks from the drawer?
- **PQ2.** What is the probability that we pull out three socks or have matching pair?
- **PQ3.** What is the probability that we draw five socks and they are all black?

Statistics

Instead, if we have no knowledge about the type of socks in the drawers, then we enter into the realm of statistics. Statistics helps us to infer properties about the population on the basis of the random sample.

Questions that would be statistical in nature are:

- **SQ1**: A random sample of 10 socks from the drawer produced one blue, four red, five black socks. What is the total population of black, blue or red socks in the drawer?
- **SQ2**: We randomly sample 10 socks, and write down the number of black socks and then return the socks to the drawer. The process is done for five times. The mean number of socks for each of these trial is 7. What is the true number of black socks in the drawer?
- etc.

Probability vs. Statistics

In other words:

- In probability, we are given a model and asked what kind of data we are likely to see.
- In statistics, we are given data and asked what kind of model is likely to have generated it.

Example 4.1: Measles Study

- A study on health is concerned with the incidence of childhood measles in parents of childbearing age in a city. For each couple, we would like to know how likely, it is that either the mother or father or both have had childhood measles.
- The current census data indicates that 20% adults between the ages 17 and 35 (regardless of sex) have had childhood measles.
  - This give us the probability that an individual in the city has had childhood measles.

Defining Random Variable

Definition 4.1: Random Variable

A random variable is a rule that assigns a numerical value to an outcome of interest.

Example 4.2: In "measles Study", we define a random variable X as the number of parents in a married couple who have had childhood measles.

This random variable can take values of $0, 1\ and\ 2$.

Note:

- Random variable is not exactly the same as the variable defining a data.
- The probability that the random variable takes a given value can be computed using the rules governing probability.
  - For example, the probability that $X = 1$ means either mother or father but not both has had measles is 0.32. Symbolically, it is denoted as **P(X=1) = 0.32**

Probability Distribution

Definition 4.2: Probability distribution

A probability distribution is a definition of probabilities of the values of random variable.

Example 4.3: Given that 0.2 is the probability that a person (in the ages between 17 and 35) has had childhood measles. Then the probability distribution is given by

X	Probability
0	0.64
1	0.32
2	0.04

?

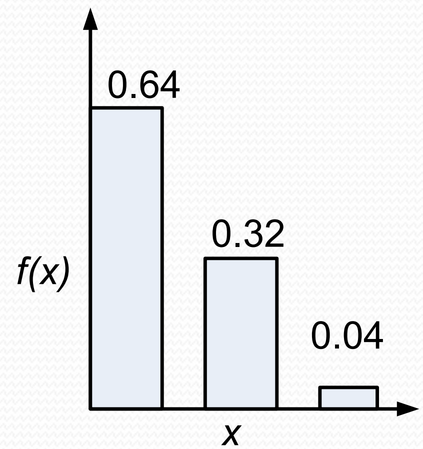
Probability Distribution

- In data analytics, the probability distribution is important with which many statistics making inferences about population can be derived .
  - In general, a probability distribution function takes the following form

x	x_1 $\quad x_2 \dots \dots \dots \dots x_n$
$f(x) = P(X = x)$	$f(x_1)$ $\quad f(x_2) \dots \dots \dots f(x_n)$

Example: Measles Study

x	0	1	2
$f(x)$	0.64	0.32	0.04



Taxonomy of Probability Distributions

Discrete probability distributions

- Binomial distribution
- Multinomial distribution
- Poisson distribution
- Hypergeometric distribution

Continuous probability distributions

- Normal distribution
- Standard normal distribution
- Gamma distribution
- Exponential distribution
- Chi square distribution
- Lognormal distribution
- Weibull distribution

Usage of Probability Distribution

- Distribution (discrete/continuous) function is widely used in simulation studies.
  - A simulation study uses a computer to simulate a real phenomenon or process as closely as possible.
  - The use of simulation studies can often eliminate the need of costly experiments and is also often used to study problems where actual experimentation is impossible.

Examples 4.4:

1) A study involving testing the effectiveness of a new drug, the number of cured patients among all the patients who use such a drug approximately follows a binomial distribution.

2) Operation of ticketing system in a busy public establishment (e.g., airport), the arrival of passengers can be simulated using Poisson distribution.

Discrete Probability Distributions

Binomial Distribution

- In many situations, an outcome has only two outcomes: success and failure.
  - Such outcome is called dichotomous outcome.
- An experiment when consists of repeated trials, each with dichotomous outcome is called Bernoulli process. Each trial in it is called a Bernoulli trial.

Example 4.5: Firing bullets to hit a target.

- Suppose, in a Bernoulli process, we define a random variable $X \equiv$ the number of successes in trials.
- Such a random variable obeys the binomial probability distribution, if the experiment satisfies the following conditions:
  1) The experiment consists of n trials.
  2) Each trial results in one of two mutually exclusive outcomes, one labelled a *"success"* and the other a *"failure"*.
  3) The probability of a success on a single trial is equal to $\boldsymbol{p}$. The value of p remains constant throughout the experiment.
  4) The trials are independent.

Defining Binomial Distribution

Definition 4.3: **Binomial distribution**

The function for computing the probability for the binomial probability distribution is given by

$$f(x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

$$for\ x = 0, 1, 2, \ldots, n$$

Here, $f(x) = P(X = x)$, where X denotes "the number of success" and $X = x$ denotes the number of success in x trials.

Binomial Distribution

Example 4.6: Measles study

X = having had childhood measles a success

p = 0.2, the probability that a parent had childhood measles

n = 2, here a couple is an experiment and an individual a trial, and the number of trials is two.

Thus,

$$P(x = 0) = \frac{2!}{0!(2-0)!}(0.2)^0(0.8)^{2-0} \quad = \mathbf{0.64}$$

$$P(x = 1) = \frac{2!}{1!\,(2-1)!}(0.2)^1(0.8)^{2-1} = \mathbf{0.32}$$

$$P(x = 2) = \frac{2!}{2!\,(2-2)!}(0.2)^2(0.8)^{2-2} = \mathbf{0.04}$$

Binomial Distribution

Example 4.7: Verify with real-life experiment

Suppose, 10 pairs of random numbers are generated by a computer (Monte-Carlo method)

15 38 68 39 49 54 19 79 38 14

If the value of the digit is 0 or 1, the outcome is "had childhood measles", otherwise, (digits 2 to 9), the outcome is "did not".

For example, in the first pair (i.e., 15), representing a couple and for this couple, $x = 1$. The frequency distribution, for this sample is

x	0	1	2
$f(x)=P(X=x)$	0.7	0.3	0.0

Note: This has close similarity with binomial probability distribution!

The Multinomial Distribution

The binomial experiment becomes a multinomial experiment, if we let each trial has more than two possible outcome.

Definition 4.4: **Multinomial distribution**

If a given trial can result in the $\boldsymbol{k}$ outcomes $E_1, E_2, \ldots \ldots, E_k$ with probabilities $p_1, p_2, \ldots \ldots, p_k$, then the probability distribution of the random variables $X_1, X_2, \ldots \ldots, X_k$ representing the number of occurrences for $E_1, E_2, \ldots \ldots, E_k$ in n independent trials is

$$f(x_1, x_2, \ldots \ldots, x_k) = \binom{n}{x_1, x_2, \ldots \ldots, x_k} {p_1}^{x_1} {p_2}^{x_2} \ldots \ldots {p_k}^{x_k}$$

$$\text{where } \binom{n}{x_1, x_2, \ldots \ldots, x_k} = \frac{n!}{x_1! x_2! \ldots \ldots x_k!}$$

$$\sum_{i=1}^{k} x_i = n \text{ and } \sum_{i=1}^{k} p_i = 1$$

The Hypergeometric Distribution

- Collection of samples with two strategies
  - With replacement
  - Without replacement
- A necessary condition of the binomial distribution is that all trials are **independent** to each other.
  - When sample is collected "with replacement", then each trial in sample collection is independent.

Example 4.8:

Probability of observing three red cards in 5 draws from an ordinary deck of 52 playing cards.

– You draw one card, note the result and then returned to the deck of cards

– Reshuffled the deck well before the next drawing is made

- The hypergeometric distribution *does not require* independence and is based on the sampling done **without** replacement.

The Hypergeometric Distribution

- In general, the hypergeometric probability distribution enables us to find the probability of selecting x successes in n trials from N items.

Properties of Hypergeometric Distribution

- A random sample of size n is selected without replacement from N items.
- k of the N items may be classified as success and $N - k$ items are classified as failure.

Let X denotes a hypergeometric random variable defining the number of successes.

Definition 4.5: Hypergeometric Probability Distribution

The probability distribution of the hypergeometric random variable X, the number of successes in a random sample of size n selected from N items of which k are labelled success and $N - k$ labelled as failure is given by

$$f(x) = P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$\max(0, n - (N - k)) \leq x \leq \min(n, k)$$

Multivariate Hypergeometric Distribution

The hypergeometric distribution can be extended to treat the case where the N items can be divided into k classes $A_1, A_2, \dots \dots, A_k$ with a_1 elements in the first class A_1, … and a_k elements in the k^{th} class. We are now interested in the probability that a random sample of size n yields x_1 elements from A_1, x_2 elements from $A_2, \dots \dots, x_k$ elements from A_k.

Definition 4.6: Multivariate Hypergeometric Distribution

If N items are partitioned into k classes $a_1, a_2, \dots \dots, a_k$ respectively, then the probability distribution of the random variables $X_1, X_2, \dots \dots, X_k$, representing the number of elements selected from $A_1, A_2, \dots \dots, A_k$ in a random sample of size n, is

$$f(x_1, x_2, \dots \dots, x_k) = P(X_1 = x_1, X_2 = x_2, \dots \dots, X_k = x_k) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2} \dots \dots \binom{a_k}{x_k}}{\binom{N}{n}}$$

with $\sum_{i=1}^{k} x_i = n$ and $\sum_{i=1}^{k} a_i = N$

The Poisson Distribution

There are some experiments, which involve the occurring of the number of outcomes during a given time interval (or in a region of space).

Such a process is called Poisson process.

Example 4.9:

Number of clients visiting a ticket selling counter in a metro station.

The Poisson Distribution

Properties of Poisson process

- The number of outcomes in one time interval is independent of the number that occurs in any other disjoint interval [Poisson process has no memory]
- The probability that a single outcome will occur during a very short interval is proportional to the length of the time interval and does not depend on the number of outcomes occurring outside this time interval.
- The probability that more than one outcome will occur in such a short time interval is negligible.

Definition 4.7: **Poisson distribution**

The probability distribution of the Poisson random variable X, representing the number of outcomes occurring in a given time interval t, is

$$f(x, \lambda t) = P(X = x) = \frac{e^{-\lambda t}.(\lambda t)^x}{x!}, x = 0, 1, \dots \dots$$

where λ is the average number of outcomes per unit time and $e = 2.71828 \dots$

Descriptive measures

Given a random variable X in an experiment, we have denoted $f(x) = P(X = x)$, the probability that $X = x$. For discrete events $f(x) = 0$ for all values of x except $x = 0, 1, 2, \dots\dots$

Properties of discrete probability distribution

1. $0 \leq f(x) \leq 1$
2. $\sum f(x) = 1$
3. $\mu = \sum x . f(x)$ [is the mean]
4. $\sigma^2 = \sum (x - \mu)^2 . f(x)$ [is the variance]

In $2, 3$ and 4, summation is extended for all possible discrete values of x.

Note: For discrete uniform distribution, $f(x) = \frac{1}{n}$ with $x = 1, 2, \dots \dots, n$

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

and $\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$

Descriptive measures

1. **Binomial distribution**

   The binomial probability distribution is characterized with p (the probability of success) and n (is the number of trials). Then

$$\mu = n.p$$

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

2. **Hypergeometric distribution**

   The hypergeometric distribution function is characterized with the size of a sample (n), the number of items (N) and k labelled success. Then

$$\mu = \frac{nk}{N}$$

$$\sigma^{2} = \frac{N-n}{N-1}.\mathrm{n}.\frac{k}{N}(1-\frac{k}{N})$$

Descriptive measures

3. **Poisson Distribution**

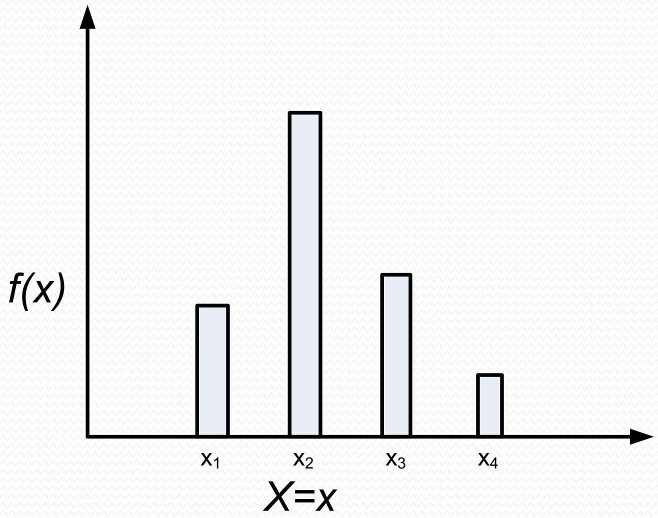
The Poisson distribution is characterized with λt where $\lambda = the\ mean\ of\ outcomes$ and $t = time\ interval.$

$$\mu = \lambda t$$

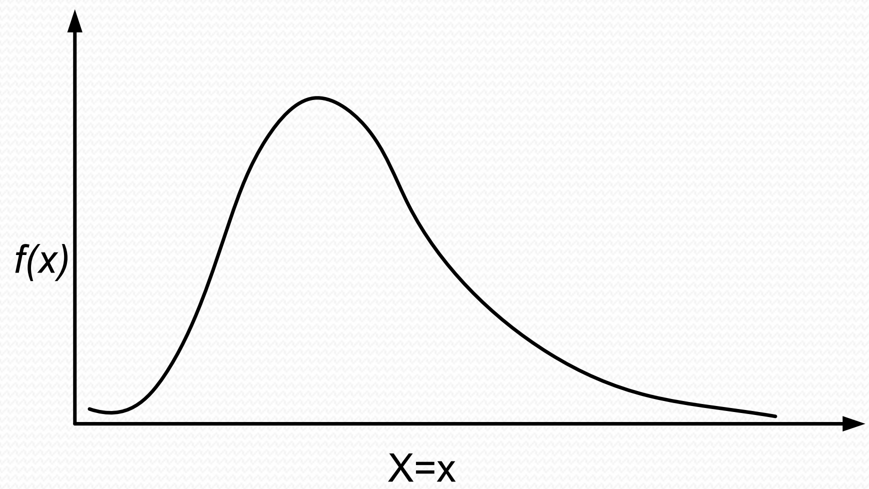
$$\sigma^2 = \lambda t$$

Continuous Probability Distributions

Continuous Probability Distributions



Discrete Probability distribution



Continuous Probability Distribution

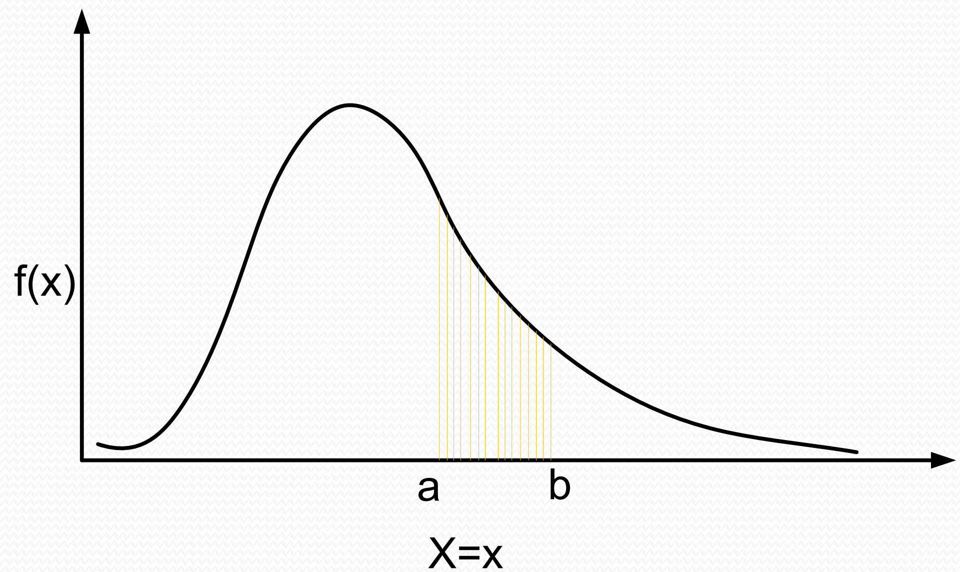
Continuous Probability Distributions

- When the random variable of interest can take any value in an interval, it is called continuous random variable.
  - Every continuous random variable has an infinite, uncountable number of possible values (i.e., any value in an interval)
- Consequently, continuous random variable differs from discrete random variable.

Properties of Probability Density Function

The function $f(x)$ is a probability density function for the continuous random variable X, defined over the set of real numbers R, if

1. $f(x) \geq 0$, for all $x \in R$
2. $\int_{-\propto}^{\propto} f(x)dx = 1$
3. $P(a \leq X \leq b) = \int_a^b f(x)\,dx$
4. $\mu = \int_{-\propto}^{\propto} xf(x)\,dx$
5. $\sigma^2 = \int_{-\propto}^{\propto} (x-\mu)^2 f(x)dx$



Continuous Uniform Distribution

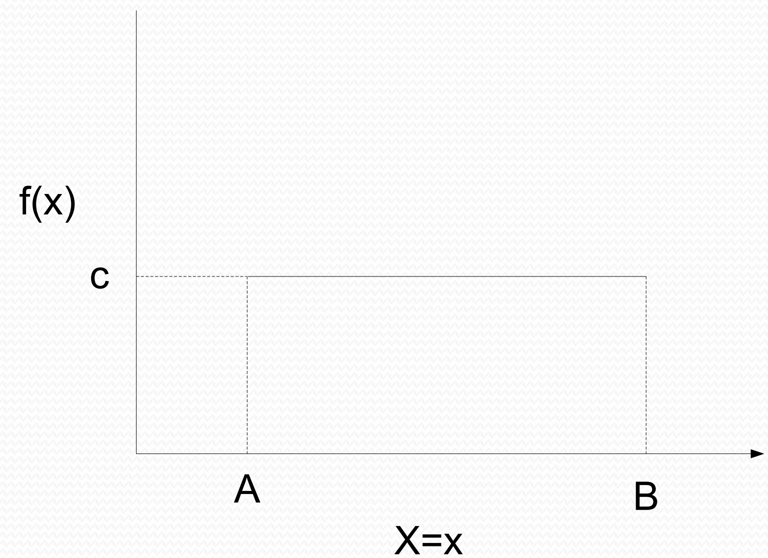
- One of the simplest continuous distribution in all of statistics is the continuous uniform distribution.

Definition 4.8: Continuous Uniform Distribution

The density function of the continuous uniform random variable X on the interval $[A, B]$ is:

$$f(x: A, B) = \begin{cases} \dfrac{1}{B - A} & A \leq x \leq B \\ 0 & Otherwise \end{cases}$$

Continuous Uniform Distribution



Note:

a) $\int_{\infty}^{-\infty} f(x)dx = \frac{1}{B-A} \times (B - A) = 1$

b) $P(c < x < d) = \frac{d-c}{B-A}$ where both c and d are in the interval (A,B)

c) $\mu = \frac{A+B}{2}$

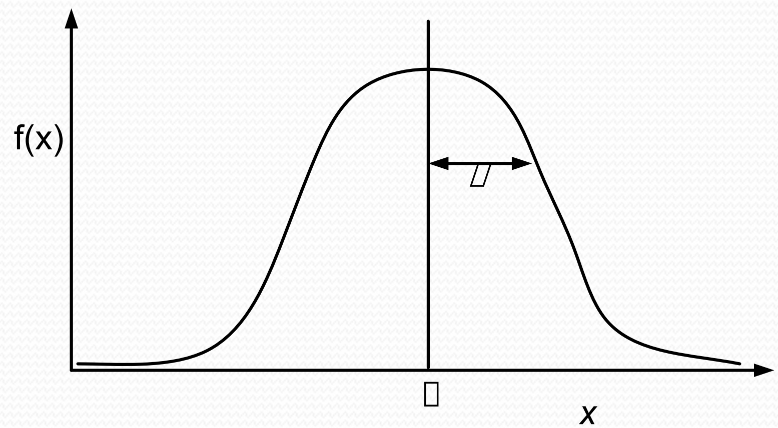
d) $\sigma^2 = \frac{(B-A)^2}{12}$

Normal Distribution

- The most often used continuous probability distribution is the normal distribution; it is also known as **Gaussian** distribution.
- Its graph called the normal curve is the bell-shaped curve.
- Such a curve approximately describes many phenomenon occur in nature, industry and research.
  - Physical measurement in areas such as meteorological experiments, rainfall studies and measurement of manufacturing parts are often more than adequately explained with normal distribution.
- A continuous random variable X having the bell-shaped distribution is called a normal random variable.

Normal Distribution

- The mathematical equation for the probability distribution of the normal variable depends upon the two parameters μ and σ, its mean and standard deviation.



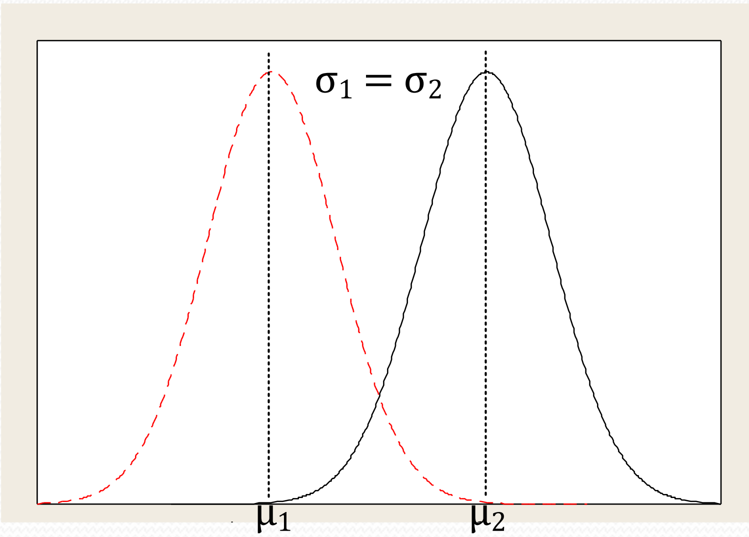
Definition 4.9: Normal distribution

The density of the normal variable x with mean μ and variance σ^2 is

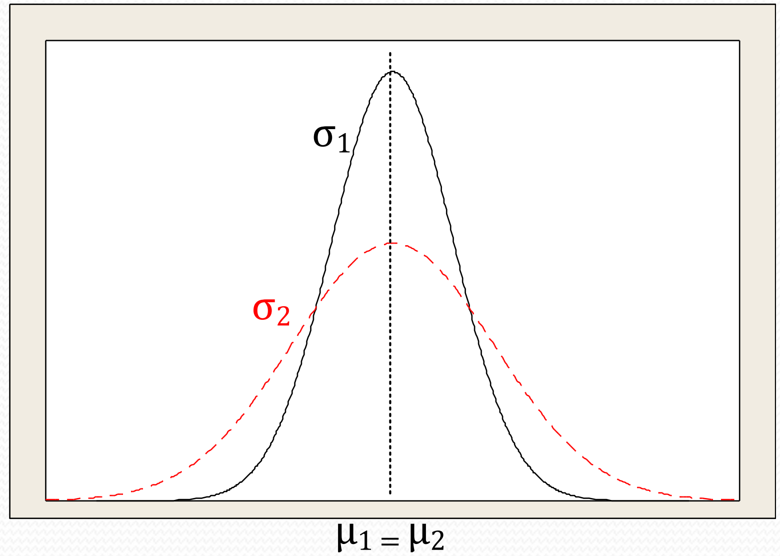
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty$$

where $\pi = 3.14159 \ldots$ and $e = 2.71828 \ldots$, the Naperian constant

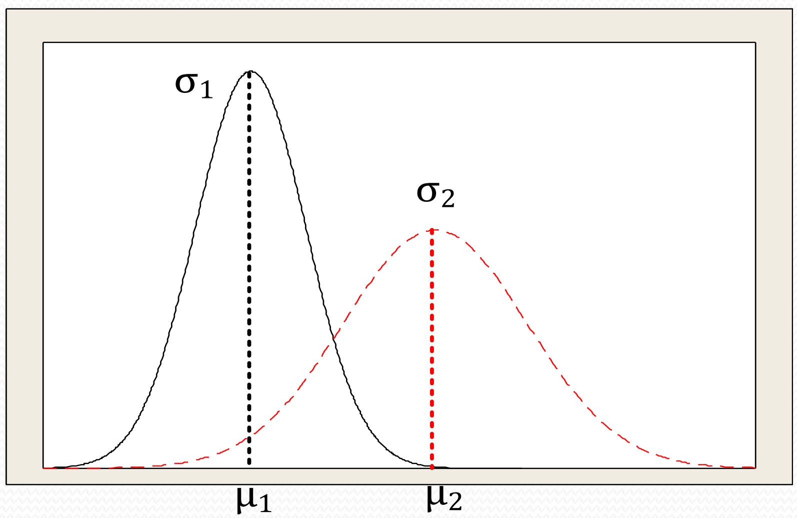
Normal Distribution



Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 = \sigma_2$



Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$



Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$

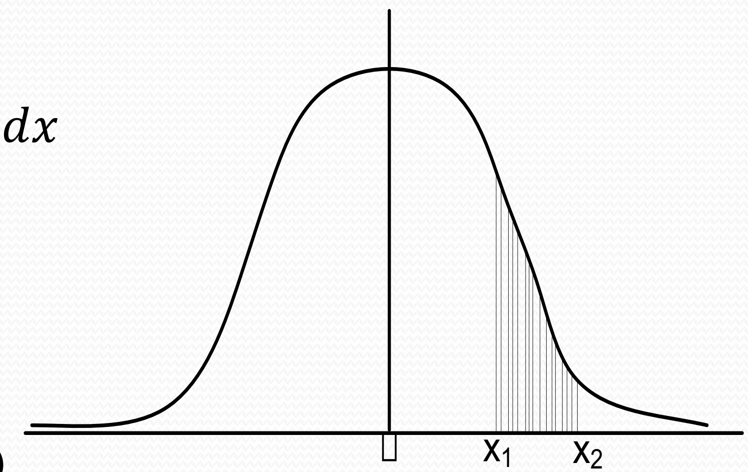
Properties of Normal Distribution

- The curve is symmetric about a vertical axis through the mean μ.
- The random variable x can take any value from $-\infty \; to \; \infty$.
- The most frequently used descriptive parameter s define the curve itself.
- The mode, which is the point on the horizontal axis where the curve is a maximum occurs at $x = \mu$.
- The total area under the curve and above the horizontal axis is equal to 1.

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

- $\mu = \int_{-\infty}^{\infty} x.f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x.e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$
- $\sigma^2 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 . e^{-\frac{1}{2}[(x-\mu)/\sigma^2]} dx$
- $P(x_1 < x < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$

denotes the probability of x in the interval (x_1, x_2).



Standard Normal Distribution

- The normal distribution has computational complexity to calculate $P(x_1 < x < x_2)$ for any two (x_1, x_2) and given μ and σ
- To avoid this difficulty, the concept of z-transformation is followed.

$$\boxed{\mathrm{z} = \frac{x-\mu}{\sigma} \qquad \text{[Z-transformation]}}$$

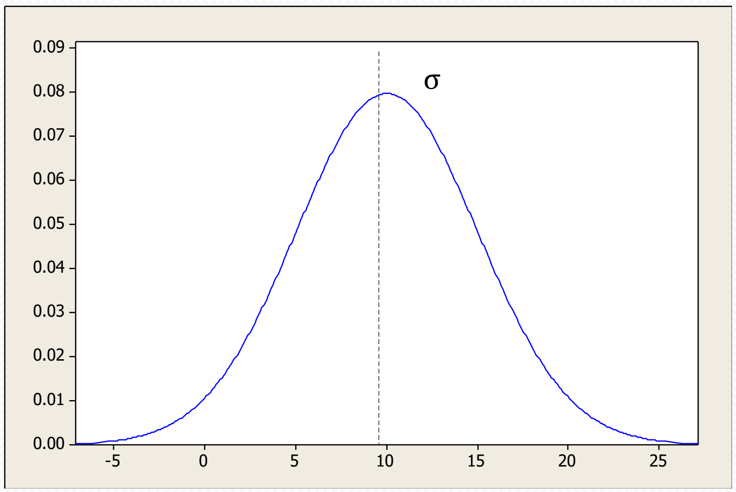
- X: Normal distribution with mean μ and variance σ^2.
- Z: Standard normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$.
- Therefore, if *f(x)* assumes a value, then the corresponding value of $f(z)$ is given by

$$f(x{:}\,\mu,\sigma): \; P(x_1 < x < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$
$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2} dz$$
$$= f(z{:}\,0,\sigma)$$

Standard Normal Distribution

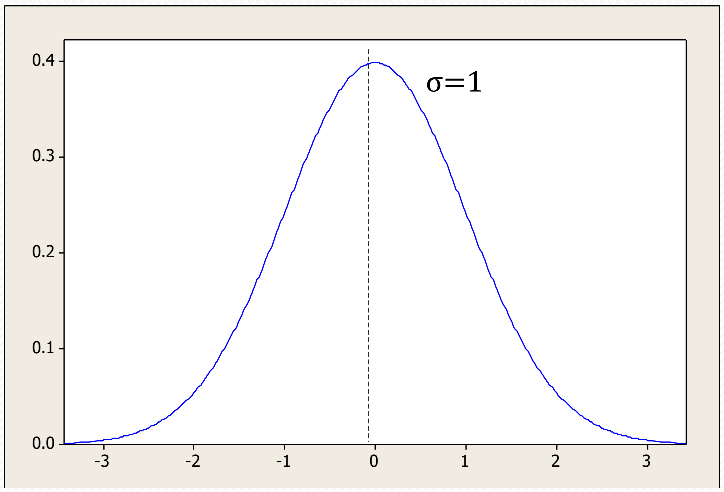
Definition 4.10: Standard normal distribution

The distribution of a normal random variable with mean 0 and variance 1 is called a standard normal distribution.



$x=\mu$

$f(x: \mu, \sigma)$



$\mu=0$

$f(z: 0, 1)$

Gamma Distribution

The gamma distribution derives its name from the well known gamma function in mathematics.

Definition 4.11: Gamma Function

$$\Gamma(\alpha) = \int_0^{\alpha} x^{\alpha-1}\, e^{-x} dx \quad \text{for } \alpha > 0$$

Integrating by parts, we can write,

$$\Gamma(\alpha) = (\alpha - 1)\int_0^{\alpha} x^{\alpha-2}\, e^{-x} dx$$
$$= (\alpha - 1)\Gamma(\alpha - 1)$$

Thus Γ function is defined as a recursive function.

Gamma Distribution

When $\alpha = n$, we can write,

$$\Gamma(n) = (n-1)(n-2) \ldots \ldots \ldots . \Gamma(1)$$

$$= (n-1)(n-2) \ldots \ldots \ldots .3.2.1$$

$$= (n-1)!$$

Further, $\Gamma(1) = \int_0^\infty e^{-x} dx = 1$

Note:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$ [An important property]

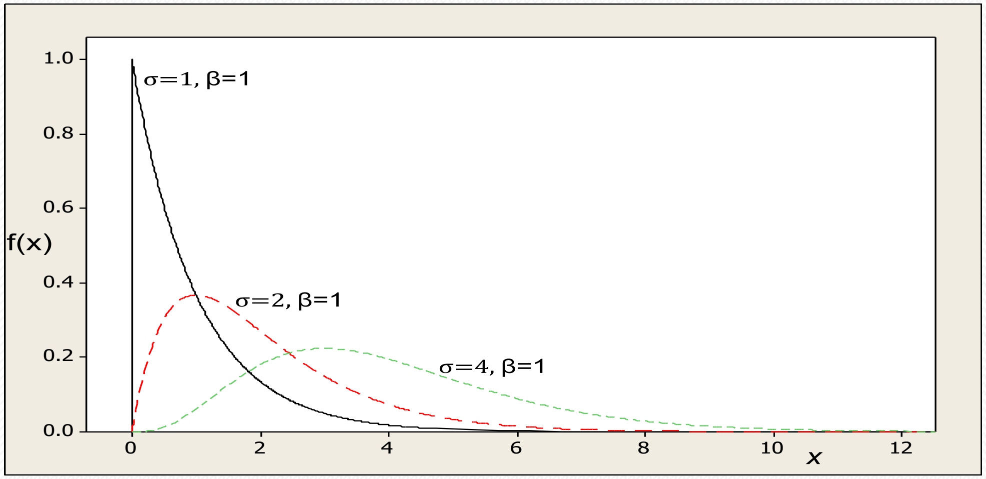
Gamma Distribution

Definition 4.12: **Gamma Distribution**

The continuous random variable x has a gamma distribution with parameters α and β such that:

$$f(x:\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^{\alpha}\,\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & x > 0 \\ 0 & Otherwise \end{cases}$$

where $\alpha > 0$ and β>0



Exponential Distribution

Definition 4.13: Exponential Distribution

The continuous random variable x has an exponential distribution with parameter β , where:

$$f(x:\beta) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}} & \text{where } \beta > 0 \\ 0 & \end{cases}$$

Note:

1) The mean and variance of gamma distribution are

$$\mu = \alpha\beta$$
$$\sigma^2 = \alpha\beta^2$$

2) The mean and variance of exponential distribution are

$$\mu = \beta$$
$$\sigma^2 = \beta^2$$

Chi-Squared Distribution

Definition 4.14: Chi-squared distribution

The continuous random variable x has a Chi-squared distribution with v degrees of freedom, is given by

$$f(x:v) = \begin{cases} \dfrac{1}{2^{\frac{v}{2}}\,\Gamma(v/2)} x^{v/2-1} e^{-\frac{x}{2}}, & x > 0 \\ 0 & Otherwise \end{cases}$$

where v is a positive integer.

- The Chi-squared distribution plays an important role in statistical inference .
- The mean and variance of Chi-squared distribution are:

$$\mu = v \ \text{ and } \ \sigma^2 = 2v$$

Lognormal Distribution

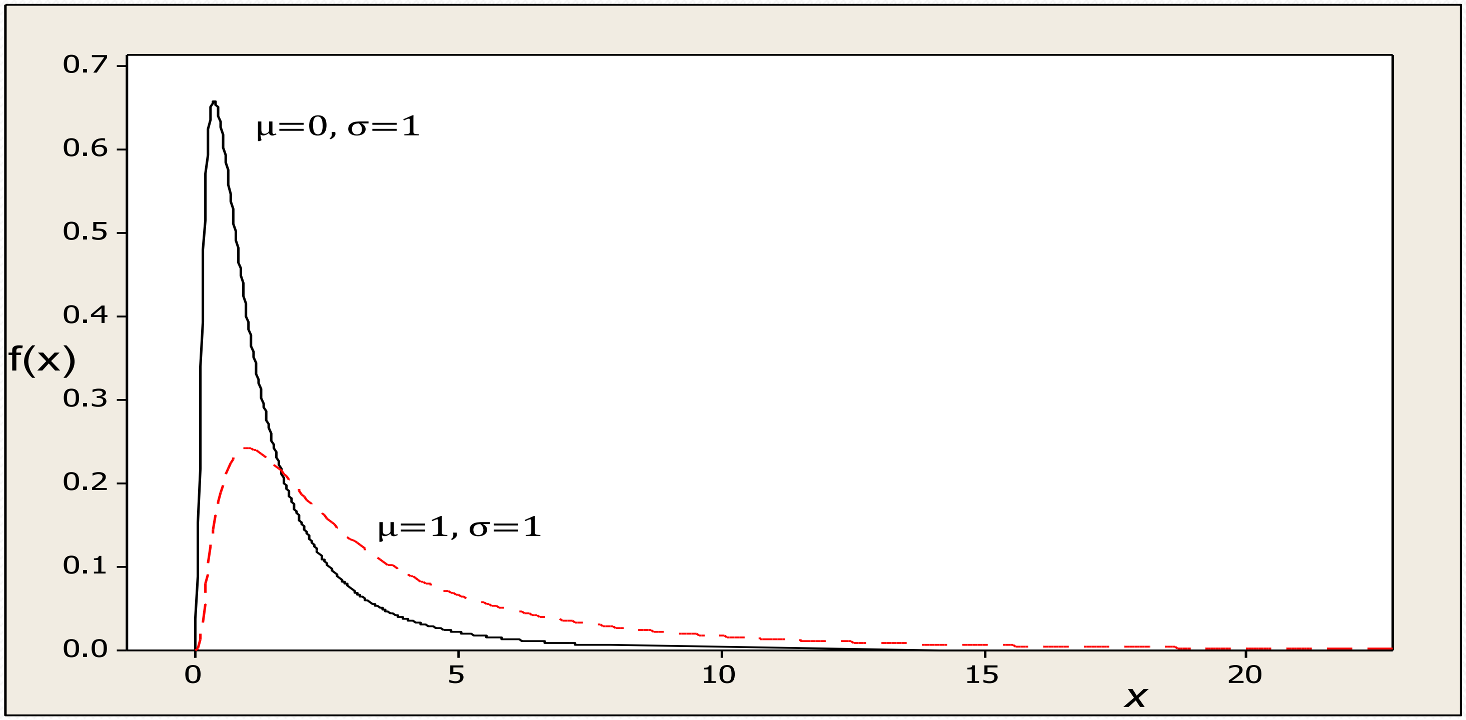
The lognormal distribution applies in cases where a natural log transformation results in a normal distribution.

Definition 4.15: Lognormal distribution

The continuous random variable x has a lognormal distribution if the random variable $y = \ln(x)$ has a normal distribution with mean μ and standard deviation σ. The resulting density function of x is:

$$f(x: \mu, \sigma) = \begin{cases} \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[\ln(x)-\mu]^2} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Lognormal Distribution



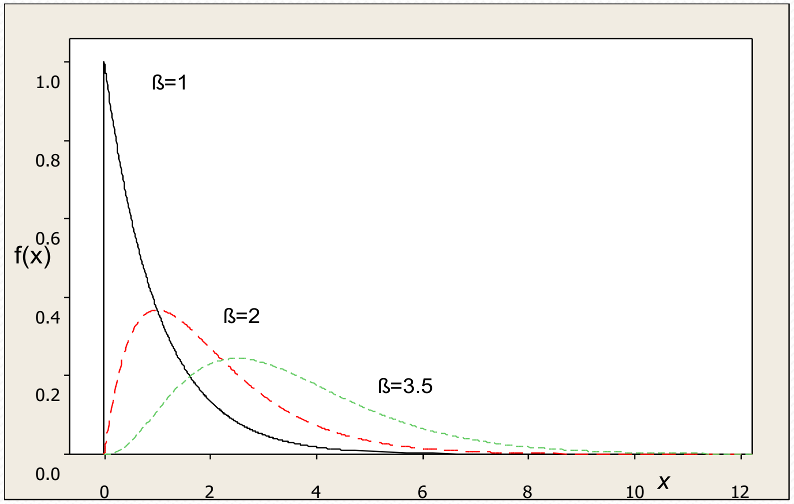
Weibull Distribution

Definition 4.16: Weibull Distribution

The continuous random variable x has a Weibull distribution with parameter α and β such that.

$$f(x:\alpha,\beta) = \begin{cases} \alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}} & x > 0 \\ 0 & Otherwise \end{cases}$$

where $\alpha > 0$ and $\beta > 0$



The mean and variance of Weibull distribution are:

$$\mu = \alpha^{-1/\beta}\Gamma\left(1+\frac{1}{\beta}\right)$$

$$\sigma^2 = \alpha^{-2/\beta}\{\Gamma\left(1+\frac{2}{\beta}\right) - [\Gamma(1+\frac{1}{\beta})]^2\}$$

Reference

- The detail material related to this lecture can be found in

Probability and Statistics for Enginneers and Scientists (8th Ed.)
by Ronald E. Walpole, Sharon L. Myers, Keying Ye (Pearson), 2013.

Any question?

You may post your question(s) at the “Discussion Forum” maintained in the course Web page!

Questions of the day…

1. Give some examples of random variables? Also, tell the range of values and whether they are with continuous or discrete values.

2. In the following cases, what are the probability distributions are likely to be followed. In each case, you should mention the random variable and the parameter(s) influencing the probability distribution function.
   a) In a retail source, how many counters should be opened at a given time period.
   b) Number of people who are suffering from cancers in a town?

Questions of the day…

2. In the following cases, what are the probability distributions are likely to be followed. In each case, you should mention the random variable and the parameter(s) influencing the probability distribution function.
   c) A missile will hit the enemy's aircraft.
   d) A student in the class will secure EX grade.
   e) Salary of a person in an enterprise.
   f) Accident made by cars in a city.
   g) People quit education after i) primary ii) secondary and iii) higher secondary educations.

Questions of the day…

3. How you can calculate the mean and standard deviation of a population if the population follows the following probability distribution functions with respect to an event.
    a) Binomial distribution function.
    b) Poisson's distribution function.
    c) Hypergeometric distribution function.
    d) Normal distribution function.
    e) Standard normal distribution function.

Stochastic Process

Anand Kumar M

Stochastic Process

- Many variables develop and change in real time:
- air temperatures, stock prices, interest rates, football scores, popularity of politicians, and also, the CPU usage, the speed of internet connection, the number of concurrent users..

What is a Stochastic Process?

- Stochastic processes are random variables that evolve and change in time.
- Stochastic Process: is a family of random variables $\{X(t) \mid t \in T\}$ (T is an index set; it may be discrete or continuous)
- Values assumed by $X(t)$ are called *states.*
- Sometimes called a random process or a chance process

DEFINITION 6.1

A **stochastic process** is a random variable that also depends on time. It is therefore a function of two arguments, $X(t, \omega)$, where:

- $t \in \mathcal{T}$ is time, with $\mathcal{T}$ being a set of possible times, usually $[0, \infty)$, $(-\infty, \infty)$, $\{0, 1, 2, \ldots\}$, or $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$;
- $\omega \in \Omega$, as before, is an outcome of an experiment, with Ω being the whole sample space.

Values of $X(t, \omega)$ are called *states*.

- At any fixed time t, we see a random variable $X_t(\omega)$, a function of a random outcome.
- On the other hand, if we fix ω, we obtain a function of time $X_\omega(t)$.

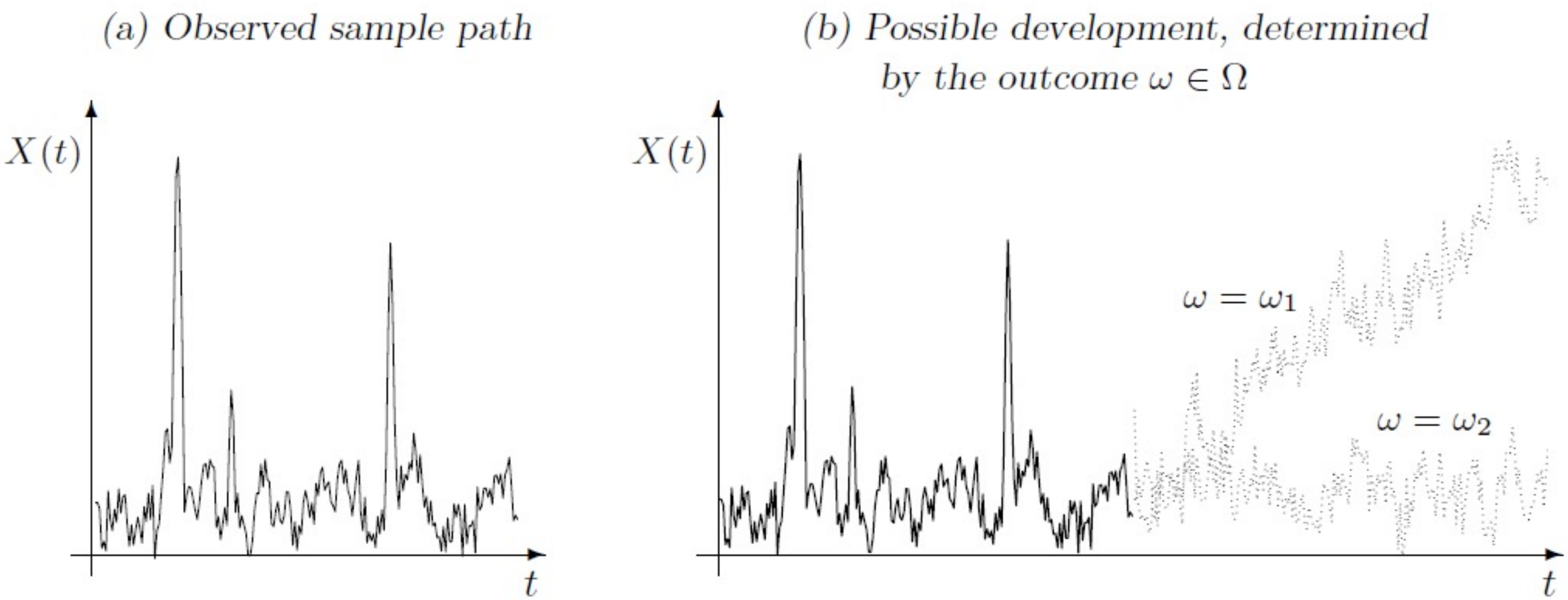


FIGURE 6.1: *Sample paths of CPU usage stochastic process.*

Stochastic Process Characterization

- Discrete and continuous cases:
  - States X(t) (i.e. time t) may be discrete/continuous
  - State space I may be discrete/continuous

DEFINITION 6.2

Stochastic process $X(t,\omega)$ is **discrete-state** if variable $X_t(\omega)$ is discrete for each time t, and it is a **continuous-state** if $X_t(\omega)$ is continuous.

DEFINITION 6.3

Stochastic process $X(t,\omega)$ is a **discrete-time process** if the set of times $\mathcal{T}$ is discrete, that is, it consists of separate, isolated points. It is a **continuous-time process** if $\mathcal{T}$ is a connected, possibly unbounded interval.

Classification of Stochastic Processes

- Four classes of stochastic processes:

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">time, Index set T</td></tr>
<tr><td>Discrete</td><td>Continuous</td></tr>
<tr><td rowspan="2">State/Sample Space
I</td><td>Discrete</td><td>Discrete-time stochastic chain</td><td>Continuous-time stochastic chain</td></tr>
<tr><td>Continuous</td><td>Discrete-time continuous-state process</td><td>Continuous-time continuous-state process</td></tr>
</table>

Examples

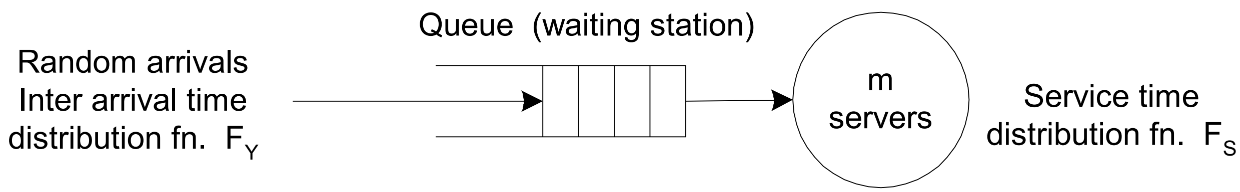
Example 6.3. The *actual* air temperature $X(t, \omega)$ at time t is a continuous-time, continuous-state stochastic process. Indeed, it changes smoothly and never jumps from one value to another. However, the temperature $Y(t, \omega)$ reported on a radio every 10 minutes is a discrete-time process. Moreover, since the reported temperature is usually rounded to the nearest degree, it is also a discrete-state process. $\Diamond$

Example 6.4. In a printer shop, let $X(n, \omega)$ be the amount of time required to print the n-th job. This is a discrete-time, continuous-state stochastic process, because $n = 1, 2, 3, \ldots$, and $X \in (0, \infty)$.

Let $Y(n, \omega)$ be the number of pages of the n-th printing job. Now, $Y = 1, 2, 3, \ldots$ is discrete; therefore, this process is discrete-time and discrete-state.

^

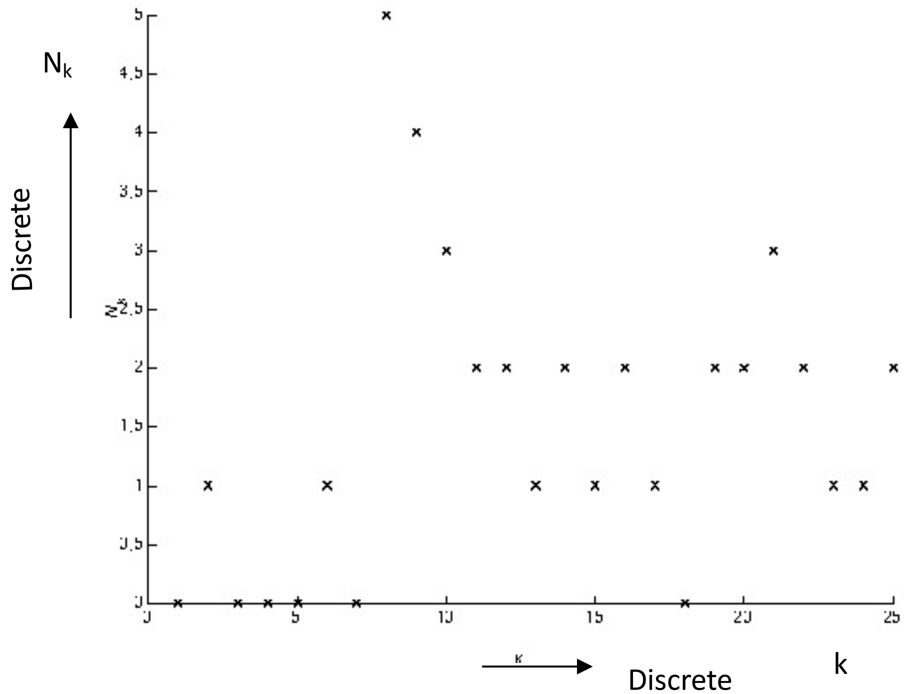
Example: a Queuing System



- Interarrival times $Y_1, Y_2, \ldots$ (common dist. Fn. F_Y)
- Service times: $S_1, S_2, \ldots$ (iid with a common cdf F_S)
- Notation for a queuing system: $F_Y/F_S/m$
- Some interarrival/service time distributions types are:
  - M: Memoryless (i.e., EXP)
  - D: Deterministic
  - E_k: k-stage Erlang etc.
  - H_k: k-stage Hyper exponential distribution
  - G: General distribution
  - GI: General independent inter arrival times
- M/M/1 → Memoryless interarrival/service times with a single server

Discrete/Continuous Stochastic Processes

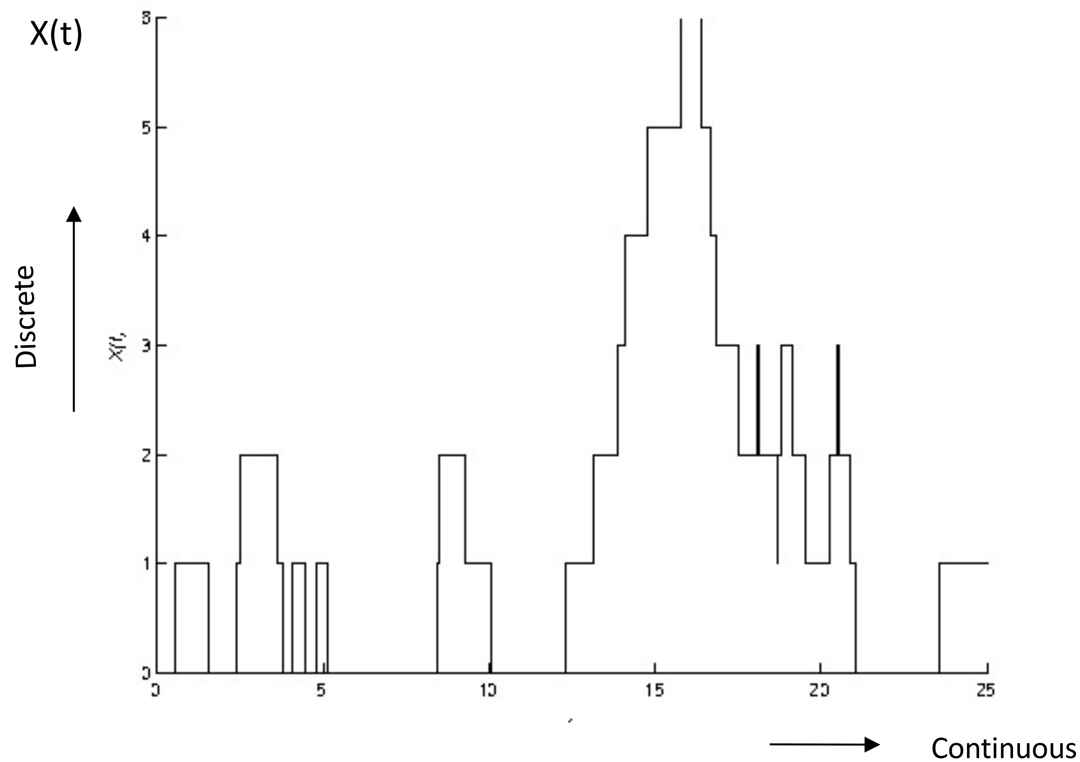
- N_k: Number of jobs waiting in the system at the time of k^{th} job's departure → Stochastic process $\{N_k \mid k=1,2,\ldots\}$:
  - Discrete time, discrete state



Continuous Time, Discrete Space

- *X(t)*: Number of jobs in the system at time *t*. {*X(t)* | *t ϵ T*} forms a *continuous-time, discrete-state* stochastic process, with,

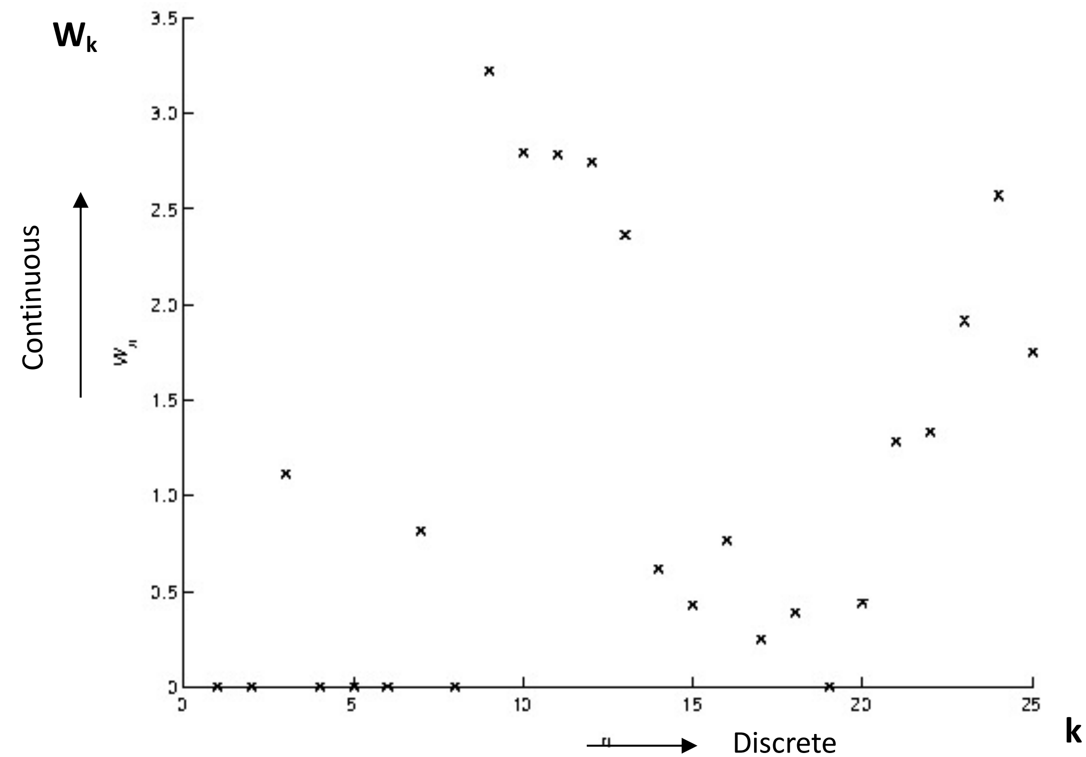
$$I = \{0, 1, 2, \ldots\} \text{ and } T = \{t | 0 \leq t < \infty\}$$



Discrete Time, Continuous Space

- W_k: waiting time for the k^{th} job. Then $\{W_k \mid k \in T\}$ forms a *Discrete-time, Continuous-state* stochastic process, where,

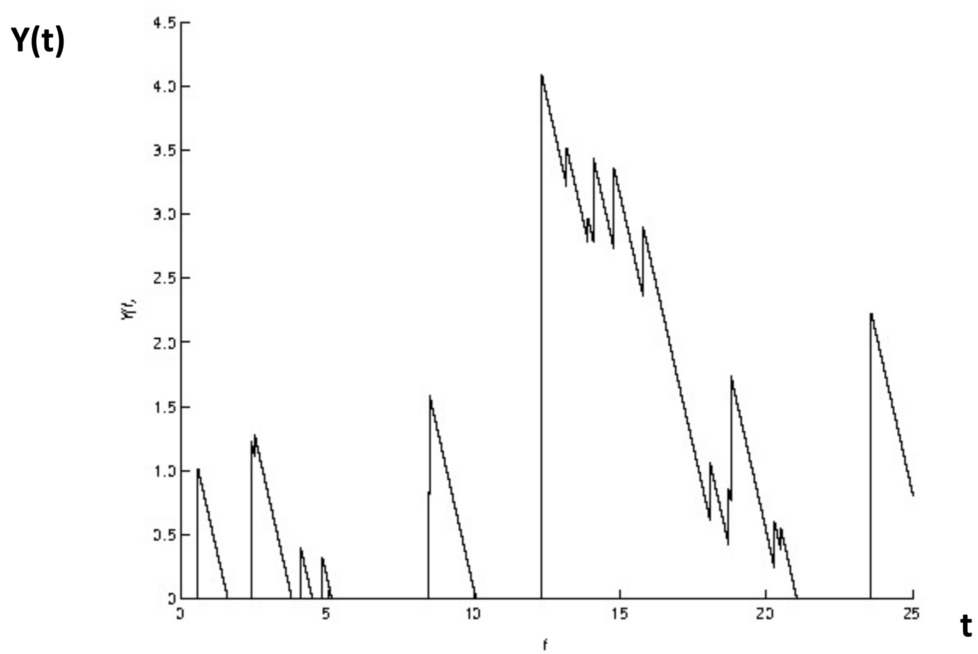
  $I = \{x | 0 \leq x < \infty\}$ and $T = \{0, 1, 2, \ldots\}$



Continuous Time, Continuous Space

- *Y(t)*: total service time for all jobs in the system at time *t. Y(t)* forms a *continuous-time, continuous-state* stochastic process, Where,

$$\{Y(t)|0 \leq t < \infty\} \text{ with } I = [0, \infty)$$



Markov Chain and Process

DEFINITION 6.4

Stochastic process $X(t)$ is **Markov** if for any $t_1 < ... < t_n < t$ and any sets $A; A_1, \ldots, A_n$

$$\begin{aligned} & \boldsymbol{P}\left\{X(t) \in A \mid X(t_1) \in A_1, \ldots, X(t_n) \in A_n\right\} \\ & \qquad = \boldsymbol{P}\left\{X(t) \in A \mid X(t_n) \in A_n\right\}. \end{aligned} \tag{6.1}$$

Let us look at the equation (6.1). It means that the conditional distribution of $X(t)$ is the same under two different conditions,

(1) given observations of the process X at several moments in the past;

(2) given only *the latest* observation of X.

If a process is Markov, then its future behavior is the same under conditions (1) and (2). In other words, knowing the present, we get no information from the past that can be used to predict the future,

$$\boldsymbol{P}\{\text{ future } \mid \text{ past, present }\} = \boldsymbol{P}\{\text{ future } \mid \text{ present }\}$$

Then, for the future development of a Markov process, only its present state is important, and it does not matter *how* the process arrived to this state.

Some processes satisfy the Markov property, and some don't.

Example 6.5 (INTERNET CONNECTIONS). Let $X(t)$ be the total number of internet connections registered by some internet service provider by the time t. Typically, people connect to the internet at random times, regardless of how many connections have already been made. Therefore, the number of connections in a minute will only depend on the current number. For example, if 999 connections have been registered by 10 o'clock, then their total number will exceed 1000 during the next minute regardless of *when* and *how* these 999 connections were made in the past. This process is *Markov*. $\Diamond$

Example 6.6 (STOCK PRICES). Let $Y(t)$ be the value of some stock or some market index at time t. If we know $Y(t)$, do we also want to know $Y(t-1)$ in order to predict $Y(t+1)$? One may argue that if $Y(t-1) < Y(t)$, then the market is rising, therefore, $Y(t+1)$ is likely (but not certain) to exceed $Y(t)$. On the other hand, if $Y(t-1) > Y(t)$, we may conclude that the market is falling and may expect $Y(t+1) < Y(t)$. It looks like knowing the past *in addition* to the present did help us to predict the future. Then, this process is *not Markov*. $\Diamond$

Markov Chain

- A Markov chain is a discrete-time, discrete-state Markov stochastic process.

- Introduce a few convenient simplifications. The time is discrete, so let us define the time
- set as T = {0, 1, 2, . . .}. We can then look at a Markov chain as a random sequence {X(0),X(1),X(2), . . .} .

Markov Chain

- The state set is also discrete, so let us enumerate the states as 1, 2, . . . , n.
- Sometimes we'll start enumeration from state 0, and sometimes we'll deal with a Markov chain with infinitely many (discrete) states, then we'll have $n = \infty$.
- The Markov property means that only the value of $X(t)$ matters for predicting $X(t + 1)$,
- so the conditional probability
- $p_{ij}(t) = P\{X(t + 1) = j \mid X(t) = i\}$

DEFINITION 6.6

Probability $p_{ij}(t)$ in (6.2) is called a **transition probability**. Probability

$$p_{ij}^{(h)}(t) = \boldsymbol{P}\left\{X(t+h) = j \mid X(t) = i\right\}$$

of moving from state i to state j by means of h transitions is an h-**step transition probability**.

DEFINITION 6.7

A Markov chain is **homogeneous** if all its transition probabilities are independent of t. Being homogeneous means that transition from i to j has the same probability at any time. Then $p_{ij}(t) = p_{ij}$ and $p_{ij}^{(h)}(t) = p_{ij}^{(h)}$.

Example 6.7 (Weather forecasts). In some town, each day is either sunny or rainy. A sunny day is followed by another sunny day with probability 0.7, whereas a rainy day is followed by a sunny day with probability 0.4.

It rains on Monday. Make forecasts for Tuesday, Wednesday, and Thursday.

Example 6.8 (WEATHER, CONTINUED). Suppose now that it does not rain yet, but meteorologists predict an 80% chance of rain on Monday. How does this affect our forecasts?

Matrix Approach

All one-step transition probabilities p_{ij} can be conveniently written in an $n \times n$ *transition probability matrix*

$$\begin{array}{c|c} P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} & \begin{array}{c} \text{From} \\ \text{state:} \\ 1 \\ 2 \\ \vdots \\ n \end{array} \\ \hline \text{To state:} \quad 1 \quad 2 \quad \cdots \quad n & \end{array}$$

$$\textbf{2-step transition probability matrix} \quad \boxed{P^{(2)} = P \cdot P = P^2}$$

Each probability $P_h(j)$ is obtained when the entire row P_0 (initial distribution of X) is multiplied by the entire j-th column of matrix P. Hence,

$$\textbf{Distribution of } X(h) \quad \boxed{P_h = P_0 P^h} \tag{6.4}$$

$$= \sum_{k=1}^{n} p_{ik}^{(h-1)} p_{kj} = \left(p_{i1}^{(h-1)}, \ldots, p_{in}^{(h-1)}\right) \begin{pmatrix} p_{1j} \\ \vdots \\ p_{nj} \end{pmatrix}.$$

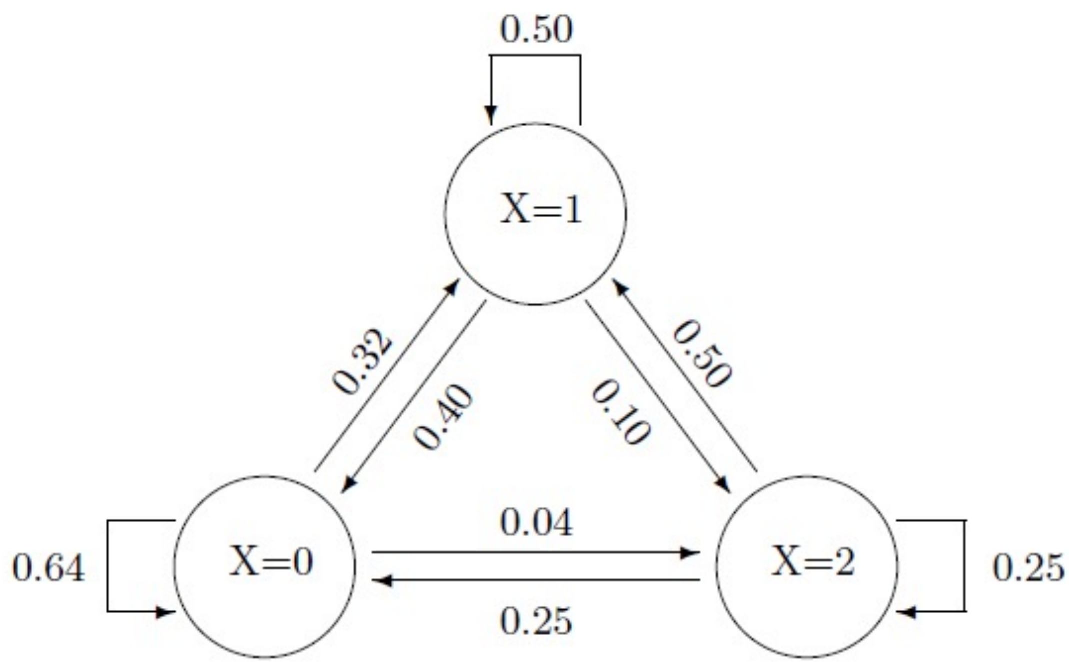
h-step transition probability matrix

$$\boxed{P^{(h)} = \underbrace{P \cdot P \cdot \ldots \cdot P}_{h \text{ times}} = P^h}$$

Each probability $P_h(j)$ is obtained when the entire row P_0 (initial distribution of X) is multiplied by the entire j-th column of matrix P. Hence,

$$\textbf{Distribution of } X(h) \qquad \boxed{P_h = P_0 P^h} \tag{6.4}$$

Example 6.9 (SHARED DEVICE). A computer is shared by 2 users who send tasks to a computer remotely and work independently. At any minute, any connected user may disconnect with probability 0.5, and any disconnected user may connect with a new task with probability 0.2. Let $X(t)$ be the number of concurrent users at time t (minutes). This is a Markov chain with 3 states: 0, 1, and 2.


0.50
X=1
0.32
0.40
0.10
0.50
0.64
X=0
0.04
0.25
X=2
0.25

Steady state distribution

DEFINITION 6.8

A collection of limiting probabilities

$$\pi_x = \lim_{h \to \infty} P_h(x)$$

is called a **steady-state distribution** of a Markov chain $X(t)$.

Steady-state distribution

$$\pi = \lim_{h \to \infty} P_h$$

is computed as a solution of

$$\begin{cases} \pi P &= \pi \\ \sum_x \pi_x &= 1 \end{cases}$$

Steady state

What is the steady state of a Markov chain? Suppose the system has reached its steady state, so that the current distribution of states is $P_t = \pi$. A system makes one more transition, and the distribution becomes $P_{t+1} = \pi P$. But $\pi P = \pi$, and thus, $P_t = P_{t+1}$. We see that *in a steady state, transitions do not affect the distribution*. A system may go from one state to another, but the distribution of states does not change. In this sense, it is *steady*.

Example 6.11 (WEATHER, CONTINUED). In Example 6.7 on p. 136, the transition probability matrix of sunny and rainy days is

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}.$$

The steady-state equation for this Markov chain is

$$\pi P = \pi,$$

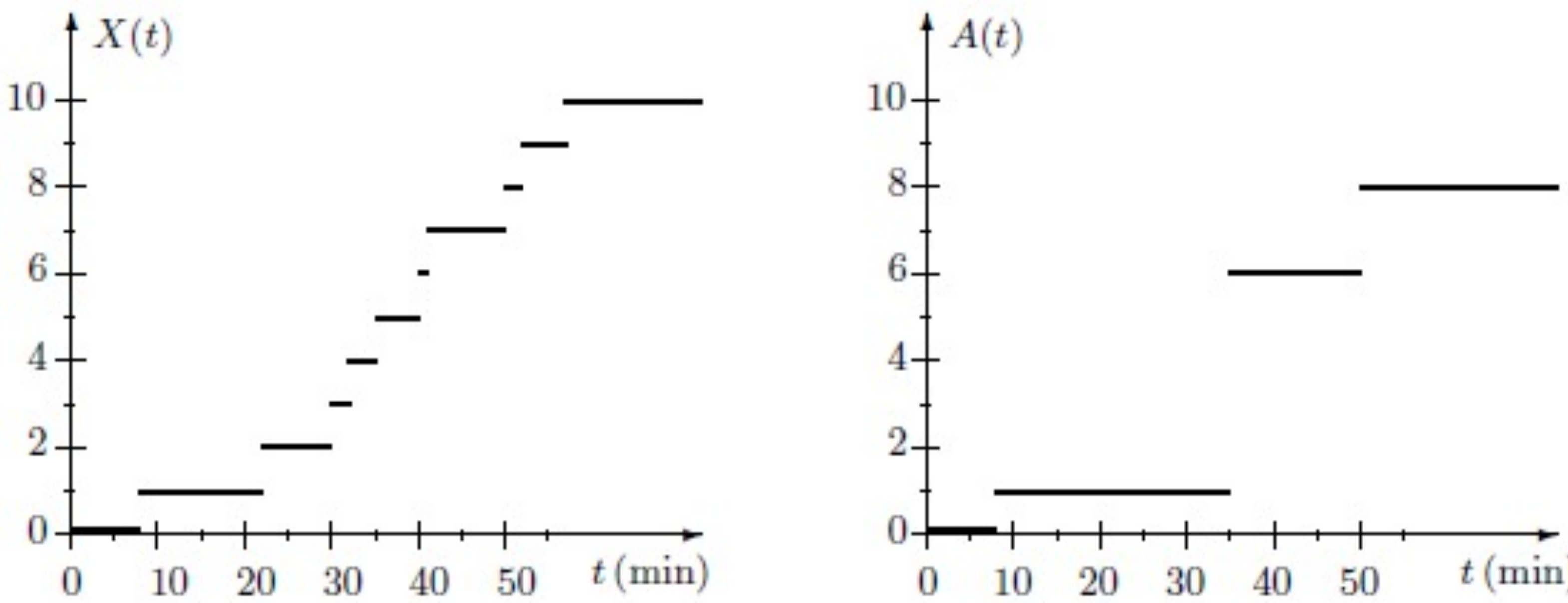
Counting Process

- These may be counts of arrived jobs, completed tasks, transmitted messages, detected errors, scored goals, and so on.

  A stochastic process X is **counting** if $X(t)$ is the number of items counted by the time t.

- As time passes, one can count additional items; therefore, sample paths of a counting process
- are always *non-decreasing*. Also, counts are *nonnegative integers*, $X(t) \in \{0, 1, 2, 3, ...\}$.
- Hence, all counting processes are *discrete-state*.

Example 6.17 (E-MAILS AND ATTACHMENTS). Figure 6.6 shows sample paths of two counting process, $X(t)$ being the number of transmitted e-mails by the time t and $A(t)$ being the number of transmitted attachments. According to the graphs, e-mails were transmitted at $t = 8, 22, 30, 32, 35, 40, 41, 50, 52$, and 57 min. The e-mail counting process $X(t)$ increments by 1 at each of these times. Only 3 of these e-mails contained attachments. One attachment was sent at $t = 8$, five more at $t = 35$, making the total of $A(35) = 6$, and two more attachments at $t = 50$, making the total of $A(50) = 8$. $\diamond$



Binomial Process

- Discrete-time Binomial process

Binomial process $X(n)$ is the number of successes in the first n independent Bernoulli trials, where $n = 0, 1, 2, \ldots$.

Arrival rate $\lambda = p/\Delta$ is the average number of successes per one unit of time. The time interval Δ of each Bernoulli trial is called a **frame**. The **interarrival time** is the time between successes.

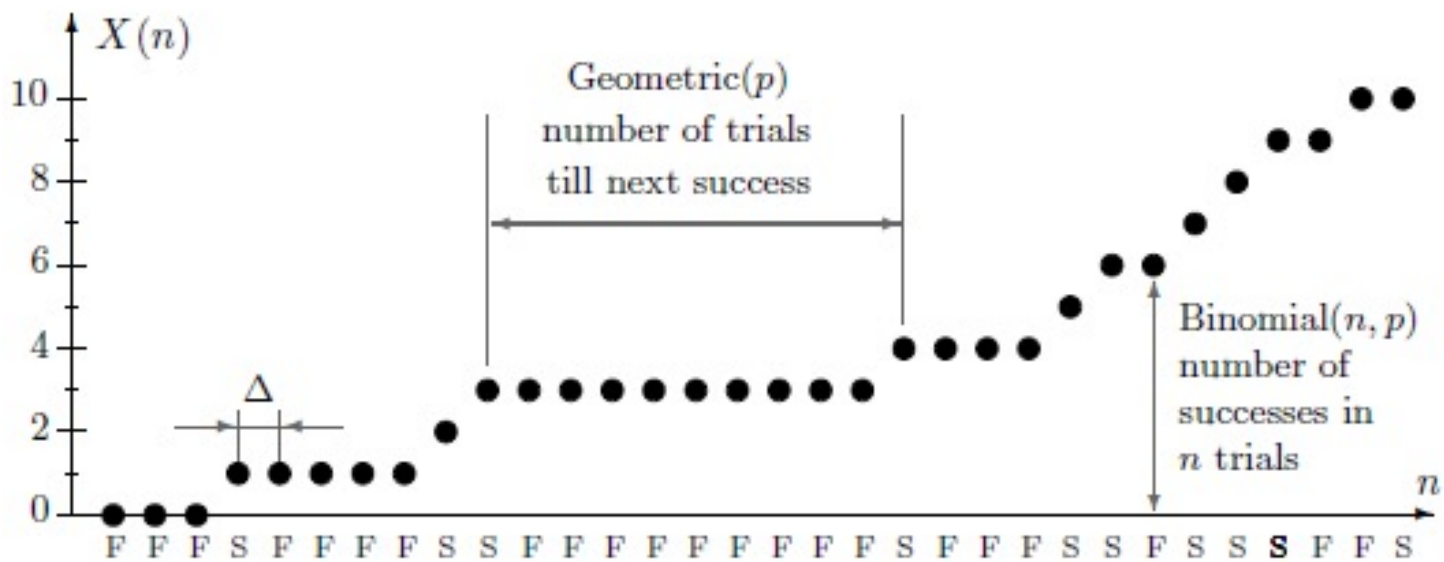


FIGURE 6.7: *Binomial process (sample path). Legend:* S = *success*, F=*failure.*

NOTATION	λ	=	arrival rate
	Δ	=	frame size
	p	=	probability of arrival (success) during one frame (trial)
	$X(t/\Delta)$	=	number of arrivals by the time t
	T	=	interarrival time

The interarrival period consists of a Geometric number of frames Y , each frame taking seconds. Hence, the interarrival time can be computed as
T = Y * Delta .

Binomial counting process

$$
\begin{aligned}
\lambda &= p/\Delta \\
n &= t/\Delta \\
X(n) &= Binomial(n, p) \\
Y &= Geometric(p) \\
T &= Y\Delta
\end{aligned}
$$

Binomial-Markov

Binomial counting process is *Markov*, with transition probabilities

$$p_{ij} = \begin{cases} p & \text{if} \quad j = i+1 \\ 1-p & \text{if} \quad j = i \\ 0 & \text{otherwise} \end{cases}$$

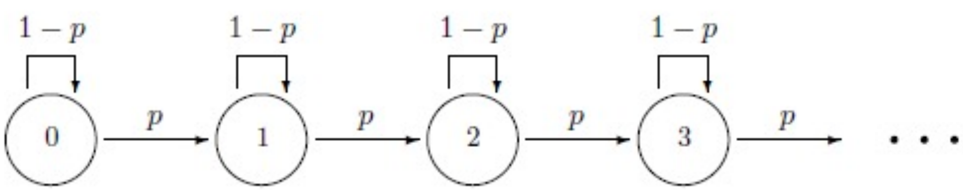


FIGURE 6.8: *Transition diagram for a Binomial count*

$$P = \begin{pmatrix} 1-p & p & 0 & 0 & \cdots \\ 0 & 1-p & p & 0 & \cdots \\ 0 & 0 & 1-p & p & \cdots \\ 0 & 0 & 0 & 1-p & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

Poisson

- continuous-time stochastic processes.
- The time variable t will run continuously through the whole interval, and thus, even during one minute there will be infinitely many moments when the process X(t) may change.
- Often a continuous-time process can be viewed as a limit of some discrete-time process
- whose frame size gradually decreases to zero, therefore allowing more frames during any fixed period of time.

Poisson Process

- A continuous time, discrete state process.
- N(t): no. of events occurring in time (0, t]. Events may be,
  1. # of packets arriving at a router port
  2. # of incoming telephone calls at a switch
  3. # of jobs arriving at file/compute server
  4. Number of component failures
- Events occurs successively and that intervals between these successive events are *iid* rvs, each following EXP( )

$$F(x) = 1 - \mathrm{e}^{-\lambda x}$$

  1. λ: arrival rate (1/ λ: average time between arrivals)
  2. λ: failure rate (1/ λ: average time between failures)

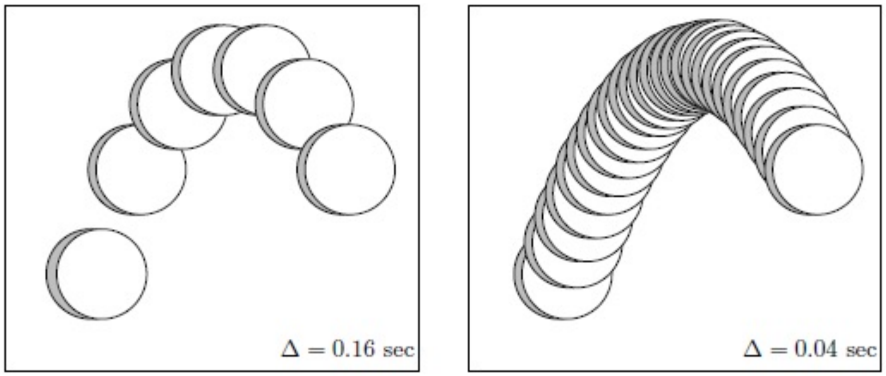


FIGURE 6.9: *From discrete motion to continuous motion: reducing the frame size Δ.*

Poisson process is a continuous-time counting stochastic process obtained from a Binomial counting process when its frame size Δ decreases to 0 while the arrival rate λ remains constant.

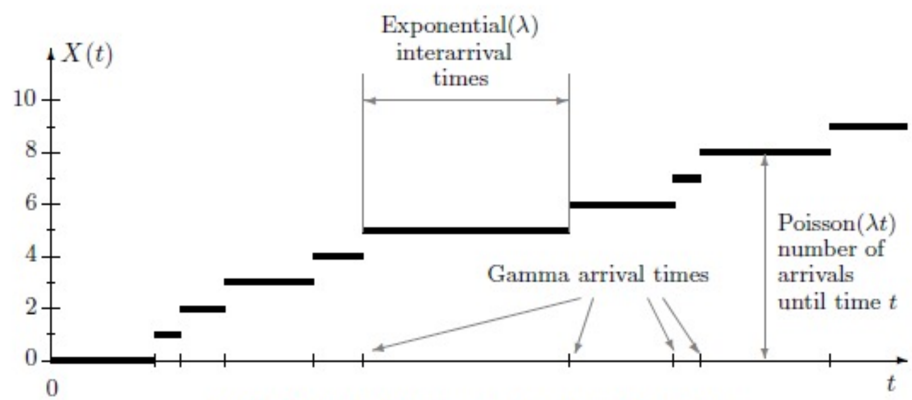
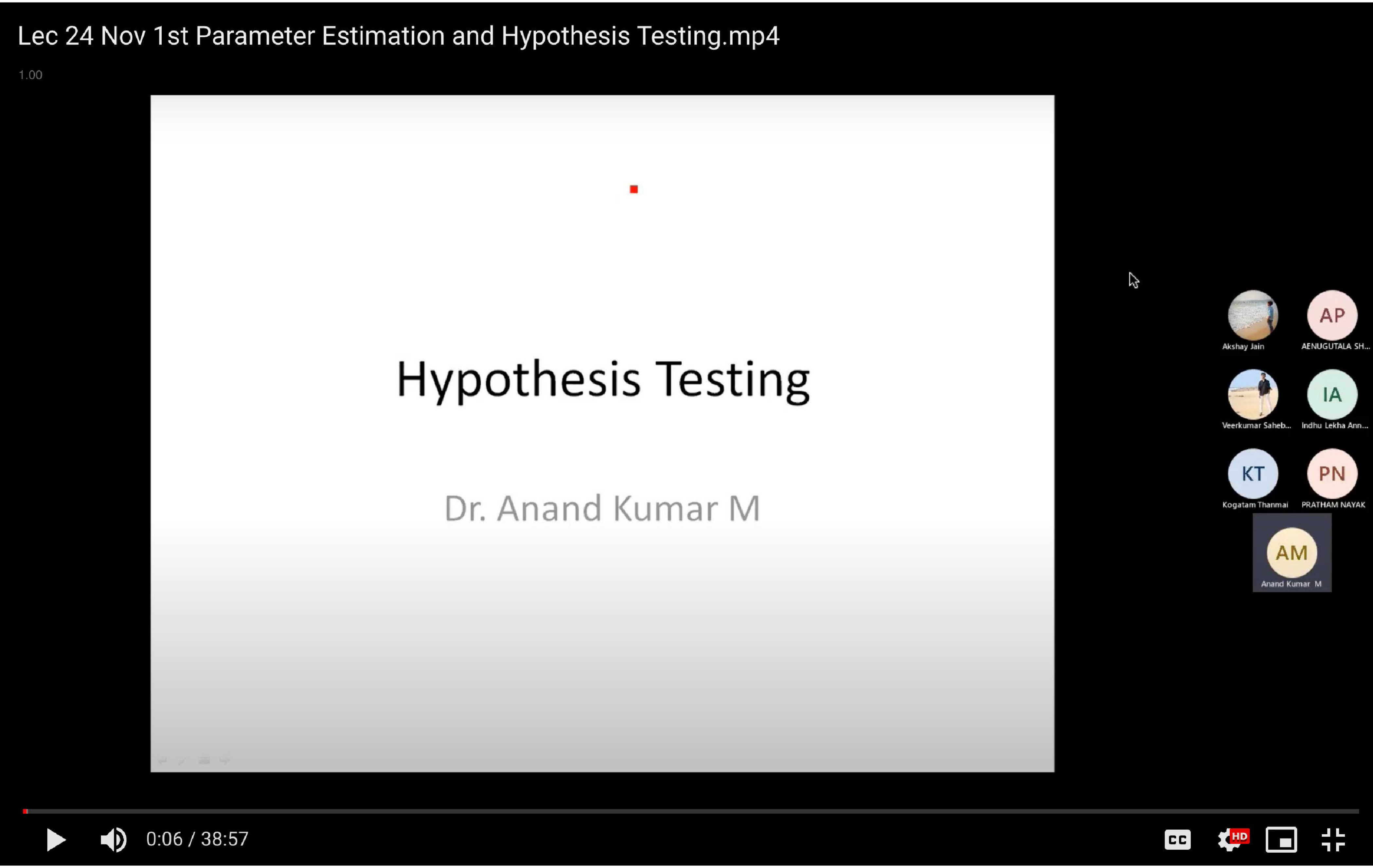


FIGURE 6.10: *Poisson process (sample path).*


Lec 24 Nov 1st Parameter Estimation and Hypothesis Testing.mp4
1.00
Hypothesis Testing
Dr. Anand Kumar M
AP
IA
KT
PN
AM
Akshay Jain
AENUGUTALA SH...
Veerkumar Saheb...
Indhu Lekha Ann...
Kogatam Thanmai
PRATHAM NAYAK
Anand Kumar M
0:06 / 38:57
HD

Lec 24 Nov 1st Parameter Estimation and Hypothesis Testing.mp4

5.00

Outline

- Parameter estimation
- Hypothesis testing

Akshay Jain
AENUGUTALA SH...
Veerkumar Saheb...
Indhu Lekha Ann...
Kogatam Thanmai
PRATHAM NAYAK
ALINA MELONY P...
Ritesh Sharma
CHARAN R
Anand Kumar M

1:52 / 38:57

Lec 24 Nov 1st Parameter Estimation and Hypothesis Testing.mp4

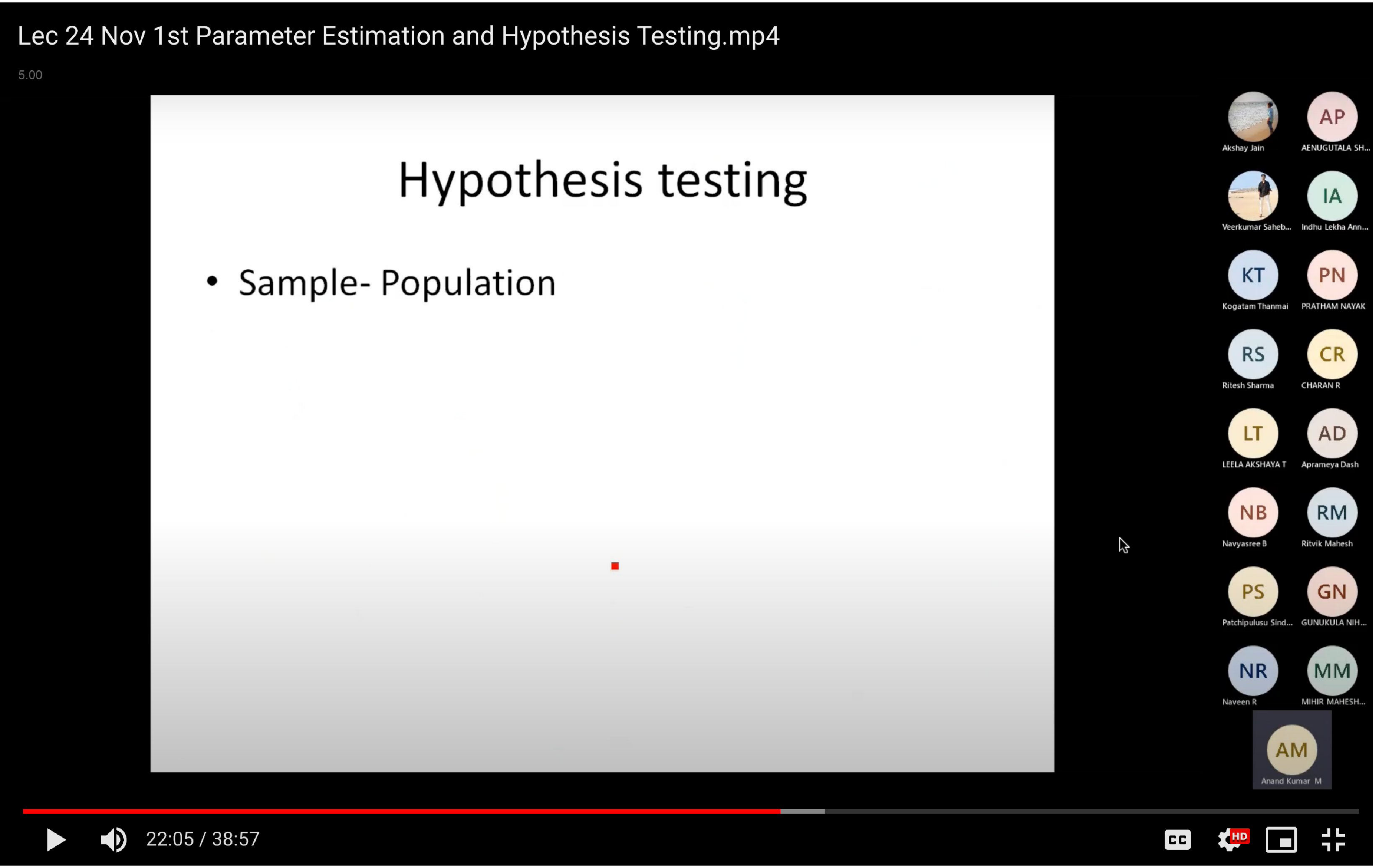
5.00

Parameter estimation

- X1..Xn   Random Sample
- Θ - set of unknown param

Akshay Jain
AENUGUTALA SH...
Veerkumar Saheb...
Indhu Lekha Ann...
Kogatam Thanmai
PRATHAM NAYAK
ALINA MELONY P...
Ritesh Sharma
CHARAN R
LEELA AKSHAYA T
Anand Kumar M

2:03 / 38:57


Lec 24 Nov 1st Parameter Estimation and Hypothesis Testing.mp4
5.00
Hypothesis testing
• Sample- Population
Akshay Jain
AP
AENUGUTALA SH...
Veerkumar Saheb...
IA
Indhu Lekha Ann...
KT
Kogatam Thanmai
PN
PRATHAM NAYAK
RS
Ritesh Sharma
CR
CHARAN R
LT
LEELA AKSHAYA T
AD
Aprameya Dash
NB
Navyasree B
RM
Ritvik Mahesh
PS
Patchipulusu Sind...
GN
GUNUKULA NIH...
NR
Naveen R
MM
MIHIR MAHESH...
AM
Anand Kumar M
22:05 / 38:57
HD

Lec 25 Nov 5th Significance Levels - Test Statistics.mp4

5.00

Press esc to exit full screen

Significance Levels

- Population distribution F_Θ
- Simple Hypothesis
- Complex Hypothesis

(a) $H_0 : \theta = 1$

(b) $H_0 : \theta \leq 1$

$-\alpha$

AM

Anand Kumar M

0:07 / 32:38

Critical region

- Suppose now that in order to test a specific null hypothesis *H0, a population sample of size n — say* $X_1, \ldots, X_n$ *— is to be observed.*
- *Based on these n values, we must decide* whether or not to accept *H0.*
- *A test for H0 can be specified by defining a region C in n-dimensional space with the proviso that the hypothesis is to be rejected if the random* sample *X1, . . . , Xn turns out to lie in C and accepted otherwise.*
- *The region C is called* the *critical region.*

Lec 25 Nov 5th Significance Levels - Test Statistics.mp4

5.00

Critical region

- *In other words, the statistical test determined by the critical region C* is the one that,

$$\text{accepts} \quad H_0 \quad \text{if} \quad (X_1, X_2, \ldots, X_n) \notin C$$

and

$$\text{rejects} \quad H_0 \quad \text{if} \quad (X_1, \ldots, X_n) \in C$$

For instance, a common test of the hypothesis that θ, the mean of a normal population with variance 1, is equal to 1 has a critical region given by

$$C = \{(X_1, \ldots, X_n) : |\bar{X} - 1| > 1.96/\sqrt{n}\} \qquad (8.2.1)$$

AM

Anand Kumar M

8:00 / 32:38

Lec 25 Nov 5th Significance Levels - Test Statistics.mp4

5.00

Types of errors

- It is important to note when developing a procedure for testing a given null hypothesis *H0 that, in any test, two different types of errors can result.*
- *The first of these, called a type I error, is said to result if the test incorrectly calls for rejecting H0 when it is indeed correct.*
- The second, called a *type II error, results if the test calls for accepting H0 when it is false*.

AM

Anand Kumar M

10:02 / 32:38

Lec 25 Nov 5th Significance Levels - Test Statistics.mp4

5.00

Level of significance of the test

- The objective of a statistical test of *H0 is not to explicitly* determine whether or not *H0 is true but rather to determine if its validity is consistent* with the resultant data.
- Hence, with this objective it seems reasonable that *H0 should* only be rejected if the resultant data are very unlikely when *H0 is true*.
- *The classical way* of accomplishing this is to specify a value *α and then require the test to have the property* that whenever ***H0 is true its probability of being rejected is never greater than α***.
- *The* value *α, called the level of significance of the test, is usually set in advance, with commonly* chosen values being *α = .1, .05, .005.*

AM

Anand Kumar M

12:38 / 32:38

Lec 25 Nov 5th Significance Levels - Test Statistics.mp4

5.00

Tests Concerning the Mean of a Normal Population

Since $\overline{X} = \sum_{i=1}^{n} X_i/n$ is a natural point estimator of μ, it seems reasonable to accept H_0 if $\overline{X}$ is not too far from μ_0. That is, the critical region of the test would be of the form

$$C = \{X_1, \ldots, X_n : |\overline{X} - \mu_0| > c\} \tag{8.3.1}$$

for some suitably chosen value c.

If we desire that the test has significance level α, then we must determine the critical value c in Equation 8.3.1 that will make the type I error equal to α. That is, c must be such that

$$P_{\mu_0}\{|\overline{X} - \mu_0| > c\} = \alpha \tag{8.3.2}$$

where we write P_{μ_0} to mean that the preceding probability is to be computed under the assumption that $\mu = \mu_0$. However, when $\mu = \mu_0$, $\overline{X}$ will be normally distributed with mean μ_0 and variance σ^2/n and so Z, defined by

$$Z \equiv \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}$$

AM

Anand Kumar M

16:45 / 32:38

Lec 25 Nov 5th Significance Levels - Test Statistics.mp4

5.00

$$P\left\{|Z| > \frac{c\sqrt{n}}{\sigma}\right\} = \alpha$$

or, equivalently,

$$2P\left\{Z > \frac{c\sqrt{n}}{\sigma}\right\} = \alpha$$

where Z is a standard normal random variable. However, we know that

$$P\{Z > z_{\alpha/2}\} = \alpha/2$$

and so

$$\frac{c\sqrt{n}}{\sigma} = z_{\alpha/2}$$

or

$$c = \frac{z_{\alpha/2}\sigma}{\sqrt{n}}$$

Thus, the significance level α test is to reject H_0 if $|\overline{X} - \mu_0| > z_{\alpha/2}\sigma/\sqrt{n}$ and accept otherwise; or, equivalently, to

$$\begin{aligned} &\text{reject} \quad H_0 \quad \text{if} \quad \frac{\sqrt{n}}{\sigma}|\overline{X} - \mu_0| > z_{\alpha/2} \\ &\text{accept} \quad H_0 \quad \text{if} \quad \frac{\sqrt{n}}{\sigma}|\overline{X} - \mu_0| \le z_{\alpha/2} \end{aligned} \qquad (8.3.3)$$

AM

Anand Kumar M

26:10 / 32:38

Lec 25 Nov 5th Significance Levels - Test Statistics.mp4

5.00

EXAMPLE 8.3a It is known that if a signal of value μ is sent from location A, then the value received at location B is normally distributed with mean μ and standard deviation 2. That is, the random noise added to the signal is an $N(0, 4)$ random variable. There is reason for the people at location B to suspect that the signal value $\mu = 8$ will be sent today. Test this hypothesis if the same signal value is independently sent five times and the average value received at location B is $\overline{X} = 9.5$.

$$\frac{\sqrt{n}}{\sigma}|\overline{X} - \mu_0| = \frac{\sqrt{5}}{2}(1.5) = 1.68$$

Since this value is less than $z_{.025} = 1.96$, *the hypothesis is accepted*

AM

Anand Kumar M

27:46 / 32:38

Lec 25 Nov 5th Significance Levels - Test Statistics.mp4

5.00

EXAMPLE 8.3b In Example 8.3a, suppose that the average of the 5 values received is $\overline{X} = 8.5$. In this case,

$$\frac{\sqrt{n}}{\sigma}|\overline{X} - \mu_0| = \frac{\sqrt{5}}{4} = .559$$

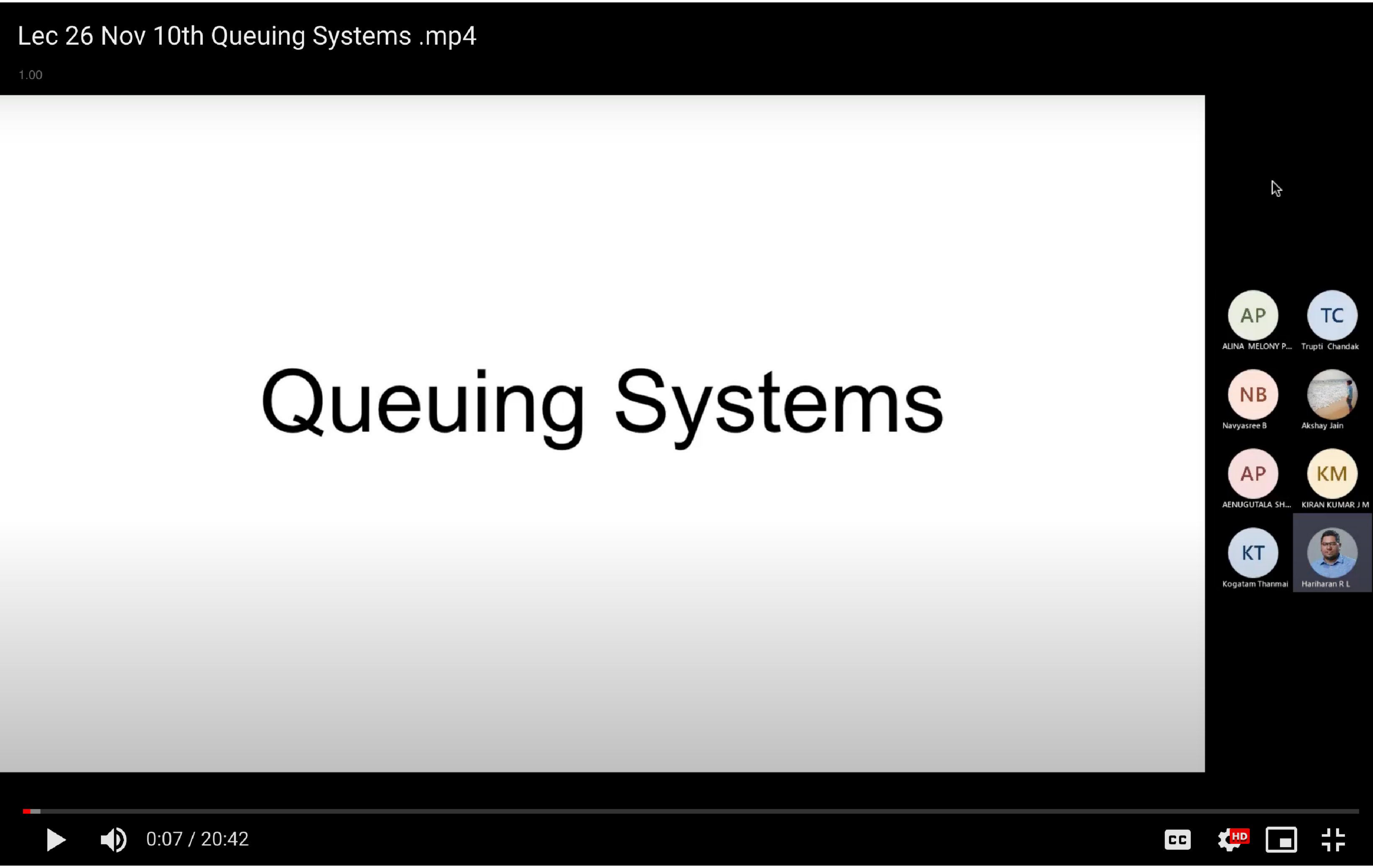
Since

$$\begin{aligned} P\{|Z| > .559\} &= 2P\{Z > .559\} \\ &= 2 \times .288 = .576 \end{aligned}$$

AM

Anand Kumar M

27:54 / 32:38


Lec 26 Nov 10th Queuing Systems .mp4
1.00
Queuing Systems
AP
ALINA MELONY P...
TC
Trupti Chandak
NB
Navyasree B
Akshay Jain
AP
AENUGUTALA SH...
KM
KIRAN KUMAR J M
KT
Kogatam Thanmai
Hariharan R L
0:07 / 20:42
HD

Lec 26 Nov 10th Queuing Systems .mp4

1.00

Introduction

A queuing system is
a facility consisting of one
or several servers
designed to perform
certain tasks or process
certain jobs and a queue
of jobs waiting to be processed

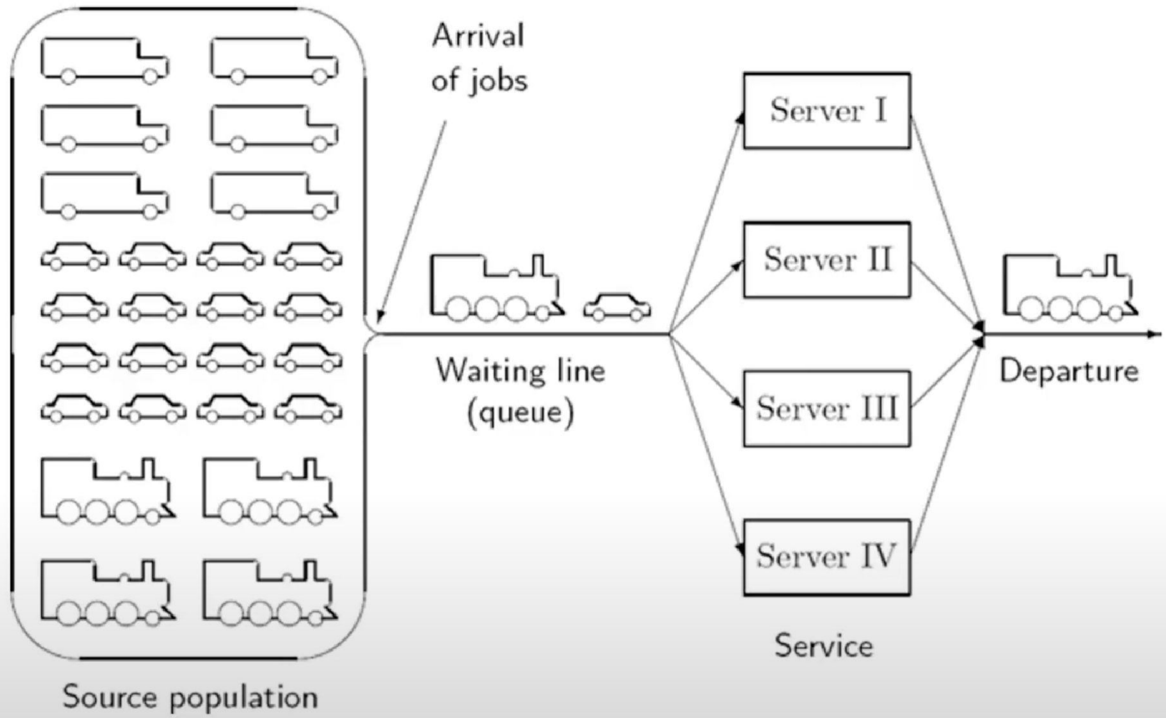


*Image from the book 'Probability and Statistics for Computer Scientists' By Michael Baron

AP ALINA MELONY P...
TC Trupti Chandak
NB Navyasree B
Akshay Jain
AP AENUGUTALA SH...
KM KIRAN KUMAR J M
KT Kogatam Thanmai
Hariharan R L

0:17 / 20:42

Lec 26 Nov 10th Queuing Systems .mp4

1.00

Examples of Queuing Systems

- a personal or shared computer executing tasks sent by its users;
- an internet service provider whose customers connect to the internet, browse, and disconnect;
- a printer processing jobs sent to it from different computers;
- a customer service with one or several representatives on duty answering calls from their customers;
- a TV channel viewed by many people at various times;
- a toll area on a highway, a fast food drive-through lane, or an automated teller machine (ATM) in a bank, where cars arrive, get the required service and depart;
- a medical office serving patients; and so on.

AP ALINA MELONY P...
TC Trupti Chandak
NB Navyasree B
Akshay Jain
AP AENUGUTALA SH...
KM KIRAN KUMAR J M
KT Kogatam Thanmai
NK NISHKA KOTIAN
Veerkumar Saheb...
SM Sanjeevi Meghana
AS ANAND SRIRAM
Hariharan R L

2:09 / 20:42

Lec 26 Nov 10th Queuing Systems .mp4

1.00

Parameters of a queuing system

NOTATION ||

$$
\begin{aligned}
\lambda_A &= \text{arrival rate}\\
\lambda_S &= \text{service rate}\\
\mu_A &= 1/\lambda_A = \text{mean interarrival time}\\
\mu_S &= 1/\lambda_S = \text{mean service time}\\
r &= \lambda_A/\lambda_S = \mu_S/\mu_A = \text{utilization, or arrival} - \text{to} - \text{service ratio}
\end{aligned}
$$

||

Random variables of a queuing system

NOTATION ||

$$
\begin{aligned}
X_s(t) &= \text{number of jobs receiving service at time } t\\
X_w(t) &= \text{number of jobs waiting in a queue at time } t\\
X(t) &= X_s(t) + X_w(t),\\
&\quad \text{the total number of jobs in the system at time } t\\
S_k &= \text{service time of the } k - \text{th job}\\
W_k &= \text{waiting time of the } k - \text{th job}\\
R_k &= S_k + W_k,\ \text{response time, the total time a job spend in the}\\
&\quad \text{system from its arrival until the departure}
\end{aligned}
$$

||

AP ALINA MELONY P...
TC Trupti Chandak
NB Navyasree B
Akshay Jain
AP AENUGUTALA SH...
KM KIRAN KUMAR J M
KT Kogatam Thanmai
NK NISHKA KOTIAN
Veerkumar Saheb...
SM Sanjeevi Meghana
AS ANAND SRIRAM
Hariharan R L

2:14 / 20:42

Lec 26 Nov 10th Queuing Systems .mp4

1.00

The Little's Law

- The Little's Law gives a simple relationship between the expected number of jobs, the expected response time, and the arrival rate. It is valid for any stationary queuing system.

$$\text{Little's Law} \quad \boxed{\lambda_A \mathrm{E}(R) = \mathrm{E}(X)}$$

- The Little's Law is universal, it applies to any stationary queuing system and even the system's components—the queue and the servers. Thus, we can immediately deduce the equations for the number of waiting jobs.

$$\mathrm{E}(X_w) = \lambda_A \mathrm{E}(W),$$

and for the number of jobs currently receiving service,

$$\mathrm{E}(X_s) = \lambda_A \mathrm{E}(S) = \lambda_{A\mu_S} = r.$$

We have obtained another important definition of *utilization*.

- Utilization r is the expected number of jobs receiving service at any given time.

AP ALINA MELONY P...
TC Trupti Chandak
NB Navyasree B
Akshay Jain
AP AENUGUTALA SH...
KM KIRAN KUMAR J M
KT Kogatam Thanmai
NK NISHKA KOTIAN
Veerkumar Saheb...
SM Sanjeevi Meghana
AS ANAND SRIRAM
Hariharan R L

2:19 / 20:42

Lec 26 Nov 10th Queuing Systems .mp4

1.00

Example on Little's Law

- (QUEUE IN A BANK). You walk into a bank at 10:00. Being there, you count a total of 10 customers and assume that this is the typical, average number. You also notice that on the average, customers walk in every 2 minutes. When should you expect to finish services and leave the bank?

Solution. We have $\mathbf{E}(X) = 10$ and $\mu_A = 2$ min. By the Little's Law,

$$E(R) = \frac{E(X)}{\lambda_A} = E(X)\mu_A = (10)(2) = \underline{20\ \text{min}}.$$

That is, your expected response time is 20 minutes, and you should expect to leave at 10:20.

3:44 / 20:42

Lec 26 Nov 10th Queuing Systems .mp4

1.00

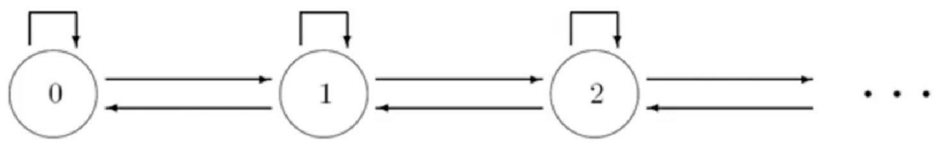
Bernoulli single-server queuing process

- The number of jobs in a queuing system, X(t), is called a ***queuing process***. In general, it is not a counting process because jobs arrive and depart, therefore, their number may increase and decrease whereas any counting process is nondecreasing.
- ***Bernoulli single-server queuing process*** is a discrete-time queuing process with the following characteristics:
  - one server
  - unlimited capacity
  - arrivals occur according to a Binomial process, and the probability of a new arrival during each frame is p_A
  - the probability of a service completion (and a departure) during each frame is p_S provided that there is at least one job in the system at the beginning of the frame
  - service times and interarrival times are independent

4:03 / 20:42

Lec 26 Nov 10th Queuing Systems .mp4

1.00



Transition diagram for a Bernoulli single-server queuing process.

- Bernoulli single-server queuing process is a *homogeneous Markov chain* because probabilities p_A and p_S never change. The number of jobs in the system increments by 1 with each arrival and decrements by 1 with each departure (as in figure). Conditions of a Binomial process guarantee that at most one arrival and at most one departure may occur during each frame. Then, we can compute all transition probabilities,

AP ALINA MELONY P...
TC Trupti Chandak
NB Navyasree B
KM KIRAN KUMAR J M
KT Kogatam Thanmai
NK NISHKA KOTIAN
Veerkumar Saheb...
SM Sanjeevi Meghana
AS ANAND SRIRAM
AP AENUGUTALA SH...
Hariharan R L

5:05 / 20:42

Lec 26 Nov 10th Queuing Systems .mp4

1.00

$$
\begin{aligned}
p_{00} &= P\{\text{no arrivals}\} &&= 1 - p_A \\
p_{01} &= P\{\text{new arrival}\} &&= p_A
\end{aligned}
$$

and for all $i \geq 1$,

$$
\begin{aligned}
p_{i,i-1} &= P\{\text{no arrivals} \cap \text{one departure}\} &&= (1-p_A)p_S \\
p_{i,i} &= \begin{array}{l} P\{\text{no arrivals} \cap \text{no departure}\} \\ +P\{\text{one arrival} \cap \text{one departure}\} \end{array} &&= (1-p_A)(1-p_S) + p_A p_S \\
p_{i,i+1} &= P\{\text{one arrival} \cap \text{no departures}\} &&= p_A(1-p_S)
\end{aligned}
$$

The transition probability matrix (of an interesting size $\infty \times \infty$) is three-diagonal,

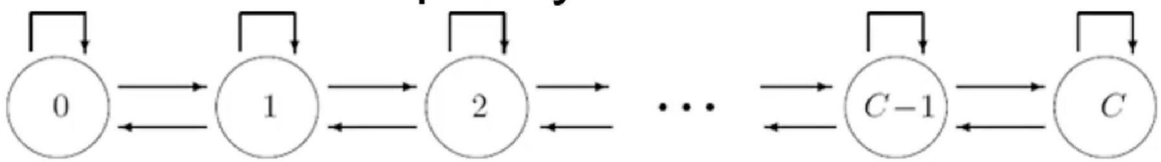
$$
P = \begin{pmatrix}
1-p_A & p_A & 0 & \cdots \\
(1-p_A)p_S & \begin{array}{c}(1-p_A)(1-p_S) \\ +p_A p_S\end{array} & p_A(1-p_S) & \cdots \\
0 & (1-p_A)p_S & \begin{array}{c}(1-p_A)(1-p_S) \\ +p_A p_S\end{array} & \cdots \\
0 & 0 & (1-p_A)p_S & \ddots \\
\vdots & \vdots & \ddots & \ddots
\end{pmatrix} \quad (7.4)
$$

5:39 / 20:42

Lec 26 Nov 10th Queuing Systems .mp4

1.00

Systems with limited capacity



Transition diagram for a Bernoulli single-server queuing process with limited capacity.

- As we see, the number of jobs in a Bernoulli single-server queuing system may potentially reach any number. However, many systems have limited resources for storing jobs. Then, there is a maximum number of jobs C that can possibly be in the system simultaneously. This number is called *capacity*.
- How does limited capacity change the behavior of a queuing system? Until the capacity C is reached, the system operates without any limitation, as if $C = \infty$. All transition probabilities are the same as in figure.
- The situation changes only when X = C. At this time, the system is full; it can accept new jobs into its queue only if some job departs. As before, the number of jobs decrements by 1 if there is a departure and no new arrival,
  - $p_{C,C-1}=(1-p_A)p_S$.
- In *all* other cases, the number of jobs remains at $X = C$. If there is no departure during some frame, and a new job arrives, this job cannot enter the system. Hence,
  - $p_{C,C}=(1-p_A)(1-p_S)+p_Ap_S+p_A(1-p_S)=1-(1-p_A)p_S$.

7:55 / 20:42

Lec 26 Nov 10th Queuing Systems .mp4

1.00

M/M/1 System

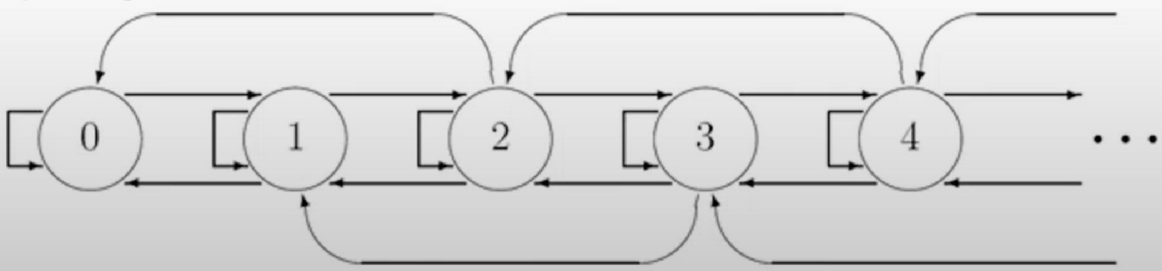
- We now turn our attention to continuous-time queuing processes. Our usual approach is to move from discrete time to continuous time gradually by reducing the frame size Δ to zero.
- A queuing system can be denoted as A/S/n/C where
  - A denotes distribution of inter arrival times
  - S denotes the distribution of service times
  - n is the number of servers
  - C is the capacity (Default capacity is ∞ (unlimited capacity))
- An M/M/1 queuing process is a continuous-time Markov queuing process with the following characteristics,
  - one server;
  - unlimited capacity;
  - Exponential interarrival times with the arrival rate λA;
  - Exponential service times with the service rate λS;
  - service times and interarrival times are independent.

AP ALINA MELONY P...
TC Trupti Chandak
NB Navyasree B
KM KIRAN KUMAR J M
KT Kogatam Thanmai
NK NISHKA KOTIAN
Veerkumar Saheb...
SM Sanjeevi Meghana
AP AENUGUTALA SH...
AS ANAND SRIRAM
NS NAVEEN SHENOY
Hariharan R L

10:32 / 20:42

Multiserver queuing systems

- We now turn our attention to queuing systems with several servers. We assume that each server can perform the same range of services; however, in general, some servers may be faster than others. Thus, the service times for different servers may potentially have different distributions.
- **Bernoulli k-server queuing process** is a discrete-time queuing process with the following characteristics:
  - k servers
  - unlimited capacity
  - arrivals occur according to a Binomial counting process; the probability of a new arrival during each frame is p_A
  - during each frame, each busy server completes its job with probability p_S independently of the other servers and independently of the process of arrivals



Lec 26 Nov 10th Queuing Systems .mp4

1.00

M/M/K systems

- An **M/M/k queuing process** is a continuous-time Markov queuing process with
  - k servers
  - unlimited capacity
  - Exponential interarrival times with the arrival rate λ_A
  - Exponential service time for each server with the service rate λ_S, independent of all the arrival times and the other servers
- **Unlimited number of servers and M/M/∞**
  - An unlimited number of servers completely eliminates the waiting time. Whenever a job arrives, there will always be servers available to handle it, and thus, the response time R consists of the service time only.

AP ALINA MELONY P...
TC Trupti Chandak
NB Navyasree B
KM KIRAN KUMAR J M
KT Kogatam Thanmai
NK NISHKA KOTIAN
AP AENUGUTALA SH...
AS ANAND SRIRAM
NS NAVEEN SHENOY
Hariharan R L

14:16 / 20:42

Lec 26 Nov 10th Queuing Systems .mp4

1.00

Simulation of Queuing Systems

- Queuing theory does not cover all the possible situations. On the other hand, we can simulate the behavior of almost any queuing system and study its properties by Monte Carlo methods.
- Markov case : A queuing system is *Markov* only when its interarrival and service times are memoryless. Then the future can be predicted from the present without relying on the past
- Monte Carlo methods of let us simulate and evaluate rather complex queuing systems far beyond Bernoulli and M/M/k. As long as we know the distributions of inter-arrival and service times, we can generate the processes of arrivals and services. To assign jobs to servers, we keep track of servers that are available each time when a new job arrives. When all the servers are busy, the new job will enter a queue.
- As we simulate the work of a queuing system, we keep records of events and variables that are of interest to us. After a large number of Monte Carlo runs, we average our records in order to estimate probabilities by long-run proportions and expected values by long-run averages

AP ALINA MELONY P...
TC Trupti Chandak
NB Navyasree B
KM KIRAN KUMAR J M
KT Kogatam Thanmai
NK NISHKA KOTIAN
AP AENUGUTALA SH...
AS ANAND SRIRAM
NS NAVEEN SHENOY
Hariharan R L

15:21 / 20:42

Lec 26 Nov 10th Queuing Systems .mp4

1.00

Examples of Queuing Systems

- a personal or shared computer executing tasks sent by its users;
- an internet service provider whose customers connect to the internet, browse, and disconnect;
- a printer processing jobs sent to it from different computers;
- a customer service with one or several representatives on duty answering calls from their customers;
- a TV channel viewed by many people at various times;
- a toll area on a highway, a fast food drive-through lane, or an automated teller machine (ATM) in a bank, where cars arrive, get the required service and depart;
- a medical office serving patients; and so on.

Lec 26 Nov 10th Queuing Systems .mp4

1.00

Summary and Conclusions

- Queuing systems are service facilities designed to serve randomly arriving jobs.
- Several classes of queuing systems were studied. We discussed discrete-time and continuous-time, Markov and non-Markov, one-server and multiserver queuing processes with limited and unlimited capacity.
- Performance of more complicated and advanced queuing systems can be evaluated by Monte Carlo methods. One needs to simulate arrivals of jobs, assignment of servers, and service times and to keep track of all variables of interest. A sample computer code for such a simulation is given.

AP ALINA MELONY P...
TC Trupti Chandak
NB Navyasree B
KM KIRAN KUMAR J M
KT Kogatam Thanmai
NK NISHKA KOTIAN
AP AENUGUTALA SH...
AS ANAND SRIRAM
NS NAVEEN SHENOY
Hariharan R L

18:13 / 20:42

Parametric and Non Parametric tests

Anand Kumar M

Outline

- **Hypothesis**
- Parametric and Non Parametric test
- Z-test
  – One Sample
  – Two Sample
- T-test
  – One Sample
  – Two Sample
- Chi Square Test

Hypothesis

- **Hypothesis: is a statement given by an individual.**
- **Usually it is required to make decisions about populations on the basis of sample information.**
- **Such decisions are called *Statistical Decisions.***
- ***In attempting to reach decisions it is often necessary to make assumption about population involved.***
- ***Such assumptions, which are not necessarily true, are called statistical hypothesis.***

Hypothesis

- **Parametric Hypothesis: A statistical hypothesis which refers only to values of unknown parameters of population.**
- **Null Hypothesis and Alternative Hypothesis:**
- **A hypothesis which is tested under the assumption that it is true is called a *null hypothesis and is denoted by H0.***
- ***The hypothesis which differs from the given Null Hypothesis H0 and is accepted when H0 is rejected is called an alternative hypothesis and is denoted by H1***
- ***The hypothesis against which we test the null hypothesis, is an alternative hypothesis***

Simple and Composite Hypothesis

- **A parametric hypothesis which describes a distribution completely is called a *simple hypothesis* *otherwise it is called composite hypothesis. For example;***
- In case of Normal Distribution N (μ, σ2), μ = 5, σ = 3 is simple hypothesis whereas μ = 5 is a composite hypothesis as nothing have been said about σ.
- Similarly, μ < 5, σ = 3 is a composite hypothesis.
- Let H0: μ = 5 be the null hypothesis, then
- H1: μ ≠ 5 is two sided composite alternative hypothesis.
- H1: μ < 5 is one sided (Left) composite alternative hypothesis.
- H1: μ > 5 is one sided (Right) composite alternative hypothesis

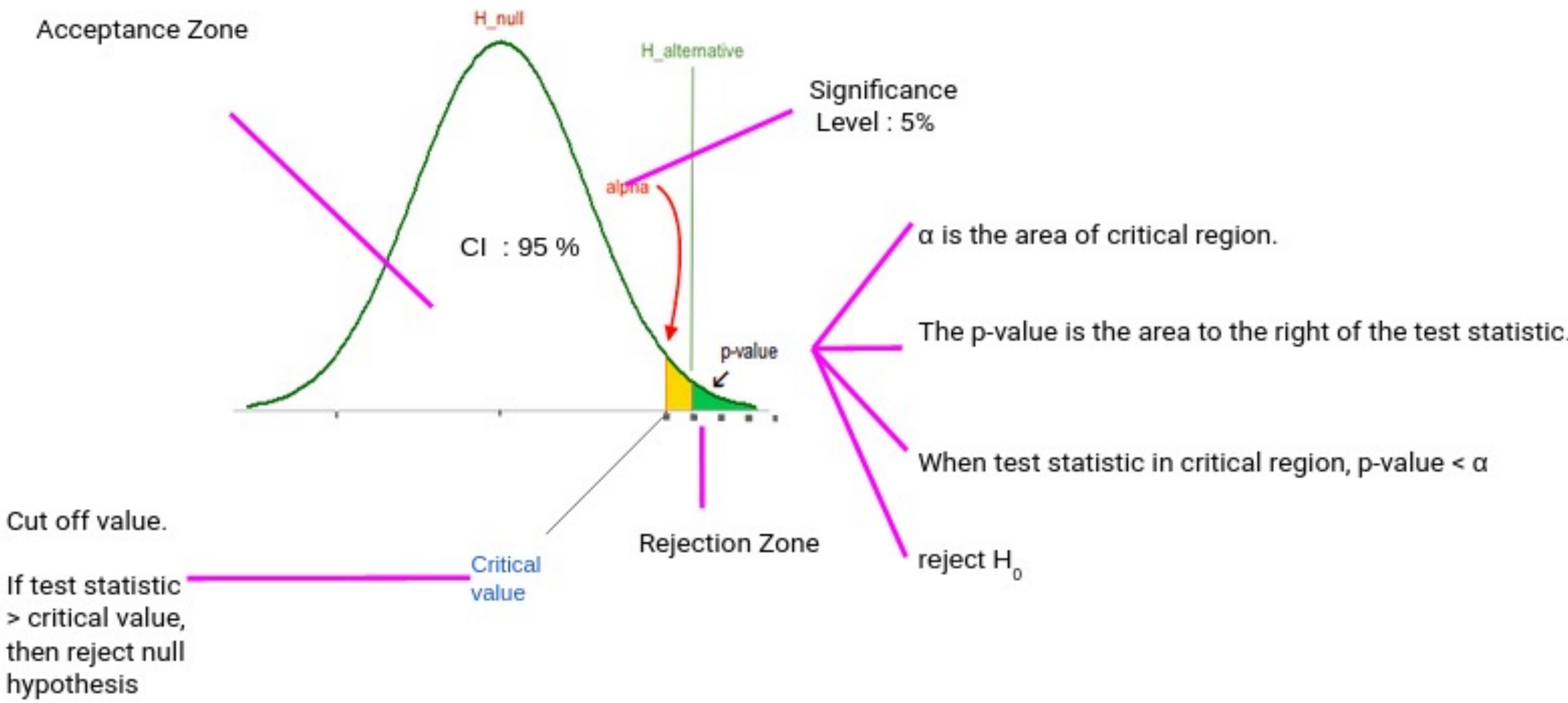
Tests of Significance

- Tests of Significance: Procedure which enables us to decide, on the basis of sample information whether to accept or reject the hypothesis or
- to determine whether observed sampling results differ significantly from expected results are called *tests of significance, rules of decisions or tests of hypothesis.*
- Level of Significance: The probability level below which we reject the hypothesis is called *level of significance. The levels of significance usually employed in testing of hypothesis are 5% and 1%.*

Critical Region and Acceptance Region

- **A region (corresponding to a statistic t) is called the sample space.**
- **The part of sample space which amounts to rejection of null hypothesis H0, is called *critical region or region of rejection.***
- If X=(x1,x2,…,xn) is the random vector observed and Wc is the critical region (which corresponds the rejection of the hypothesis according to a prescribed test procedure) of the sample space W, then Wa = W – Wc of the sample space is called the *acceptance region.*

Critical Value, p-value

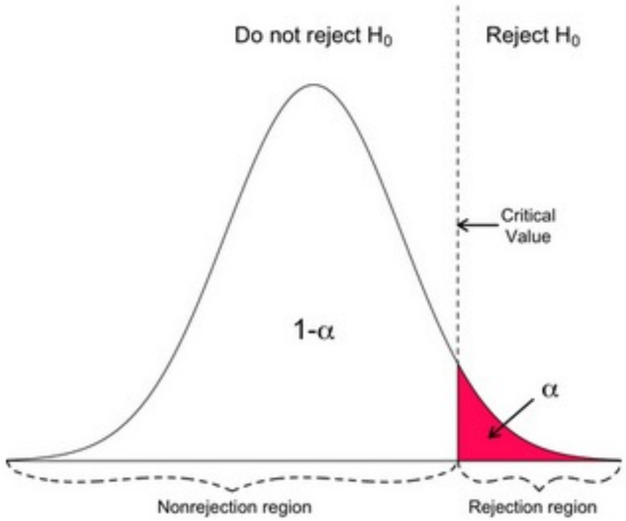


https://www.analyticsvidhya.com/blog/2020/06/statistics-analytics-hypothesis-testing-z-test-t-test/

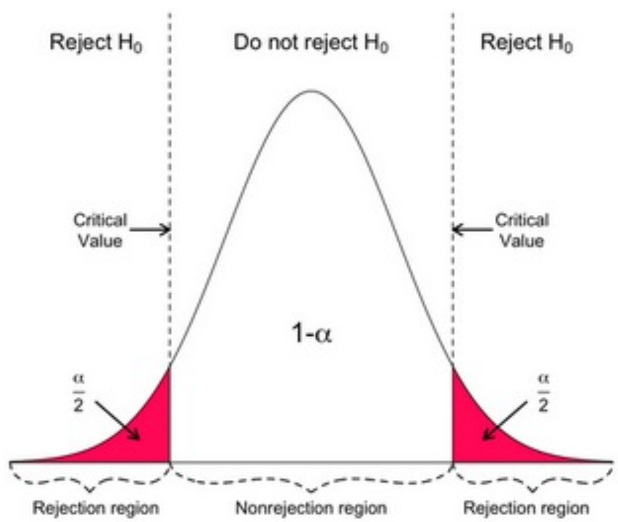
Critical Value, p-value

- Typically, we set the Significance level at 10%, 5%, or 1%.
- If our test score lies in the Acceptance Zone we fail to reject the Null Hypothesis.
- If our test score lies in the critical zone, *we reject the Null Hypothesis and accept the Alternate Hypothesis*.
- Critical Value is the cut off value between Acceptance Zone and Rejection Zone.
- We compare our test score to the critical value and if the test score is greater than the critical value, that means our test score lies in the Rejection Zone and we reject the Null Hypothesis.
- On the opposite side, if the test score is less than the Critical Value, that means the test score lies in the Acceptance Zone and we fail to reject the null Hypothesis.

Directional and Non Directional Hyphothesis



Ex. Females will score significantly higher than Males on IQ score

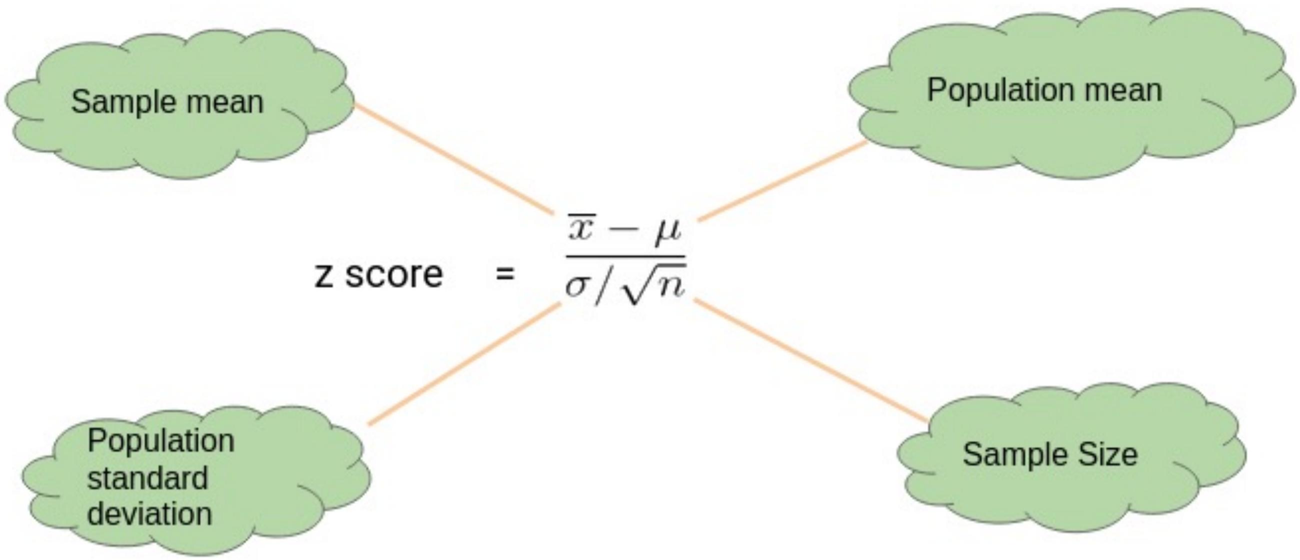


Ex. Females will score significantly different than Males on IQ score

Z Test

- z-test is the statistical calculation that can be used to compare population averages to a sample's. The z-test will tell you how far, in **standard deviations** terms, a data point is from the average of a data set.
- z tests are a statistical way of testing a hypothesis when either:
- We know the population variance, or We do not know the population variance but our sample size is large $n \geq 30$
- *If we have a sample size of less than 30 and do not know the population variance, then we must use a t-test.*
-

One Sample Z test



Example

- Let's say we need to determine if girls on average score higher than 600 in the exam. We have the information that the standard deviation for girls' scores is 100. So, we collect the data of 20 girls by using random samples and record their marks. Finally, we also set our α value (significance level) to be 0.05.

Score
650
730
510
670
480
800
690
530
590
620
710
670
640
780
650
490
800
600
510
700

In this example:
Mean Score for Girls is 641
The size of the sample is 20
The population mean is 600
Standard Deviation for Population is 100

Example

$$\text{z score} = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{641 - 600}{100/\sqrt{20}}$$

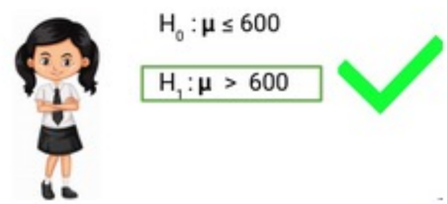
$$= 1.8336$$

$$\text{p value} = .033357.$$

Critical Value = 1.645

Z score > Critical Value

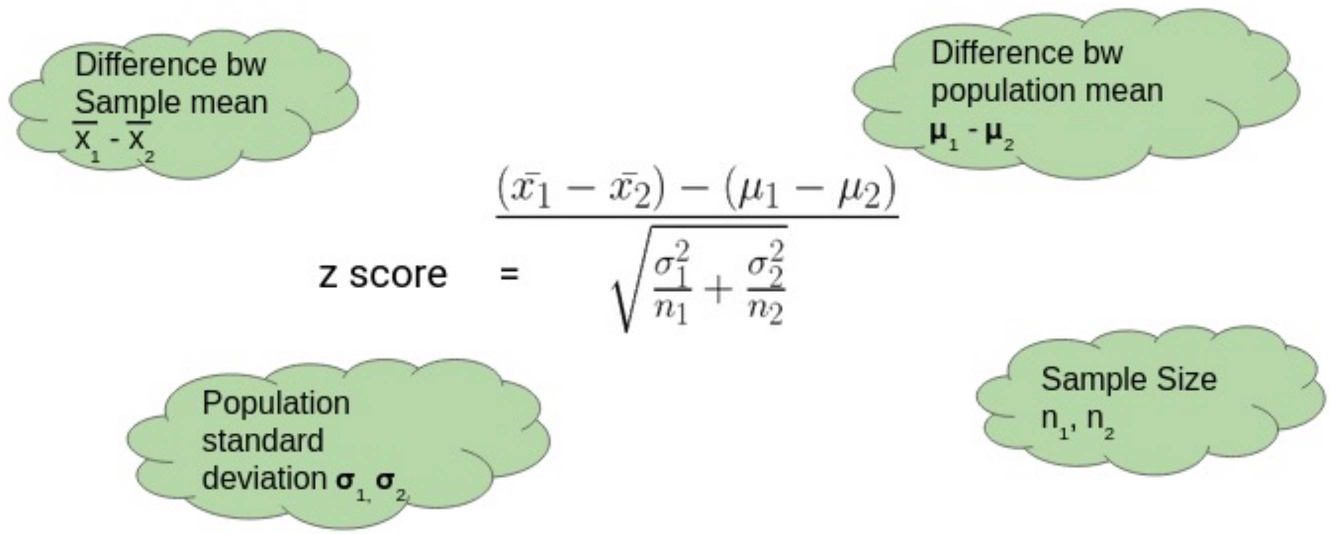
P value < 0.05



- **Since the P-value is less than 0.05, we can reject the null hypothesis** and conclude based on our result that Girls on average scored higher than 600.

Two Sample Z Test

- We perform a Two Sample Z test when we want to compare **the mean of two samples.**



Example

- Here, let's say we want to know if *Girls on average score 10 marks* more than the boys. We have the information that the standard deviation for girls' Score is **100 and for boys' score is 90**. Then we collect the data of 20 girls and 20 boys by using random samples and record their marks. Finally, we also set our α value (significance level) to be 0.05.

Mean Score for Girls (Sample Mean) is 641
Mean Score for Boys (Sample Mean) is 613.3
Standard Deviation for the Population of Girls' is 100
Standard deviation for the Population of Boys' is 90
Sample Size is 20 for both Girls and Boys
Difference between Mean of Population is 10

Score
650
730
510
670
480
800
690
530
590
620
710
670
640
780
650
490
800
600
510
700

Score
630
720
462
631
440
783
673
519
543
579
677
649
632
768
615
463
781
563
488
650

Example

$$\text{z score} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{(641 - 613.3) - (10)}{\sqrt{\frac{100^2}{20} + \frac{90^2}{20}}}$$

$$= 0.588$$

P value = 0.278

Critical Value = 1.645

Z score < Critical Value

P value > 0.05

$H_0 : \mu_1 - \mu_2 \leq 10$

$\boxed{H_1 : \mu_1 - \mu_2 > 10}$

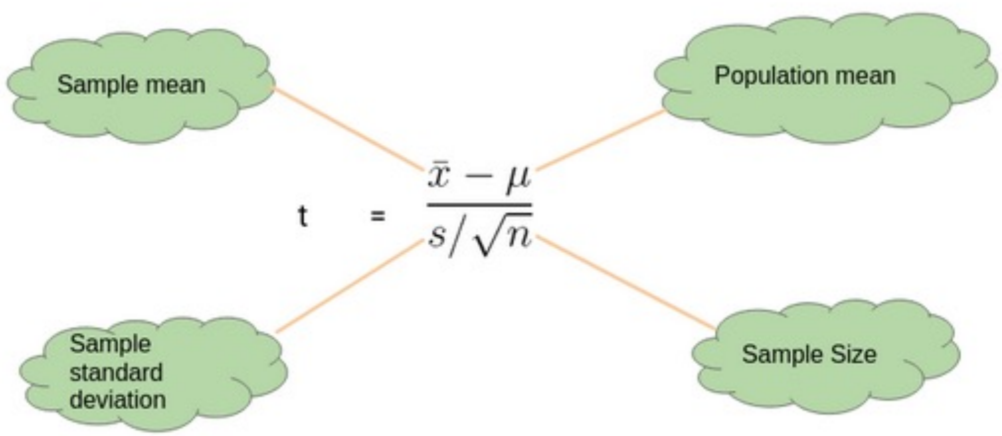
Thus, we can **conclude based on the P-value that we fail to reject the Null Hypothesi**s

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451

T-test

- **T-tests** are also calculations that can be used to test a hypothesis, but they are very useful when we need to determine if there is a statistically significant comparison between the 2 independent sample groups.
- In other words, a t-test asks whether the comparison between the averages of 2 groups is unlikely to have occurred due to random chance.
- Usually, t-tests are more appropriate when dealing with problems with a limited **sample size** (i.e., $n < 30$).

One-Sample t-test

- We perform a One-Sample t-test when we want to **compare a sample mean with the population mean**.
- The difference from the Z Test is that we do **not have the information on Population Variance** here. We use the **sample standard deviation** instead of population standard deviation in this case.
- 

Example

- Let’s say we want to determine if on average girls score more than 600 in the exam. We do not have the information related to variance (or standard deviation) for girls’ scores. To a perform t-test, we randomly collect the data of 10 girls with their marks and choose our α value (significance level) to be 0.05 for Hypothesis Testing.
-

Mean Score for Girls is 606.8
The size of the sample is 10
The population mean is 600
Standard Deviation for the sample is 13.14

Girls_Score
587
602
627
610
619
622
605
608
596
592

Example

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{606.8 - 600}{13.14/\sqrt{10}}$$

$$= 1.64$$

Critical Value = 1.833

t score < Critical Value

P value = 0.0678

P value > 0.05

$H_0 : \mu \leq 600$

$\boxed{H_1 : \mu > 600}$

P-value is greater than 0.05 thus we fail to reject the null hypothesis

Two-Sample t-Test

Difference bw
Sample mean
$\overline{x}_1 - \overline{x}_2$

Difference bw
population mean
$\mu_1 - \mu_2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Sample standard
deviation s_1, s_2

Sample Size
n_1, n_2

Example

- Here, let’s say we want to determine if on average, boys score 15 marks more than girls in the exam. We do not have the information related to variance (or standard deviation) for girls’ scores or boys’ scores. To perform a t-test. we randomly collect the data of 10 girls and boys with their marks. We choose our α value (significance level) to be 0.05 as the criteria for Hypothesis Testing.

Mean Score for Boys is 630.1
Mean Score for Girls is 606.8
Difference between Population Mean 15
Standard Deviation for Boys’ score is 13.42
Standard Deviation for Girls’ score is 13.14

Girls_Score
587
602
627
610
619
622
605
608
596
592

Boys_Score
626
643
647
634
630
649
625
623
617
607

Example

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\frac{(630.1 - 606.8) - (15)}{\sqrt{\frac{(13.42)^2}{10} + \frac{(13.14)^2}{10}}}$$

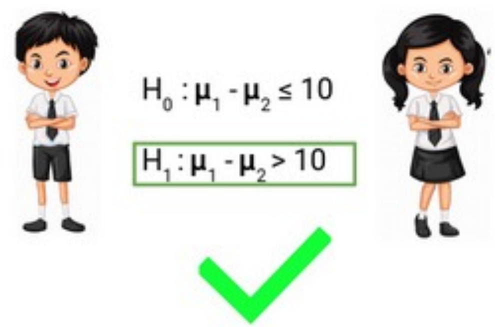
Critical Value = 1.833

t = 2.23

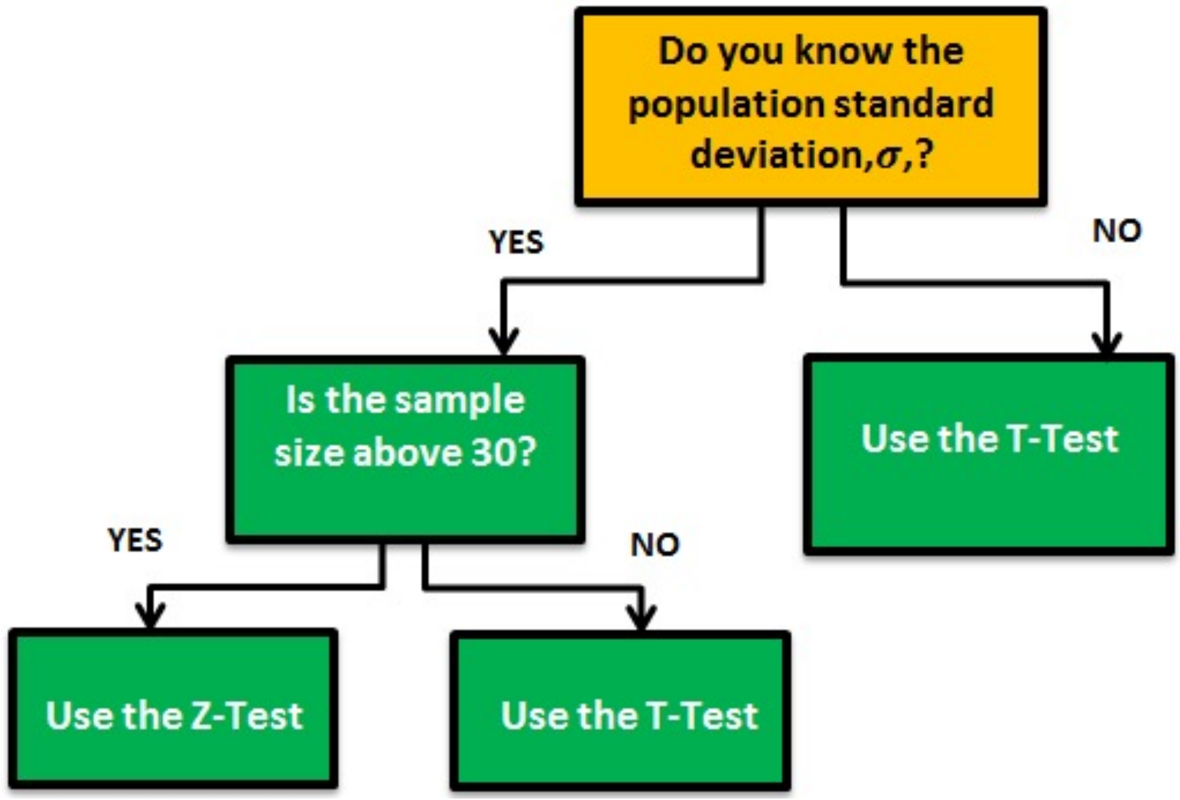
P value = 0.019

Critical Value > t score

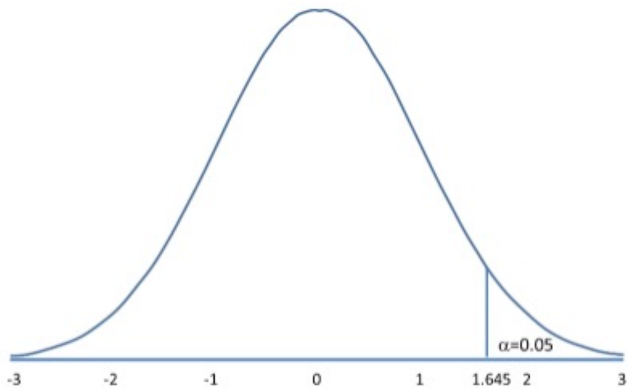
P value < 0.05



P-value is less than 0.05 so we can reject the null hypothesis


Do you know the population standard deviation,σ,?
YES
NO
Is the sample size above 30?
Use the T-Test
YES
NO
Use the Z-Test
Use the T-Test

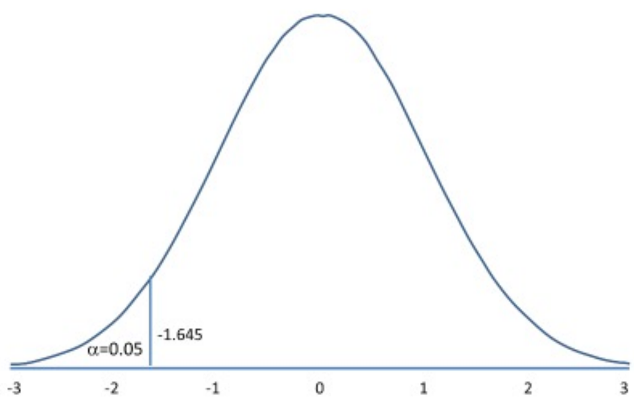
Critical value from Significance Level



Rejection Region for Upper-Tailed Z Test (H_1: $\mu > \mu_0$) with $\alpha=0.05$

The decision rule is: Reject H_0 if $Z \geq 1.645$.

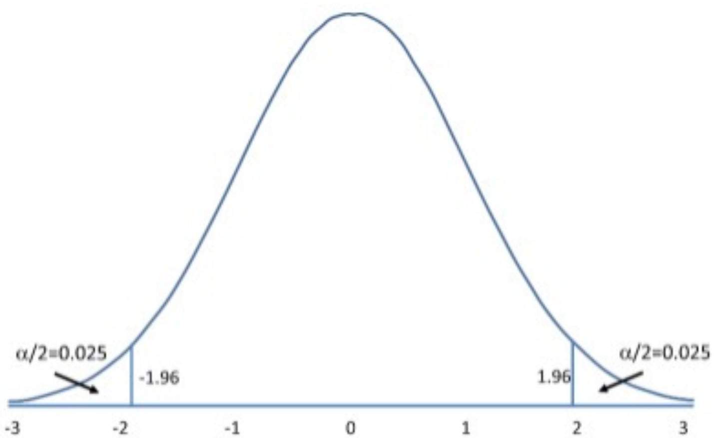
Upper-Tailed Test	
α	Z
0.10	1.282
0.05	1.645
0.025	1.960
0.010	2.326
0.005	2.576
0.001	3.090
0.0001	3.719



Rejection Region for Lower-Tailed Z Test (H_1: $\mu < \mu_0$) with $\alpha = 0.05$

The decision rule is: Reject H_0 if $Z \leq 1.645$.

Lower-Tailed Test	
a	Z
0.10	-1.282
0.05	-1.645
0.025	-1.960
0.010	-2.326
0.005	-2.576
0.001	-3.090
0.0001	-3.719



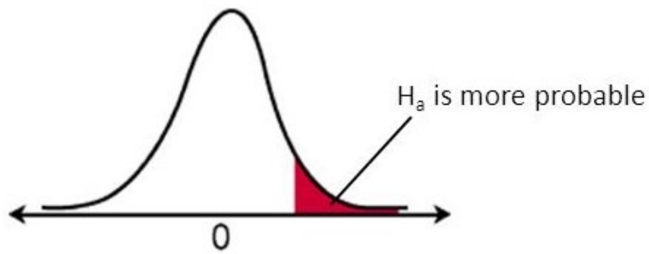
Rejection Region for Two-Tailed Z Test (H_1: $\mu \neq \mu_0$) with α =0.05

The decision rule is: Reject H_0 if $Z \leq -1.960$ or if $Z \geq 1.960$.

Two-Tailed Test	
α	**Z**
0.20	1.282
0.10	1.645
0.05	1.960
0.010	2.576
0.001	3.291
0.0001	3.819

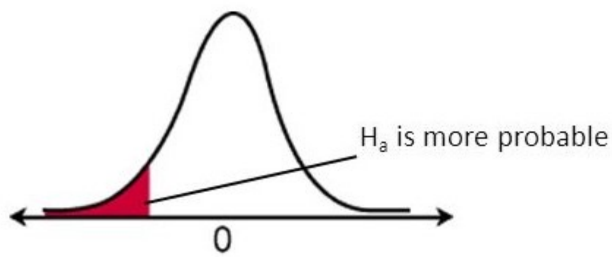
z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	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	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	$\boxed{0.9236}$	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	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

Right tail/Left tail/Two tail



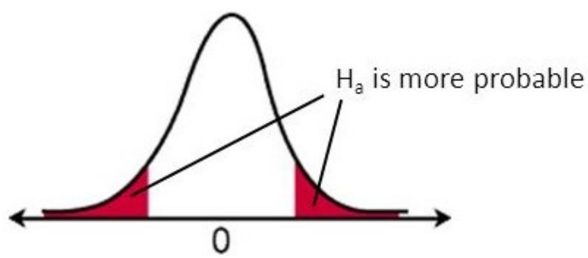
Right-tail test

H_a: $\mu >$ value



Left-tail test

H_a: $\mu <$ value



Two-tail test

H_a: $\mu \neq$ value