



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 results
from the sample
are generalized
to the population

The sample
is selected from
the population

THE SAMPLE
The individuals selected to
participate in the research study

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

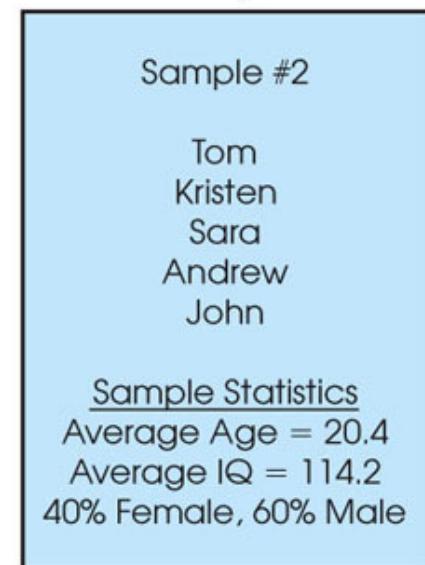
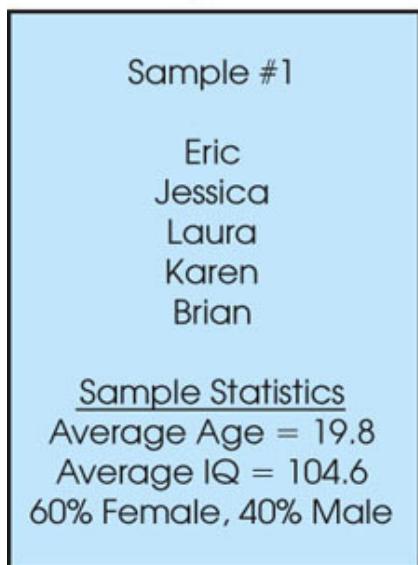
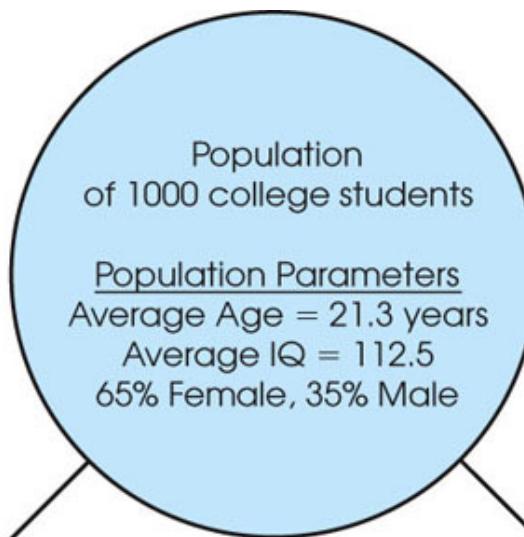
- **Descriptive statistics** are methods for organizing and summarizing data.
- For example, tables or graphs are used to organize data, and descriptive values such as the average score are used to summarize data.
- A descriptive value for a population is called a **parameter** and a descriptive value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

Inferential Statistics

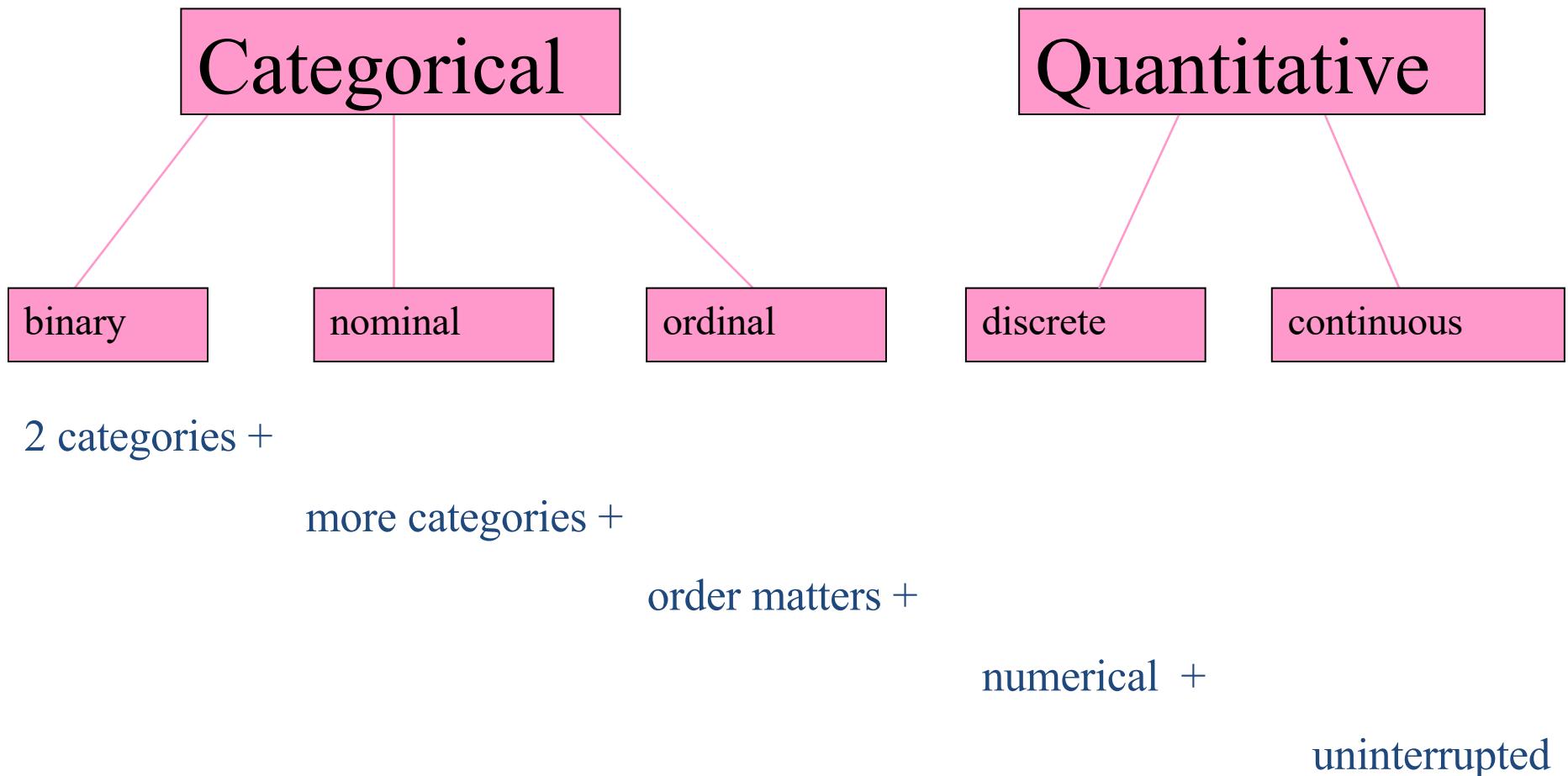
- **Inferential statistics** are methods for using sample data to make general conclusions (inferences) about populations.
- Because a sample is typically only a part of the whole population, sample data provide only limited information about the population. As a result, sample statistics are generally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



Types of Variables: Overview



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

- Discrete Numbers – 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

- Continuous Variables - 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)

Categorical variables

- Bar Chart

Continuous variables

- Box Plot
- Histogram

Describing Dataset – Descriptive

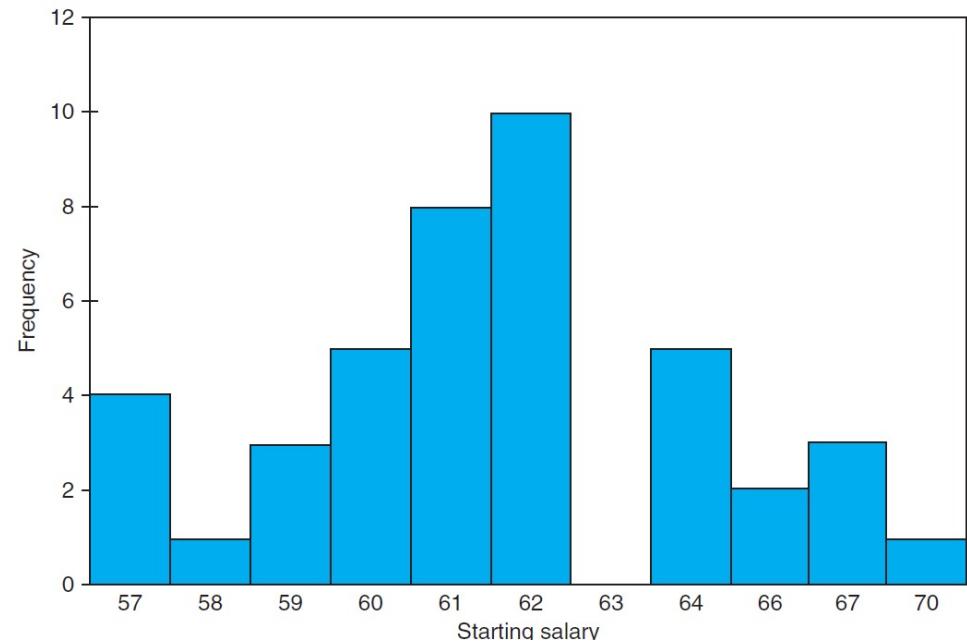
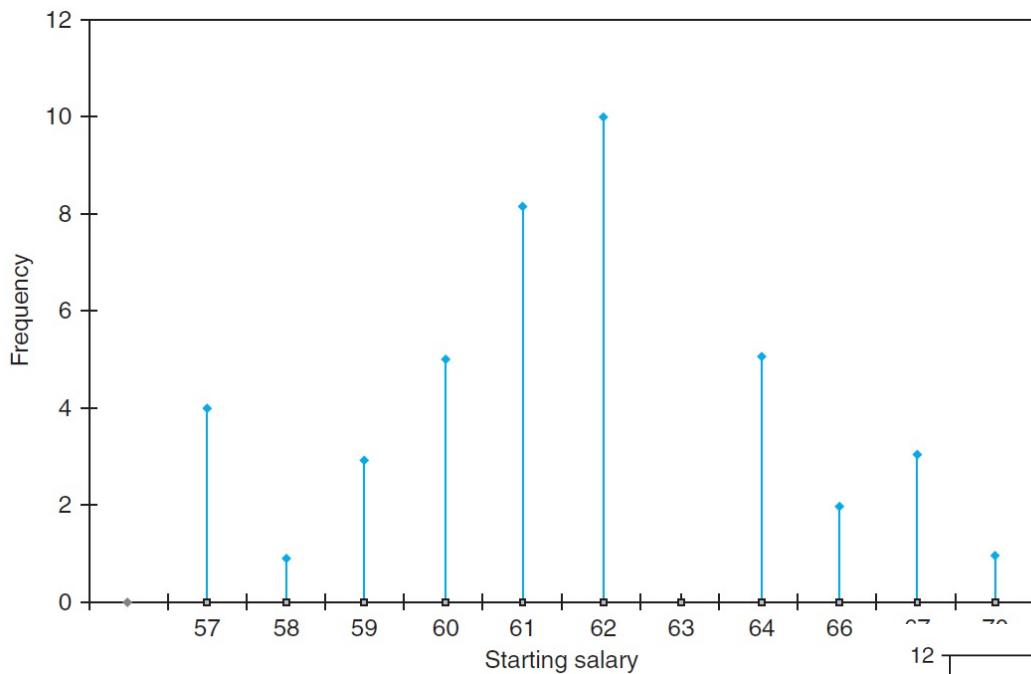
- Freq Table
- Relative Freq Table
- Data Grouping
- Ogives
- Stem and Leaf

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

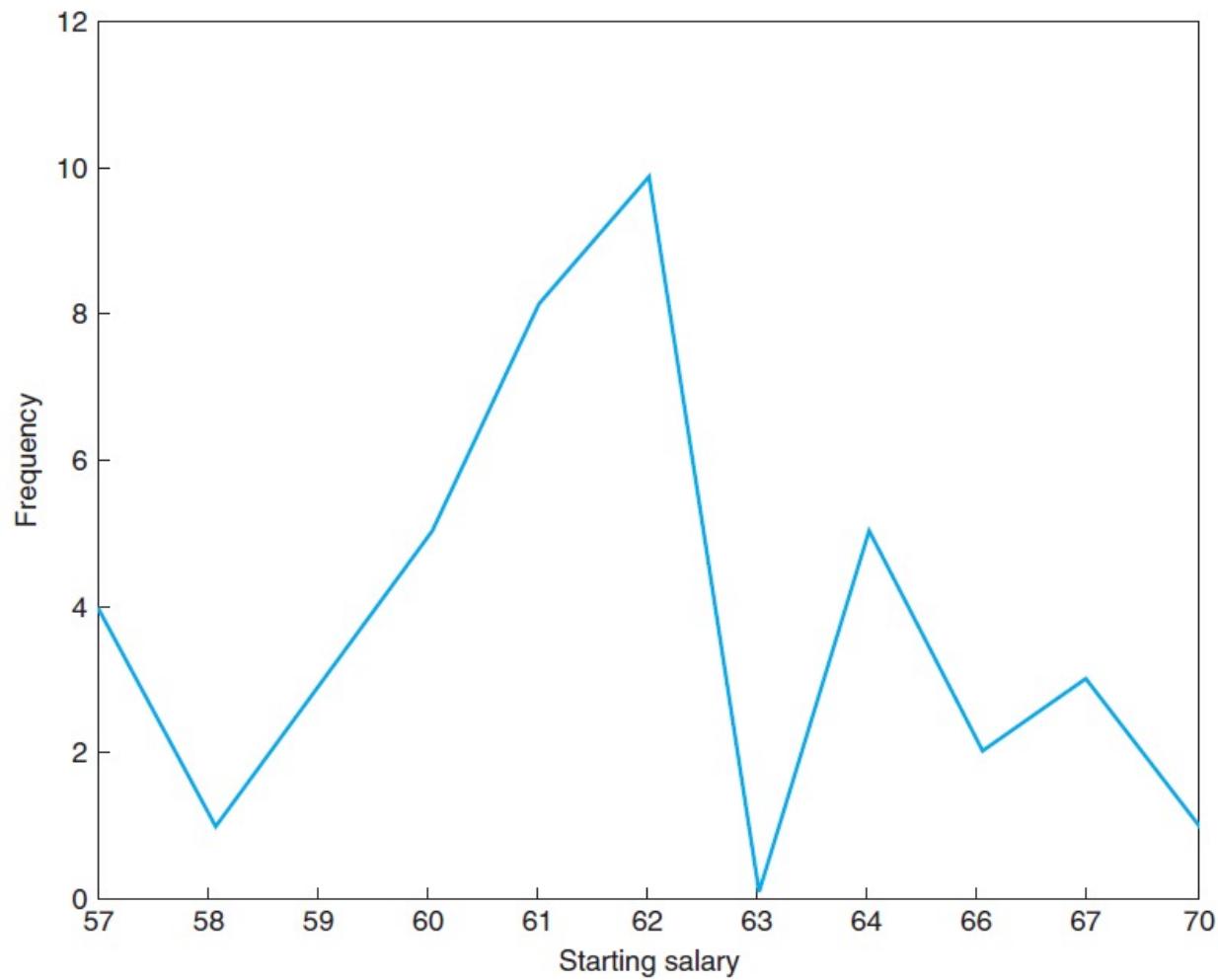
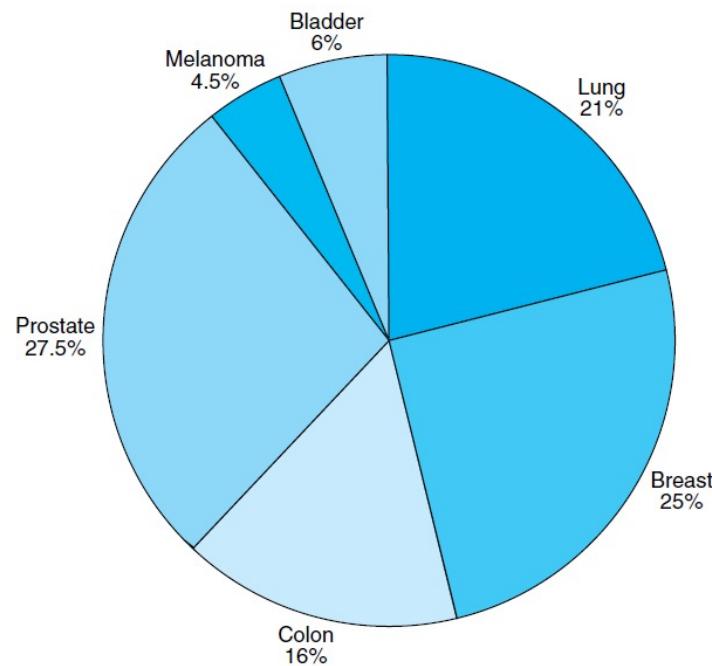


FIGURE 2.3 Frequency polygon for starting salary data.

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

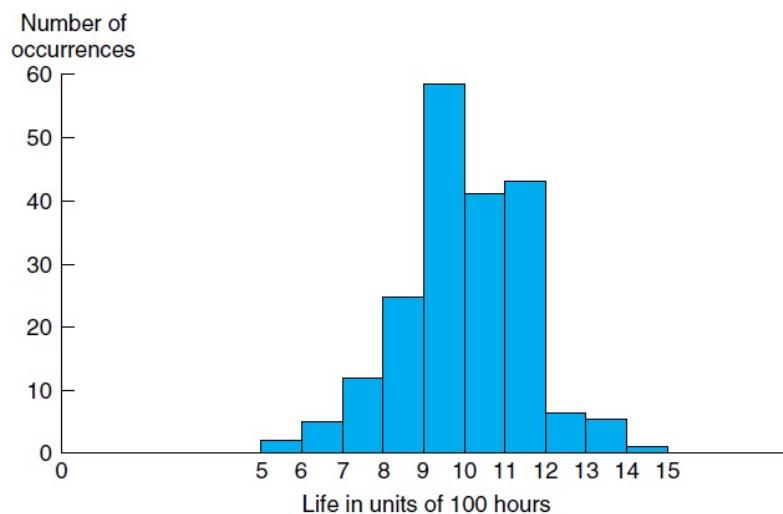
Data Grouping -Histograms

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

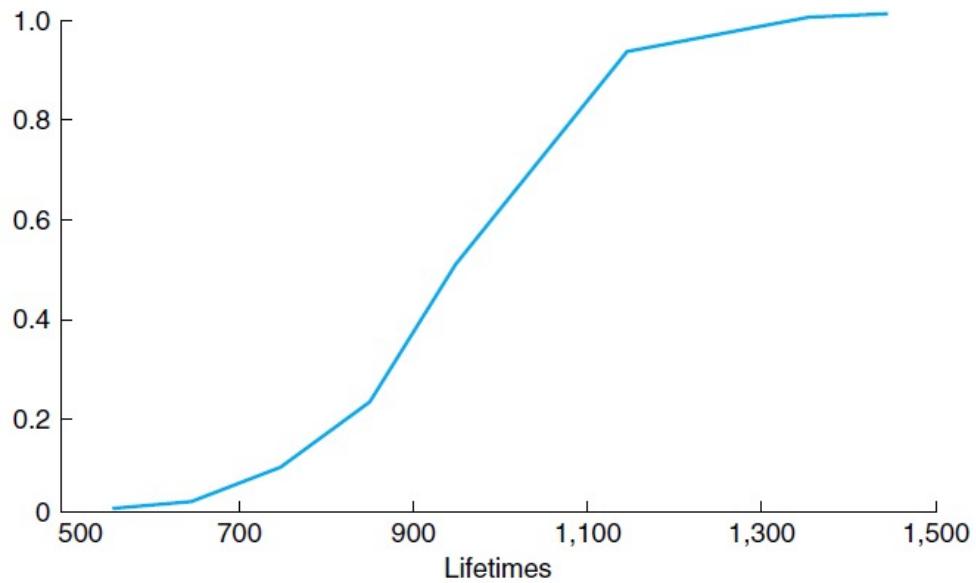
TABLE 2.4 *A Class Frequency Table*

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 x_1, x_2, \dots, x_n .
- *The sample mean is the arithmetic average of these values.*

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

Mean: example

Some data:

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

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

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

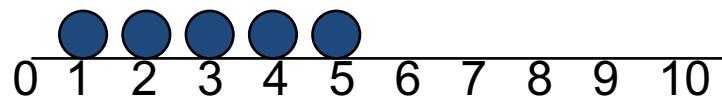
$$\bar{y} = -8/10$$

it follows that

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

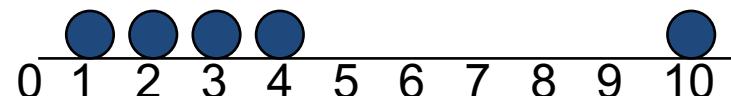
Mean

- The mean is affected by extreme values (outliers)



Mean = 3

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



Mean = 4

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

Mean?

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

Central Tendency

- Median – the exact middle value

Calculation:

- 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

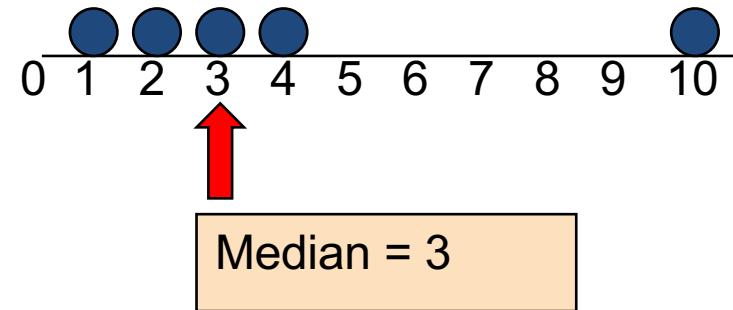
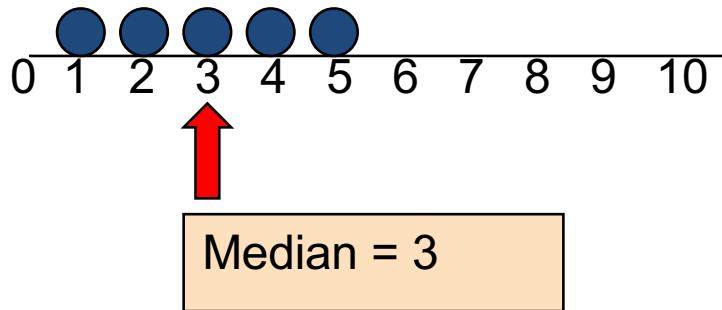
Some data:

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

$$\text{Median} = (22+23)/2 = 22.5$$

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.

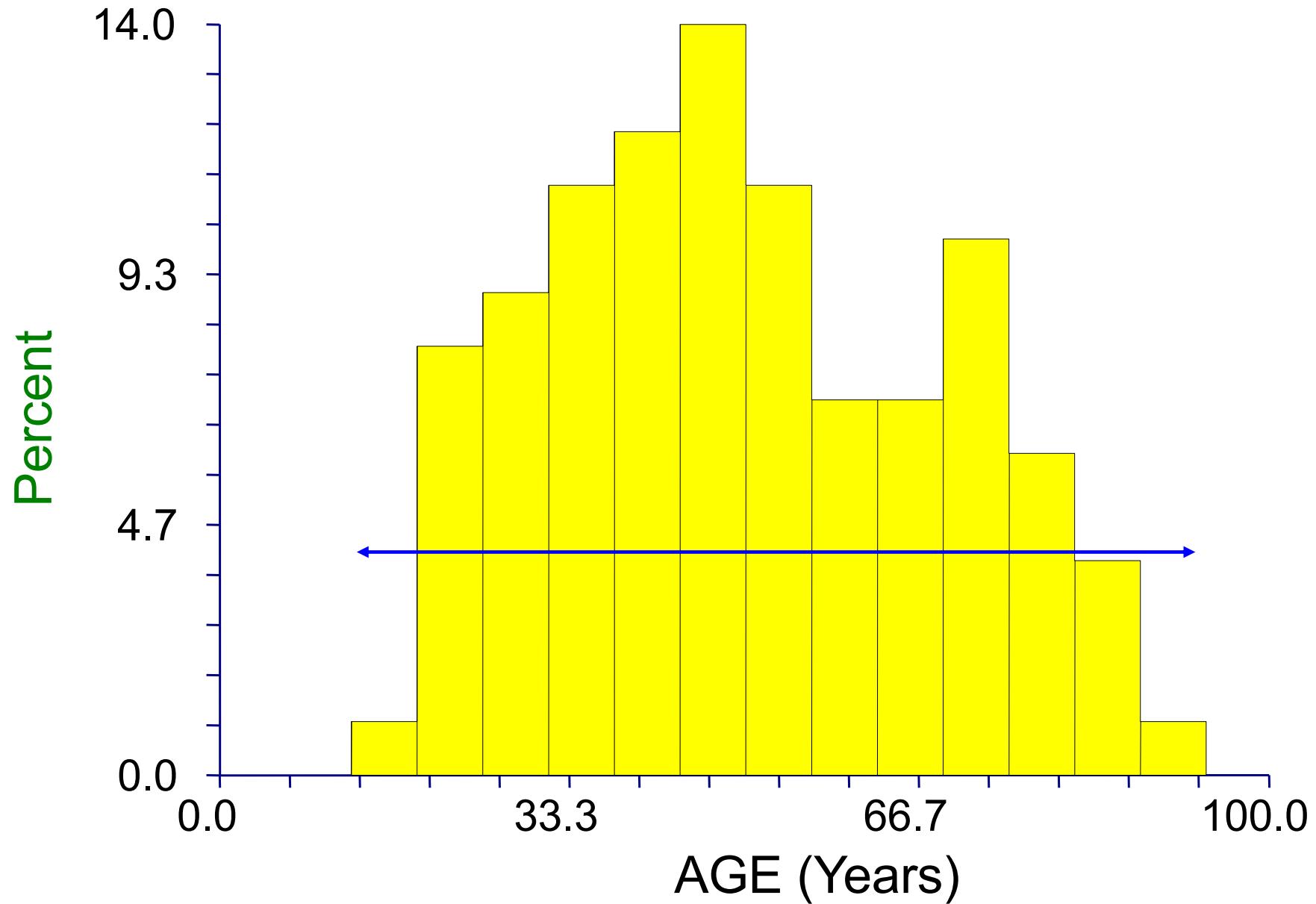
Measures of Variation/Dispersion

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

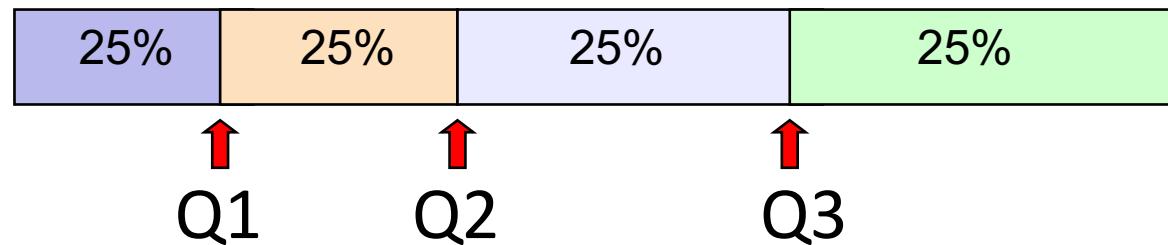
Range

- Difference between the largest and the smallest observations.

Range of age: 94 years-15 years = 79 years



Quartiles

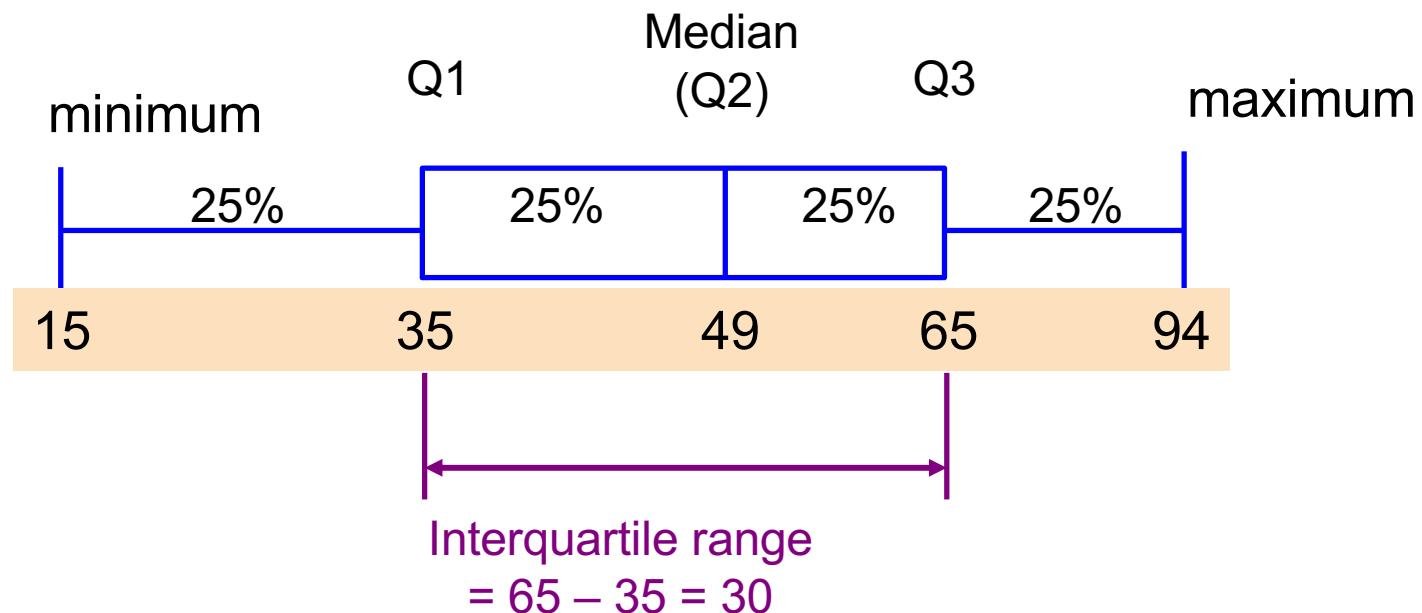


- The first quartile, Q_1 , is the value for which 25% of the observations are smaller and 75% are larger
- Q_2 is the same as the median (50% are smaller, 50% are larger)
- Only 25% of the observations are greater than the third quartile

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.

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

Standard Deviation

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

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

Calculation Example: Sample Standard Deviation

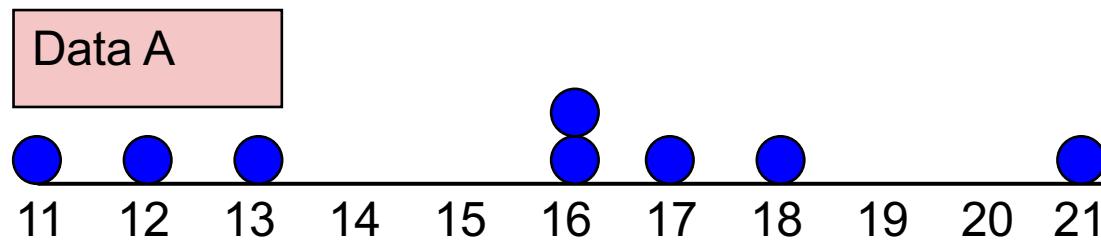
Age data (n=8) : 17 19 21 22 23 23 23 38

$$n = 8 \quad \text{Mean} = X = 23.\overline{25}$$

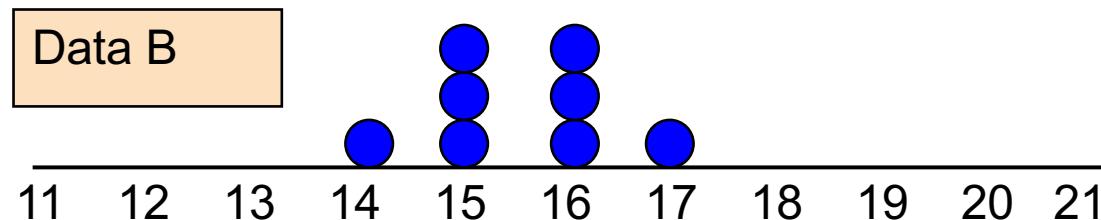
$$\begin{aligned} S &= \sqrt{\frac{(17 - 23.25)^2 + (19 - 23.25)^2 + \dots + (38 - 23.25)^2}{8 - 1}} \\ &= \sqrt{\frac{280}{7}} = 6.3 \end{aligned}$$

standard deviation gets bigger when numbers
are more spread out.

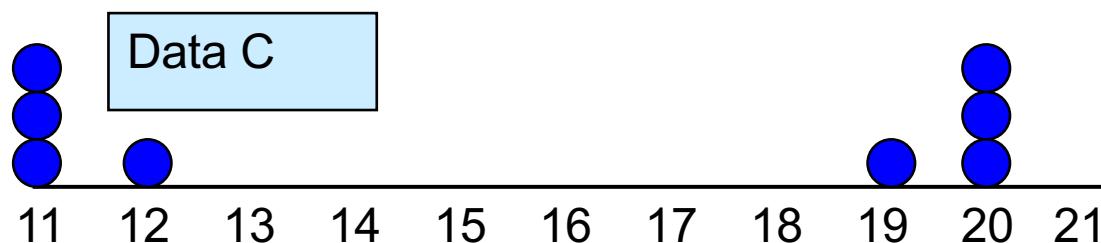
Comparing Standard Deviations



Mean = 15.5
 $S = 3.338$

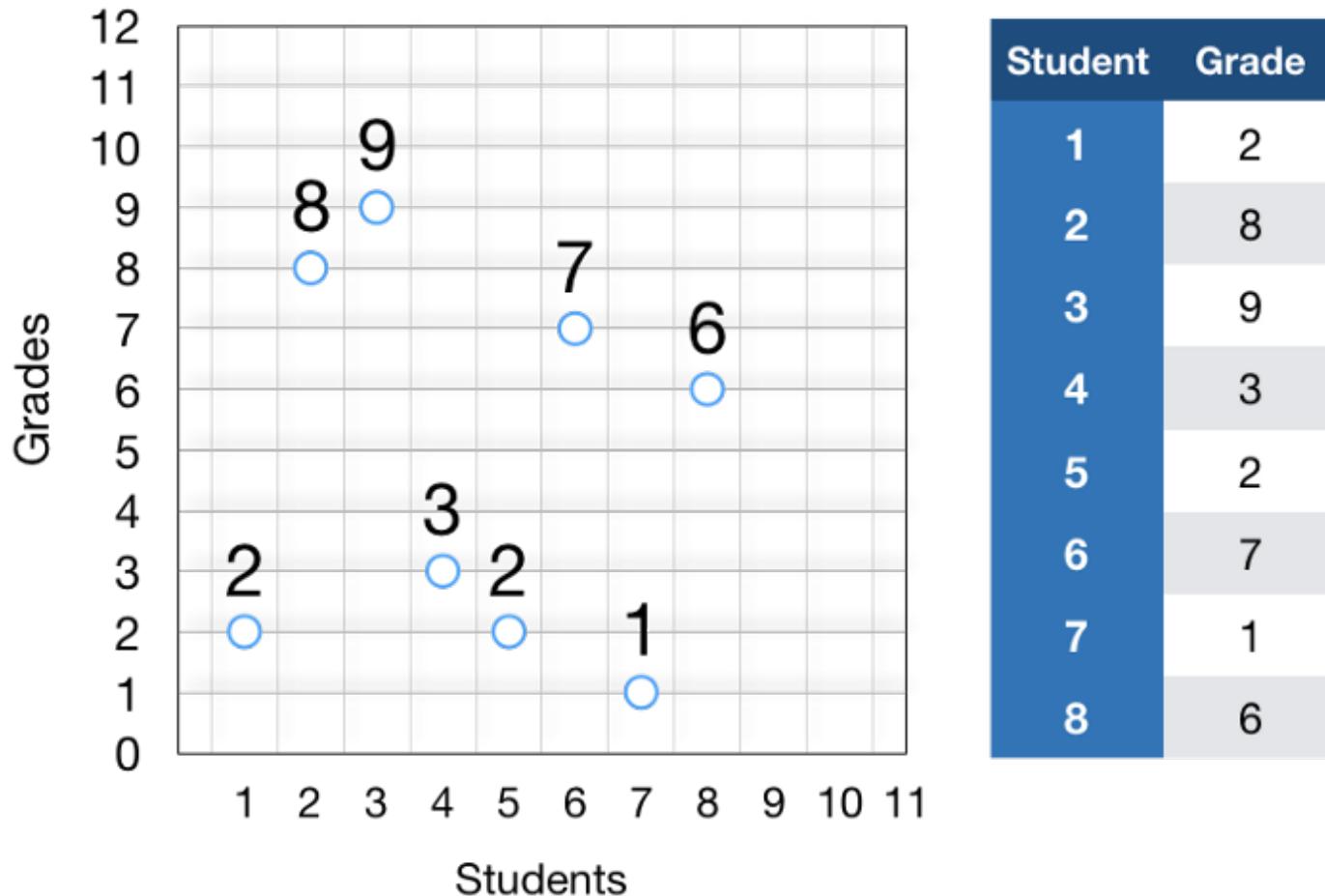


Mean = 15.5
 $S = 0.926$

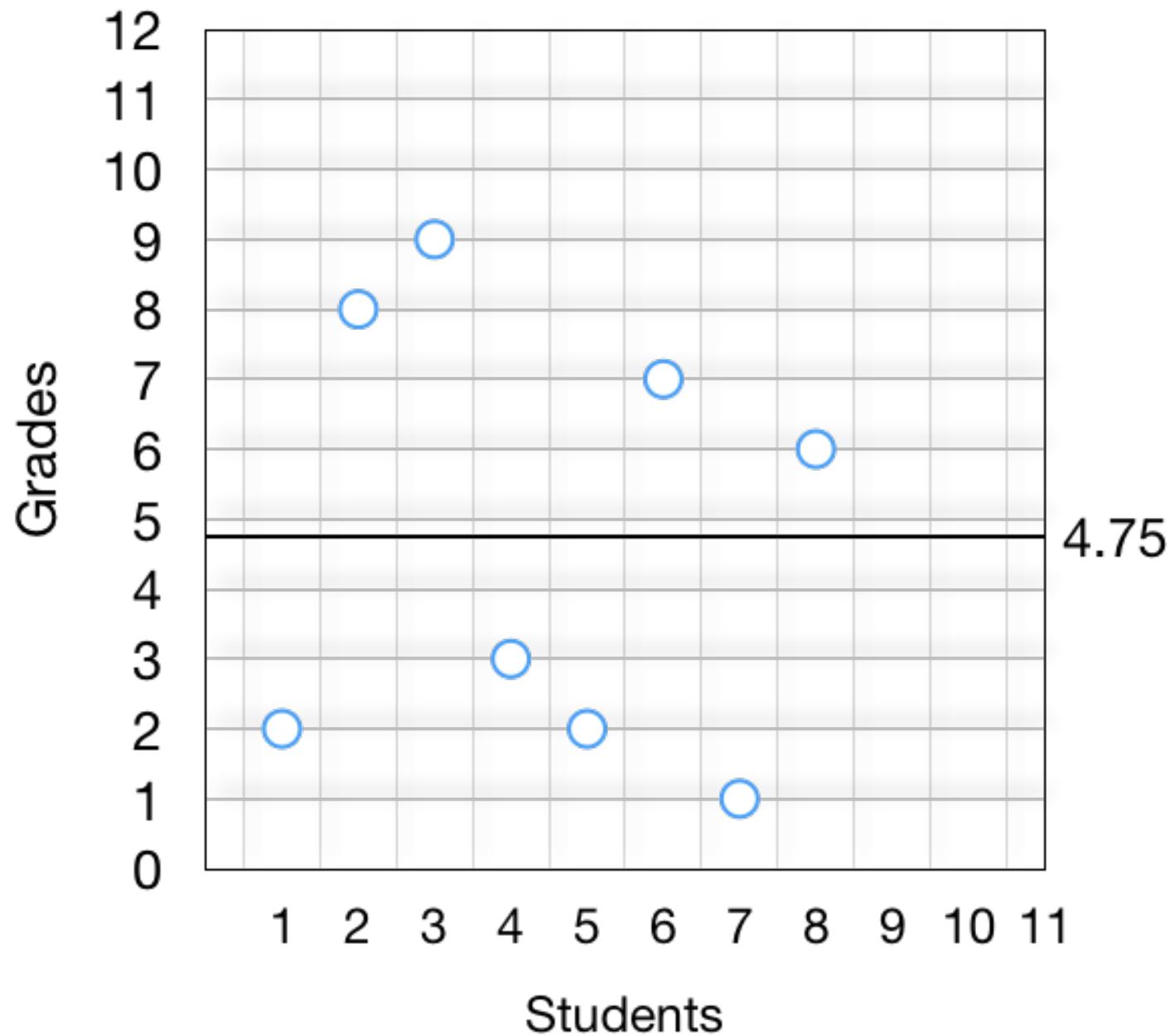


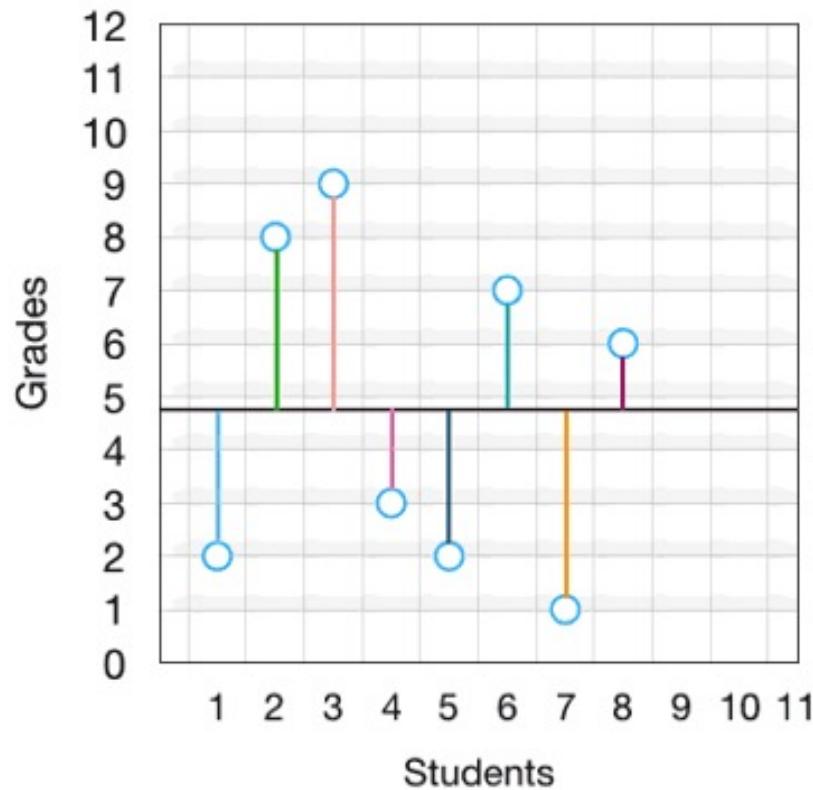
Mean = 15.5
 $S = 4.570$

Visual Meaning

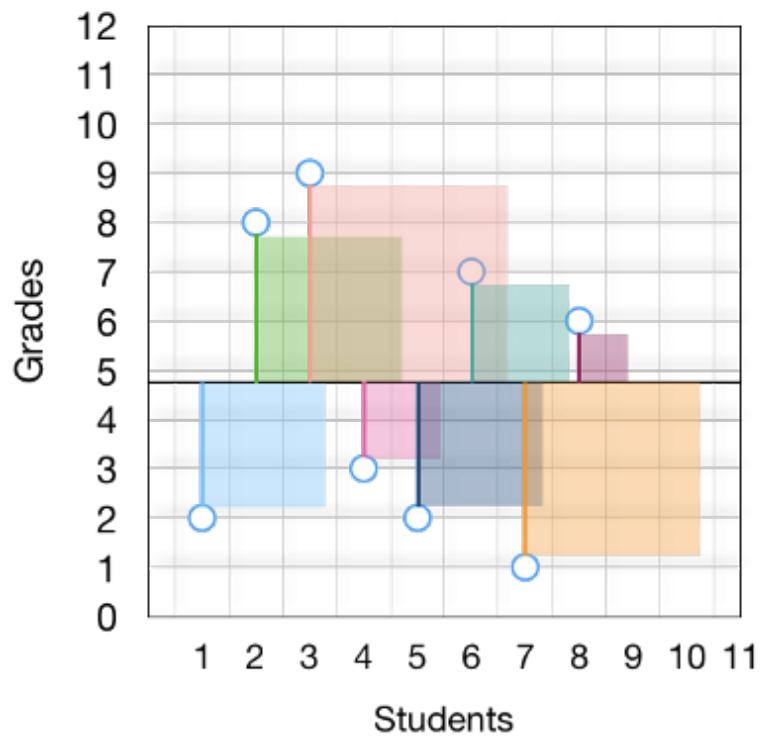


<https://towardsdatascience.com/a-visual-interpretation-of-the-standard-deviation-30f4676c291c>





$$\begin{aligned}x - \bar{x} = & \\(2-4.75) + (8-4.75) & \\+ (9-4.75) + (3-4.75) & \\+ (2-4.75) + (7-4.75) & \\+ (1-4.75) + (6-4.75) &\end{aligned}$$



$$\begin{aligned}\sum (x_n - \bar{x})^2 &= \\ 7.5625 + 10.5625 &+ 18.0625 + 3.0625 \\ + 7.5625 + 5.0625 &+ 14.0625 + 1.5625 \\ = 67.5\end{aligned}$$

$$\text{Mean} \left(\begin{array}{c} \text{teal} \\ \text{blue} \\ \text{light blue} \\ \text{orange} \\ \text{pink} \\ \text{green} \\ \text{purple} \\ \text{pink} \end{array} \right) = 8.45$$

$$\sqrt{8.45 \quad 2.91} = 2.91$$

Bienaym  -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
$(1 - 1/1^2) = 0\%$	$k=1 \ (\mu \pm 1\sigma)$
$(1 - 1/2^2) = 75\%$	$k=2 \ (\mu \pm 2\sigma)$
$(1 - 1/3^2) = 89\%$	$k=3 \ (\mu \pm 3\sigma)$

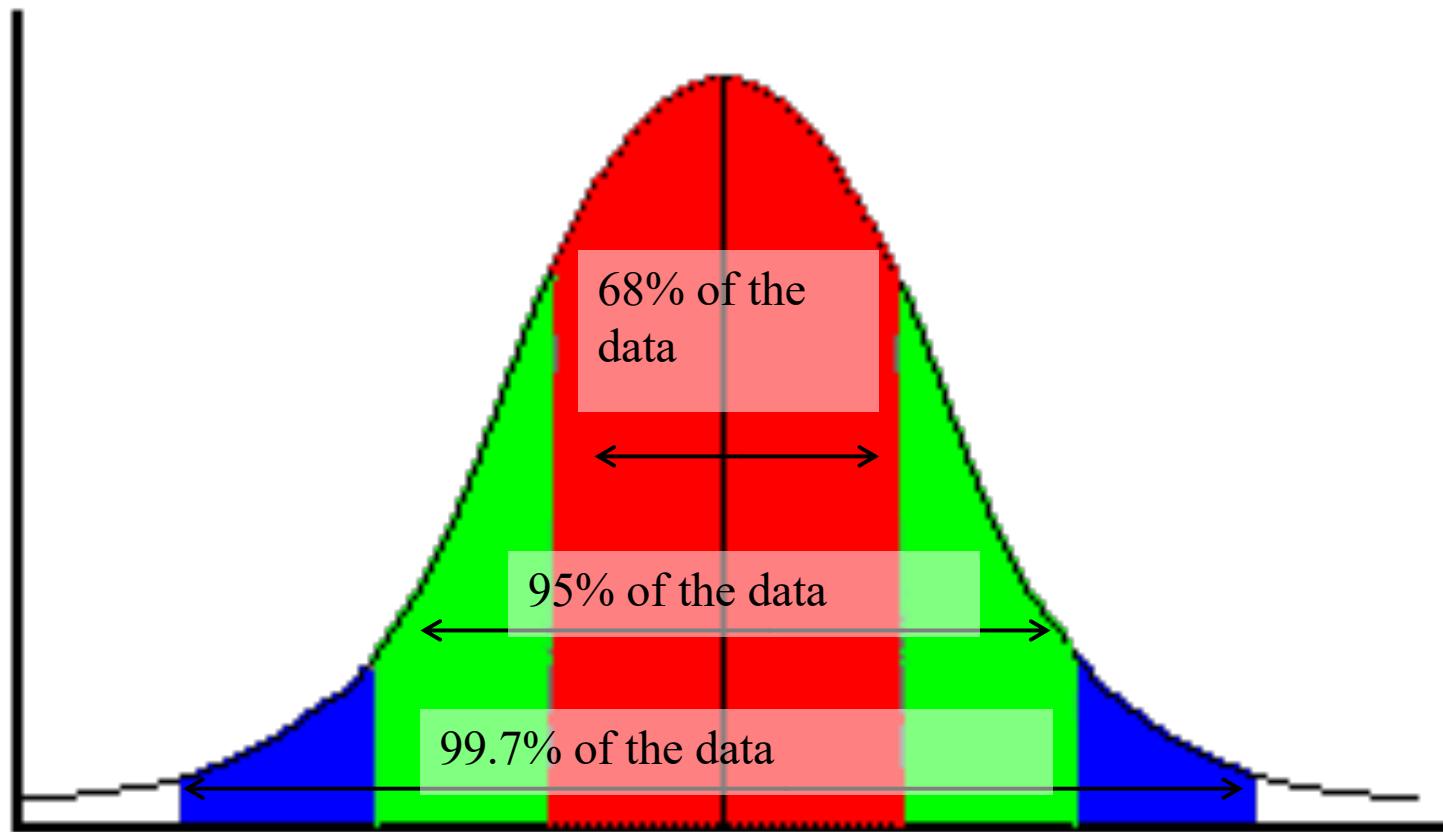
Symbol Clarification

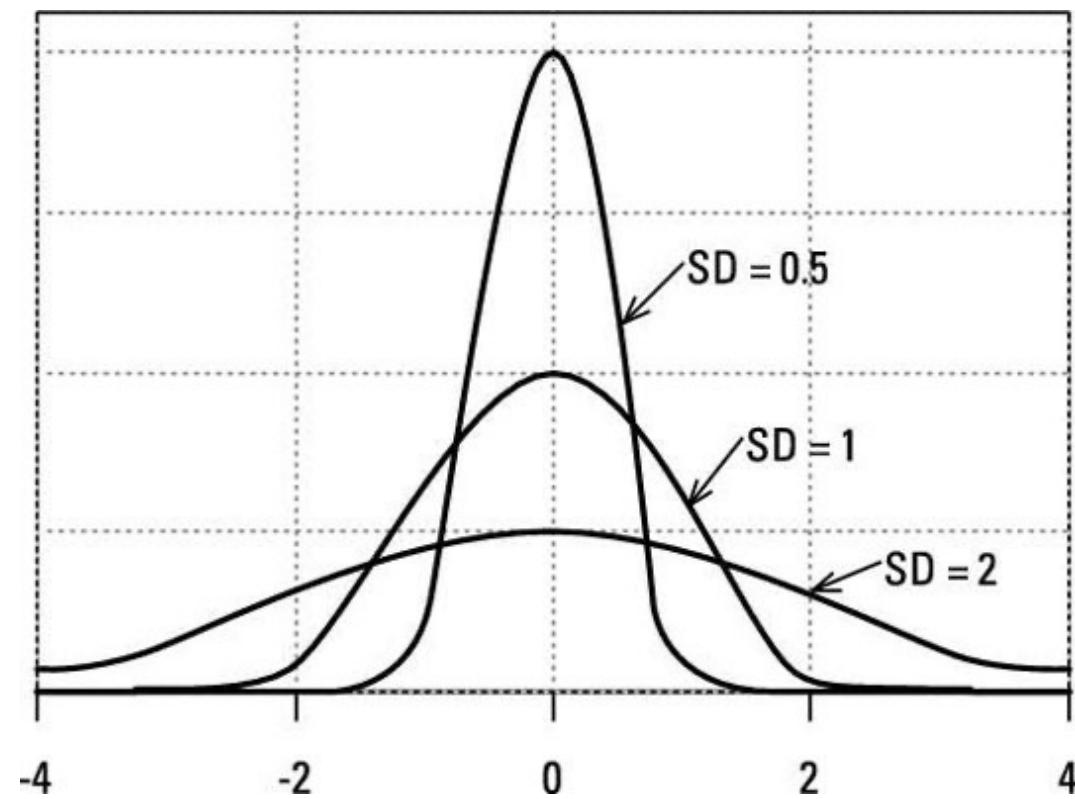
- s = Sample standard deviation (example of a “sample statistic”)
- σ = Standard deviation of the entire population (example of a “population parameter”) or from a theoretical probability distribution
- \bar{x} = Sample mean
- μ = Population or theoretical mean

**The beauty of the normal curve:

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

68-95-99.7 Rule





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 $S = \{\text{Head}, \text{Tail}\}$

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

$$S = \{\square \cdot, \square \circ, \square \bullet, \square \square, \square \bullet \bullet, \square \square \square\}$$

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 $S = \{\text{Head, Tail}\}$
2. Rolling a die – outcomes

$$S = \{\begin{array}{|c|} \hline \bullet \\ \hline \end{array}, \begin{array}{|c|c|} \hline \bullet & \\ \hline & \bullet \\ \hline \end{array}, \begin{array}{|c|c|} \hline \bullet & \\ \hline & \bullet \\ \hline \end{array}, \begin{array}{|c|c|c|} \hline \bullet & & \\ \hline & \bullet & \\ \hline & & \bullet \\ \hline \end{array}, \begin{array}{|c|c|c|} \hline \bullet & & \\ \hline & \bullet & \\ \hline & & \bullet \\ \hline \end{array}, \begin{array}{|c|c|c|} \hline \bullet & & \\ \hline & \bullet & \\ \hline & & \bullet \\ \hline \end{array}\}$$

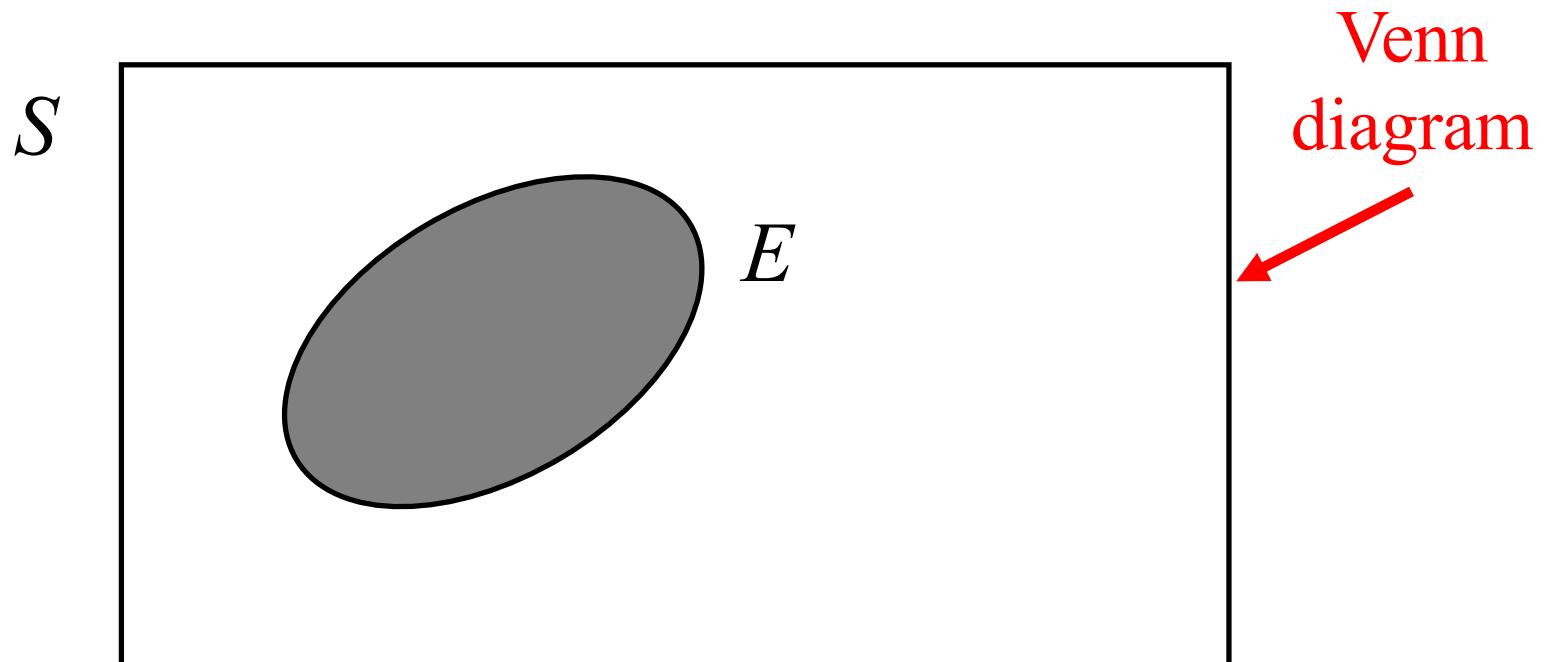
$$= \{1, 2, 3, 4, 5, 6\}$$

Events -E

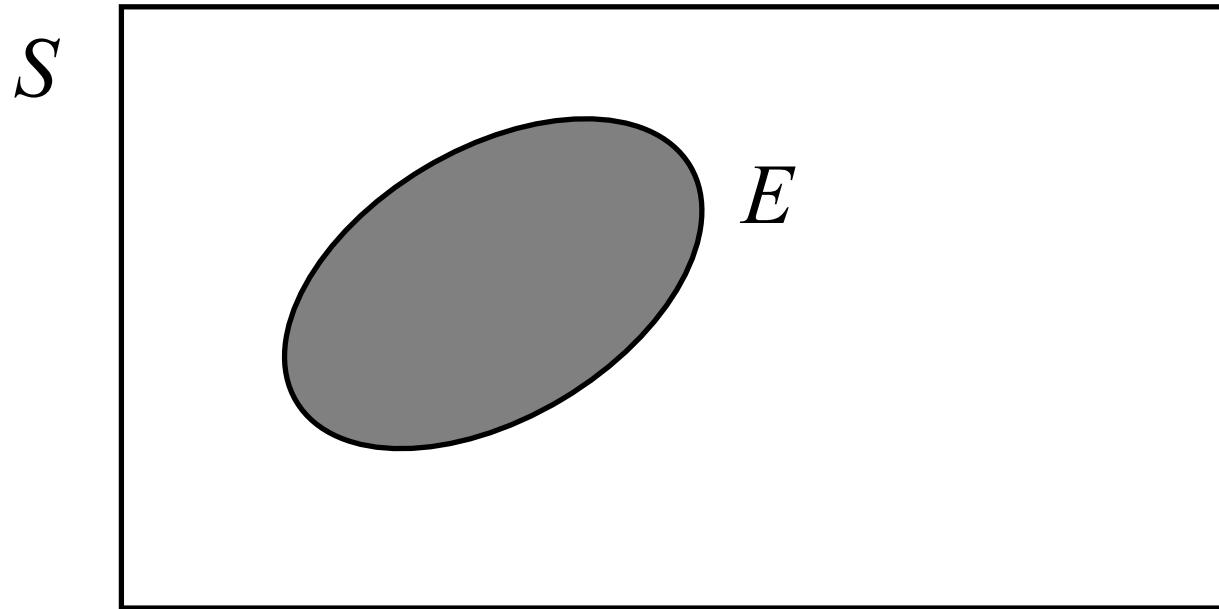
- Events are sets:
 - Can describe in words
 - Can describe in notation
 - Can describe with Venn diagrams
-
- Experiment: toss a coin 3 times.
 - Event: You get 2 or more heads = { HHH, HHT, HTH, THH}

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

$$S = \{\begin{array}{|c|} \hline \bullet \\ \hline \end{array}, \begin{array}{|c|} \hline \circ \\ \hline . \\ \hline \end{array}, \begin{array}{|c|c|} \hline \circ & . \\ \hline . & \bullet \\ \hline \end{array}, \begin{array}{|c|c|} \hline \circ & \circ \\ \hline \circ & \bullet \\ \hline \end{array}, \begin{array}{|c|c|c|} \hline \circ & \circ & . \\ \hline \circ & . & \bullet \\ \hline \end{array}, \begin{array}{|c|c|c|} \hline \circ & \circ & \circ \\ \hline \circ & \circ & \bullet \\ \hline \end{array}\}$$
$$= \{1, 2, 3, 4, 5, 6\}$$

E = the event that an even number is rolled

$$= \{2, 4, 6\}$$
$$= \{\begin{array}{|c|} \hline \circ \\ \hline . \\ \hline \end{array}, \begin{array}{|c|c|} \hline \circ & \circ \\ \hline \circ & \bullet \\ \hline \end{array}, \begin{array}{|c|c|c|} \hline \circ & \circ & \circ \\ \hline \circ & \circ & \bullet \\ \hline \end{array}\}$$

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

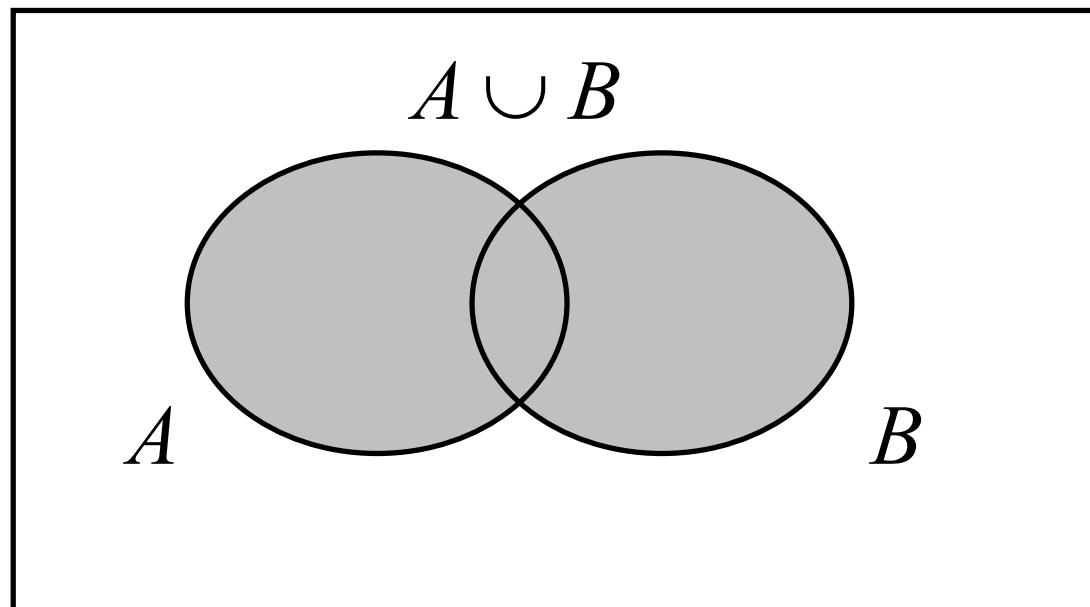
- **Experiment:** a repeatable procedure
- **Sample space:** set of all possible outcomes S (or Ω).
- **Event:** a subset of the sample space.
- **Probability function,** $P(\omega)$: gives the probability for each outcome $\omega \in S$
- Probability is between 0 and 1
- Total probability of all possible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

Set operations on Events

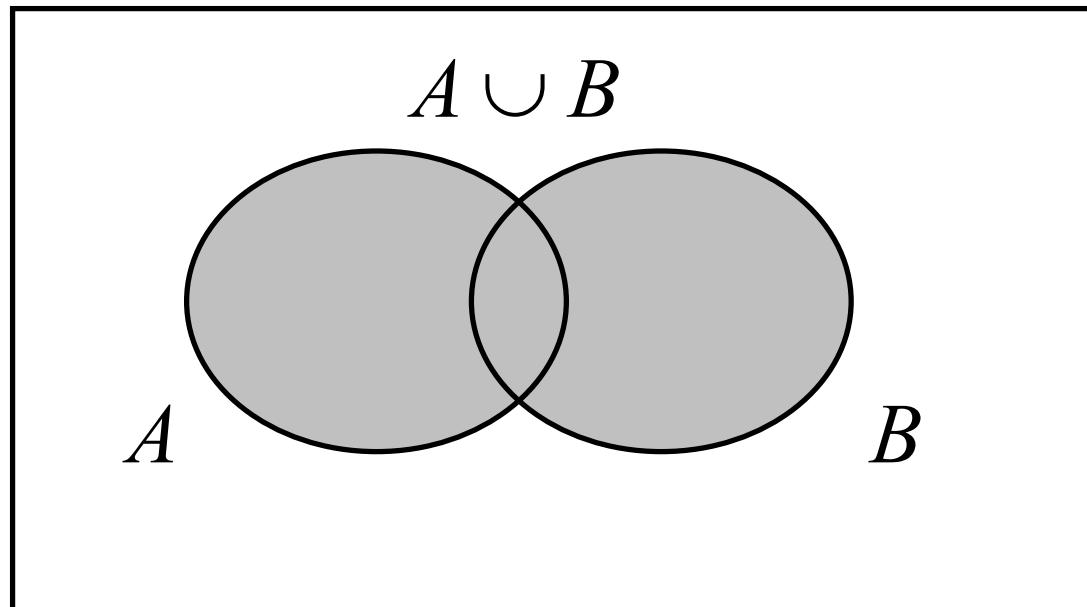
Union

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

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



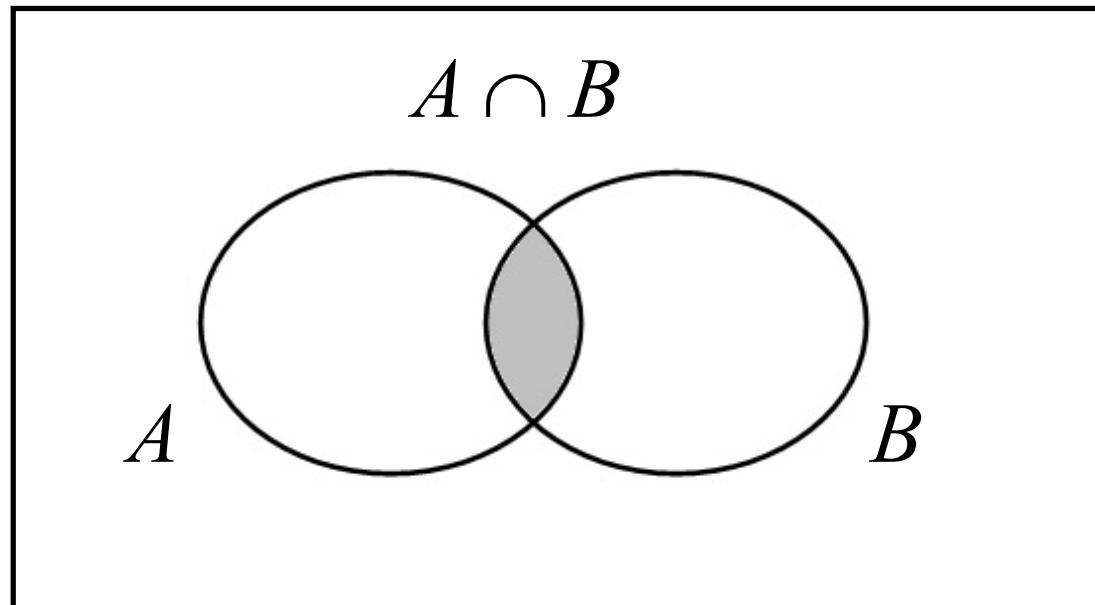
The event $A \cup B$ **occurs** if the event A **occurs or**
the event B **occurs**.



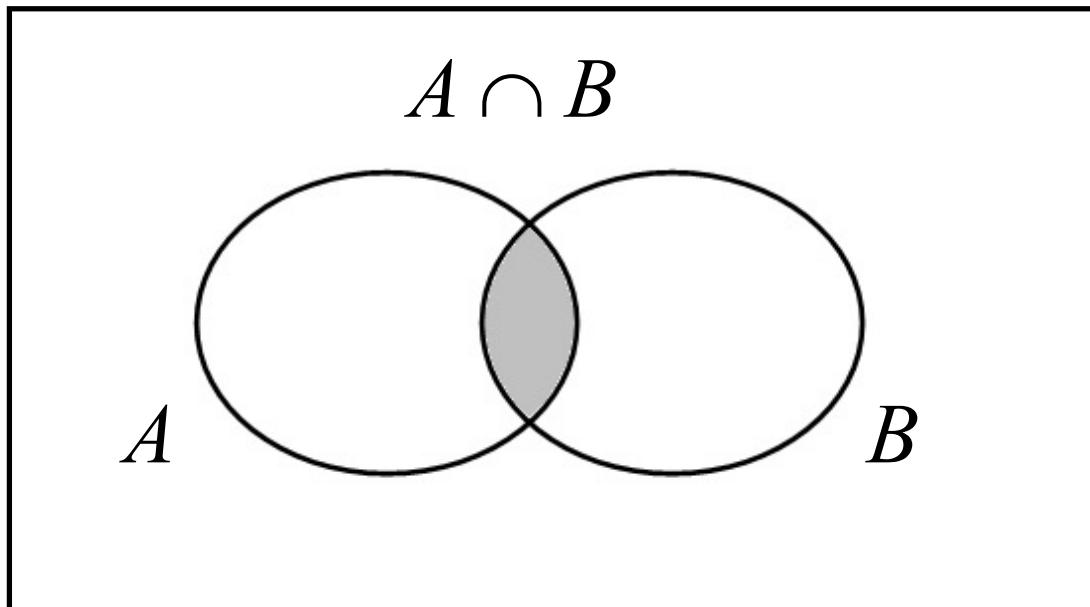
Intersection

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

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



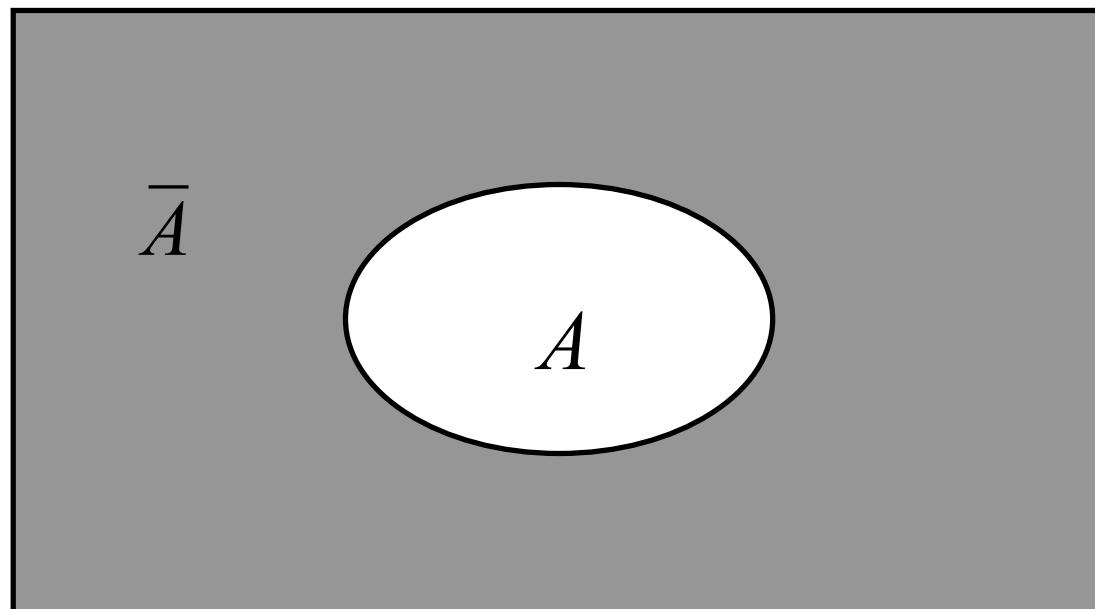
The event $A \cap B$ occurs if the event **A occurs and**
the event **B occurs**.



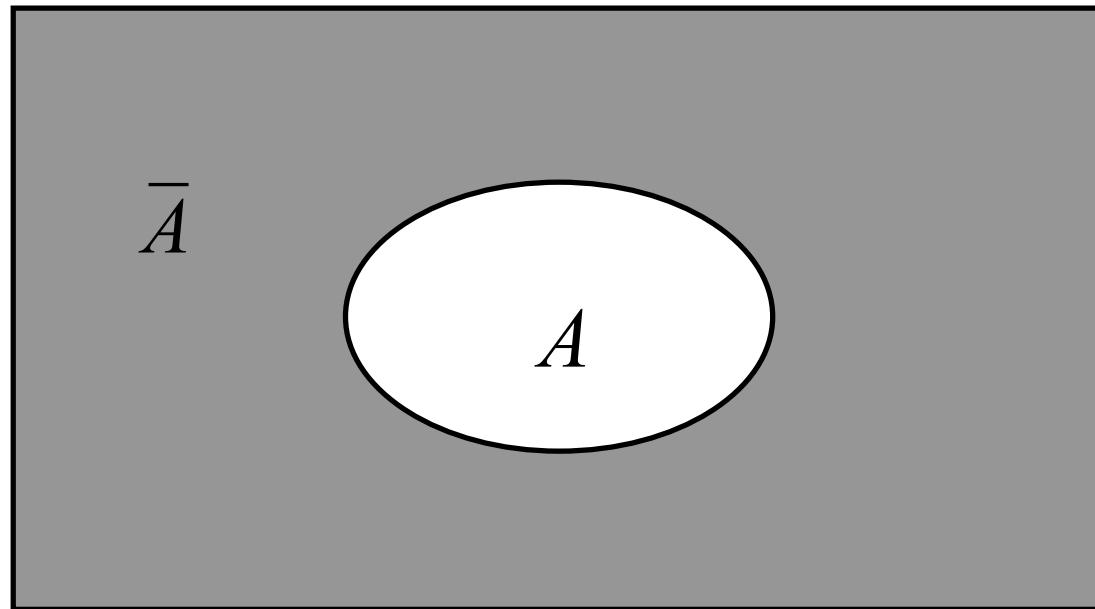
Complement

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

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



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



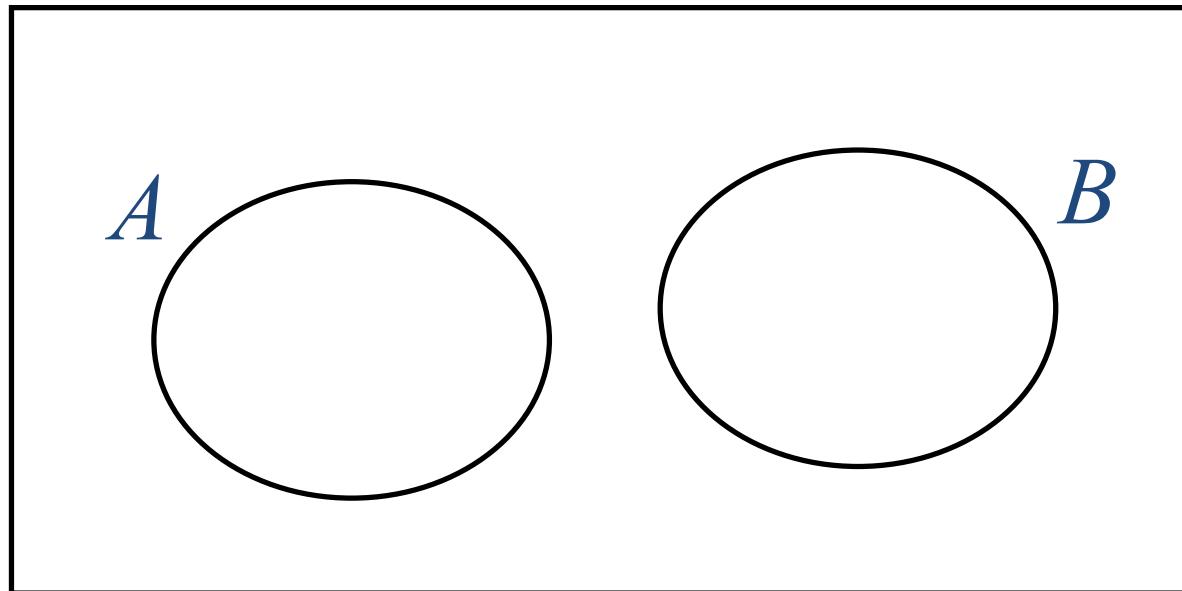
In problems you will recognize that you are working with:

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:

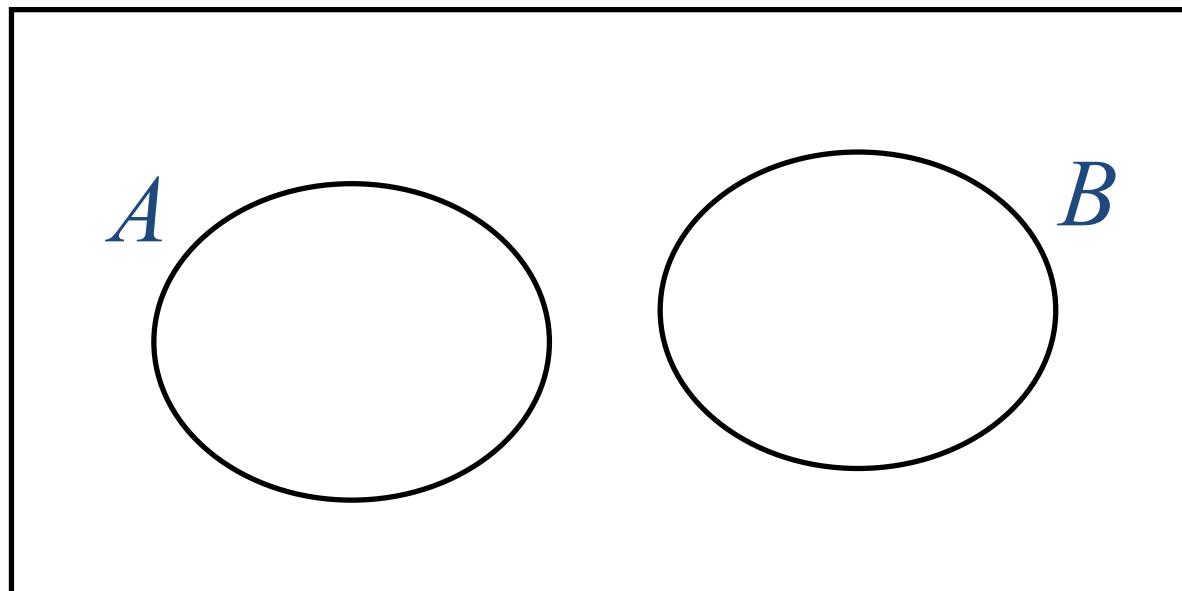
$$A \cap B = \emptyset$$



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

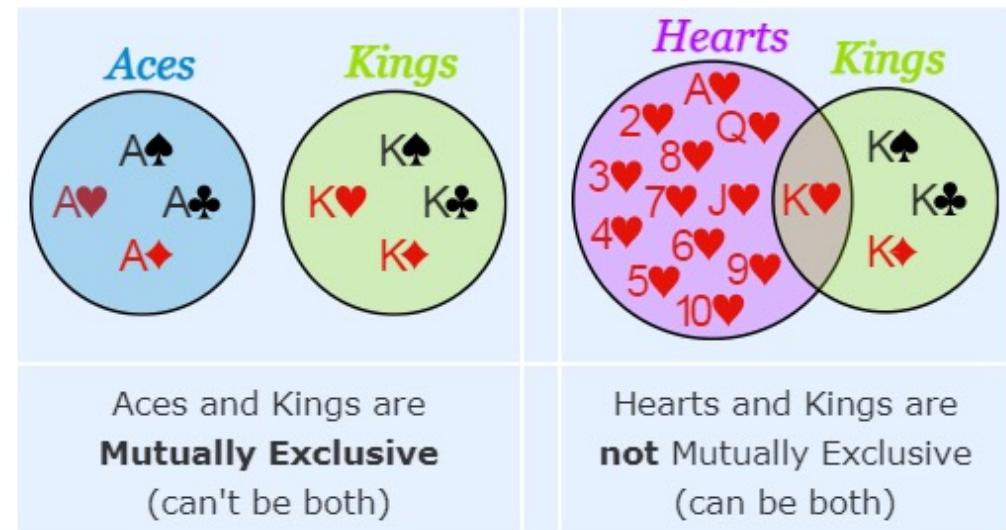
1. They have no outcomes in common.

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 $S = \{o_1, o_2, o_3, \dots, o_N\}$ has a finite number, N , of outcomes.

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

Then for any event E

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

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)

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

i.e.

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

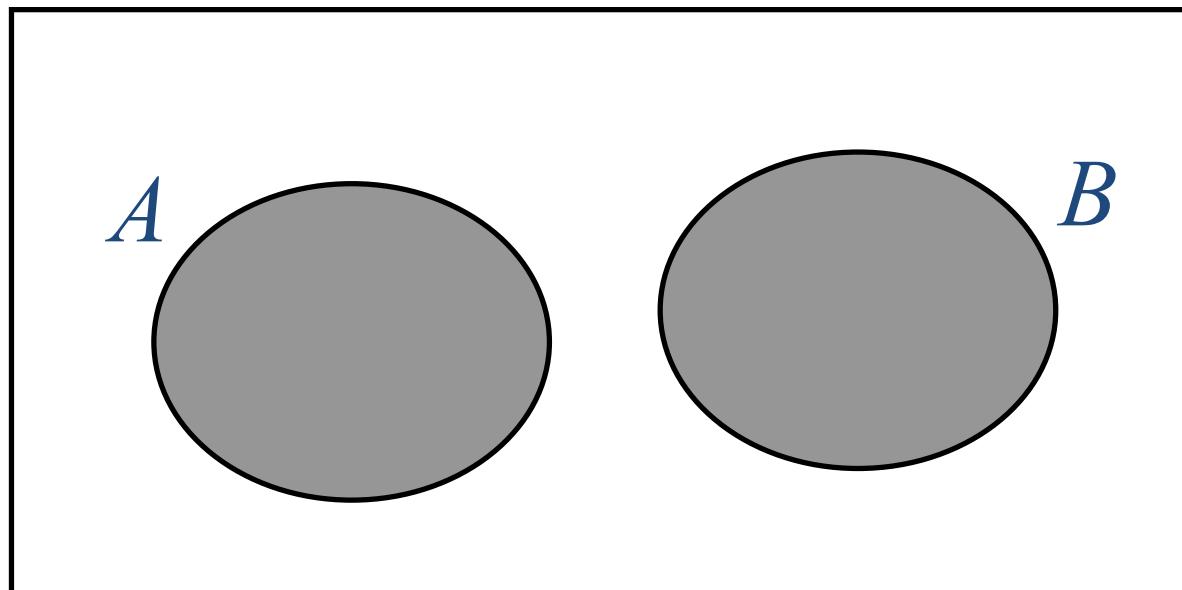
if $A \cap B = \phi$

(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.

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

A

B

Rule The additive rule (In general)

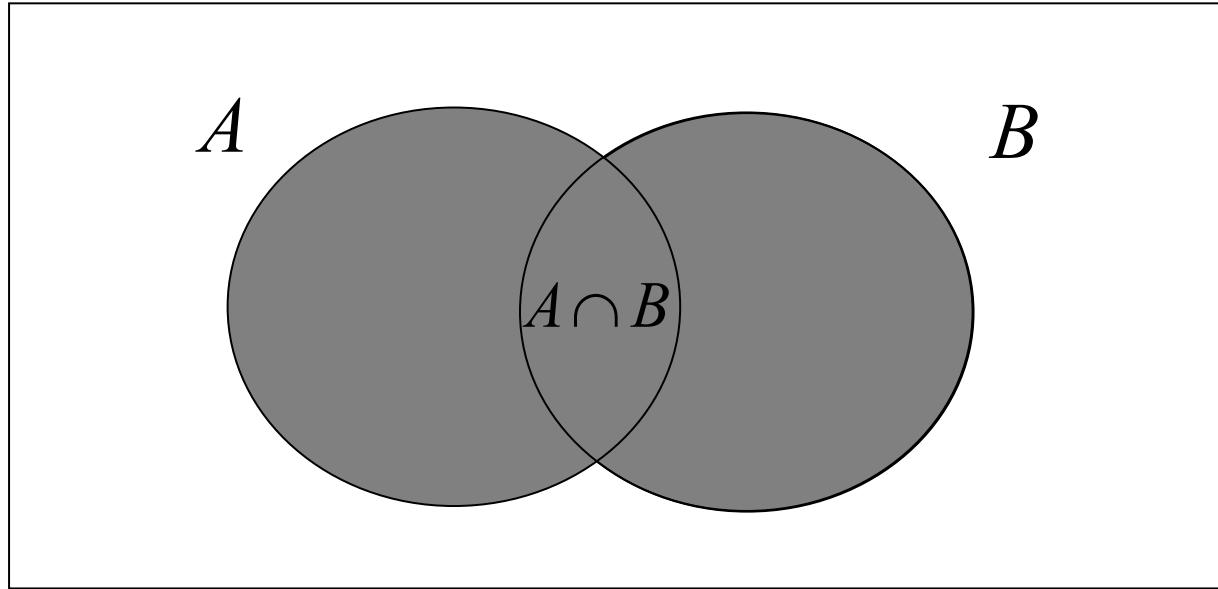
$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

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” $\equiv \cup$.

$$\begin{aligned}P[A \cup B] &= P[A] + P[B] - P[A \cap B] \\&= 0.20 + 0.35 - 0.08 = 0.47\end{aligned}$$

Rule for complements

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

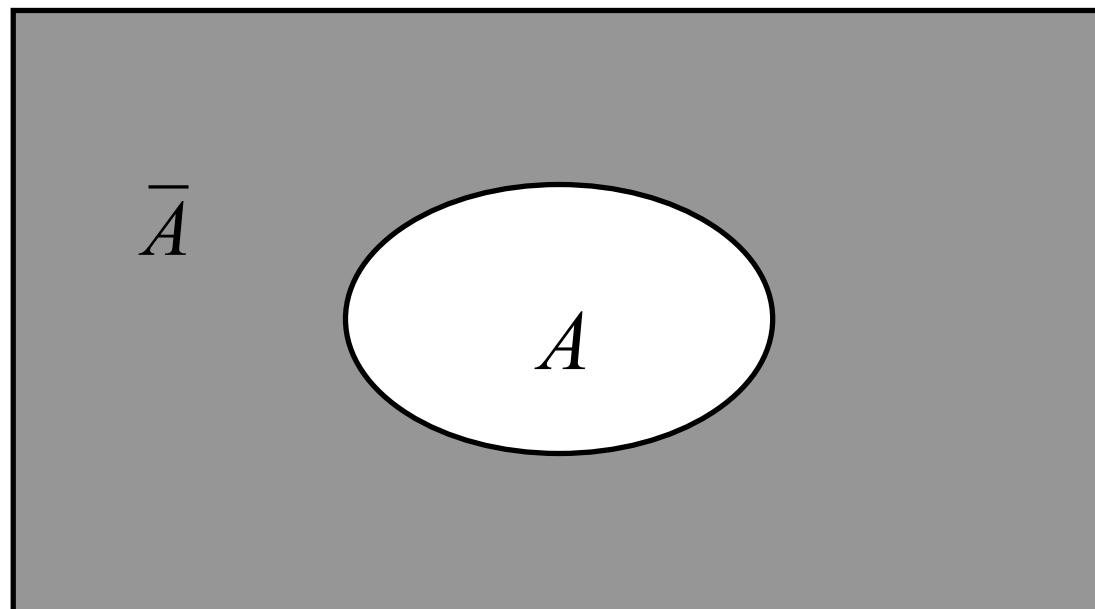
or

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

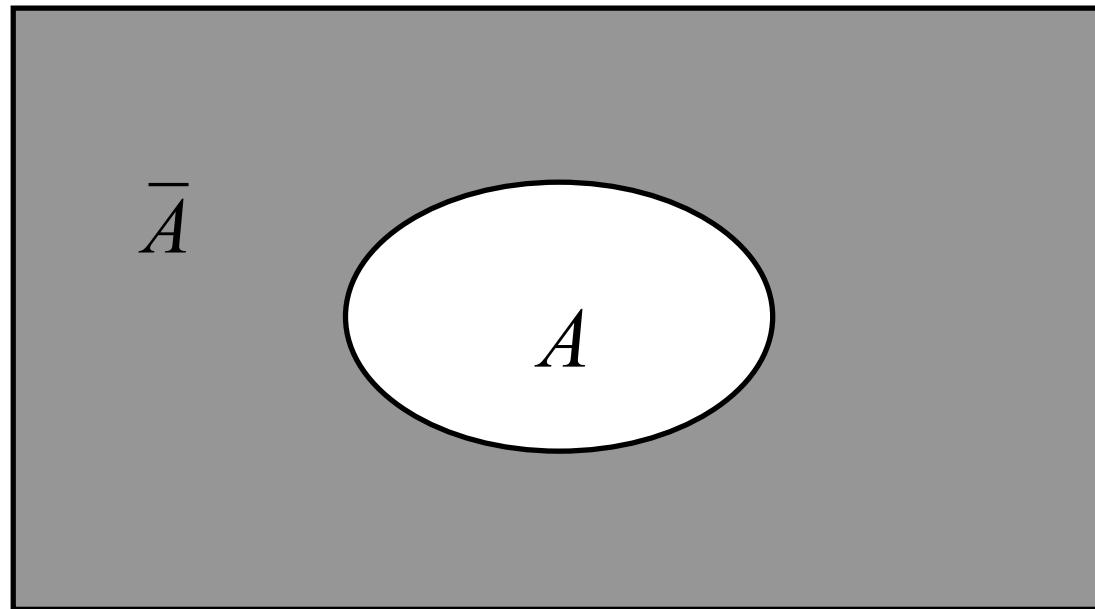
Complement

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

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



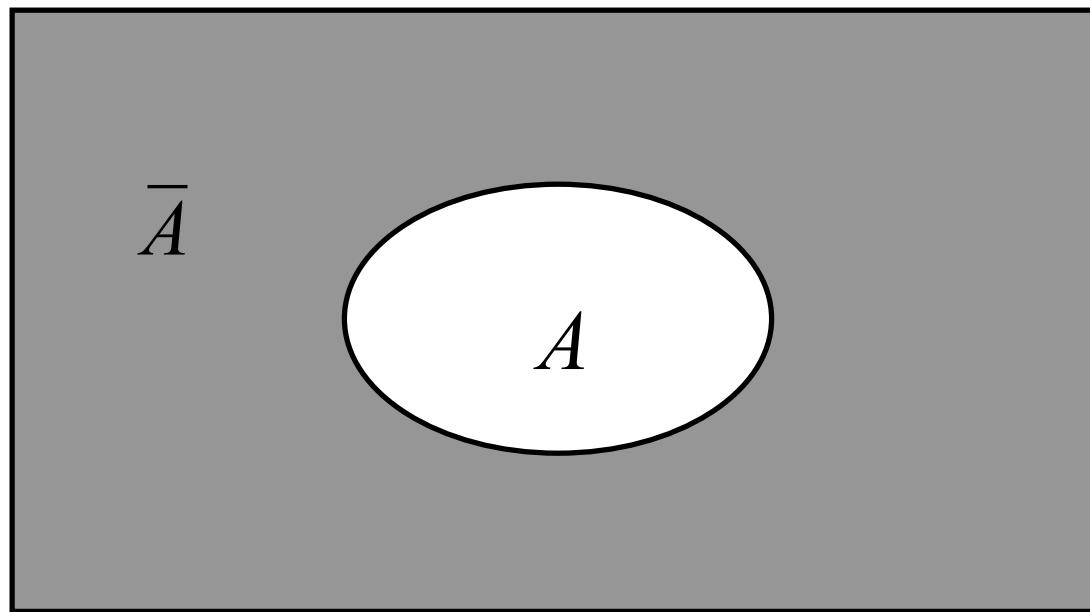
The event \bar{A} occurs if the event A does not occur



Logic:

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

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



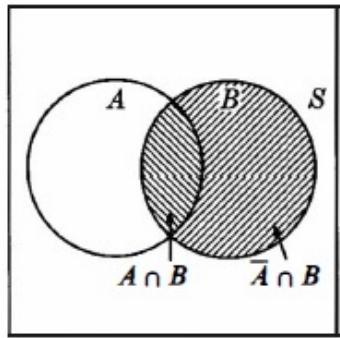
$$\text{thus } 1 = P[S] = P[A] + P[\bar{A}]$$

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

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

$$\begin{aligned} A \cup B &= A \cup (B \cap \bar{A}), \\ B &= (A \cap B) \cup (B \cap \bar{A}). \end{aligned}$$

Hence

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

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 $\{t \mid t \geq 0\}$. Let the three events A, B, and C be defined as follows:
- $A = \{t \mid t < 100\}$; $B = \{t \mid 50 \leq t \leq 200\}$; $C = \{t \mid t > 150\}$.
- $A \cup B$, $A \cap B$, $B \cup C$, $B \cap C$, $A \cap C$, $A \cup C$ and A, C

EXAMPLE

- $A \cup B = \{t \mid t \leq 200\}$; $A \cap B = \{t \mid 50 \leq t < 100\}$;
- $B \cup C = \{t \mid t \geq 50\}$; $B \cap C = \{t \mid 150 \leq t \leq 200\}$;
 $A \cap C = \text{null}$;
- $A \cup C = \{t \mid t < 100 \text{ or } t > 150\}$; $\underline{A} = \{t \mid t \geq 100\}$; $\underline{C} = \{t \mid t \leq 150\}$.

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:

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

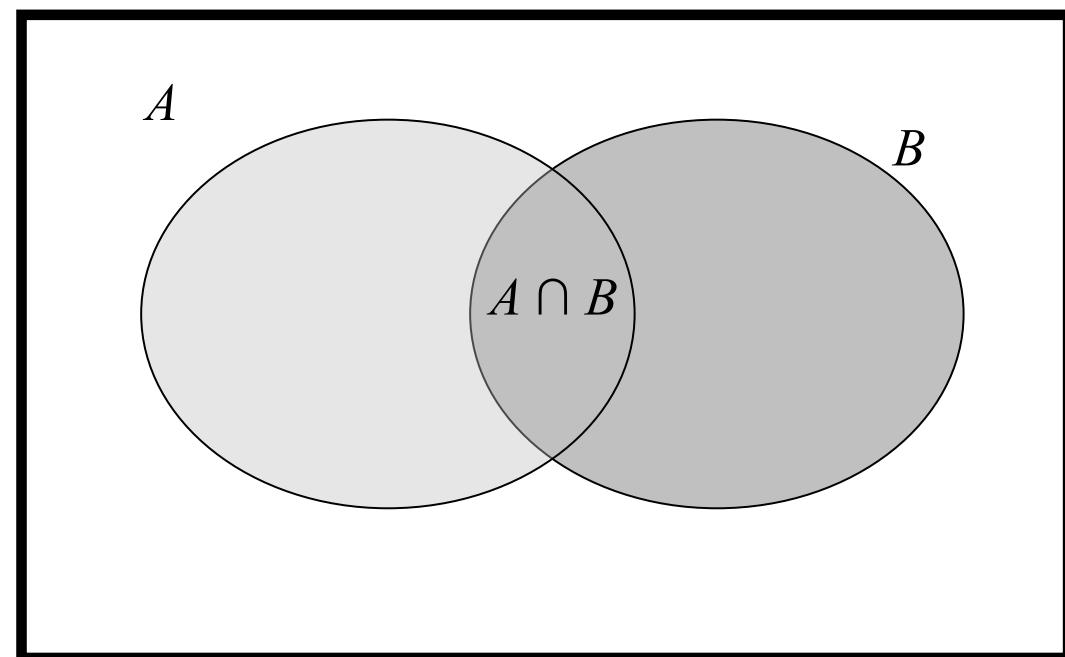
Rationale:

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 $A \cap B$. Hence the new probability of A is:

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



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

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

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

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

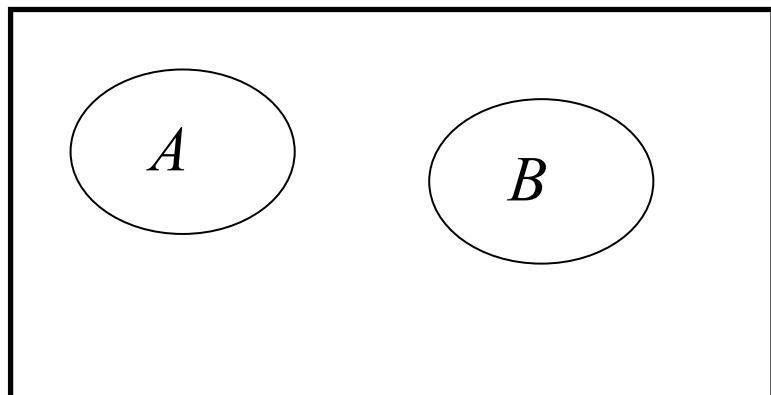
Thus in the case of independence the conditional probability of an event is not affected by the knowledge of the other event

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

mutually exclusive

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

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



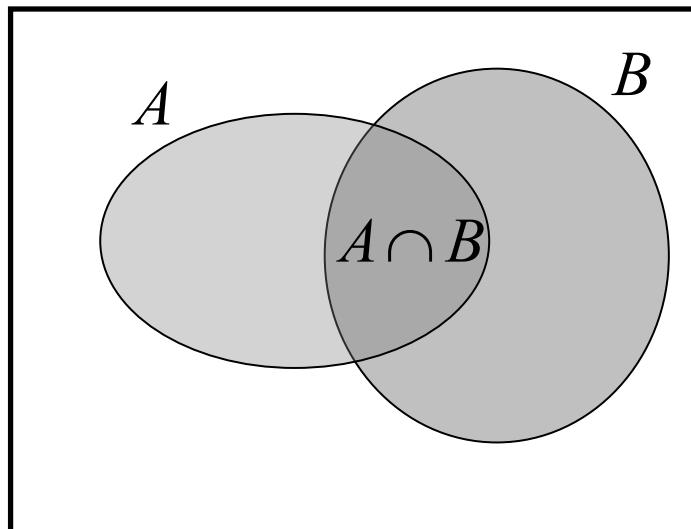
Mutually exclusive events are highly dependent otherwise. A and B **cannot** occur simultaneously. If one event occurs the other event does not occur.

Independent events

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

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

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

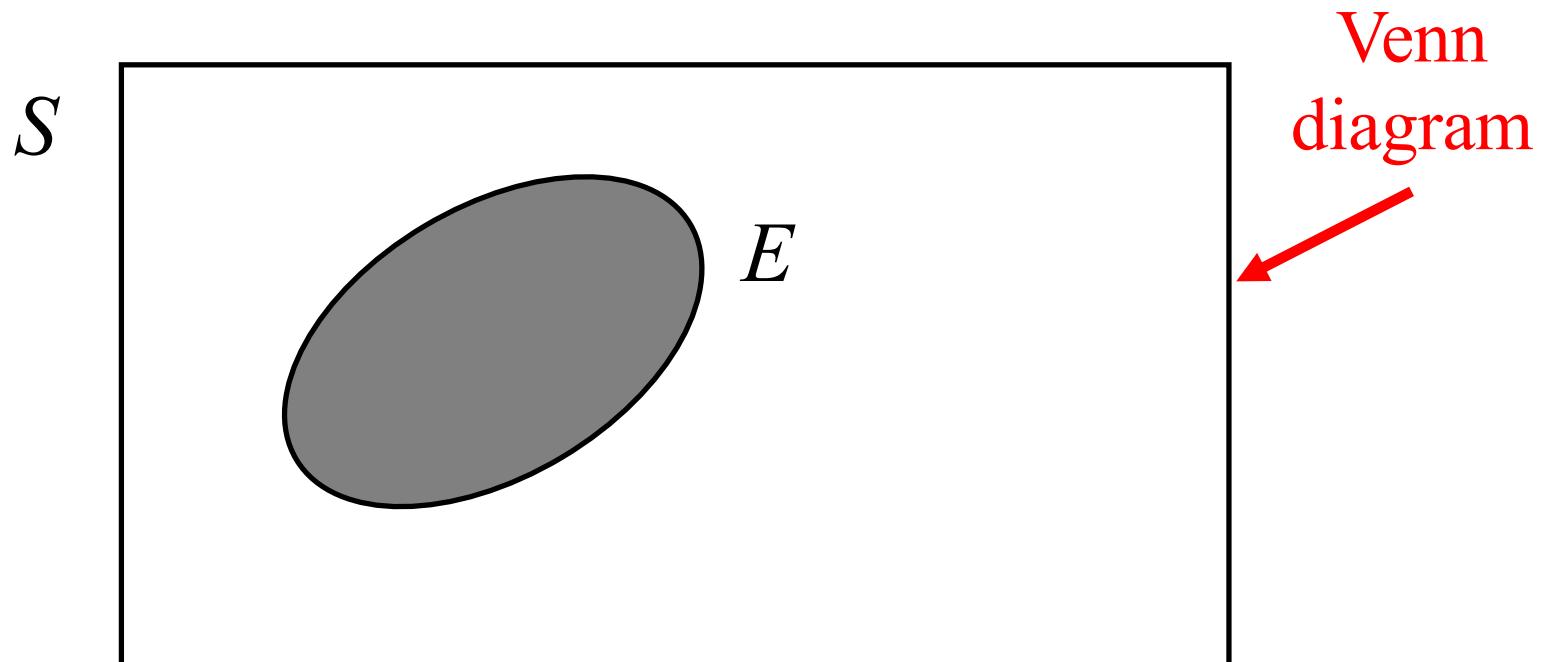
Models for **random** phenomena

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 $S = \{o_1, o_2, o_3, \dots o_N\}$ has a finite number, N , of outcomes.

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

Then for any event E

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

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

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

and

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

The Rule for complements

for any event E

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

Conditional probability

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

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.
 - b.
 - c. $P(\overline{A})$
 - d. $P(\overline{B})$
 - $P(A \text{ or } B)$
 - $P(A \text{ and } B)$

 - 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

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

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:
 $P(H \text{ and } H) = P(H) \cdot P(H)$

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

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:
$$P(H \text{ and } 4) = P(H) \cdot P(4)$$

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

Example.....

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

$$P(Q \text{ and } A) = P(Q) \cdot P(A)$$

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

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

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

b. Selecting a blue ball and
then a white ball

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

c. Selecting a red ball and
then a blue ball

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

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:
 $P(S \text{ and } S \text{ and } S) = P(S) \cdot P(S) \cdot P(S)$
 $P(\text{Stress}) = (0.46)^3 = 0.097$

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

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

- 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:
$$P(H \text{ and } C) = P(H) \cdot P(C|H)$$

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

Example.....

- 3 cards are drawn from a deck and NOT 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.....

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

b. Getting an ace, king, queen.....

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

c. Getting a club, spade, and heart.....

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

d. Getting 3 clubs.....

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

Back to Conditional Probability - Remember.....

$$P(A \text{ and } B) = P(A) \cdot P(B/A)$$

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

“Given”

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

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

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.....

$$P(\text{Deck and Garage}) = .42$$

$$P(\text{Deck}) = .60$$

Find $P(G/D)$

- Answer:

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

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

Independent \neq mutually exclusive

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

Conceptually, once A has happened, $\sim A$ 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:

$$P(\text{bridge}) = .83$$

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

$$P(\text{champ}/\text{bridge}) = .75$$

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

Find $P(\text{champ and bridge})$

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

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 $A = \text{'at least three heads'}$ $B = \text{'first toss is tails'}$.

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

Note: $n(A) =$ 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

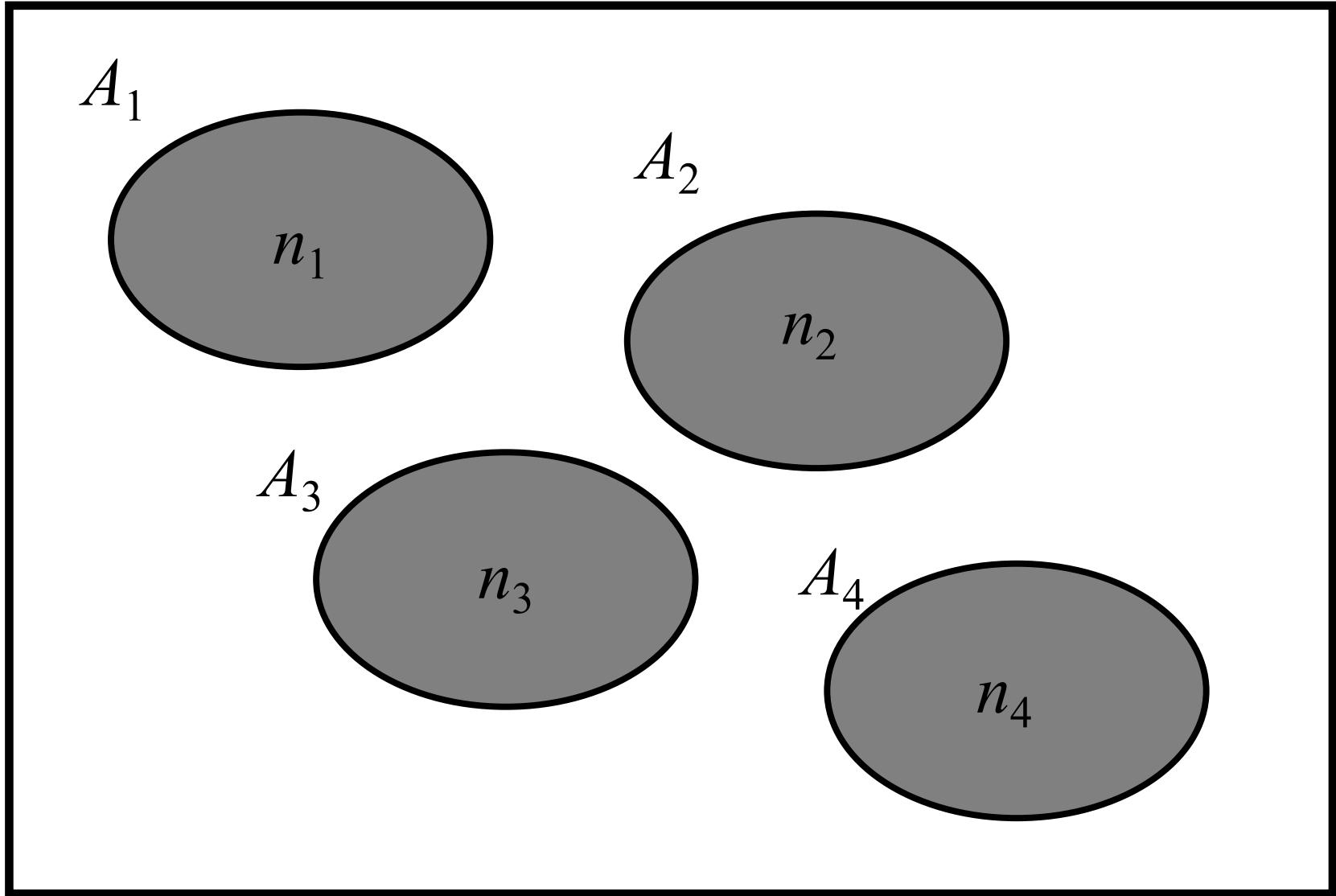
Suppose we carry out have a sets A_1, A_2, A_3, \dots
and that any pair are mutually exclusive

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

$n_i = n(A_i)$ = the number of elements in A_i .

Let $A = A_1 \cup A_2 \cup A_3 \cup \dots$

Then $N = n(A) =$ the number of elements in A
 $= n_1 + n_2 + n_3 + \dots$



Rule 2

Suppose we carry out two operations in sequence

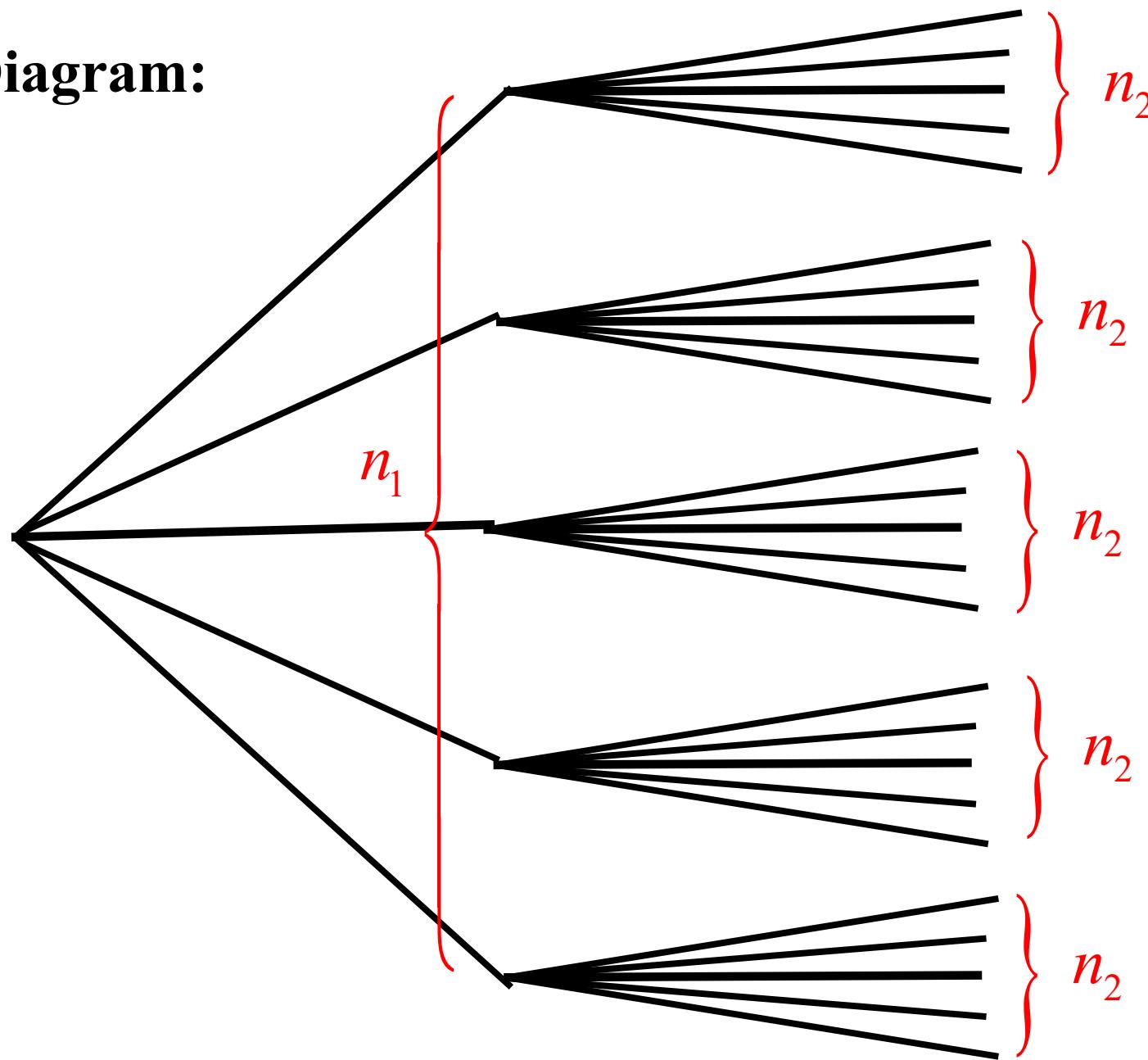
Let

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 $N = n_1 n_2$ = 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 many 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

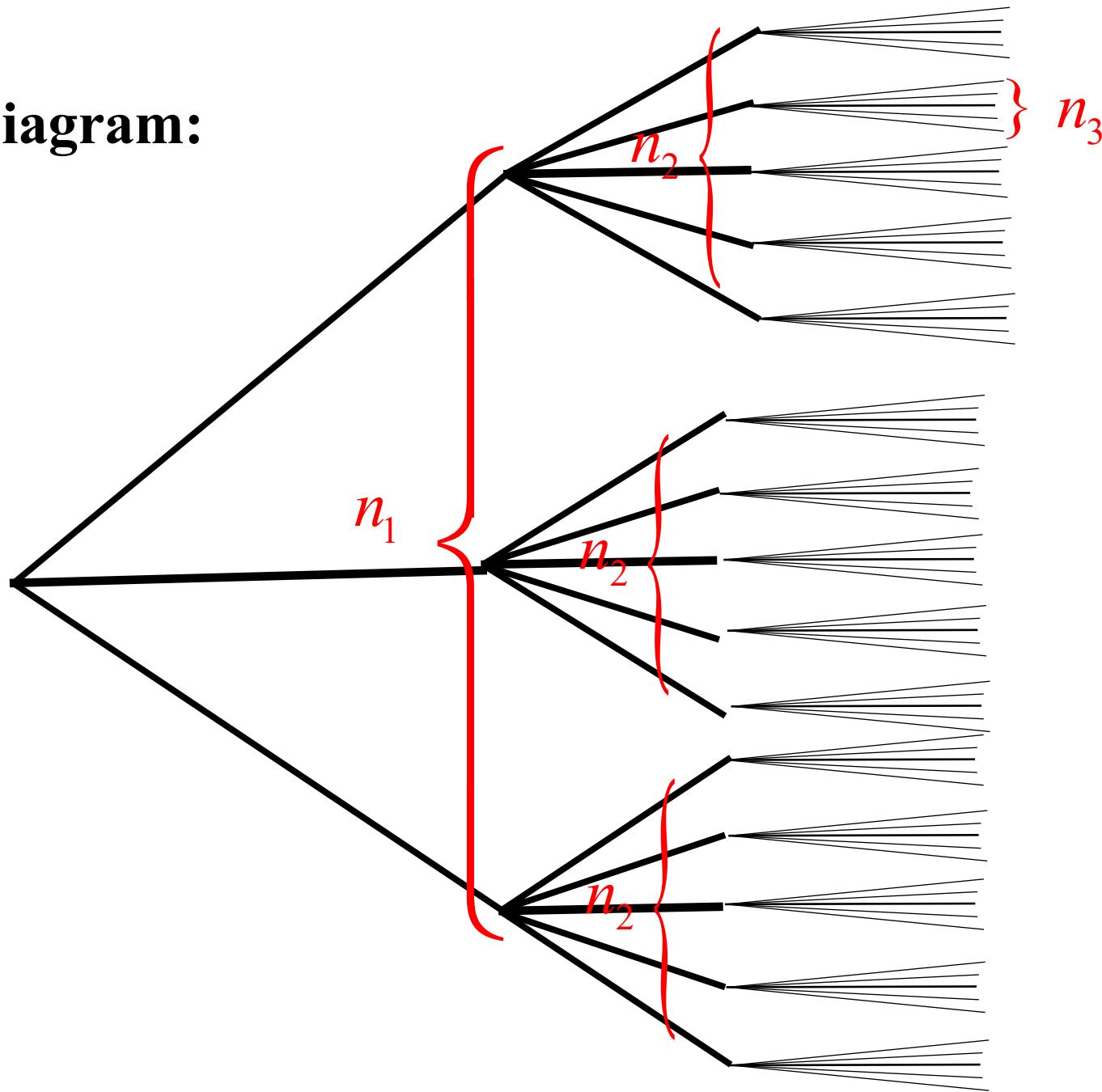
Let

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. $i = 2, 3, \dots, k$

Then $N = n_1 n_2 \dots n_k$ = 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:

$$N = n(n - 1) \dots (n - k + 1) \dots (3)(2)(1) = n!$$

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:

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

This number is denoted by the symbol

$${}_n P_k = n(n - 1) \dots (n - k + 1) = \frac{n!}{(n - k)!}$$

Definition: $0! = 1$

This definition is consistent with

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

for $k = n$

$${}_nP_n = \frac{n!}{0!} = \frac{n!}{1} = n!$$

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

Solution:
$${}_5 P_3 = \frac{5!}{(5-3)!} = 5(4)(3) = 60$$

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: Essentially 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 ($\leq n$):** 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
$${}_n P_k = \frac{n!}{(n-k)!} = n_1 k!$$

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

The number:

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

is denoted by the symbol

$${}_n C_k \quad \text{or} \quad \binom{n}{k} \quad \text{read “}n \text{ choose } k\text{”}$$

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

${}_n C_k$ is also called a **binomial coefficient**.

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

Summary of counting rules

Rule 1

$$n(A_1 \cup A_2 \cup A_3 \cup \dots) = n(A_1) + n(A_2) + n(A_3) + \dots$$

if the sets A_1, A_2, A_3, \dots are pairwise mutually exclusive

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

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

$$N = n_1 n_2 \dots n_k$$

= 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. $i = 2, 3, \dots, k$

Basic counting formulae

1. Orderings

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

2. Permutations

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

3. Combinations

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

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



Bayes' Rule: derivation

- Definition:

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

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

$P(B|A)$



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:

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

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:

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

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”

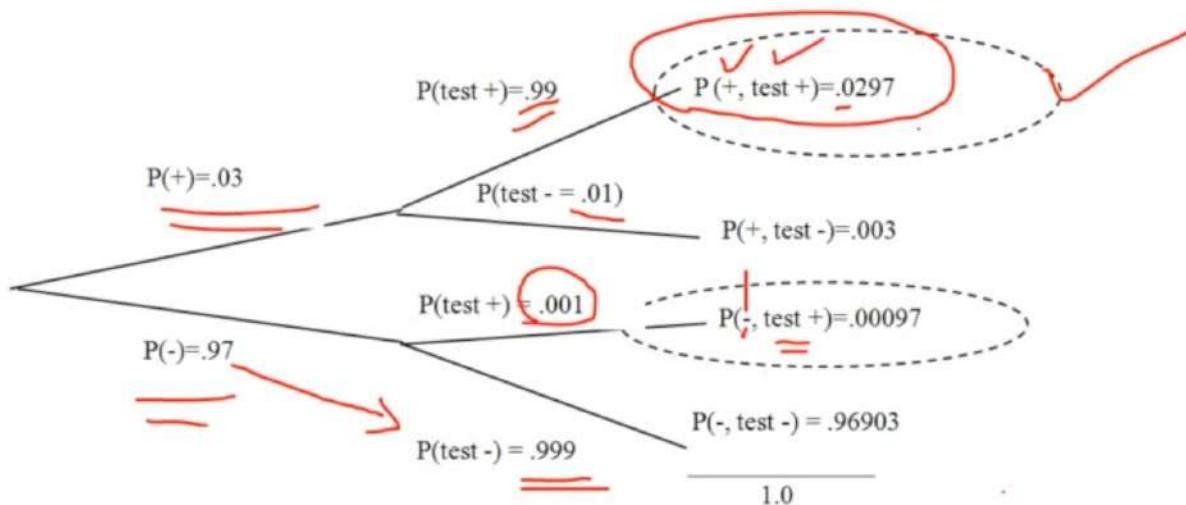
Bayes' Rule:

- 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.

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

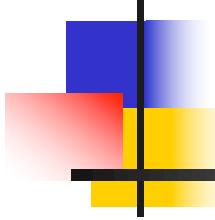
Answer: using Bayes' rule

$$P(true+ / test+) = \frac{P(test+ / true+)P(true+)}{P(test+ / true+)P(true+) + P(test+ / true-)P(true-)} =$$
$$\frac{.99(.03)}{.99(.03) + .001(.97)} = 96.8\%$$

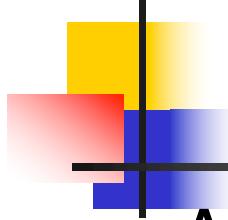
Practice problem

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?

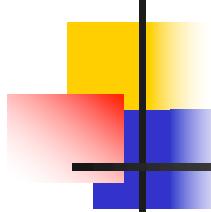


Probability Distributions



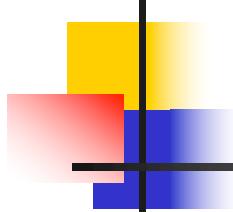
Random Variable

- 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 also a random variable (the percentage will be slightly differently every time you poll).
- Roughly, probability is how frequently we expect different outcomes to occur if we repeat the experiment over and over (“frequentist” view)



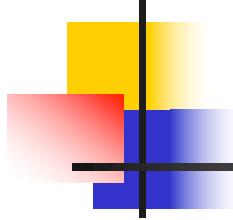
Random variables can be discrete or continuous

- **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.

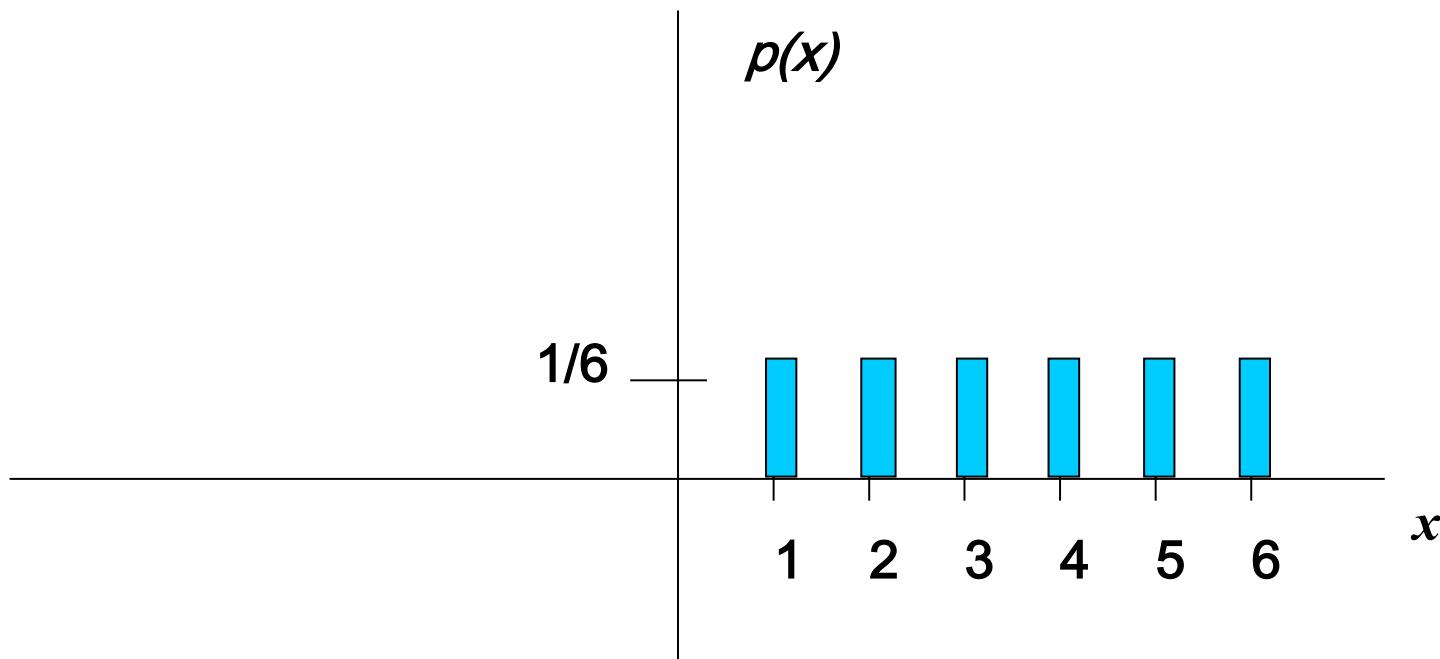


Probability functions

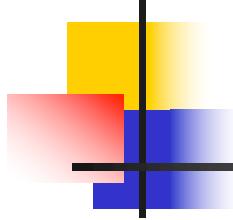
- A probability function maps the possible values of x against their respective probabilities of occurrence, $p(x)$
- $p(x)$ is a number from 0 to 1.0.
- The area under a probability function is always 1.



Discrete example: roll of a die



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

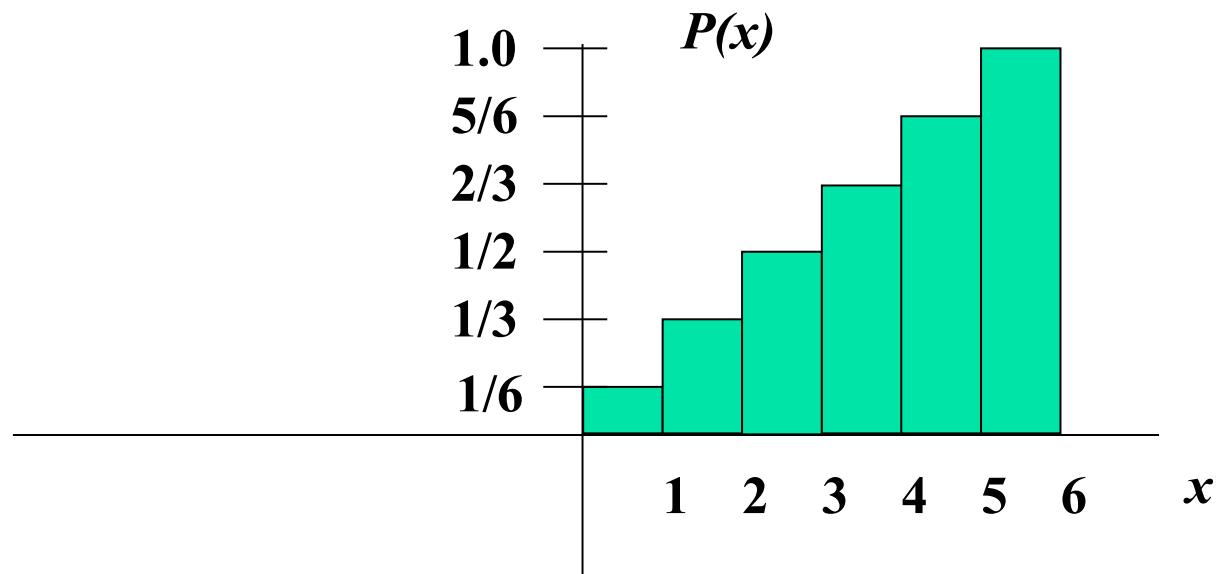


Probability mass function (pmf)

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	$p(x=6)=1/6$

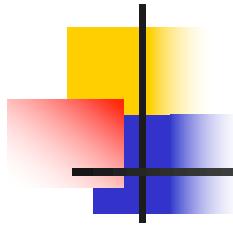
1.0

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	$P(x \leq 6) = 6/6$



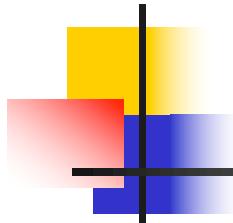
Examples

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?

$$P(x \geq 5) = 1 - P(x \leq 4) = 1 - 2/3 = 1/3$$



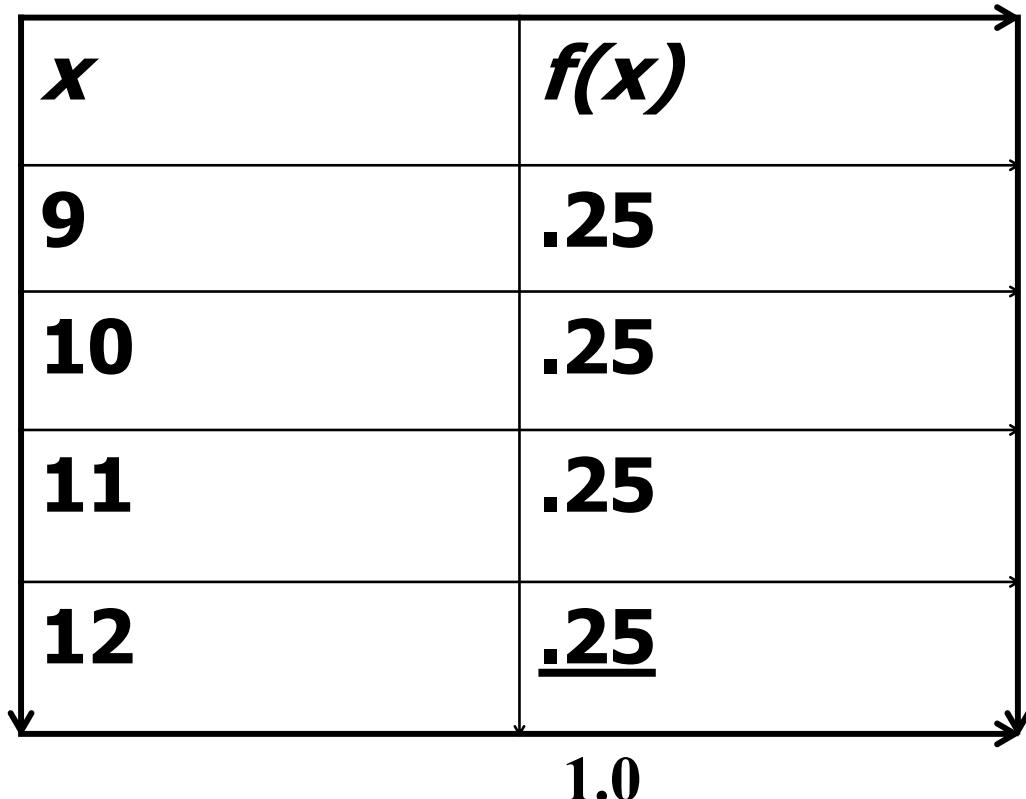
Practice Problem

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 $x=9, 10, 11, 12$



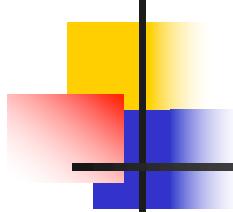
Yes, probability function!

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.



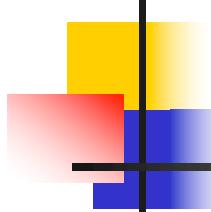
Answer (c)

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	<u>13/25</u>

$24/25$

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



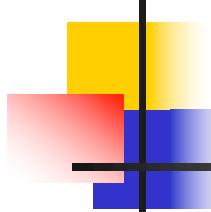
Practice Problem:

- 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:

- exactly 14 ships arrive $p(x=14) = .1$
- At least 12 ships arrive $p(x \geq 12) = (.2 + .1 + .1) = .4$
- At most 11 ships arrive $p(x \leq 11) = (.4 + .2) = .6$



Practice Problem:

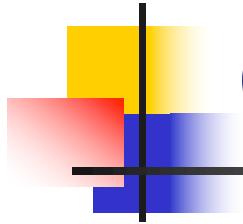
You are lecturing to a group of 1000 students. You ask them to each randomly pick an integer between 1 and 10. Assuming, their picks are truly random:

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

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

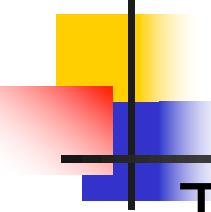
- What percentage of the students would you expect picked a number less than or equal to 6?

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



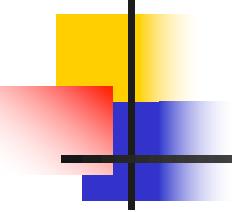
Important discrete distributions in epidemiology...

- 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

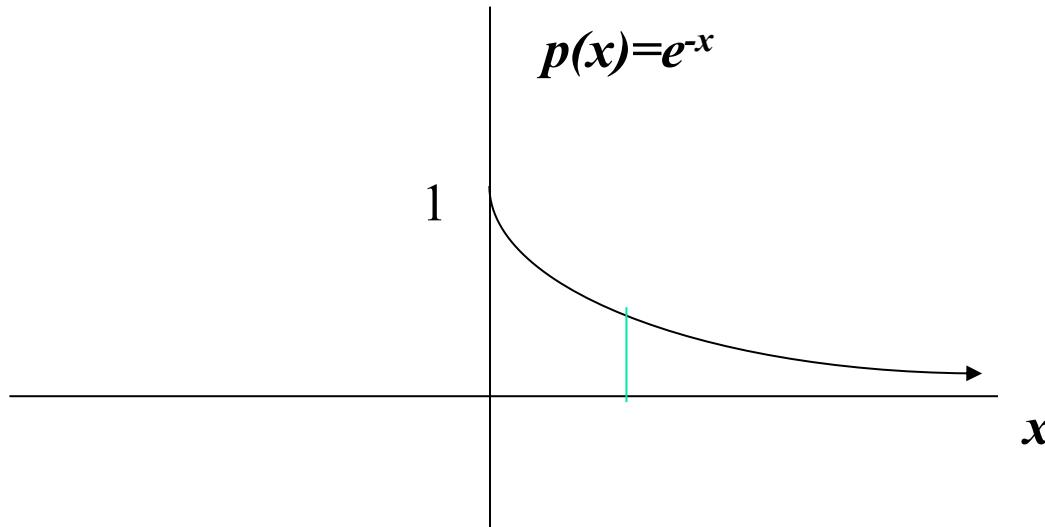
- For example, recall the negative exponential function (in probability, this is called an “exponential distribution”):

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

- This function integrates to 1:

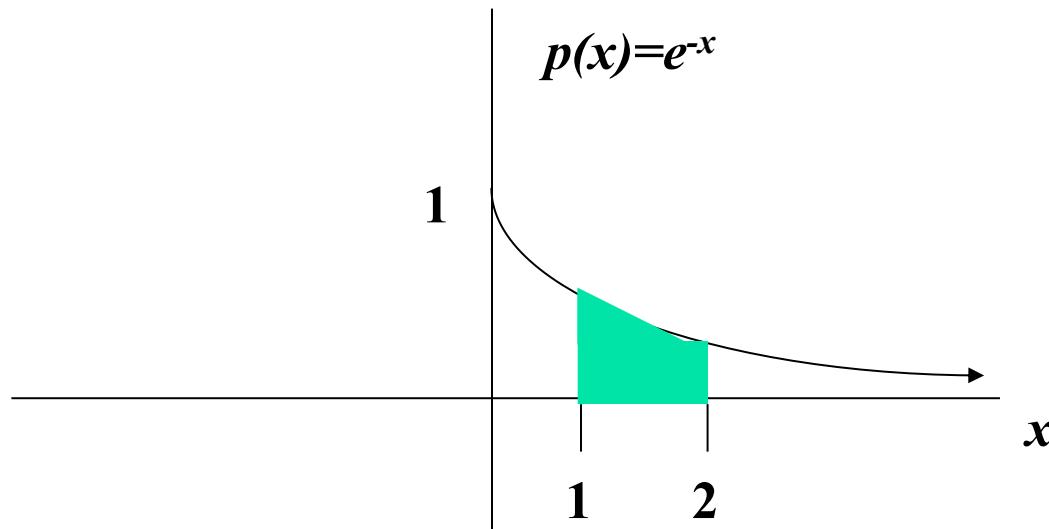
$$\int_0^{+\infty} e^{-x} = -e^{-x} \Big|_0^{+\infty} = 0 + 1 = 1$$

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:



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

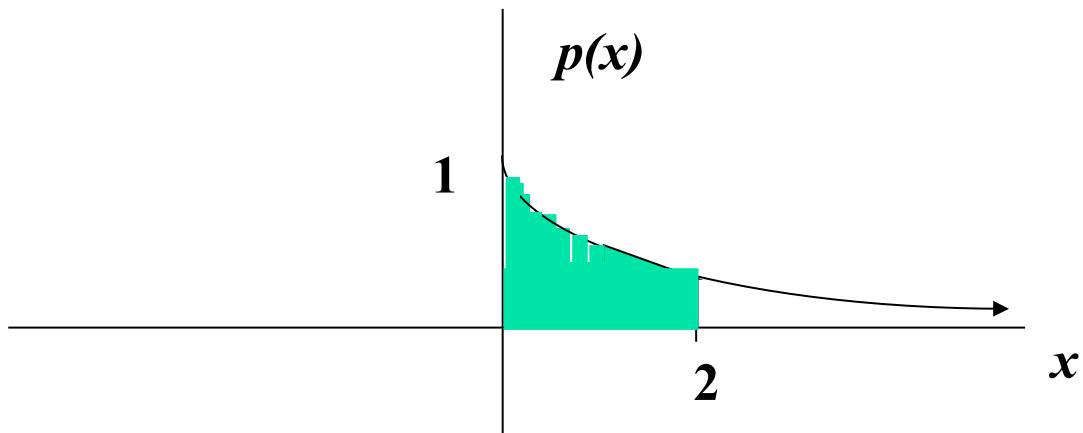
Cumulative distribution function

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

The CDF here = $P(x \leq A) =$

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

Example

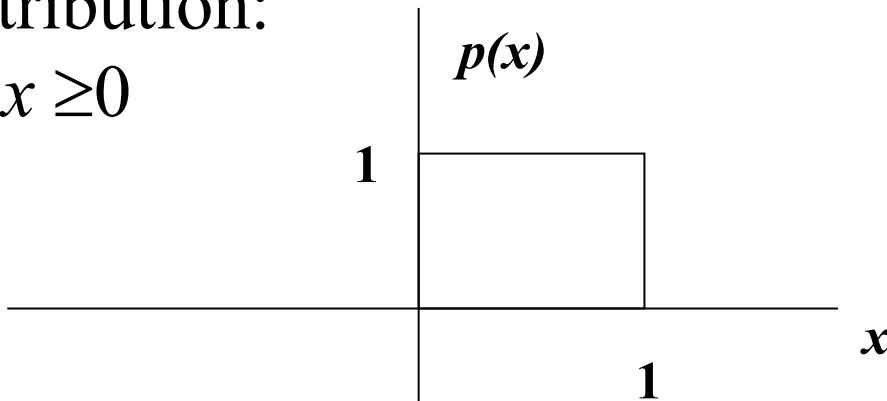


$$P(x \leq 2) = 1 - e^{-2} = 1 - .135 = .865$$

Example 2: Uniform distribution

The uniform distribution: all values are equally likely

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

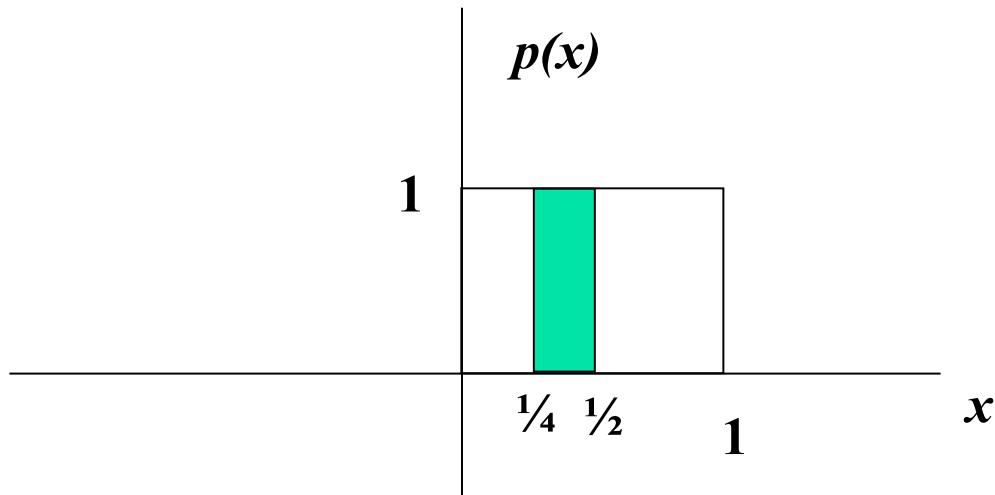


We can see it's a probability distribution because it integrates to 1 (the area under the curve is 1):

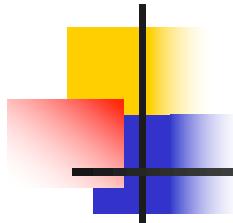
$$\int_0^1 1 = x \Big|_0^1 = 1 - 0 = 1$$

Example: Uniform distribution

What's the probability that x is between $\frac{1}{4}$ and $\frac{1}{2}$?



$$P(\frac{1}{2} \geq x \geq \frac{1}{4}) = \frac{1}{4}$$



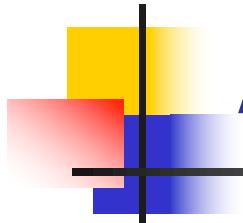
Practice Problem

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):

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

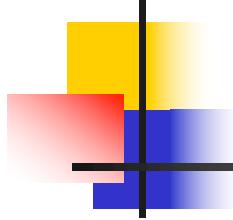
The probability of dying within 1 year can be calculated using the cumulative distribution function:

Cumulative distribution function is:

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

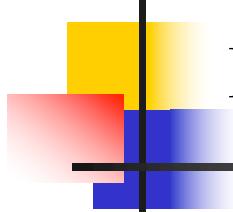
The chance of surviving past 1 year is: $P(x \geq 1) = 1 - P(x \leq 1)$

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

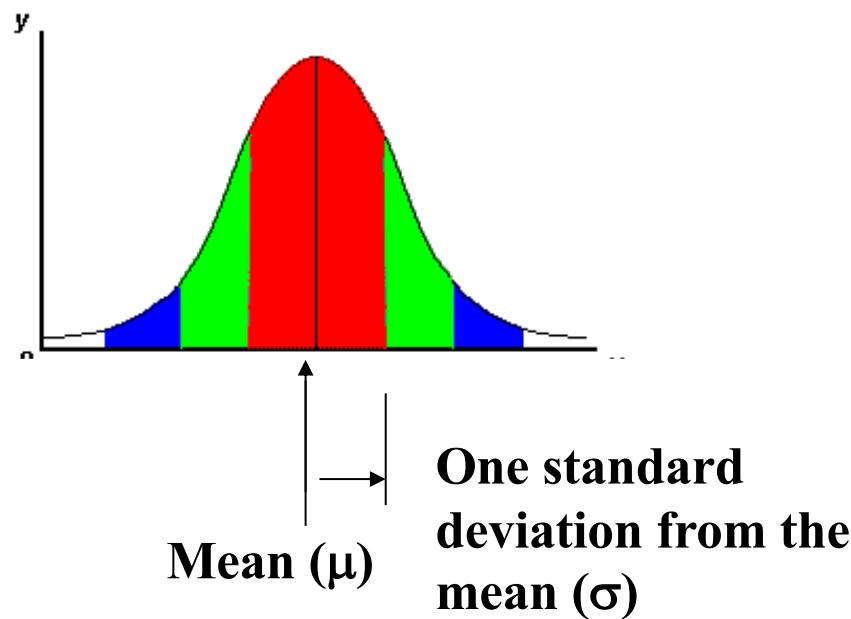


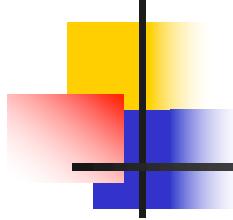
Expected Value and Variance

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



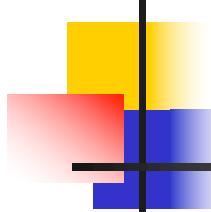
For example, bell-curve (normal) distribution:





Expected value, or mean

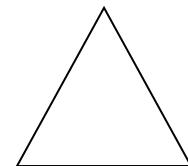
- If we understand the underlying probability function of a certain phenomenon, then we can make informed decisions based on how we expect x to behave on-average over the long-run...(so called “frequentist” theory of probability).
- Expected value is just the weighted average or mean (μ) of random variable 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 .



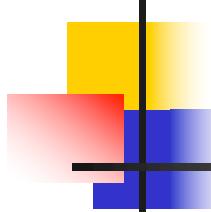
Example: expected value

- 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:

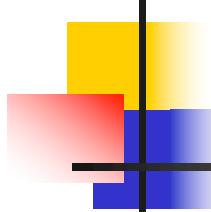
$$E(X) = \int_{\text{all } x} x_i p(x_i) dx$$

Empirical Mean is a special case of Expected Value...

Sample mean, for a sample of n subjects: $=$

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

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



Expected value, formally

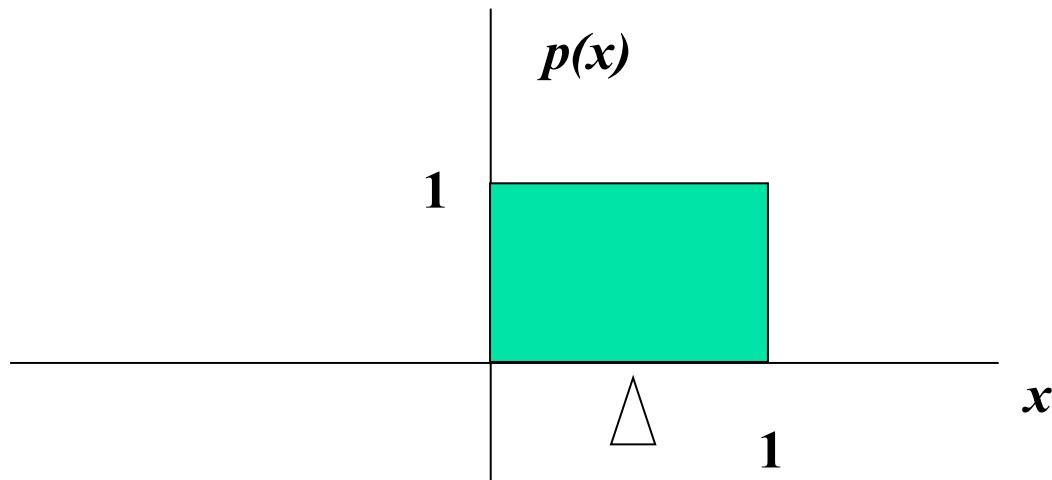
Discrete case:

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

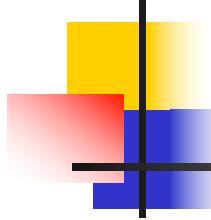
Continuous case:

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

Extension to continuous case: uniform distribution

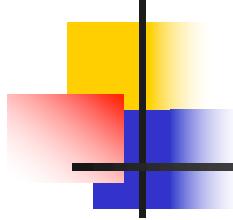


$$E(X) = \int_0^1 x(1)dx = \frac{x^2}{2} \Big|_0^1 = \frac{1}{2} - 0 = \frac{1}{2}$$



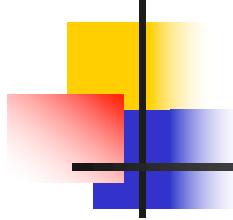
Symbol Interlude

- $E(X) = \mu$
 - 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

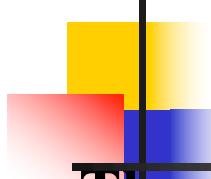
Calculate the probability of winning in 1 try:

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

“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	7.2×10^{-8}



Expected Value

The probability function

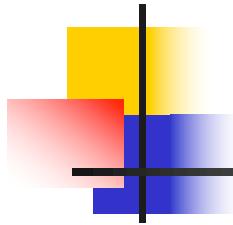
$x$$	$p(x)$
-1	.9999999928
+ 2 million	7.2×10^{-8}

Expected Value

$$\begin{aligned} E(X) &= P(\text{win}) * \$2,000,000 + P(\text{lose}) * -\$1.00 \\ &= 2.0 \times 10^6 * 7.2 \times 10^{-8} + .9999999928 (-1) = .144 - .9999999928 = -\$.86 \end{aligned}$$

Negative expected value is never good!

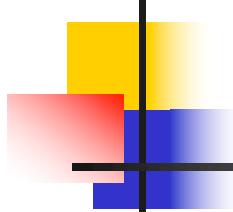
You shouldn't play if you expect to lose money!



Expected Value

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

$$520 \times (-.86) = -\$447.20$$



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.

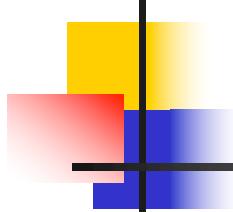
$$E(X) = 1(18/38) - 1(20/38) = -\$0.053$$

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

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

$$E(X) = 10(18/38) - 10(20/38) = -\$0.53$$

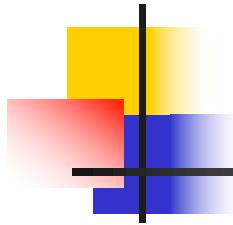
If the cost is \$10 per game, the casino wins an average of 53 cents per game. If 10,000 games are played in a night, that's a cool \$5300.



**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...

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

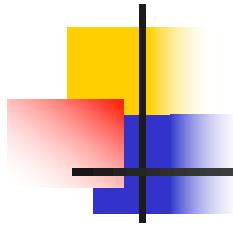


$$E(c) = c$$

$$E(c) = c$$

Example: If you cash in soda cans in CA, you always get 5 cents per can.

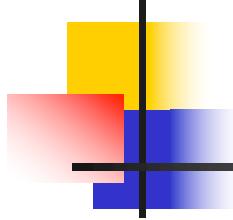
Therefore, there's no randomness. You always expect to (and do) get 5 cents.



$$E(cX) = cE(X)$$

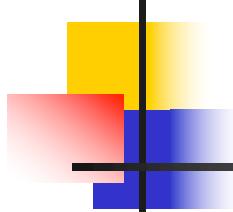
$$E(cX) = cE(X)$$

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!)


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

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

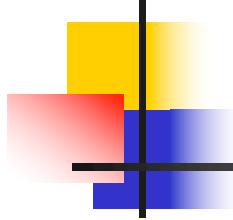
Example, if the casino throws in a free drink worth exactly \$5.00 every time you play a game, you always expect to (and do) gain an extra \$5.00 regardless of the outcome of the game.


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

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

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

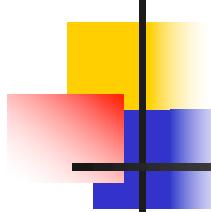
NOTE: This works even if X and Y are dependent!! Does not require independence!! Proof left for later...



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)

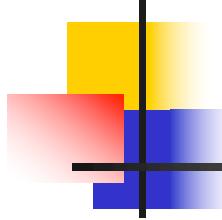
- a. Suppose a particular disease has a prevalence of 10% in a third-world population and you have 500 blood samples to screen. If you pool 20 samples at a time (25 lots), how many tests do you expect to have to run (assuming the test is perfect!)?

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

$$E(X) = P(\text{pooled lot is negative})(1) + P(\text{pooled lot is positive}) (21)$$

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

$$E(\text{total number of tests}) = 25 * 18.56 = 464$$

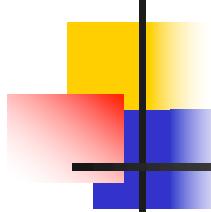


Answer (b)

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

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

$$50 \text{ lots} * 7.5 = 375$$

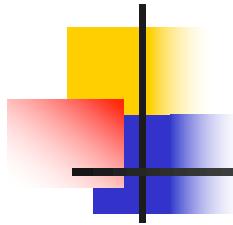


Answer (c)

c. 5 samples at a time?

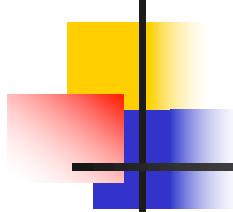
$$E(X) = (.90)^5 (1) + [1 - .90^5] (6) = 59\% (1) + 41\% (6) = 3.05 \text{ average per lot}$$

$$100 \text{ lots} * 3.05 = 305$$



Practice Problem

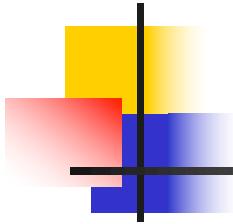
If X is a random integer between 1 and 10, what's the expected value of X ?



Answer

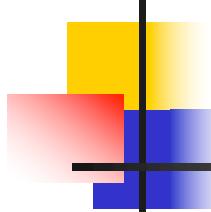
If X is a random integer between 1 and 10, what's the expected value of X ?

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



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?

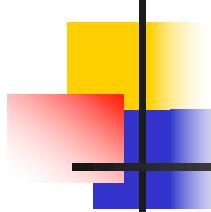


Deal or No Deal...

- 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

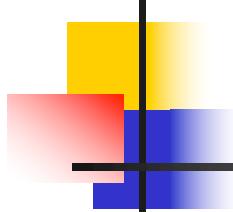
$$\mu = E(X) = 200,000$$



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?

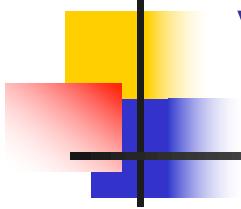


Variance/standard deviation

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

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

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



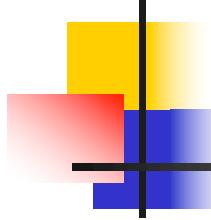
Variance, formally

Discrete case:

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

Continuous case:

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



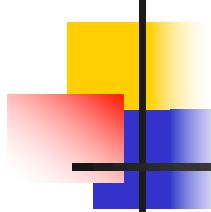
Similarity to empirical variance

The variance of a sample: $s^2 =$

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

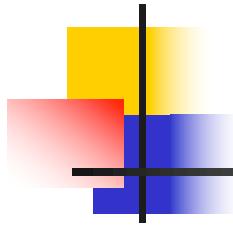


Division by $n-1$ reflects the fact that we have lost a “degree of freedom” (piece of information) because we had to estimate the sample mean before we could estimate the sample variance.



Symbol Interlude

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

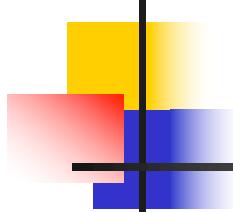


Variance: Deal or No Deal

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

$$\begin{aligned}\sigma^2 &= \sum_{\text{all } x} (x_i - \mu)^2 p(x_i) = \\ &= (1 - 200,000)^2 (.5) + (400,000 - 200,000)^2 (.5) = 200,000^2 \\ \sigma &= \sqrt{200,000^2} = 200,000\end{aligned}$$

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 = -\$0.053.

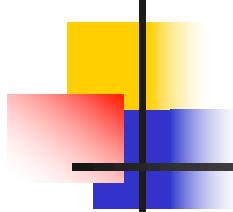
What's the variance of X ?



Answer

$$\begin{aligned}\sigma^2 &= \sum_{\text{all } x} (x_i - \mu)^2 p(x_i) \\&= (+1--.053)^2(18/38) + (-1--.053)^2(20/38) \\&= (1.053)^2(18/38) + (-1+.053)^2(20/38) \\&= (1.053)^2(18/38) + (-.947)^2(20/38) \\&= .997 \\ \sigma &= \sqrt{.997} = .99\end{aligned}$$

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!



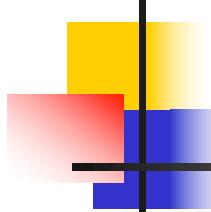
Handy calculation formula!

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:

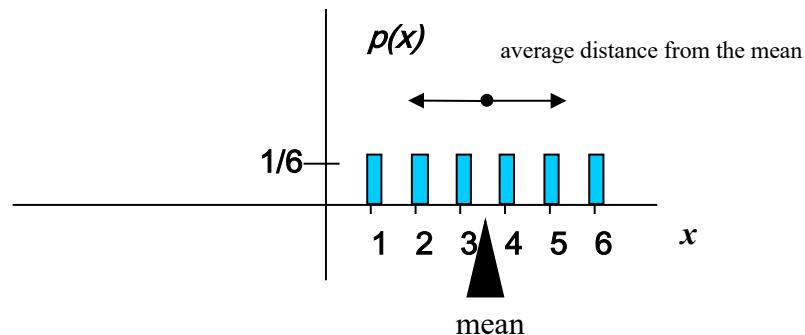
$$E(x-\mu)^2 =$$

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

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

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	

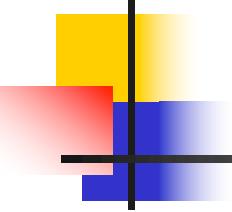


$$E(x) = \sum_{\text{all } x} x_i p(x_i) = (1)\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{21}{6} = 3.5$$

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

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

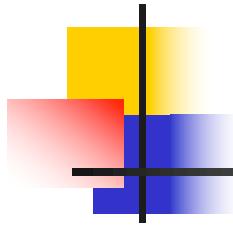
$$\sigma_x = \sqrt{2.92} = 1.71$$



**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

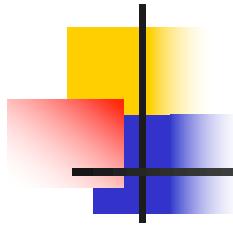
- $\text{Var}(c) = 0$
- $\text{Var}(c+X) = \text{Var}(X)$
- $\text{Var}(cX) = c^2\text{Var}(X)$
- $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$ ***ONLY IF X and Y are independent!!!!***
- $\{\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$ ***IF X and Y are not independent}***



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

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

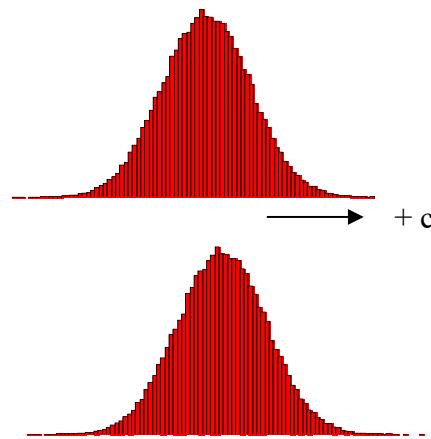
Constants don't vary!

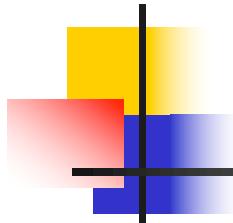


$$\text{Var}(c + X) = \text{Var}(X)$$

$$\text{Var}(c + X) = \text{Var}(X)$$

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.

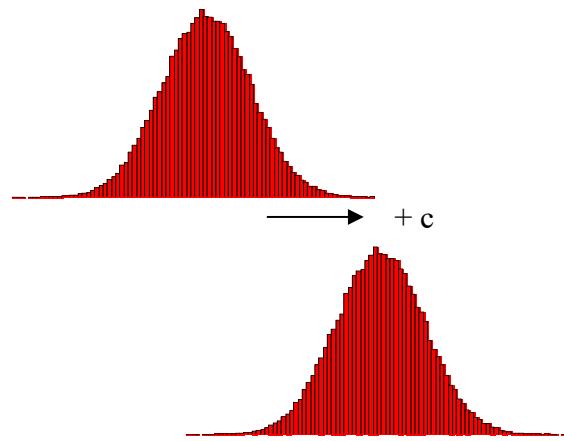


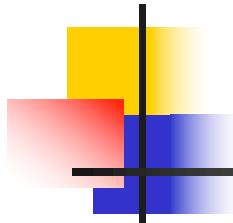


$$\text{Var}(c + X) = \text{Var}(X)$$

$$\text{Var}(c + X) = \text{Var}(X)$$

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.

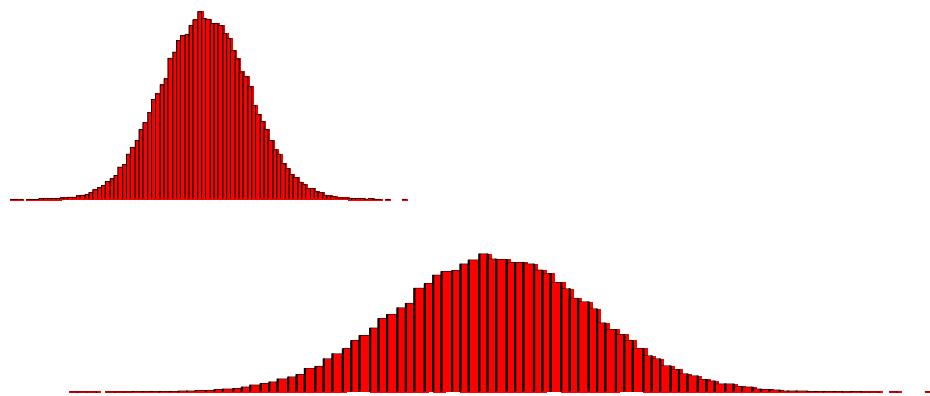


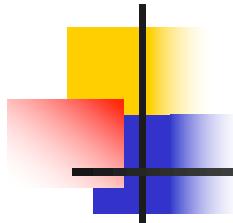


$$\text{Var}(cX) = c^2 \text{Var}(X)$$

$$\text{Var}(cX) = c^2 \text{Var}(X)$$

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.





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

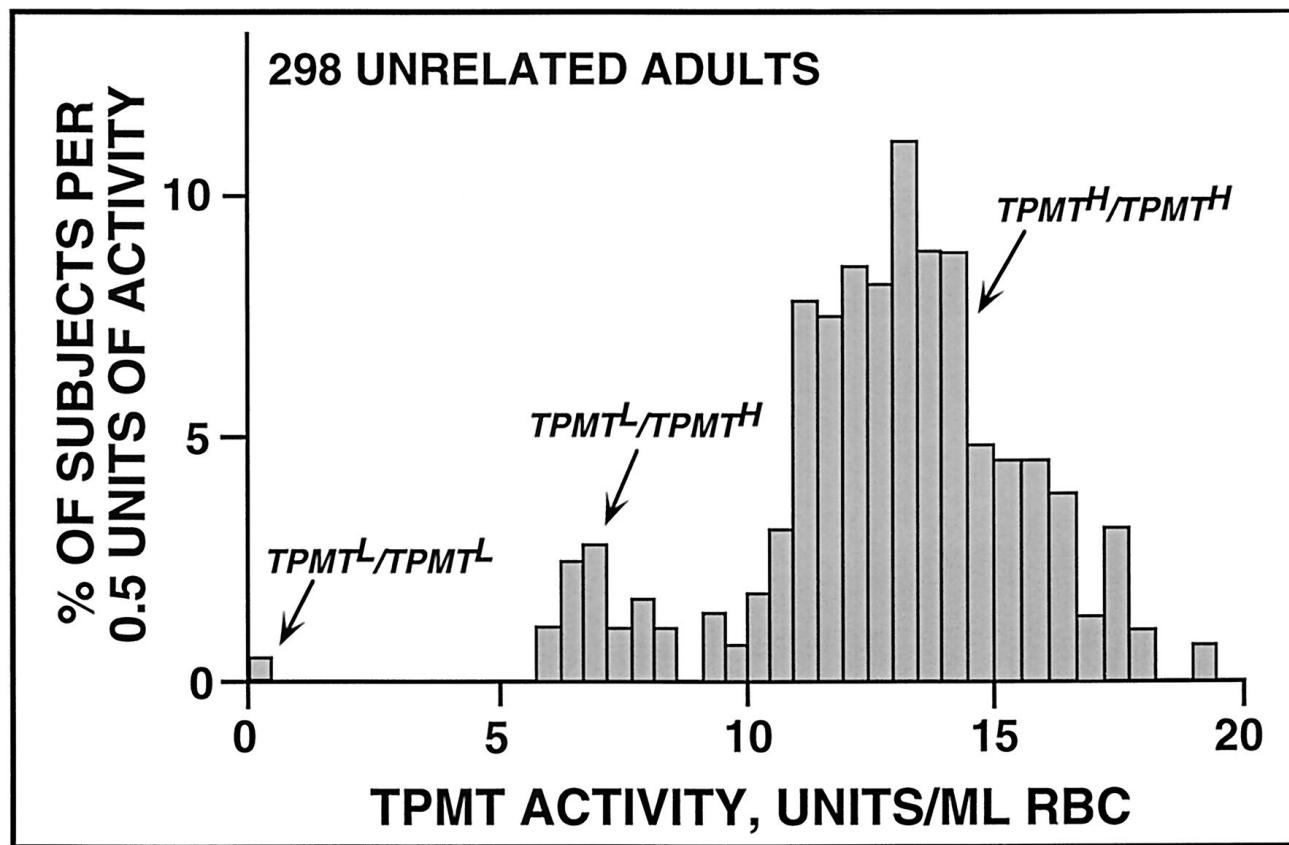
$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$ ***ONLY IF X and Y are independent!!!!!!***

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

Example of $\text{Var}(X+ Y) = \text{Var}(X)$ + $\text{Var}(Y)$: TPMT

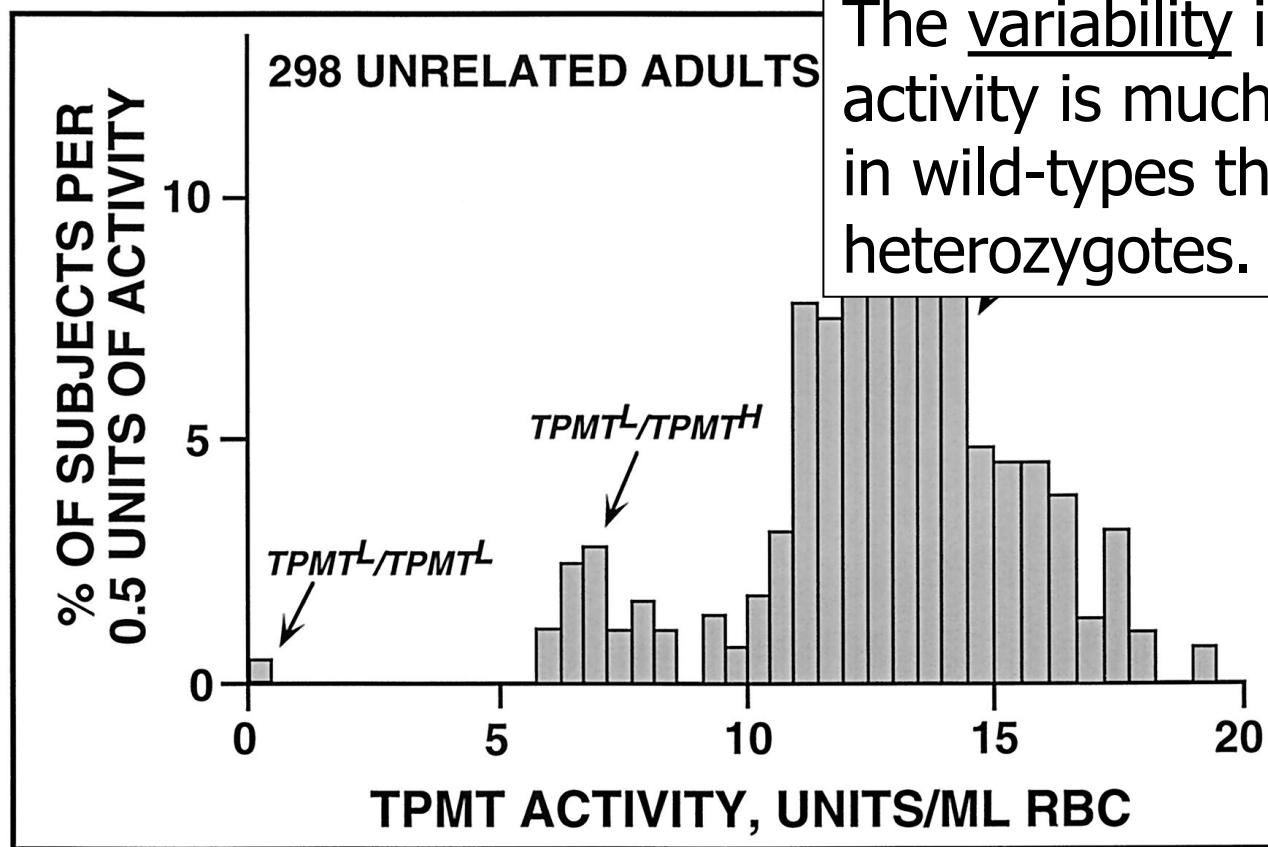
- TPMT metabolizes the drugs 6-mercaptopurine, azathioprine, and 6-thioguanine (chemotherapy drugs)
- People with TPMT⁻/ TPMT⁺ have reduced levels of activity (10% prevalence)
- People with TPMT⁻/ TPMT⁻ have no TPMT activity (prevalence 0.3%).
- They cannot metabolize 6-mercaptopurine, azathioprine,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

TPMT activity by genotype



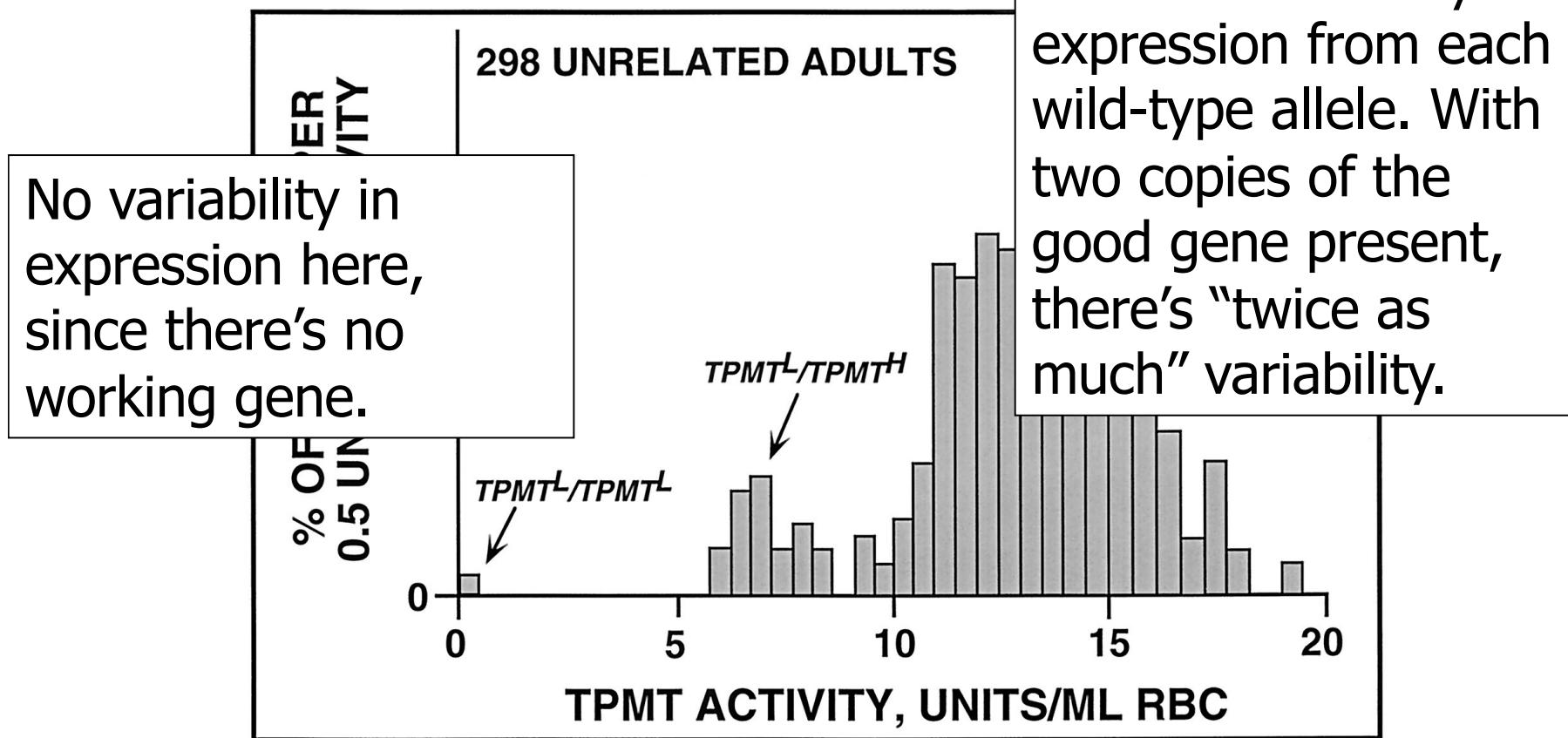
Weinshilboum R. Drug Metab Dispos. 2001 Apr;29(4 Pt 2):601-5

TPMT activity by genotype

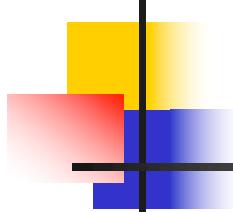


Weinshilboum R. Drug Metab Dispos. 2001 Apr;29(4 Pt 2):601-5

TPMT activity by genotype



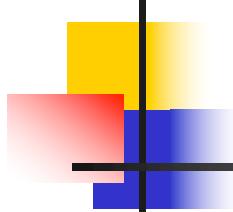
Weinshilboum R. Drug Metab Dispos. 2001 Apr;29(4 Pt 2):601-5



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

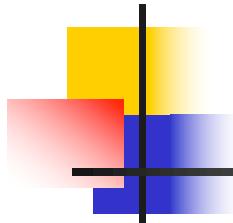
x^2	100	121	144	169	196
$P(x)$.4	.2	.2	.1	.1

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

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

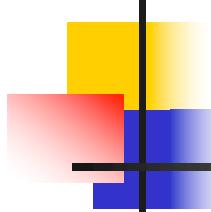
$$stddev(x) = \sqrt{1.81} = 1.35$$

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→

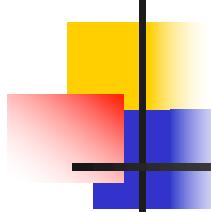
Think of tossing 1 coin. $E(X=\text{number of heads}) = (1)P(\text{heads}) + (0)P(\text{tails})$

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

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)

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

$$100 E(X_1) = 50$$



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.

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

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

$$\text{Var}(X) = .5 - .5^2 = .5 - .25 = .25$$

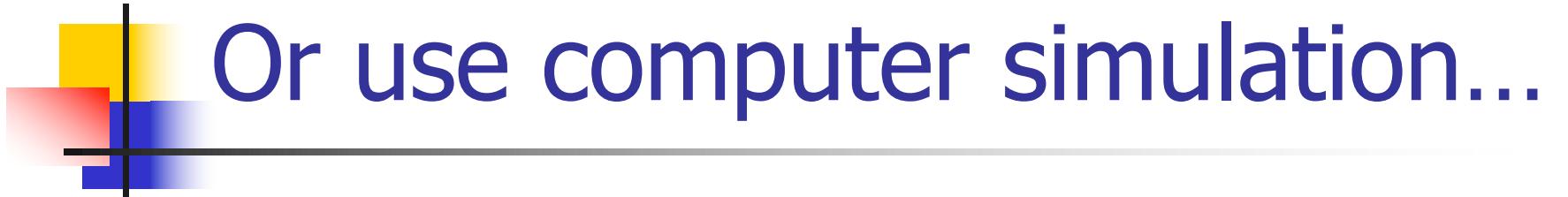
Then, using our rule: $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$ (coin tosses are independent!)

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

$$100 \text{ Var}(X_1) = 100 (.25) = 25$$

$$\text{SD}(X) = 5$$

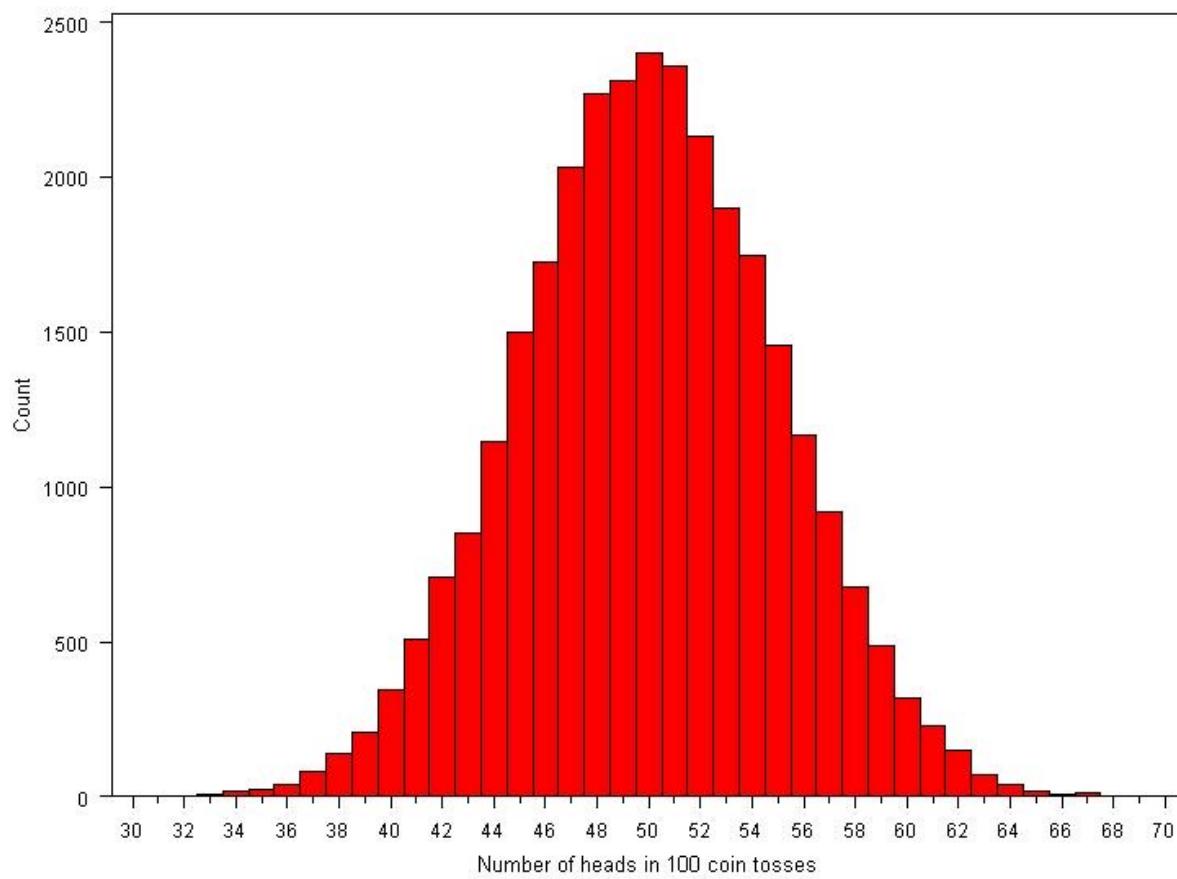
Interpretation: When we toss a coin 100 times, we expect to get 50 heads plus or minus 5.



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...

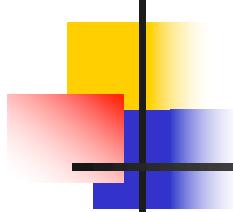


Mean = 50

Std. dev = 5

Follows a normal distribution

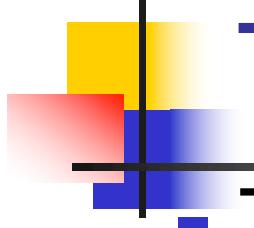
∴ 95% of the time, we get between 40 and 60 heads...



Covariance: joint probability

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

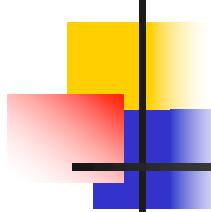
$$\sigma_{xy} = \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) P(x_i, y_i)$$



The Sample Covariance

■ The sample covariance:

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



Interpreting Covariance

- **Covariance** between two random variables:

$\text{cov}(X,Y) > 0 \rightarrow X \text{ and } Y \text{ are positively correlated}$

$\text{cov}(X,Y) < 0 \rightarrow X \text{ and } Y \text{ are inversely correlated}$

$\text{cov}(X,Y) = 0 \rightarrow X \text{ and } Y \text{ are independent}$

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 which 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 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 $x = 0, 1, 2, \dots, n$

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 exact arrangement?

$$P(\text{heads}) \times P(\text{heads}) \times P(\text{heads}) \times P(\text{tails}) \times P(\text{tails}) = (1/2)^3 \times (1/2)^2$$

Another way to get exactly 3 heads: THHHT

$$\begin{aligned} \text{Probability of this exact outcome} &= (1/2)^1 \times (1/2)^3 \times \\ &(1/2)^1 = (1/2)^3 \times (1/2)^2 \end{aligned}$$

Binomial distribution

In fact, $(1/2)^3 \times (1/2)^2$ 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:

$(1/2)^3 \times (1/2)^2 + (1/2)^3 \times (1/2)^2 + (1/2)^3 \times (1/2)^2 + \dots$
for as many unique arrangements as there are—but
how many are there??

$$\binom{5}{3}$$

ways to
arrange 3
heads in
5 trials

Outcome	Probability
THHHT	$(1/2)^3 \times (1/2)^2$
HHHTT	$(1/2)^3 \times (1/2)^2$
TTHHH	$(1/2)^3 \times (1/2)^2$
HTTHH	$(1/2)^3 \times (1/2)^2$
HHTTH	$(1/2)^3 \times (1/2)^2$
THTHH	$(1/2)^3 \times (1/2)^2$
HTHTH	$(1/2)^3 \times (1/2)^2$
HHTHT	$(1/2)^3 \times (1/2)^2$
THHTH	$(1/2)^3 \times (1/2)^2$
<u>HTHHT</u>	<u>$(1/2)^3 \times (1/2)^2$</u>
10 arrangements	$x (1/2)^3 \times (1/2)^2$

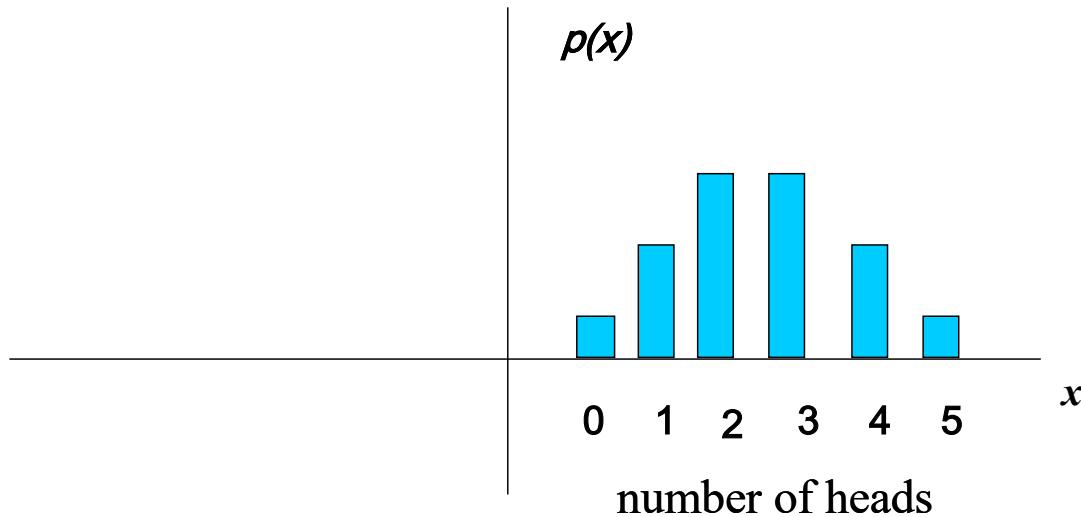
The probability
of each unique
outcome (note:
they are all
equal)

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

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

$$10 \times (\frac{1}{2})^5 = 31.25\%$$

Binomial distribution function:
X= the number of heads tossed in 5 coin tosses



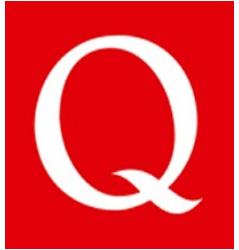
Example: A component has a 20% chance of being a dud. If five are selected from a large batch, what is the probability that more than one is a dud?

Answer:

Let X = number of duds in selection of 5

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

$$\begin{aligned} P(\text{More than one dud}) &= P(X > 1) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) \\ &= 1 - C_0^5 0.2^0 (1 - 0.2)^5 - C_1^5 0.2^1 (1 - 0.2)^4 \\ &= 1 - 1 \times 1 \times 0.8^5 - 5 \times 0.2 \times 0.8^4 \\ &= 1 - 0.32768 - 0.4096 \approx 0.263. \end{aligned}$$



Binomial or not?

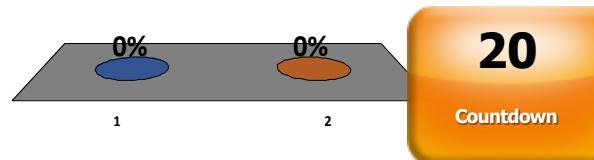
A mixed box of 10 screws contains 5 that are galvanized and 5 that are non-galvanized.

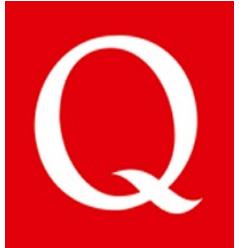
Three screws are picked at random without replacement. I want galvanized screws, so consider picking a galvanized screw to be a success.

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

$$\text{If } X \sim B(n, p), \quad \mu = E(X) = np \quad \sigma^2 = \text{var}(X) = np(1 - p)$$

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

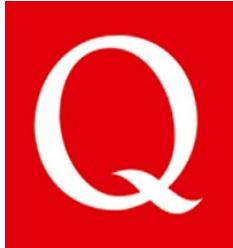
$$\mu_1 = \sum_k kP(X = k) = 1 \times p + 0 \times (1 - p) = p$$

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).

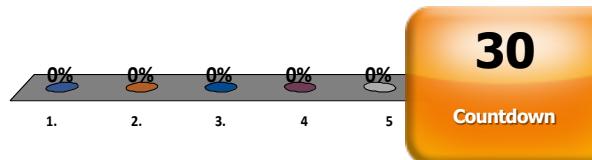
An opinion poll asks 1000 people if they will vote for Le Pen (*YES*) or not (*NO*). The expected number of Le Pen voters (*YESs*) in the poll is therefore $\mu = np = 200$

Mean and variance of a binomial distribution

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

What is the standard deviation (approximately)?

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



Binomial Distribution Summary

Discrete random variable $X \sim B(n, p)$ 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

$$\mu = np$$

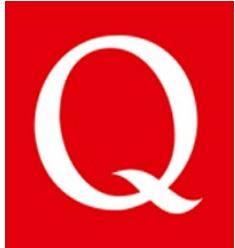
$$\sigma^2 = np(1 - p)$$

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” =

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

n = number of trials
 X = # successes out of n trials
 p = probability of success
 $1-p$ = probability of failure



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 (*YESs*) in the poll is therefore $\mu = np = 200$

What is the standard deviation (approximately)?



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

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

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

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

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

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

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

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

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

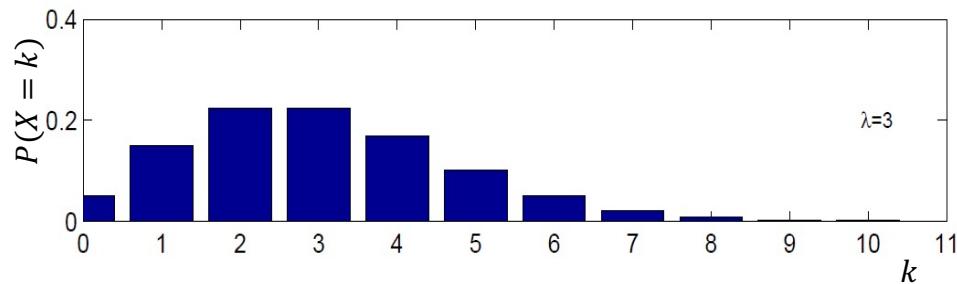
Then $X + Y$ is Poisson with average number $\lambda_X + \lambda_Y$

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.

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

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



Question from Derek Bruff

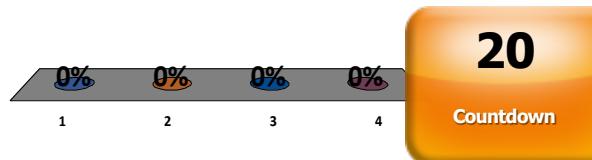
Poisson distribution

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?

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



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



Q

Poisson distribution

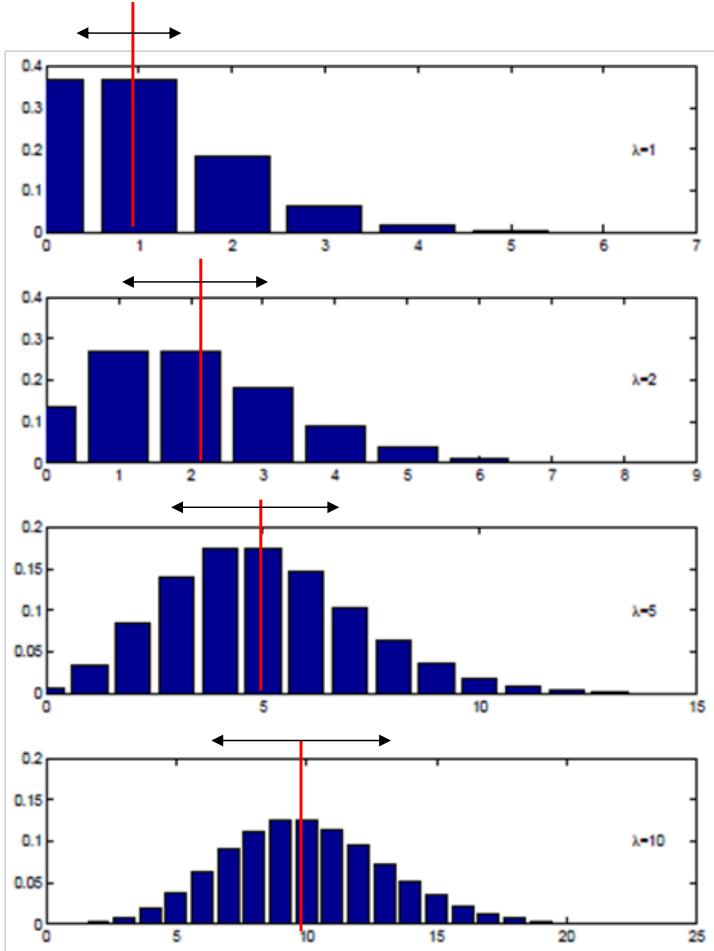
Suppose that trucks arrive at a receiving dock with an average arrival rate of 3 per hour. What is the probability exactly 5 trucks will arrive in a two-hour period?



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 $X \sim \text{Poisson}$ 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 $Y \sim \text{Poisson}$, 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

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

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

$$\begin{aligned} P(k = 0) &= \frac{e^{-\lambda} \lambda^k}{k!} = \frac{e^{-1.2t} (1.2t)^0}{0!} = e^{-1.2t} & \Rightarrow 1 - e^{-1.2t} \leq 0.05 \\ && \Rightarrow -e^{-1.2t} \leq 0.05 - 1 \\ && \Rightarrow e^{-1.2t} \geq 0.95 \\ && \Rightarrow -1.2t \geq \ln(0.95) = -0.05129 \end{aligned}$$

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

1. Binomial distribution

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

$$\mu = n \cdot p$$

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

3. Poisson Distribution

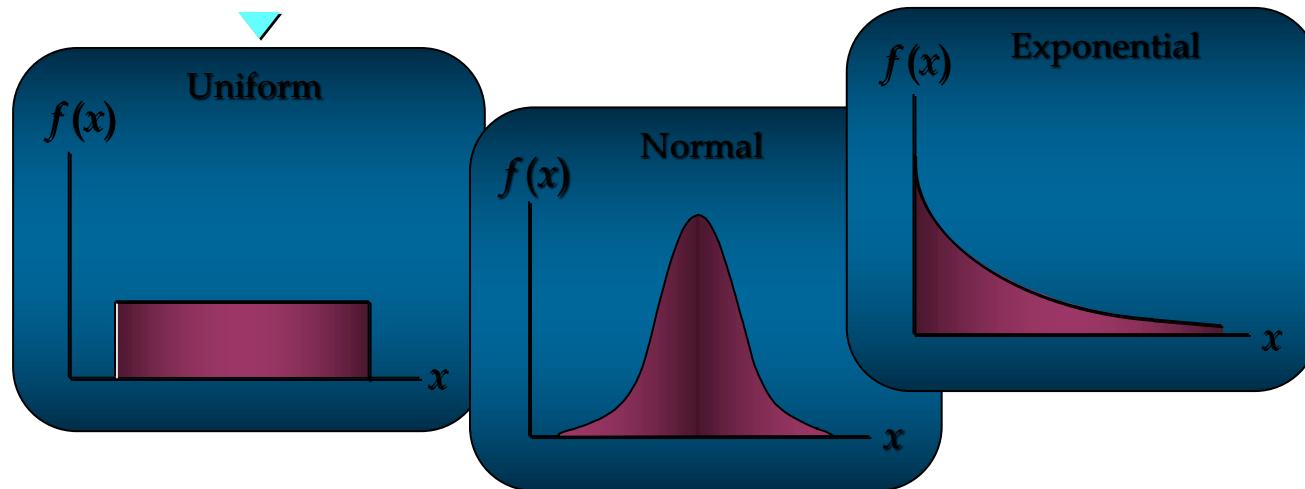
The Poisson distribution is characterized with λt where $\lambda =$ *the mean of outcomes* and $t =$ *time interval*.

$$\mu = \lambda t$$

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

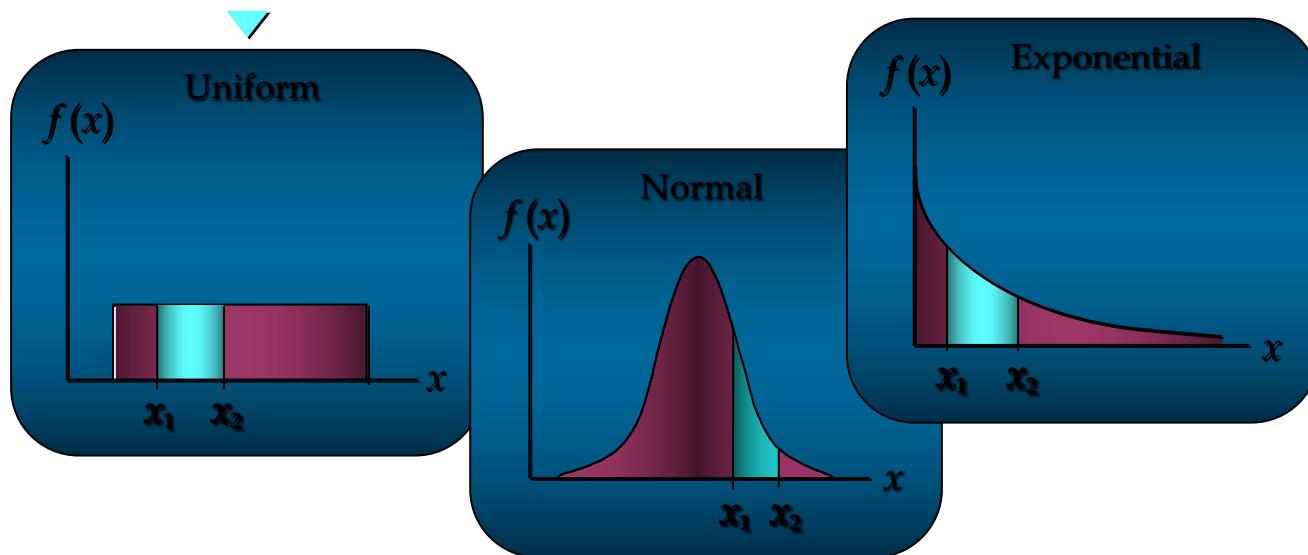
Uniform Probability Distribution

- 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:

$$\begin{aligned} > f(x) &= 1/(b - a) && \text{for } a \leq x \leq b \\ &= 0 && \text{elsewhere} \end{aligned}$$

where: a = smallest value the variable can assume

b = largest value the variable can assume

Uniform Probability Distribution

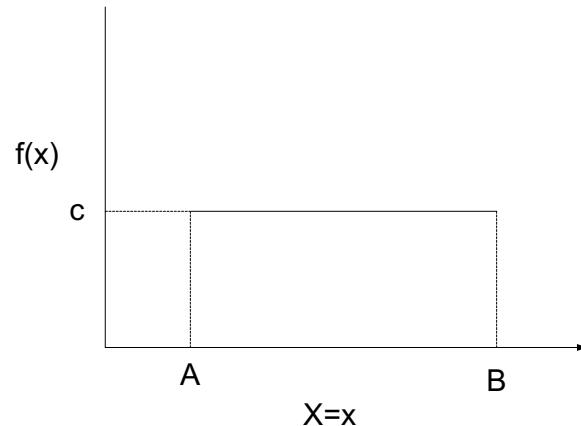
- ■ Expected Value of x

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

- ■ Variance of x

$$\text{Var}(x) = (b - a)^2/12$$

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

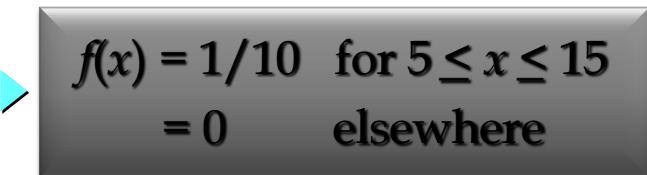
d) $\sigma^2 = \frac{(B-A)^2}{12}$

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


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

where:

x = salad plate filling weight

Uniform Probability Distribution

► ■ Expected Value of x

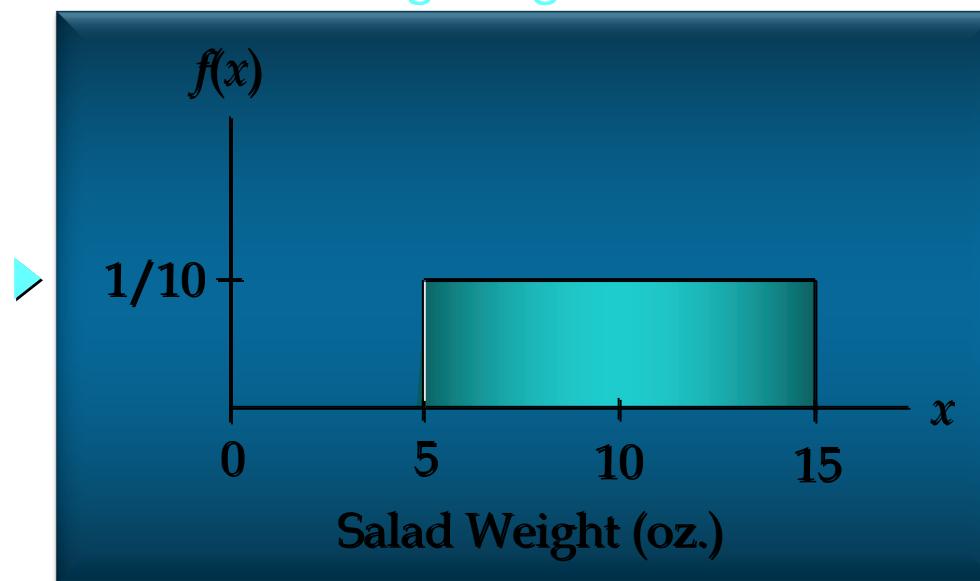
$$\begin{aligned} E(x) &= (a + b)/2 \\ &= (5 + 15)/2 \\ &= 10 \end{aligned}$$

► ■ Variance of x

$$\begin{aligned} \text{Var}(x) &= (b - a)^2/12 \\ &= (15 - 5)^2/12 \\ &= 8.33 \end{aligned}$$

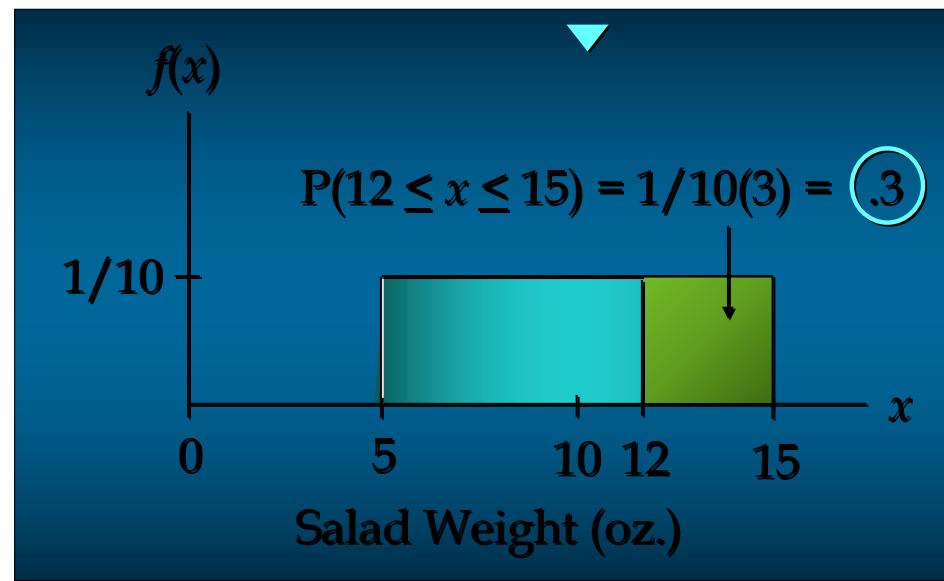
Uniform Probability Distribution

■ Uniform Probability Distribution
for Salad Plate Filling Weight

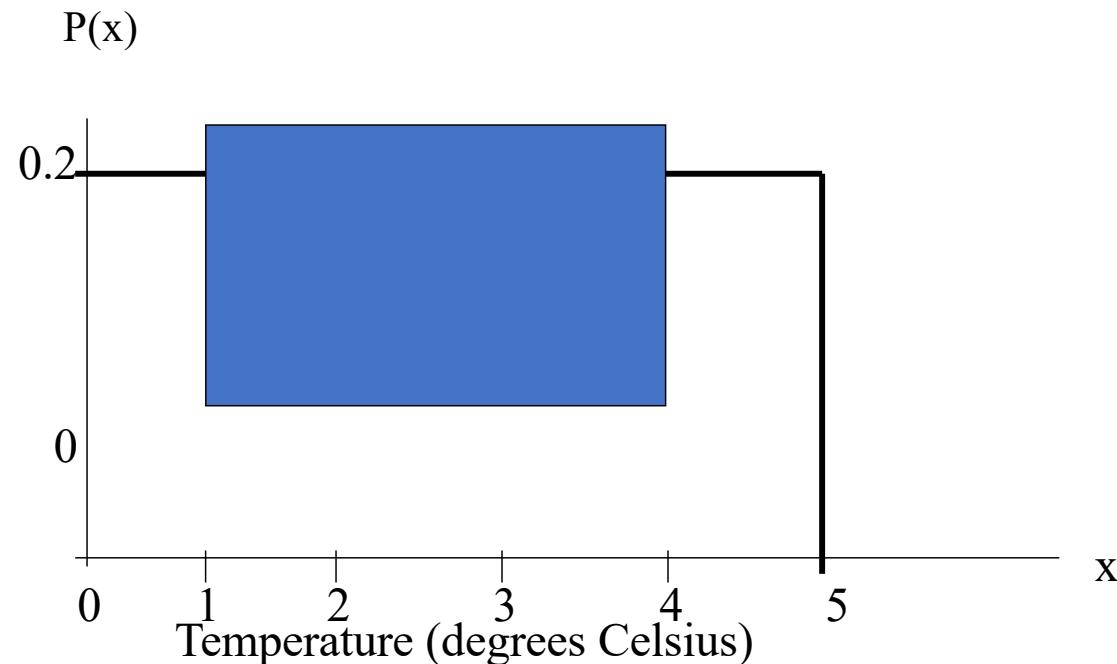


Uniform Probability Distribution

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°C and 4°C ?



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

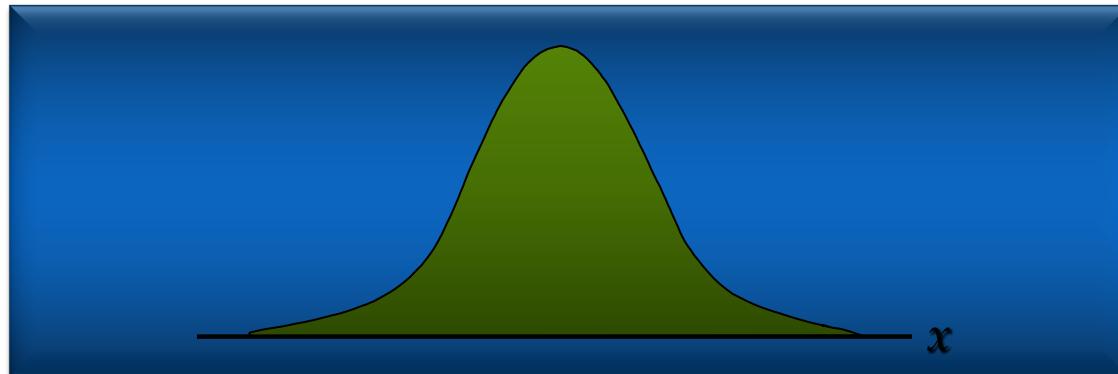
π = 3.14159

e = 2.71828

Normal Probability Distribution

■ Characteristics

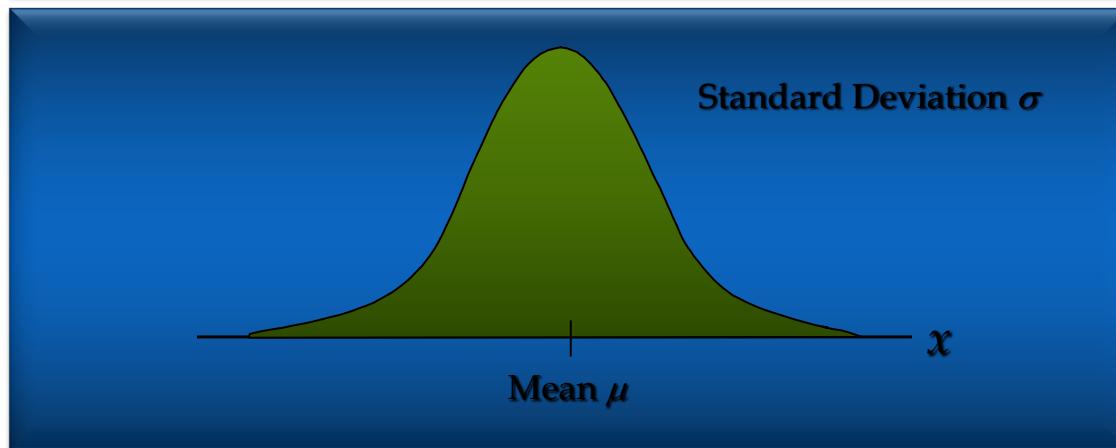
- ▶ The distribution is symmetric; its skewness measure is zero.



Normal Probability Distribution

■ Characteristics

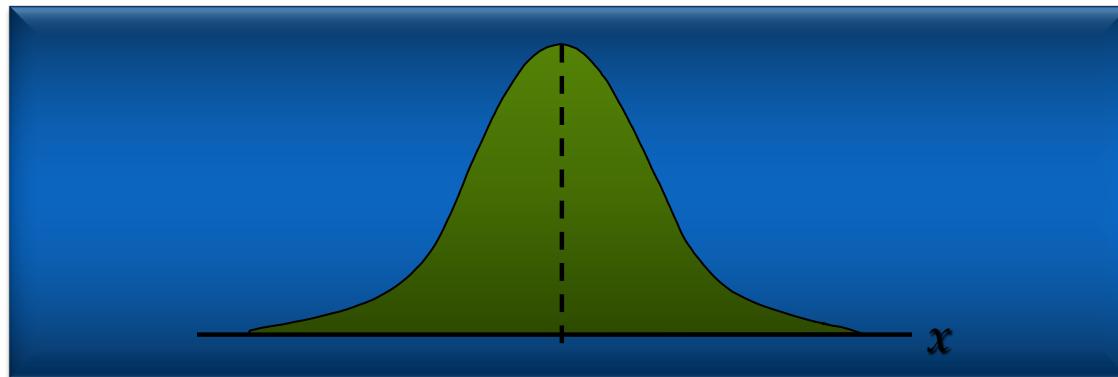
- ▶ The entire family of normal probability distributions is defined by its mean μ and its standard deviation σ .



Normal Probability Distribution

■ Characteristics

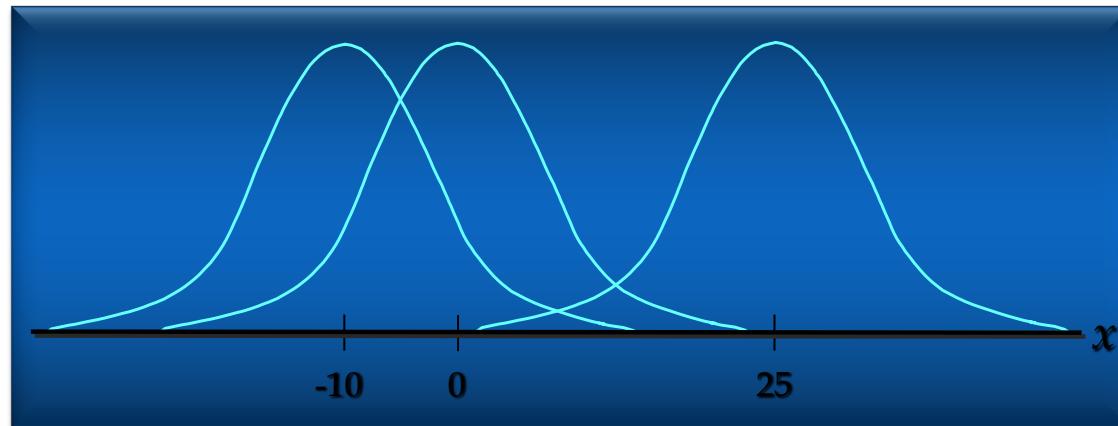
- ▶ The highest point on the normal curve is at the mean, which is also the median and mode.



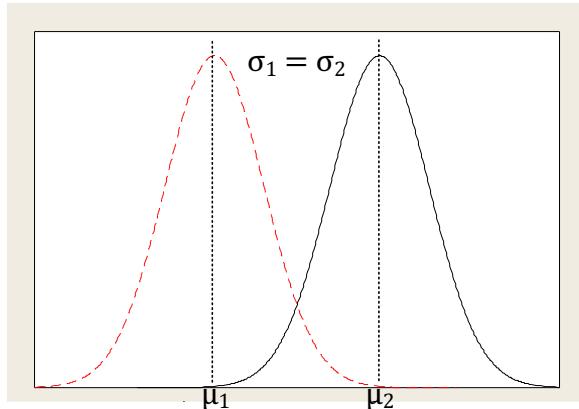
Normal Probability Distribution

■ Characteristics

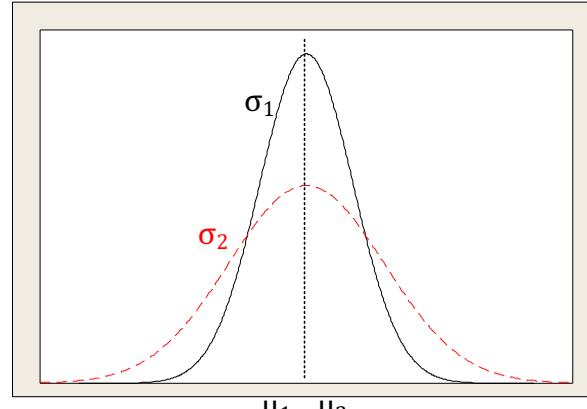
- ▶ The mean can be any numerical value: negative, zero, or positive.



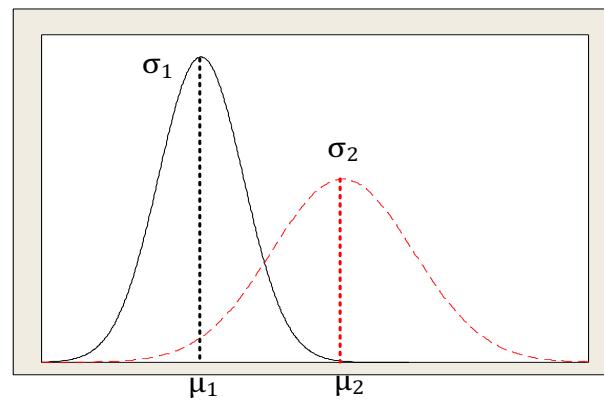
Normal Distribution



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



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

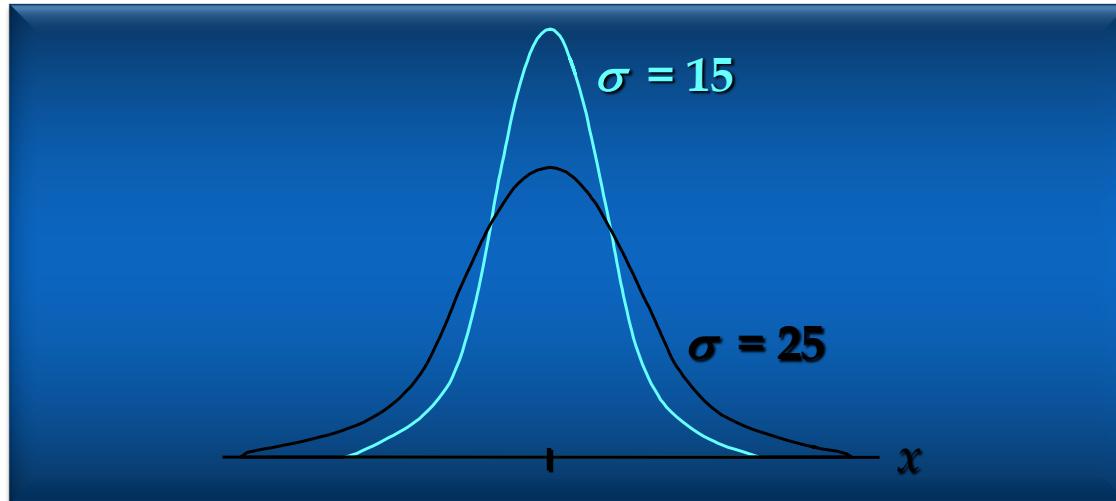


Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$

Normal Probability Distribution

■ Characteristics

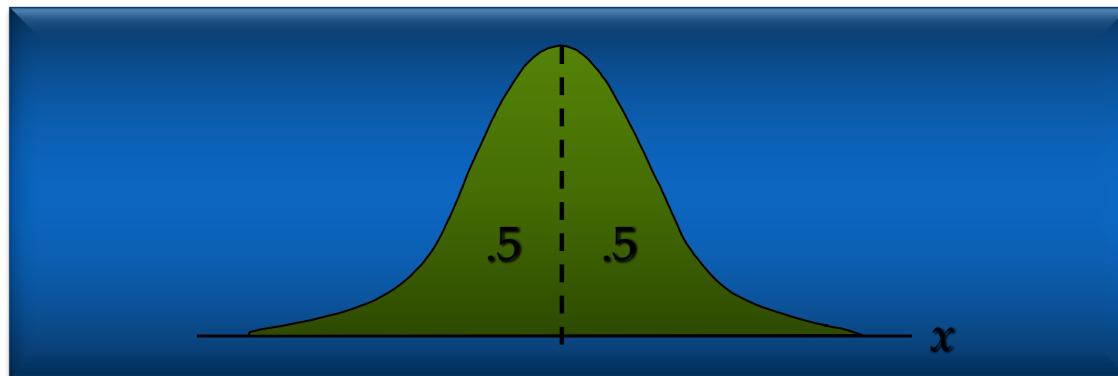
- ▶ The standard deviation determines the width of the curve: larger values result in wider, flatter curves.



Normal Probability Distribution

■ Characteristics

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



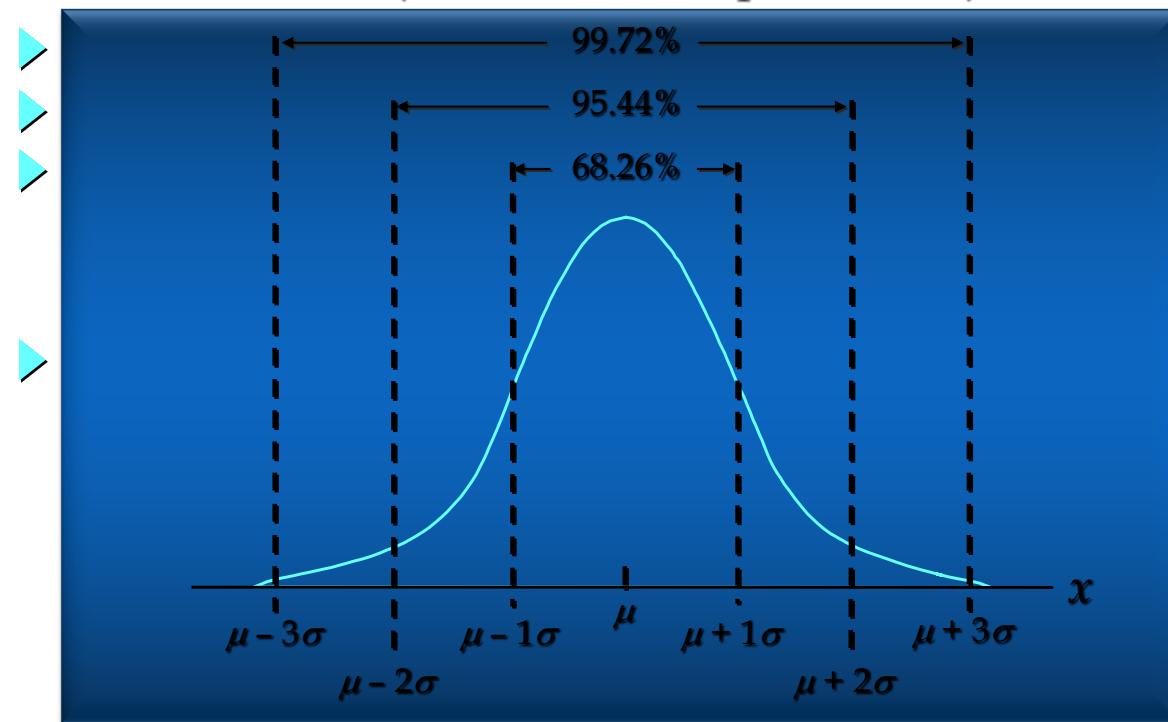
Normal Probability Distribution

■ 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.

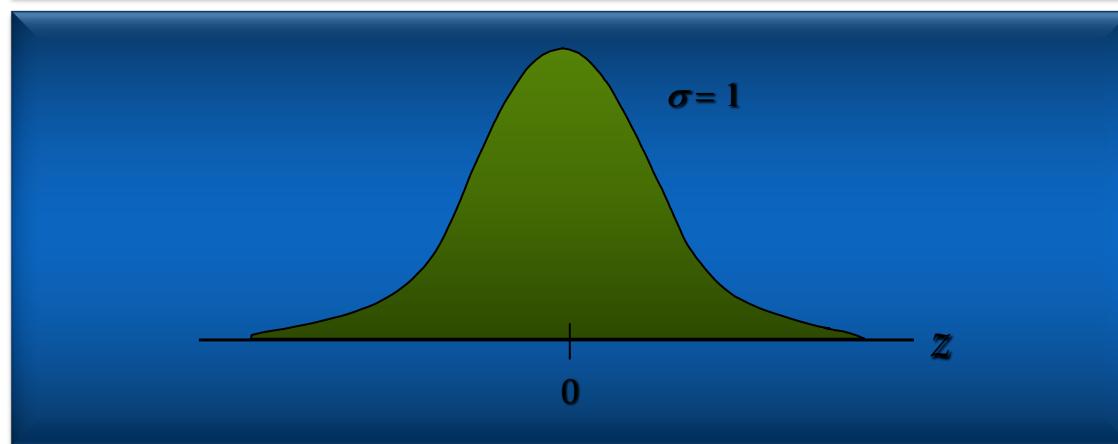
$$\mu = 0, \quad \sigma = 1$$

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

Standard Normal Probability Distribution

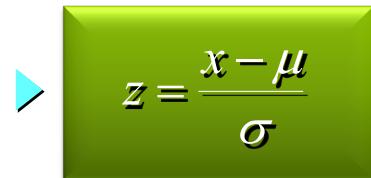
■ Characteristics

- ▶ 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.

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

- 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

$$\begin{aligned} f(x: \mu, \sigma) : P(x_1 < x < x_2) &= \frac{1}{\sigma \sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\ &= \frac{1}{\sigma \sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2} dz \\ &= f(z: 0, \sigma) \end{aligned}$$

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?

$$X \sim \text{Poisson } (\lambda=2/\text{month})$$

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

Exactly 6 cases?

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

(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 $\lambda=2$ 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 $\lambda=2$ 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?

$X \sim \text{Poisson } (\lambda=2 \text{ calls/hour})$

$$P(X \geq 1) = 1 - P(X=0)$$

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

$$\therefore P(X \geq 1) = 1 - .05 = 95\% \text{ chance}$$

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

$$E(X) = \lambda t = 2(1.5) = 3$$

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 $P(X=6)$.

$$\begin{aligned} \text{Now } P(X=6) &= \frac{e^{-\lambda} \lambda^6}{6!} \\ &= \frac{e^{-3.4} (3.4)^6}{6!} \quad (\text{mean, } \lambda = 3.4) \\ &= 0.071\,604\,409 = 0.072 \quad (\text{to 3 d.p.}). \end{aligned}$$

Example

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

$$= \frac{3}{2} e^{-0.5}$$
$$\approx 0.9098.$$

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.

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

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

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

or

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

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

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

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

$$\approx 0.0144.$$

Solution

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

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

$$= 0.9098 \quad (\text{from tables in Appendix 3 (p257), to 4 d.p.})$$

or, from the formula,

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

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

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

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

$$\approx 0.9098.$$

$$\begin{aligned}\text{(ii)} \quad P(A > 2) &= 1 - P(A \leq 2) \\ &= 1 - 0.9856 \quad (\text{from tables}) \\ &= 0.0144 \quad (\text{to 4 d. p.})\end{aligned}$$

or

$$\begin{aligned}1 - [P(A = 0) + P(A = 1) + P(A = 2)] \\ &= 1 - \left[e^{-0.5} + e^{-0.5} 0.5 + \frac{e^{-0.5} (0.5)^2}{2!} \right] \\ &= 1 - e^{-0.5} (1 + 0.5 + 0.125) \\ &= 1 - 1.625 e^{-0.5} \\ &\approx 0.0144.\end{aligned}$$

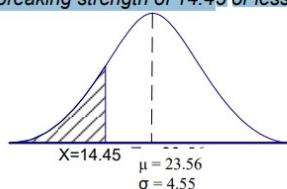
$$\text{(b)} \quad P(0 \text{ in 3 weeks}) = (e^{-0.5})^3 \approx 0.223.$$

Example

Wool fibre breaking strengths are normally distributed with mean $\mu = 23.56$ Newtons and standard deviation, $\sigma = 4.55$.

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

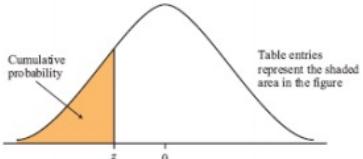
- **Draw a diagram, label and shade area required:**



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

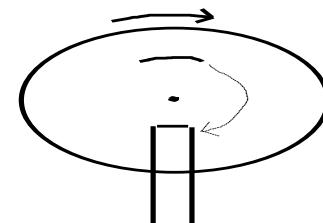


FIRST DIGIT OF z	SECOND DIGIT OF z										FIRST DIGIT OF z
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	z
-3.0	.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010	-3.0
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014	-2.9
-2.8	0.0026	0.0025	0.0024	0.0023	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

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 $= \frac{1}{7200} s = 8.33\text{ms}$.

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, $U(0, 8.33\text{ms})$

Continuous Random Variables

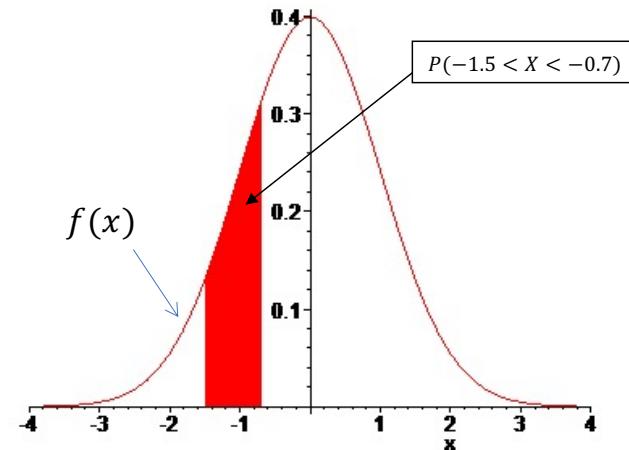
A continuous random variable is a random variable which can take values measured on a continuous scale e.g. weights, strengths, times or lengths.

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

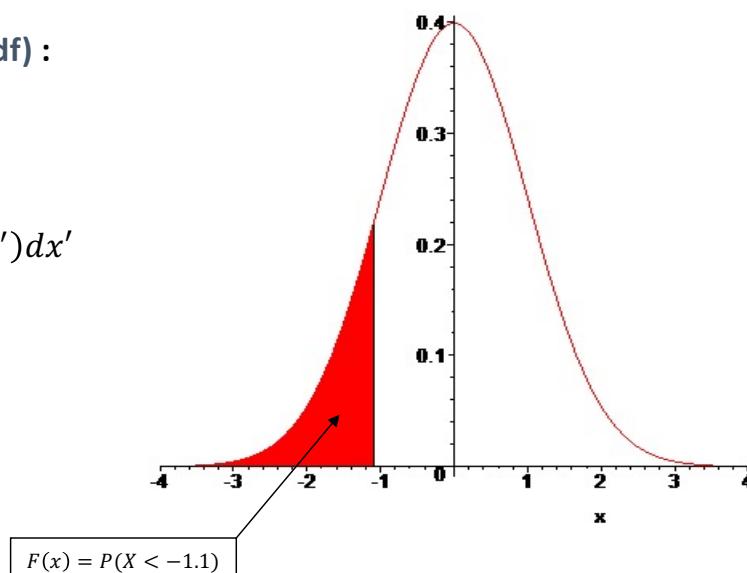
$$\int_{-\infty}^{\infty} f(x)dx = P(-\infty < X < \infty) = 1$$

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 : $\mu = \int_{-\infty}^{\infty} xf(x)dx$

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

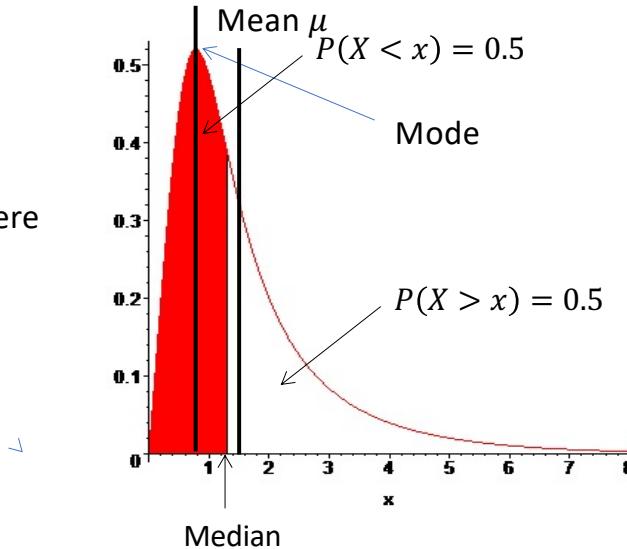
$$= \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

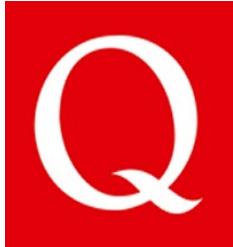
Note: the mean and variance may not be well defined for distributions with broad tails.

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

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





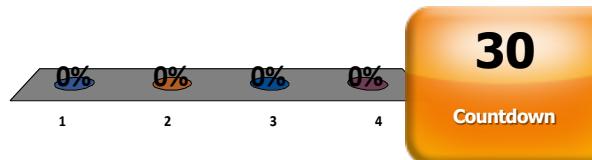
Probability density function

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 $P(X < 20) = 1$.



Question from Derek Bruff 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

Probability density function

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

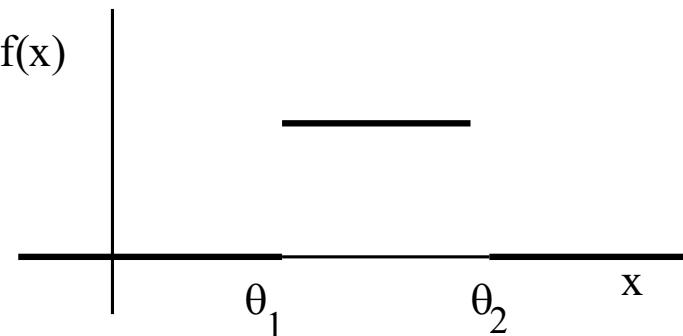
1. No portion of the graph of f can lie below the x -axis.
- **Correct**, $f(x) \geq 0$ for all probabilities to be ≥ 0
2. f is non-zero for x in the range $0 \leq x < 20$
- **Incorrect**, $f(x)$ can be zero
e.g babies must weigh more than an embryo, so at least $f(x < \text{embryo weight}) = 0$
3. The 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

$$f(x) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & \theta_1 \leq x \leq \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

$X \sim U(\theta_1, \theta_2)$, for short.



Occurrence of the Uniform distribution

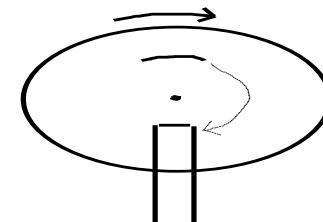
- 1) Waiting times from random arrival time until a regular event (see later)
- 2) Simulation: programming languages often have a standard routine for simulating the $U(0, 1)$ distribution. This can be used to simulate other probability distributions.

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 $= \frac{1}{7200} s = 8.33\text{ms}$.

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, $U(0, 8.33\text{ms})$

Mean and variance: for $U(\theta_1, \theta_2)$

$$\mu = \frac{(\theta_1 + \theta_2)}{2} \quad \sigma^2 = \frac{(\theta_2 - \theta_1)^2}{12}$$

Proof:

Let y be the distance from the mid-point,

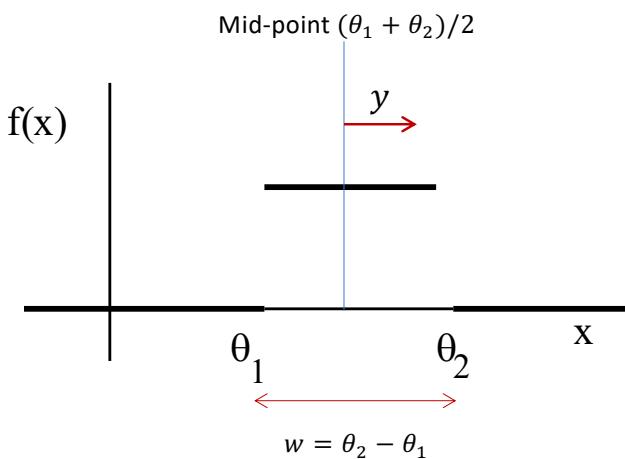
$$y = x - (\theta_2 + \theta_1)/2$$

and the width be

$$w = \theta_2 - \theta_1.$$

Then since $x = \frac{\theta_1 + \theta_2}{2} + y$, and means add

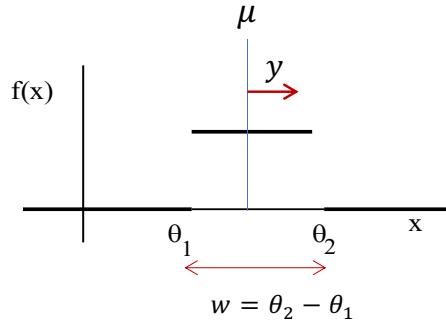
$$\begin{aligned} \mu &= \langle x \rangle = \frac{\theta_2 + \theta_1}{2} + \langle y \rangle \\ &= \frac{\theta_2 + \theta_1}{2} + \int_{-\frac{w}{2}}^{\frac{w}{2}} y f(y) dy = \frac{\theta_2 + \theta_1}{2} + \int_{-\frac{w}{2}}^{\frac{w}{2}} y \frac{1}{w} dy = \frac{\theta_1 + \theta_2}{2} + 0 \end{aligned}$$



Unsurprisingly the mean is the midpoint!

Variance:

$$\begin{aligned}
 \sigma^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \\
 &= \int_{-\frac{w}{2}}^{\frac{w}{2}} y^2 \frac{1}{w} dy \\
 &= \frac{1}{w} \left[\frac{y^3}{3} \right]_{-\frac{w}{2}}^{\frac{w}{2}} \\
 &= \frac{1}{3w} \left(\frac{w^3}{8} + \frac{w^3}{8} \right) = \frac{w^2}{12} \\
 &= \frac{(\theta_2 - \theta_1)^2}{12}
 \end{aligned}$$



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)

$\mu = 4.2 \text{ ms}$ For 2000 reads the mean total time is $\mu_{tot} = 2000 \times 4.2\text{ms} = 8.3\text{s}$.

Note: rms = Root Mean Square = standard deviation

$\sigma = 2.4 \text{ ms}$ So the variance is $\sigma_{tot}^2 = 2000 \times \sigma^2 = 2000 \times 5.8\text{ms}^2 = 0.012\text{s}^2$

$$\Rightarrow \sigma_{tot} = \sqrt{0.012 \text{ s}^2} = 0.11\text{s}$$