# Predictive Systems

## Unit 1: Introduction and Concepts


May 2020

# Who am I?

Mario Alejandro Campos Soberanis (mario.campos@upy.edu.mx)

Master in Computer Science (UADY)
Director of Research and Academic Relations at SoldAI
Experience as Webmaster, Chief technology officer and Research engineer at SoldAI
Research interest: conversational systems, Automatic reasoning and Biologically inspired algorithms

# Course Syllabus

- Unit 1: Basic concepts: 12/06/2020
  - Basic concepts
  - Probability meassure
  - Probability distributions
  - Predictive analysis process

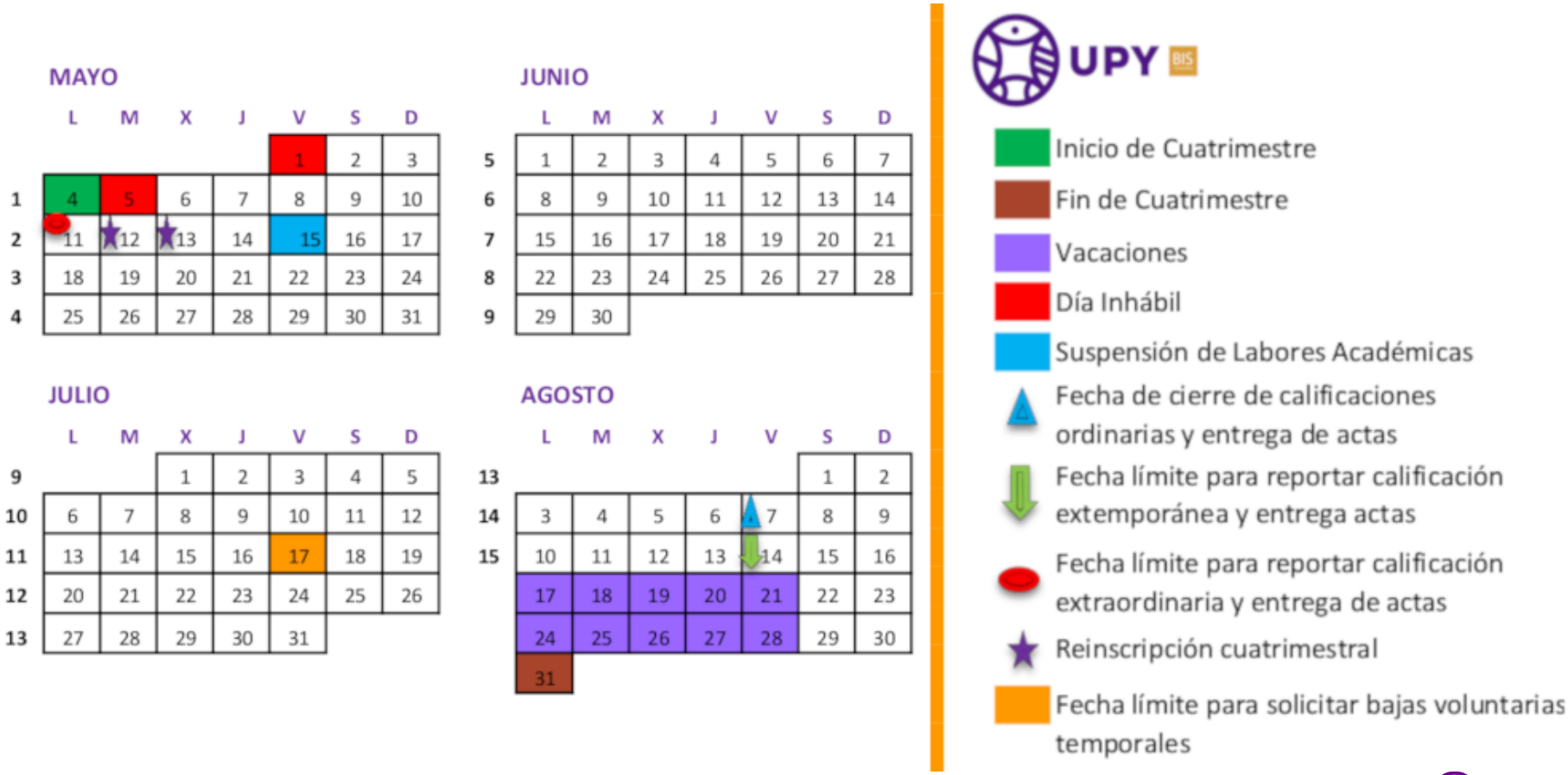
# Course Syllabus

- Unit 2: Probabilistic prediction methods: 16/07/2020
  - Probabilistic formulation
  - Bayes theorem
  - Bayesian nets
  - Bayesian networks message messages
  - Markov models

# Course Syllabus

- Unit 3: Predictive systems applications: 07/08/2020
  - Predictive systems applications models
    - Latent Dirichlet Allocation
    - Gaussian Mixture Model
    - Author-Topic Model
  - Technologies and tools for predictive systems

# Calendar


MAYO
L M X J V S D
1 2 3
1 4 5 6 7 8 9 10
2 11 12 13 14 15 16 17
3 18 19 20 21 22 23 24
4 25 26 27 28 29 30 31
JUNIO
L M X J V S D
5 1 2 3 4 5 6 7
6 8 9 10 11 12 13 14
7 15 16 17 18 19 20 21
8 22 23 24 25 26 27 28
9 29 30
JULIO
L M X J V S D
9 1 2 3 4 5
10 6 7 8 9 10 11 12
11 13 14 15 16 17 18 19
12 20 21 22 23 24 25 26
13 27 28 29 30 31
AGOSTO
L M X J V S D
13 1 2
14 3 4 5 6 7 8 9
15 10 11 12 13 14 15 16
17 18 19 20 21 22 23
24 25 26 27 28 29 30
31
UPY BIS
Inicio de Cuatrimestre
Fin de Cuatrimestre
Vacaciones
Día Inhábil
Suspensión de Labores Académicas
Fecha de cierre de calificaciones ordinarias y entrega de actas
Fecha límite para reportar calificación extemporánea y entrega actas
Fecha límite para reportar calificación extraordinaria y entrega de actas
Reinscripción cuatrimestral
Fecha límite para solicitar bajas voluntarias temporales

# About the course

- Homework
  - Individual
  - Teams (2 persons)

# About the course

- Homework
  - Individual
  - Teams (2 persons)
- Evaluation
  - Participation
  - Assignments (40 %)
  - Exam and projects (60 %)

# About assignments

- Deadline weekly (if an assignment is requested on monday the deadline is next monday before 23:59:59 email/schoology time)

# About assignments

- Deadline weekly (if an assignment is requested on monday the deadline is next monday before 23:59:59 email/schoology time)
- Each day of delay the assignment value reduces 20 % of it's value:
  - 1 day delay: 80 %
  - 2 days delay: 60 %
  - 3 days delay: 40 %
  - 4 days delay: 20 %
  - 5 days delay: Better luck next time!!

# About assignments

- Format
  - Reports/Essays/Presentations: PDF
  - Programming assignments: Jupyter Notebok (.ipynb)
  - Projects: Python code (.py)

# About assignments

- Format
  - Reports/Essays/Presentations: PDF
  - Programming assignments: Jupyter Notebok (.ipynb)
  - Projects: Python code (.py)
- Naming Individual:
  PS_{*homework_no*}_{*last_name*}_{*first_name*}.{*file_extension*}
  Team:
  PS_{*homework_no*}_{*team*}_{*last_names*}.{*file_extension*}
  examples: PS_01_Campos_Mario.pdf,
  PS_03_TeamA_Campos_Soberanis_Perez.pdf

# About assignments

If an algorithm is asked to be implemented:

# About assignments

If an algorithm is asked to be implemented:

Implement it!

# About assignments

If an algorithm is asked to be implemented:

Implement it!

Submit YOUR OWN work

# About assignments

If an algorithm is asked to be implemented:

Implement it!

Submit YOUR OWN work

**As a professional ethic is really IMPORTANT!!**

# Enrole the course

Enrole the schoology course:

7JD6-B7TG-BK9G5

# What is predictive modelling?

Predictive modelling is an art; its a science of unearthing the story impregnated into silos of data.

## Predictive modelling

Predictive modelling is an ensemble of statistical algorithms coded in a statistical tool, which when applied on historical data, outputs a mathematical function. It can in-turn be used to predict outcomes based on some inputs from the future to drive a goal in business context or enable better decision making in general.

# What we need to perform predictive modelling?

# Statistics!!

Statistics are important to understand data.

- Which volume of data we have?
- How is the data distributed?
- Is it centered with little variance or does it varies widely?
- Are two of the variables dependent on or independent of each other?

# Case of study

To explain the concepts associated with probability, let's de fine a case of study:

# Case of study

To explain the concepts associated with probability, let's de fine a case of study:

Let's say an UPY student is going to bet his bus money in a dice game. Before he accept the game he wants to be sure that the dice is fair (all the numbers have the same probability to be rolled). He ask you to help him to find out the dice is fair.

# Case of study

To explain the concepts associated with probability, let's de fine a case of study:

Let's say an UPY student is going to bet his bus money in a dice game. Before he accept the game he wants to be sure that the dice is fair (all the numbers have the same probability to be rolled). He ask you to help him to find out the dice is fair.

What would you do to find if the dice is fair?

# Experiment

## Experiment

An experiment is a process by which an *observation* is made.

## Example

Roll the dice to note which face ends up

# Experiment

The observation or measurement generated by an experiment may or may not produce a numerical value. Here are some examples of experiments:

- Recording a test grade
- Measuring daily rainfall
- Interviewing a householder to obtain his or her opinion on a greenbelt zoning ordinance
- Testing a printed circuit board to determine whether it is a defective product or an acceptable product
- Tossing a coin and observing the face that appears

# Event

## Simple Event

A simple event is the outcome that is observed on a single repetition of the experiment.

## Example

After roll the dice the face with the number six is shown up

# Sample space

## Sample space

The set of possible outcomes of a probabilistic experiment. Is called the sample, event, or possibility space.

## Example

All the posible outcomes to roll a dice:

- $E_1$: Observe the number 1
- $E_2$: Observe the number 2
- $E_3$: Observe the number 3
- $E_4$: Observe the number 4
- $E_5$: Observe the number 5
- $E_6$: Observe the number 6

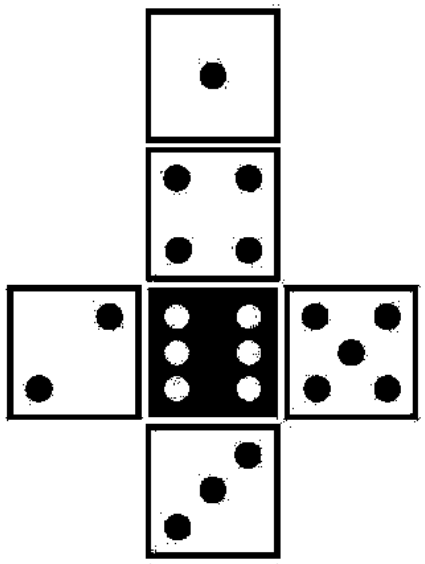
# Sample space

## Discrete sample space

Sample space which contains either a finite or a countable number of distinct sample points.

## Example

$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$



# Event

## Event

An event in a discrete sample space S, is a set of simple points, thus any subset of S.

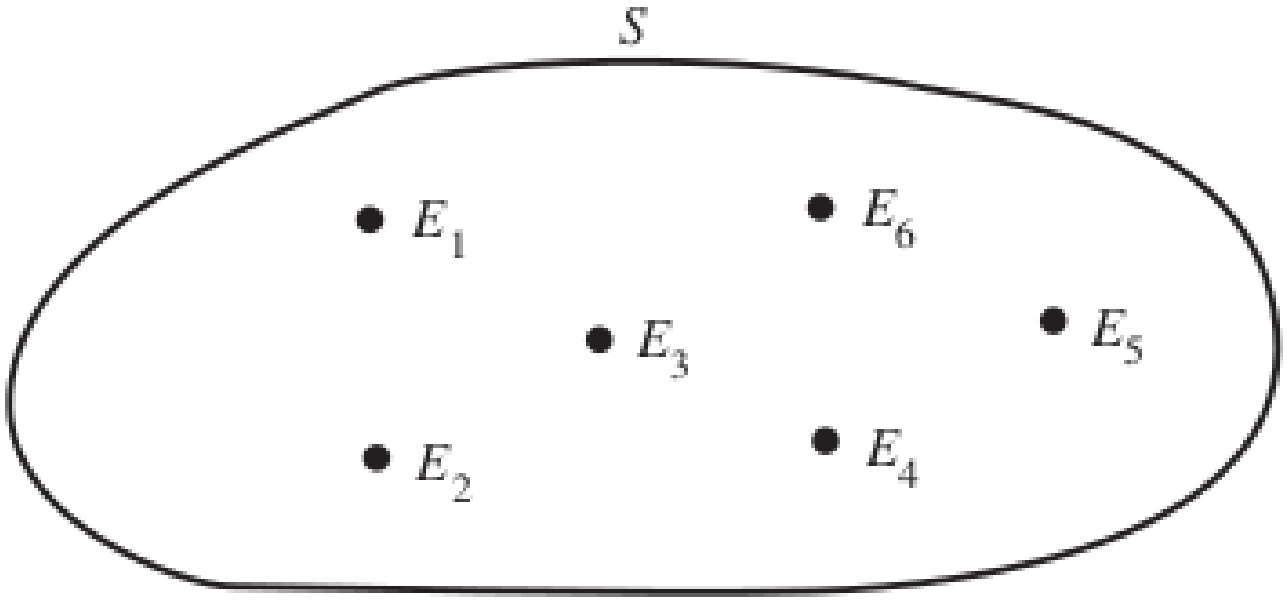
## Example

- $E_1$: Observe the number 1, $E_2$: Observe the number 2
- $E_3$: Observe the number 3, $E_4$: Observe the number 4
- $E_5$: Observe the number 5, $E_6$: Observe the number 6
- $A$ : Observe an odd number
- $B$ : Observe a number minor to 5

$A = \{E_1, E_3, E_5\}$
$B = \{E_1, E_2, E_3, E_4\}$

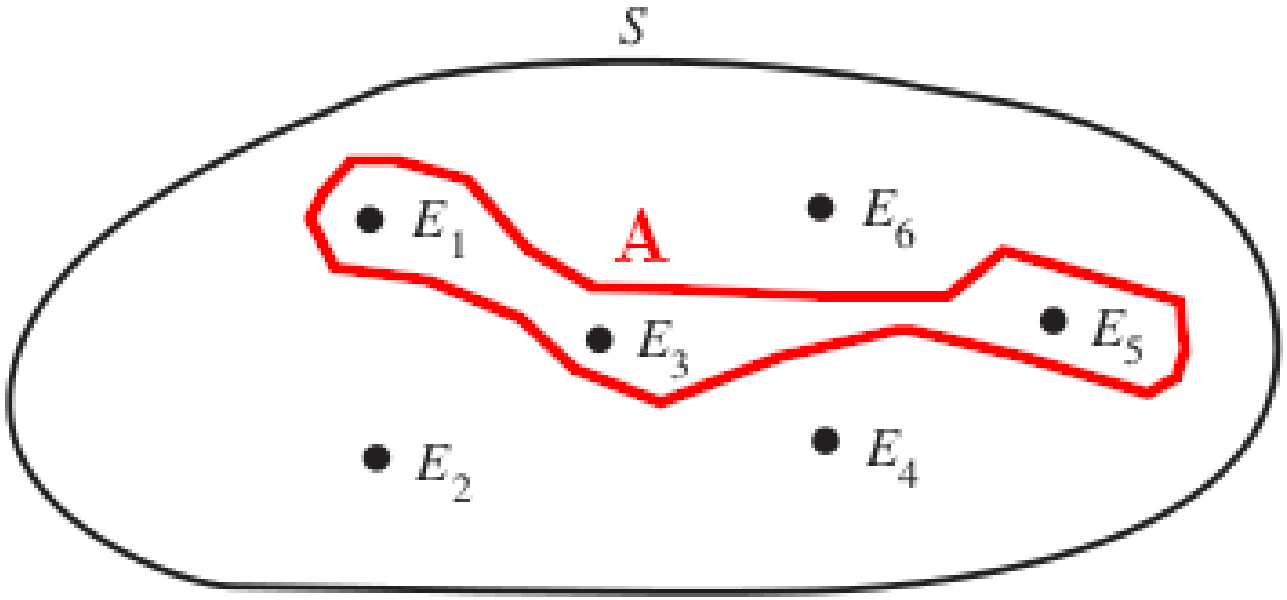
# Event

## Sample space



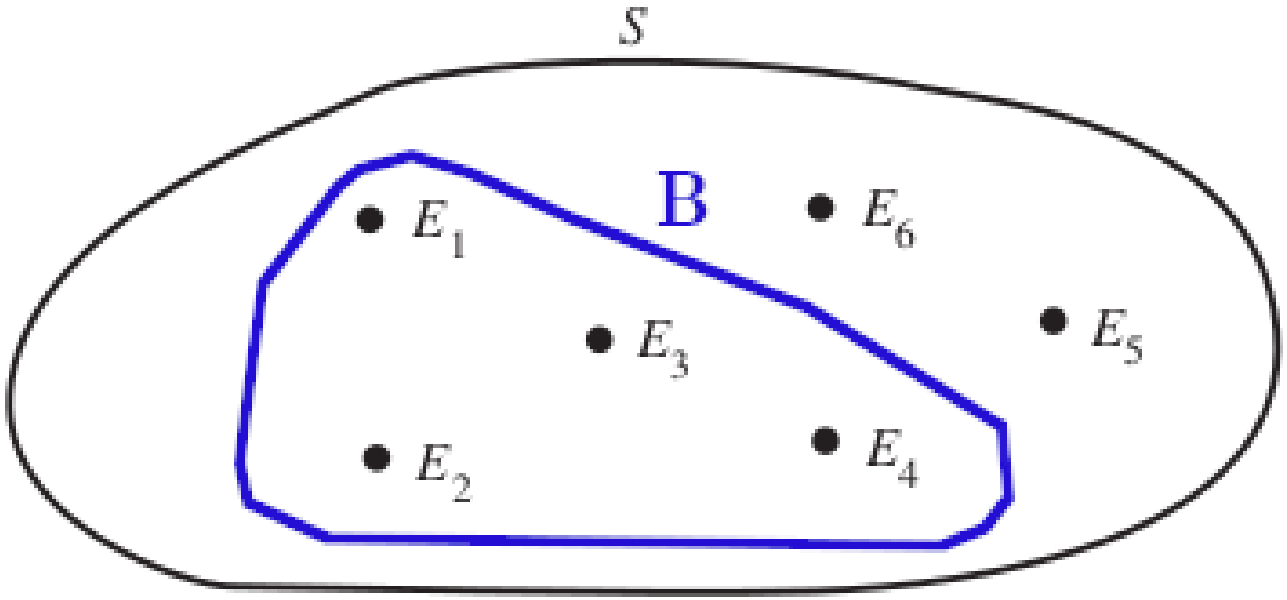
# Event

$$A = \{E_1, E_3, E_5\}$$



# Event

$$B = \{E_1, E_2, E_3, E_4\}$$



# Event

An event is said to ocurred if any of it's sample points is observed. When an event has more than one sample point is said to be *composed* otherwise is said to be *simple*.

## Example

$A = \{E_1, E_3, E_5\}$ is composed
$E_1$ is simple

# Mutually exclusive events

## Mutually exclusive events

Two events are mutually exclusive if, when one event occurs, the others cannot, and vice versa.

## Example

When the number in the dice is 1 it coannot be 2 at the same time.

# Probability

Let $S$ be a sample space associated with an experiment. To each event $A$ in $S$ we assign a number, $P(A)$, called $A$ probability so the following axioms are fullfilled:

## Probability definition

Axiom 1: $P(A) \geq 0$
Axiom 2: $P(S) = 1$
Axiom 3: If $A_1, A_2, A_3, \ldots,$ are a sequence of events mutually exclusive in $S$ ($A_i \cap A_j = \varnothing$ if $i \neq j$), then:

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = \sum_{i=1}^{\infty} P(A_i)$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_i) = \frac{1}{|S|} = \frac{1}{6}$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_i) = \frac{1}{|S|} = \frac{1}{6}$$

$$P(E_i) \geq 0$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_i) = \frac{1}{|S|} = \frac{1}{6}$$

$$P(E_i) \geq 0$$

$$P(S) = 1$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_i) = \frac{1}{|S|} = \frac{1}{6}$$

$$P(E_i) \geq 0$$

$$P(S) = 1$$

$$E_i \cap E_j = \varnothing, \forall i \neq j$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_i) = \frac{1}{|S|} = \frac{1}{6}$$

$$P(E_i) \geq 0$$

$$P(S) = 1$$

$$E_i \cap E_j = \varnothing, \forall i \neq j$$

$$P(E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5 \cup E_6) = \sum_{i=1}^{6} P(A_i) = 1$$

# Probability meassure

$$P(E_i) = \begin{cases} \frac{2}{6} & \text{if } i \text{ is odd} \\ 0 & \text{otherwise.} \end{cases}$$

# Probability meassure

$$P(E_i) = \begin{cases} \frac{2}{6} & \text{if } i \text{ is odd} \\ 0 & \text{otherwise.} \end{cases}$$

Is a valid probability meassure?

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_1) = P(E_3) = P(E_5) = \frac{2}{6}$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_1) = P(E_3) = P(E_5) = \frac{2}{6}$$

$$P(E_2) = P(E_4) = P(E_6) = 0$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_1) = P(E_3) = P(E_5) = \frac{2}{6}$$

$$P(E_2) = P(E_4) = P(E_6) = 0$$

$$P(E_i) \geq 0$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_1) = P(E_3) = P(E_5) = \frac{2}{6}$$

$$P(E_2) = P(E_4) = P(E_6) = 0$$

$$P(E_i) \geq 0$$

$$P(S) = 1$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_1) = P(E_3) = P(E_5) = \frac{2}{6}$$

$$P(E_2) = P(E_4) = P(E_6) = 0$$

$$P(E_i) \geq 0$$

$$P(S) = 1$$

$$P(E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5 \cup E_6) = \sum_{i=1}^{6} P(A_i) = \frac{2}{6} + 0 + \frac{2}{6} + 0 + \frac{2}{6} + 0 = 1$$

# Probability meassure

$$P(E_i) = \begin{cases} \frac{3}{6} & \text{if } i \text{ is odd} \\ -\frac{1}{3} & \text{otherwise.} \end{cases}$$

# Probability meassure

$$P(E_i) = \begin{cases} \frac{3}{6} & \text{if } i \text{ is odd} \\ -\frac{1}{3} & \text{otherwise.} \end{cases}$$

Is a valid probability meassure?

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_1) = P(E_3) = P(E_5) = \frac{3}{6}$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_1) = P(E_3) = P(E_5) = \frac{3}{6}$$

$$P(E_2) = P(E_4) = P(E_6) = -\frac{1}{3}$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_1) = P(E_3) = P(E_5) = \frac{3}{6}$$

$$P(E_2) = P(E_4) = P(E_6) = -\frac{1}{3}$$

$P(E_i) \geq 0$ Here fails!!

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_1) = P(E_3) = P(E_5) = \frac{3}{6}$$

$$P(E_2) = P(E_4) = P(E_6) = -\frac{1}{3}$$

$P(E_i) \geq 0$ Here fails!!

$$P(S) = 1$$

# Probability of a fair dice

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

$$P(E_1) = P(E_3) = P(E_5) = \frac{3}{6}$$

$$P(E_2) = P(E_4) = P(E_6) = -\frac{1}{3}$$

$$P(E_i) \geq 0 \text{ Here fails!!}$$

$$P(S) = 1$$

$$P(E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5 \cup E_6) = \sum_{i=1}^{6} P(A_i) = \frac{3}{6} - \frac{1}{3} + \frac{3}{6} - \frac{1}{3} + \frac{3}{6} - \frac{1}{3} = 1$$

# Event composition method

A summary of the steps used in the event-composition method follows:

- Define the experiment.
- Visualize the nature of the sample points. Identify a few to clarify your thinking.
- Write an equation expressing the event of interest say, A as a composition of two or more events, using unions, intersections, and/or complements. (Notice that this equates point sets.) Make certain that event A and the event implied by the composition represent the same set of sample points.
- Apply the additive and multiplicative laws of probability to the compositions obtained in step 3 to find $P(A)$.

# Calculating the probability of an event

- List all the simple events in the sample space.
- Assign an appropriate probability to each simple event.
- Determine which simple events result in the event of interest.
- Sum the probabilities of the simple events that result in the event of interest.

# Useful counting rules

## *mn* rule

If an experiment is performed in $k$ stages, with $n_1$ ways to accomplish the first stage, $n_2$ ways to accomplish the second stage, ..., and $n_k$ ways to accomplish the kth stage, then the number of ways to accomplish the experiment is

$$\prod_{i=1}^{k} n_i = n_1 n_2 \dots n_k$$

# Useful counting rules

## Example

Suppose that you can order a car in one of three styles and in one of four paint colors. To find out how many options are available, you can think of first picking one of the $m = 3$ styles and then selecting one of the $n = 4$ paint colors. Using the $mn$ Rule, you have $mn = (3)(4) = 12$ possible options.

# Useful counting rules

## Counting rule for permutations

The number of ways we can arrange $n$ distinct objects, taking them $r$ at a time, is:

$$P_n^r = \frac{n!}{(n-r)!}$$

When $r = n$ we have:

$$P_n^r = n!$$

# Useful counting rules

## Example

A piece of equipment is composed of five parts that can be assembled in any order. A test is to be conducted to determine the time necessary for each order of assembly. If each order is to be tested once, how many tests must be conducted?

$$P_5^5 = 5! = 120$$

# Useful counting rules

## Counting rule for combinations

The number of distinct combinations of $n$ distinct objects, that can be formed, taking $r$ at a time, is:

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

The number of combinations is related to the permutations:

$$C_n^r = \frac{P_n^r}{r!}$$

# Useful counting rules

## Example

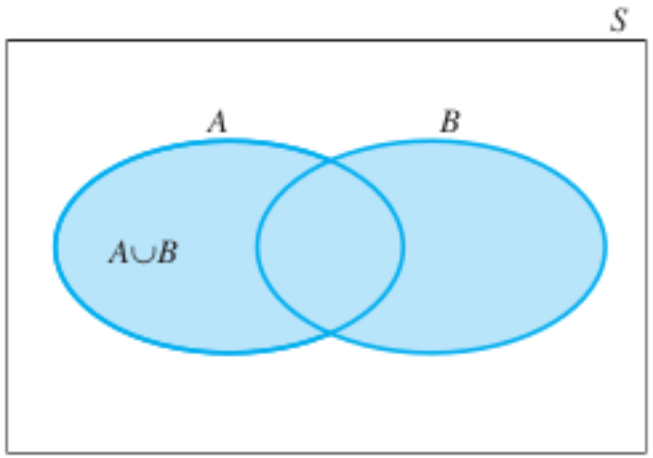
A printed circuit board may be purchased from five suppliers. In how many ways can three suppliers be chosen from the five?

$$C_5^3 = \frac{5!}{3!(5-3)!} = \frac{5!}{3!(2!)} = \frac{120}{12} = 10$$

# Event relations

## Union of events

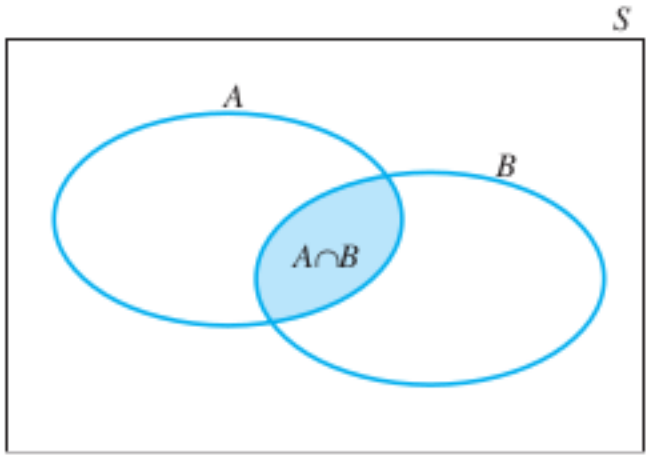
The union of events $A$ and $B$, denoted by $A \cup B$, is the event that either $A$ or $B$ or both occur.



# Event relations

## Intersection of events

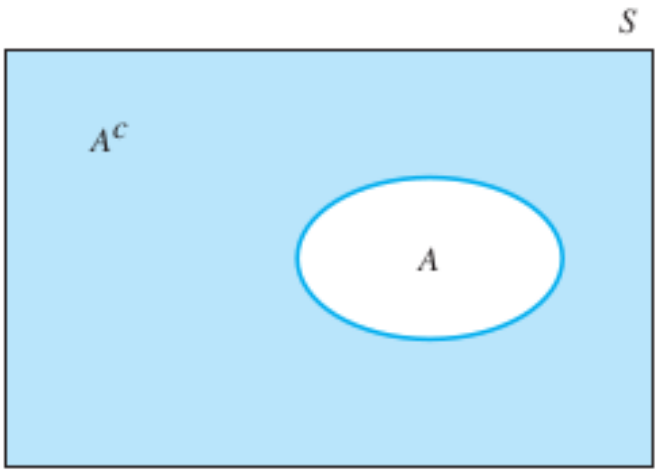
The intersections of events $A$ and $B$, denoted by $A \cap B$, is the event that event both $A$ and $B$ occur.



# Event relations

## Complement of an event

The complement of an event $A$ denoted by $A^c$, is the event that $A$ does not occur.



# Event relations

## Example

Two fair coins are tossed, and the outcome is recorded. These are the events of interest:
$A$: Observe at least one head ($H$)
$B$: Observe at least one tail ($T$)
Define the events $A$, $B$, $A \cup B$, $A \cap B$, and $A^c$ as collections of simple events, and find their probabilities.

# Event relations

$E_1 = HH \quad E_2 = HT \quad E_3 = TH \quad E_4 = TT$

Each event has a probability of $\frac{1}{4}$

$$A = \{E_1, E_2, E_3\}, P(A) = \frac{3}{4}$$

$$B = \{E_2, E_3, E_4\}, P(B) = \frac{3}{4}$$

$$A \cup B = \{E_1, E_2, E_3, E_4\}, P(A \cup B) = \frac{4}{4} = 1$$

$$A \cap B = \{E_2, E_3\}, P(A \cap B) = \frac{2}{4} = \frac{1}{2}$$

$$A^c = \{E_4\}, P(A^c) = 1 - P(A) = \frac{1}{4}$$

# Additive law of probability

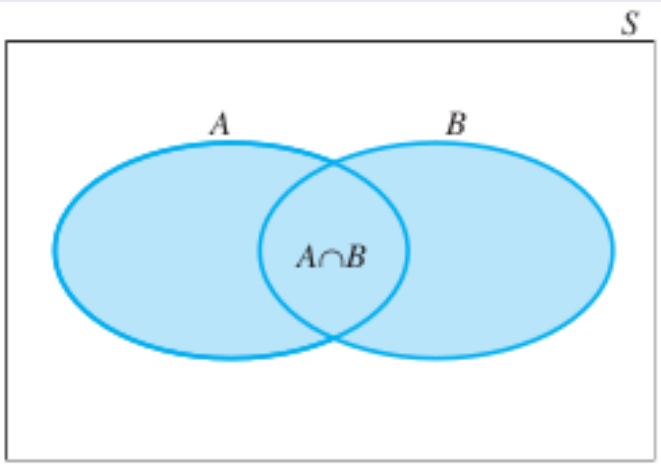
## Additive law of probability

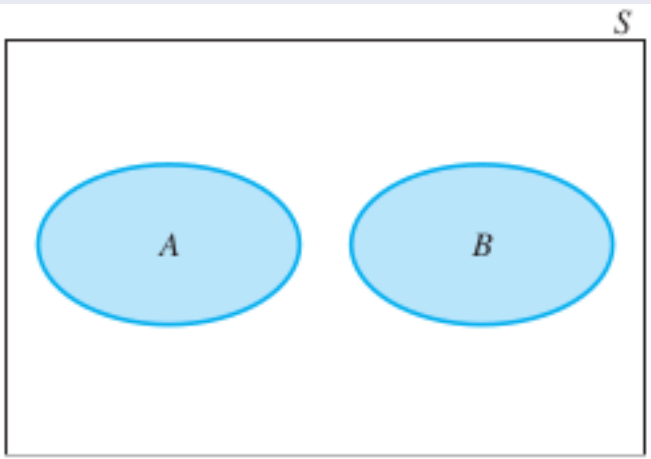
Probability of two events union is given by:

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

if $A$ and $B$ are mutually exclusive:

$$P(A \cup B) = P(A) + P(B)$$



# Conditionally independent events

## Independent events

Two events, $A$ and $B$, are said to be independent if and only if the probability of event $B$ is not influenced or changed by the occurrence of event $A$, or vice versa. If $A$ and $B$ are independents any of the following cases is fullfilled:

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

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

Otherwise the events are dependent.

# Independence of events

## Example

Suppose a researcher notes a person's gender and whether or not the person is colorblind to red and green. Define two events:
$A$ : Person is a male
$B$ : Person is colorblind
Since colorblindness is a male sex-linked characteristic, the probability that a man is colorblind will be greater than the probability that a person chosen from the general population will be colorblind. The probability of event $B$, that a person is colorblind, depends on whether or not event $A$, that the person is a male, has occurred. We say that $A$ and $B$ are dependent events.

# Independence of events

## Example

Consider tossing a single die two times, and define two events:
$A$: Observe a 2 on the first toss
$B$: Observe a 2 on the second toss
If the die is fair, the probability of event $A$ is $P(A) = \frac{1}{6}$. Consider the probability of event $B$. Regardless of whether event $A$ has or has not occurred, the probability of observing a 2 on the second toss is still $\frac{1}{6}$. We could write:
$P(B \text{ given that } A \text{ occurred}) = \frac{1}{6}$
$P(B \text{ given that } A \text{ did not occur}) = \frac{1}{6}$
Since the probability of event $B$ is not changed by the occurrence of event $A$, we say that $A$ and $B$ are independent events.

# Conditional probability

The conditional probability of $A$ given $B$ denoted $P(A|B)$, is the probability that event $A$ has occurred in a trial of a random experiment for which it is known that event $B$ has definitely ocurred. It may be computed as follows:

## Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

# Multiplicative law of probability

## Multiplicative law of probability

Probability of two events intersection is given by:

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

if $A$ and $B$ are independent:

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

# Conditional probability

Suppose that in the general population, there are 51 % men and 49 % women, and that the proportions of colorblind men and women are shown in the probability table below:

| | men | women | total |
|---|---|---|---|
| Colorblind | 0.04 | 0.002 | 0.042 |
| Not colorblind | 0.47 | 0.488 | 0.958 |
| Total | 0.51 | 0.49 | 1.00 |

# Conditional probability

If a person is drawn at random from this population and is found to be a man (event $B$), what is the probability that the man is colorblind (event $A$)?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.04}{0.51} = 0.078$$

What is the probability of being colorblind, given that the person is female? Now we are restricted to only the 49 % of the population that is female, and

$$P(A|B^c) = \frac{P(A \cap B^c)}{P(B^c)} = \frac{0.002}{0.49} = 0.004$$

# Total probability

## Law of total probability

Given a set of events $S_1, S_2, \ldots, S_k$ that are mutually exclusive and exhaustive and an event $A$, the probability of the event $A$ can be expressed as:

$$P(A) = P(S_1)P(A|S_1) + P(S_2)P(A|S_2) + \cdots + P(S_k)P(A|S_k)$$

# Total law of probability

Table below gives the fraction of U.S. adults 20 years of age and older who own five or more pairs of wearable sneakers, along with the fraction of the U.S. adult population 20 years or older in each of five age groups:

| | $20-34$ | $35-64$ | 65 and more |
|---|---|---|---|
| 5 of more pairs | 0.46 | 0.31 | 0.14 |
| Adults 20 and older | 0.29 | 0.54 | 0.17 |

Use the Law of Total Probability to determine the unconditional probability of an adult 20 years and older owning five or more pairs of wearable sneakers.

# Total law of probability

Let $A$ be the event that a person chosen at random from the U.S. adult population 20 years of age and older owns five or more pairs of wearable sneakers. Let $G_1, G_2, G_3$ represent the event that the person selected belongs to each of the three age groups, respectively. Since the groups are exhaustive, you can write the event A as:

$$P(A) = (0.29)(0.46) + (0.54)(0.31) + (0.17)(0.14) = 0.3246$$

# Bayes rule

## Bayes rule

Let $S_1, S_2, \ldots, S_k$ represent $k$ mutually exclusive and exhaustive subpopulations with prior probabilities $P(S_1), P(S_2), \ldots, P(S_k)$. If an event $A$ occurs, the posterior probability of $S_i$ given $A$ is the conditional probability:

$$P(S_i|A) = \frac{P(S_i)P(A|S_i)}{\sum_{j=1}^{k} P(S_j)P(A|S_j)}$$

for $i = 1, 2, \ldots, k$

# Conditional probability

## Example

- Find the probability that the number rolled is a five, given that it is odd.
- Find the probability that the number rolled is odd, given that it is a five.

# Random variable

## Random variable

A random variable is a real values function that maps events defined on a sample space into a set of values. Several different random variables may be defined in relation to a given experiment.

## Random variable

A variable $x$ is a random variable if the value that it assumes, corresponding to the outcome of an experiment, is a chance or random event.

## Example

Define an experiment as tossing a dice 6 consecutive times. Let $Y$ equal to the number of 4's obtained.

# Questions

What is the cardinality of the sample space?

# Questions

What is the cardinality of the sample space?

If the dice is fair what is the porbability of each event?

# Questions

What is the cardinality of the sample space?

If the dice is fair what is the porbability of each event?

Which sequence is more probably to happen:
$S_1 = \{1, 1, 1, 1, 1, 1\}$ or $S_2 = \{1, 1, 1, 2, 2, 2\}$

# Questions

What is the cardinality of the sample space?
$6^6 = 46656$

# Questions

What is the cardinality of the sample space?
$6^6 = 46656$

If the dice is fair what is the porbability of each event?

$$(\frac{1}{6})^6 = \frac{1}{6^6} = \frac{1}{46656} \simeq 0.0000214$$

# Questions

What is the cardinality of the sample space?
$6^6 = 46656$

If the dice is fair what is the porbability of each event?

$$(\frac{1}{6})^6 = \frac{1}{6^6} = \frac{1}{46656} \simeq 0.0000214$$

Which sequence is more probably to happen:
$S_1 = \{1, 1, 1, 1, 1, 1\}$ or $S_2 = \{1, 1, 1, 2, 2, 2\}$
They have the same probability

# Random variable

## Discrete random variables

A random variable $Y$ is said to be discrete if it can assume only a finite or countably infinite number of distinct values.

## Example

Define an experiment as tossing a dice 6 consecutive times. Let $Y$ equal to the number of 4's obtained.

# Random, Variable

## Probability for a discrete random variable

The probability that $Y$ takes on the value $y$, $P(Y = y)$, is the sum of the probabilities of all sample points in $S$ that are assigned the value $y$. We will sometimes denote $P(Y = y)$ by $p(y)$

## Probability distribution

The probability distribution for a discrete variable $Y$ can be represented by a formula, a table or a graph that provides $p(y) = P(Y = y)$ for all $y$. A probability distribution must fulfill:

$$0 \geq p(x) \leq 1$$

$$\sum p(x) = 1$$

# Probability distribution

Find the probability distribution for the experiment defined as:

**Example**

Tossing a dice 4 consecutive times. Let $Y$ equal to the number of 4's obtained.

# Probability distribution

Total cases = ?

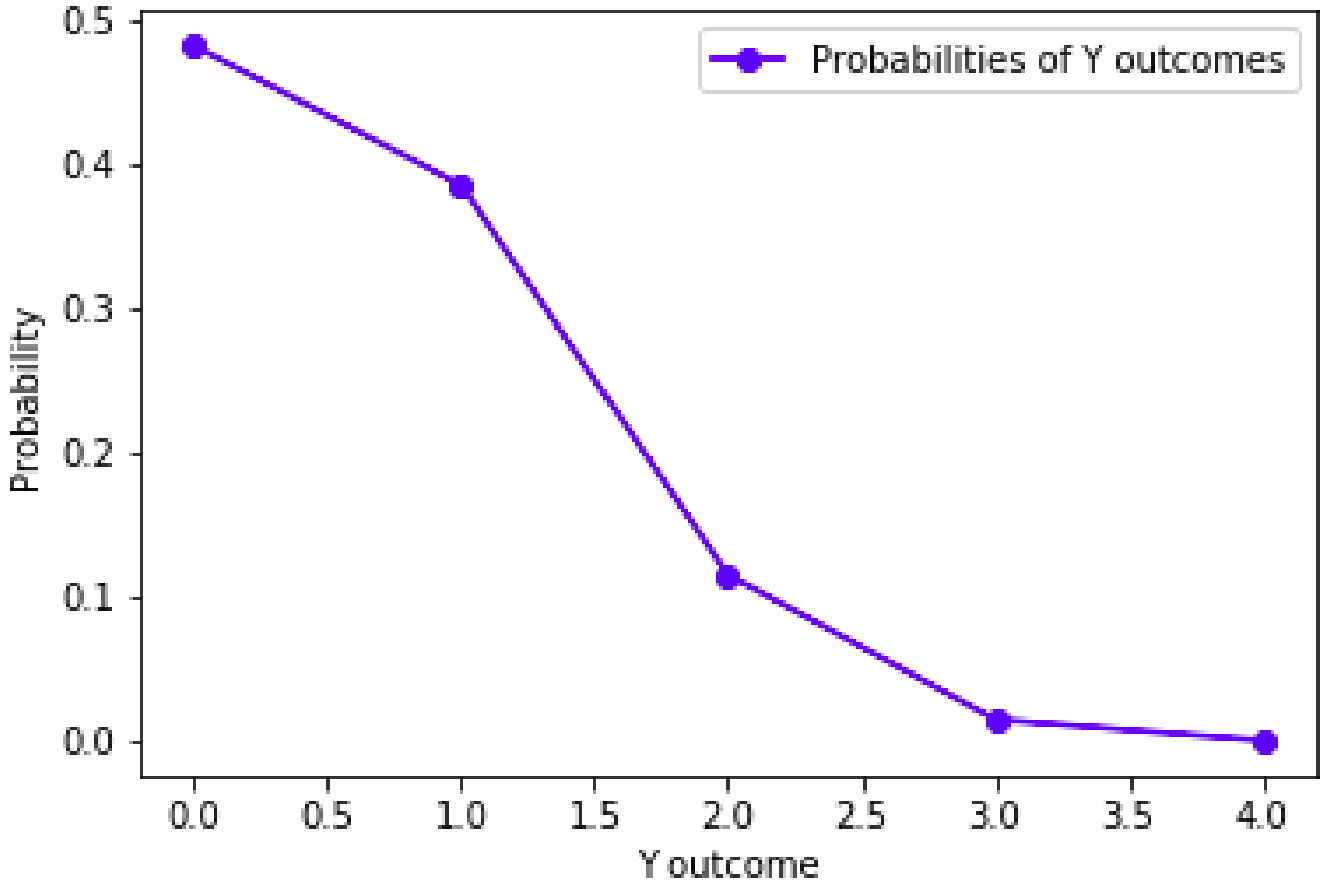
| Number of 4's | Number of cases | Probability |
|---|---|---|
| 0 | ? | ? |
| 1 | ? | ? |
| 2 | ? | ? |
| 3 | ? | ? |
| 4 | ? | ? |

# Probability distribution

Total cases $= 6^4 = 1296$

| Number of 4's | Number of cases | Probability |
|---|---|---|
| 0 | $5^4 = 625$ | 0.482 |
| 1 | $4(5^3) = 500$ | 0.385 |
| 2 | $6(5^2) = 150$ | 0.115 |
| 3 | $4(5^1) = 20$ | 0.015 |
| 4 | 1 | 0.0007 |

# Probability distribution



# Probability distribution

## Example

A supervisor in a manufacturing plant has three men and three women working for him. He wants to choose two workers for a special job. Not wishing to show any biases in his selection, he decides to select the two workers at random. Let $Y$ denote the number of women in his selection. Find the probability distribution for Y.

# Probability distribution

Ways to choose two employees out of 6:

$$C_n^k = \binom{6}{2} = \frac{6!}{2!(6-2)!} = \frac{30}{2} = 15$$

# Probability distribution

Ways to choose two employees out of 6:

$$C_n^k = \binom{6}{2} = \frac{6!}{2!(6-2)!} = \frac{30}{2} = 15$$

Probability of choose 0 women:

$$p(0) = P(Y = 0) = \frac{\binom{3}{0}\binom{3}{2}}{15} = \frac{(1)(3)}{15} = \frac{1}{5} = 0.20$$

# Probability distribution

Ways to choose two employees out of 6:

$$C_n^k = \binom{6}{2} = \frac{6!}{2!(6-2)!} = \frac{30}{2} = 15$$

Probability of choose 0 women:

$$p(0) = P(Y = 0) = \frac{\binom{3}{0}\binom{3}{2}}{15} = \frac{(1)(3)}{15} = \frac{1}{5} = 0.20$$

Probability of choose 1 woman:

$$p(1) = P(Y = 1) = \frac{\binom{3}{1}\binom{3}{1}}{15} = \frac{(3)(3)}{15} = \frac{9}{15} = \frac{3}{5} = 0.60$$

# Probability distribution

Ways to choose two employees out of 6:

$$C_n^k = \binom{6}{2} = \frac{6!}{2!(6-2)!} = \frac{30}{2} = 15$$

Probability of choose 0 women:

$$p(0) = P(Y = 0) = \frac{\binom{3}{0}\binom{3}{2}}{15} = \frac{(1)(3)}{15} = \frac{1}{5} = 0.20$$

Probability of choose 1 woman:

$$p(1) = P(Y = 1) = \frac{\binom{3}{1}\binom{3}{1}}{15} = \frac{(3)(3)}{15} = \frac{9}{15} = \frac{3}{5} = 0.60$$

Probability of choose 2 women:

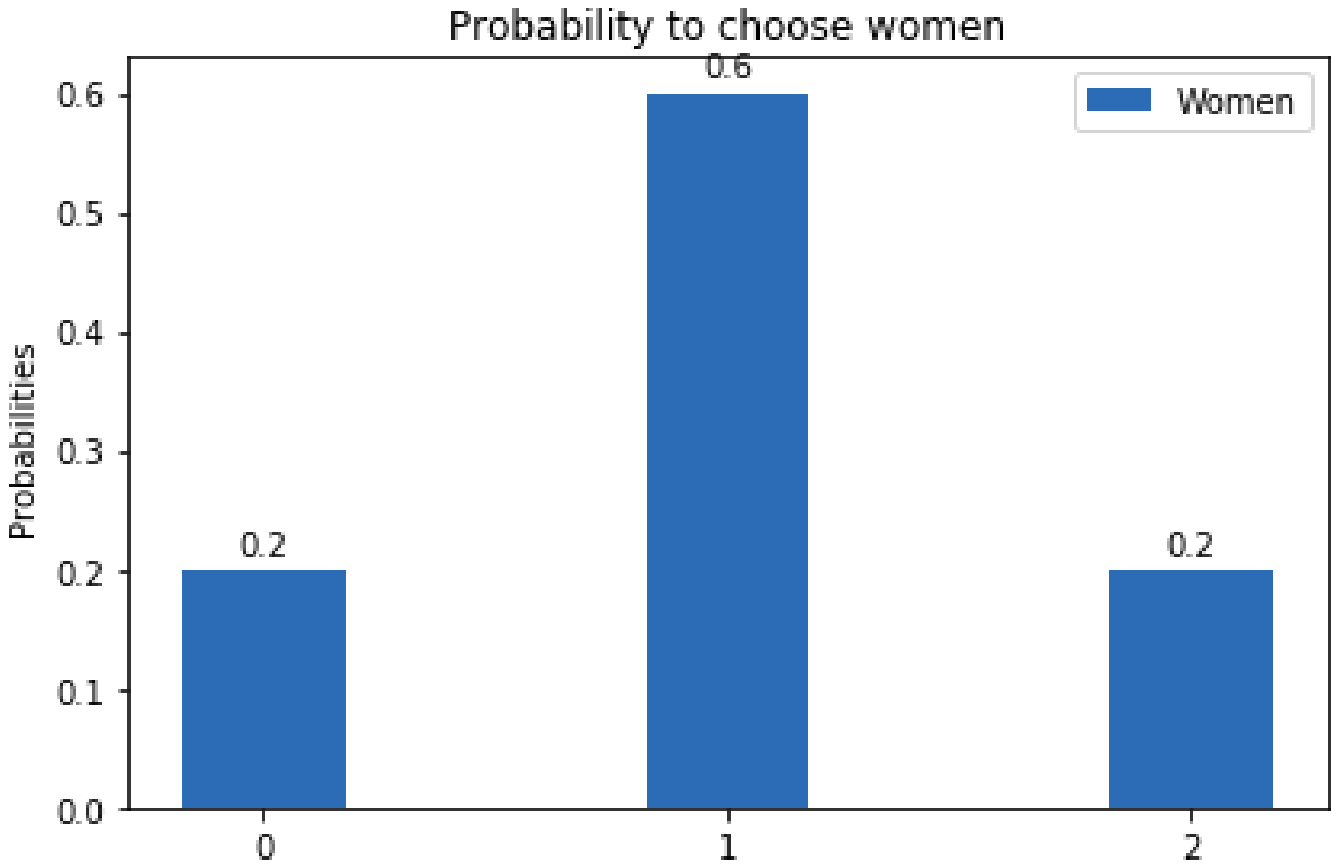
$$p(2) = P(Y = 0) = \frac{\binom{3}{2}\binom{3}{0}}{15} = \frac{(3)(1)}{15} = \frac{1}{5} = 0.20$$

# Probability distribution

## Probablity distribution as a table

| $y$ | $p(y)$ |
|---|---|
| 0 | 0.20 |
| 1 | 0.60 |
| 2 | 0.20 |

# Probability distribution



# Discrete random variable

## Expected value

Let $Y$ be a discrete random variable with the probability function $p(y)$. Then the expected value of $Y$, $E(Y)$, is defined to be:

$$E(Y) = \sum_{y} yp(y)$$

If $p(y)$ is an accurate characterization of the population frequency distribution, then $E(Y) = \mu$, the population mean.

# Discrete random variable

## Expected value of a function of a random variable

Let $Y$ be a discrete random variable with probability function $p(y)$ and $g(Y)$ be a real-valued function of $Y$. Then the expected value of $g(Y)$ is given by

$$E[g(y)] = \sum_{y} g(y)p(y)$$

# Variance of a random variable

## Variance

If $Y$ is a random variable with mean $E(Y) = \mu$, the variance of a random variable $Y$ is defined to be the expected value of $(Y - \mu)^2$. That is

$$V(Y) = E[(Y - \mu)^2]$$

The standard deviation of Y is the positive square root of $V(Y)$.

# Example of variance

## Probablity distribution as a table

| $y$ | $p(y)$ |
|---|---|
| 0 | $\frac{1}{8}$ |
| 1 | $\frac{1}{4}$ |
| 2 | $\frac{3}{8}$ |
| 3 | $\frac{1}{4}$ |

# Variance example

$$\mu = E(Y) = \sum_{0}^{3} yp(y) = 0(\frac{1}{8}) + 1(\frac{1}{4}) + 2(\frac{3}{8}) + 3(\frac{1}{4}) = 1.75$$

$$\sigma^2 = E[(Y - \mu)^2] = \sum_{0}^{3} (y - \mu)^2 p(y)$$

$$= (0-1.75)^2(\frac{1}{8}) + (1-1.75)^2(\frac{1}{4}) + (2-1.75)^2(\frac{3}{8}) + (3-1.75)^2(\frac{1}{4}) = 0.9375$$

$$\sigma = +\sqrt{\sigma^2} = \sqrt{0.9375} = 0.97$$

# Variance of a discrete variable

## Variance of a discrete variable

Let $Y$ be a discrete random variable with probability function $p(y)$ and mean $E(Y) = \mu$ then

$$V(Y) = \sigma^2 = E[(Y - \mu)^2] = E(Y^2) - \mu^2$$

# Continuous random variables

Imagine we want to meassure the rainfall during a day. Theoretically, with measuring equipment of perfect accuracy, the amount of rainfall could take on any value between 0 and 5 inches.

As a result, each of the uncountably infinite number of points in the interval $(0, 5)$ represents a distinct possible value of the amount of rainfall.

A random variable that can take on any value in an interval is called *continuous*.

# Continuous random variables

## Distribution function

Let $Y$ denote any random variable. The distribution function of $Y$, denoted by $F(y)$, is such that:

$$F(y) = P(Y \leq y) \text{ for } -\infty < y < \infty$$

# Contonuous random variables

## Properties of a distribution function

If $F(y)$ is a distribution function, then:

1. $$F(-\infty) = \lim_{y \to -\infty} F(y) = 0$$

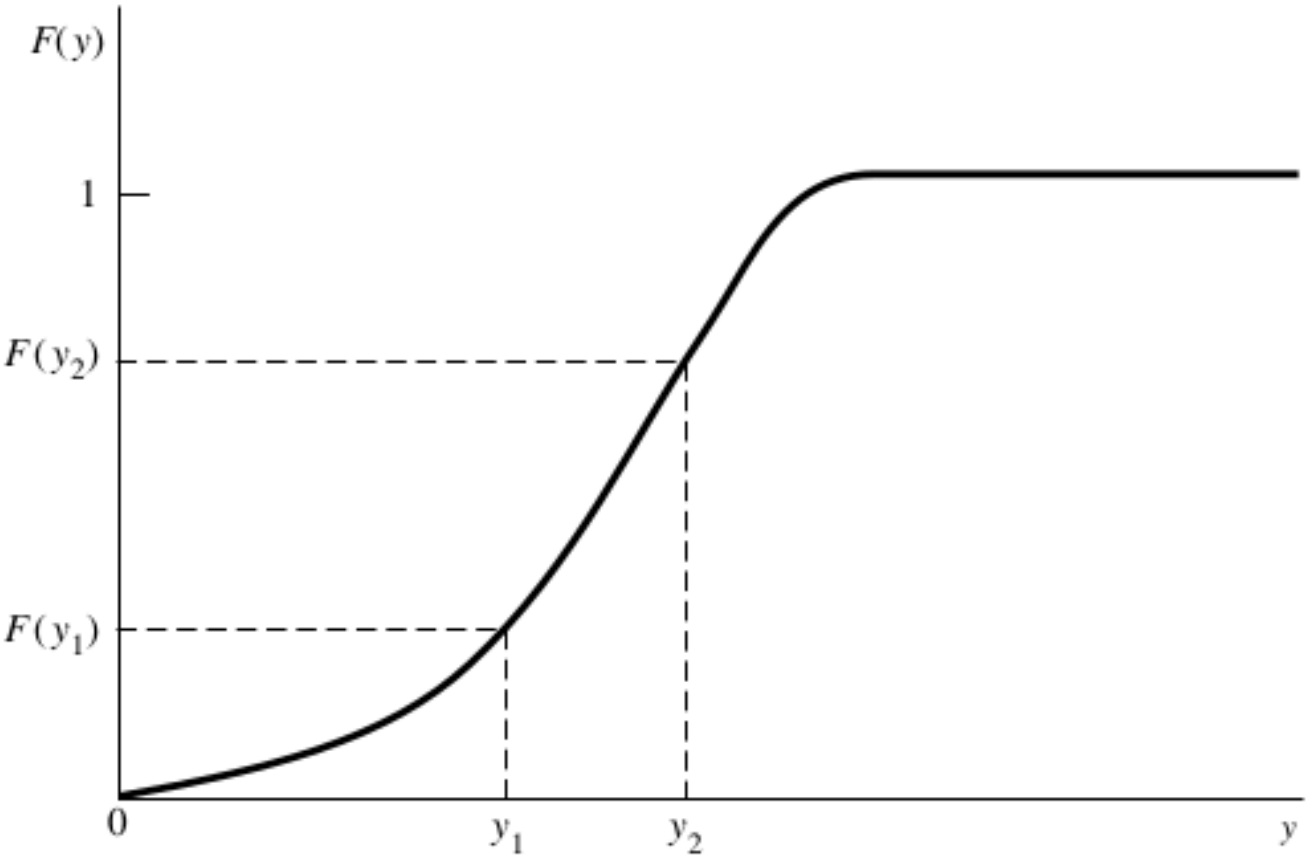
2. $$F(\infty) = \lim_{y \to \infty} F(y) = 1$$

3. $F(y)$ is a nondecreasing function of y, so if $y_1$ and $y_2$ are any values such that $y_1 < y_2$ then:

$$F(y_1) \leq F(y_2)$$

# Continuous random variable

A random variable Y with distribution function $F(y)$ is said to be continuous if $F(y)$ is continuous, for $-\infty < y < \infty$



# Continuous random variable

If $Y$ is a continuous random variable, then for any real number $y$:

$$P(Y = y) = 0$$

# Continuous random variable

If $Y$ is a continuous random variable, then for any real number $y$:

$$P(Y = y) = 0$$

Wait, ... what?

# Continuous random variable

If $Y$ is a continuous random variable, then for any real number $y$:

$$P(Y = y) = 0$$

Wait, ... what?

Why?

# Continuous random variable

If this were not true and $P(Y = y_0) = p_0 > 0$, then $F(y)$ would have a discontinuity (jump) of size $p_0$ at the point $y_0$ , violating the assumption that $Y$ was continuous.

Consider the example of measuring daily rainfall. What is the probability that we will see a daily rainfall measurement of exactly 2.193 inches? It is quite likely that we would never observe that exact value even if we took rainfall measurements for a lifetime, although we might see many days with measurements between 2 and 3 inches.

# Continuous random variable

## Probability density functiion

Let $F(y)$ be the distribution function for a continuous random variable $Y$. Then $f(y)$, given by:

$$f(y) = \frac{dF(y)}{dy} = F'(y)$$

wherever the derivative exists, is called the *probability density function* for the random variable $Y$.

# Continuous random variable

The previous definition makes possible to write the probability distribution in terms of the probability density function:

$$\int_{-\infty}^{y} f(t)d(t)$$

Where $f$ is the probability density function and t is the integration variable.

# Continuous random variable

## Properties of a Density Function

If $f(y)$ is a density function for a continuous random variable, then:

- $$f(y) \geq 0 \text{ for all } y, -\infty < y < \infty$$
- $$\int_{-\infty}^{\infty} f(y)dy = 1$$

# Continuous random variable

The next example gives the distribution function and density function for a continuous random variable.

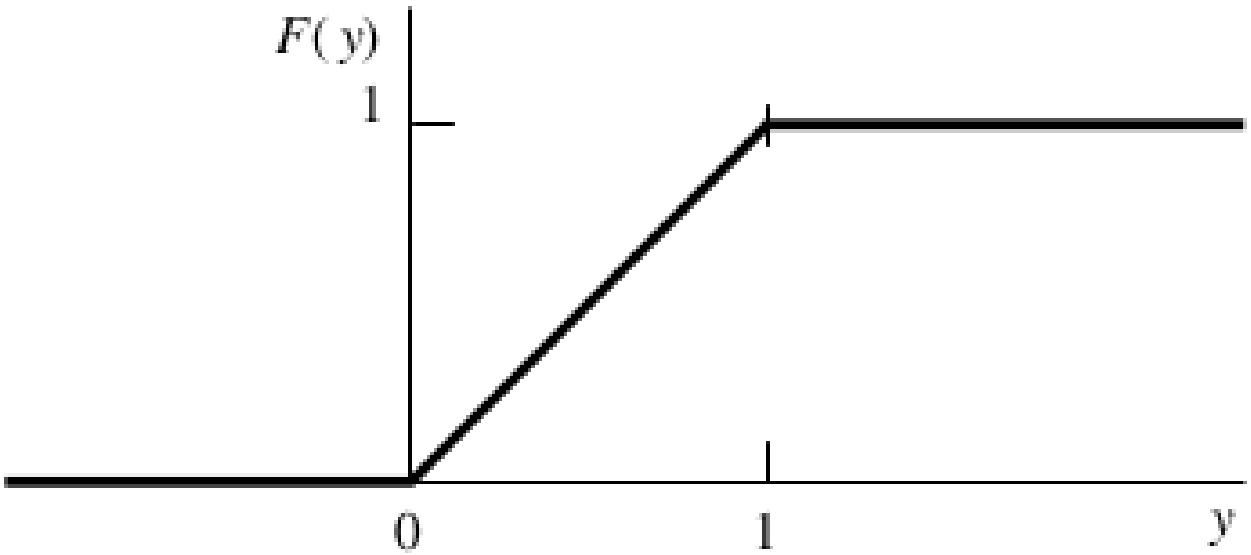
## Example

$$F(y) = \begin{cases} 0 & \text{for } y < 0 \\ y & \text{for } 0 \leq y \leq 1 \\ 1 & \text{for } y > 1 \end{cases}$$

Find the probability density function for $Y$ and graph it.

# Continuous random variable

Graph of the probability distribution:



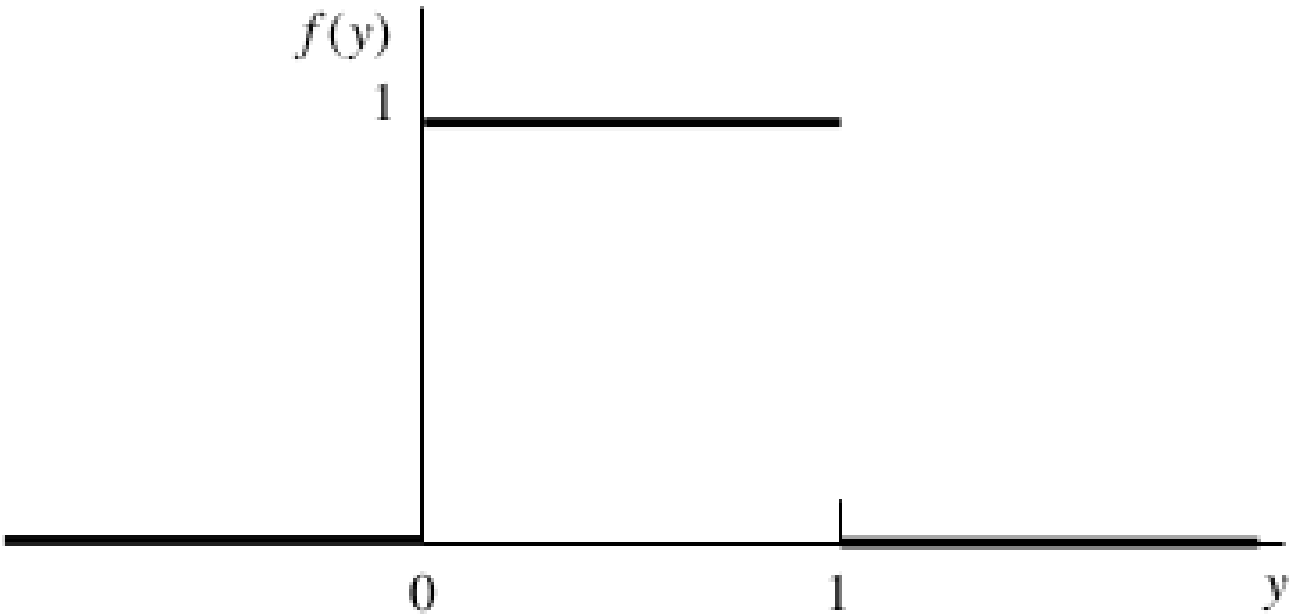
# Continuous random variable

## Example

$$f(y) = \frac{dF(y)}{dy} = \begin{cases} \frac{d(0)}{dy} = 0 & \text{for } y < 0 \\ \frac{dy)}{dy} = 1 & \text{for } 0 \leq y \leq 1 \\ \frac{d(1)}{dy} = 0 & \text{for } y > 1 \end{cases}$$

# Continuous random variable

Graph of the probability density function:



# Continuous random variable

Let $Y$ be a continuous random variable with probability density function given by:

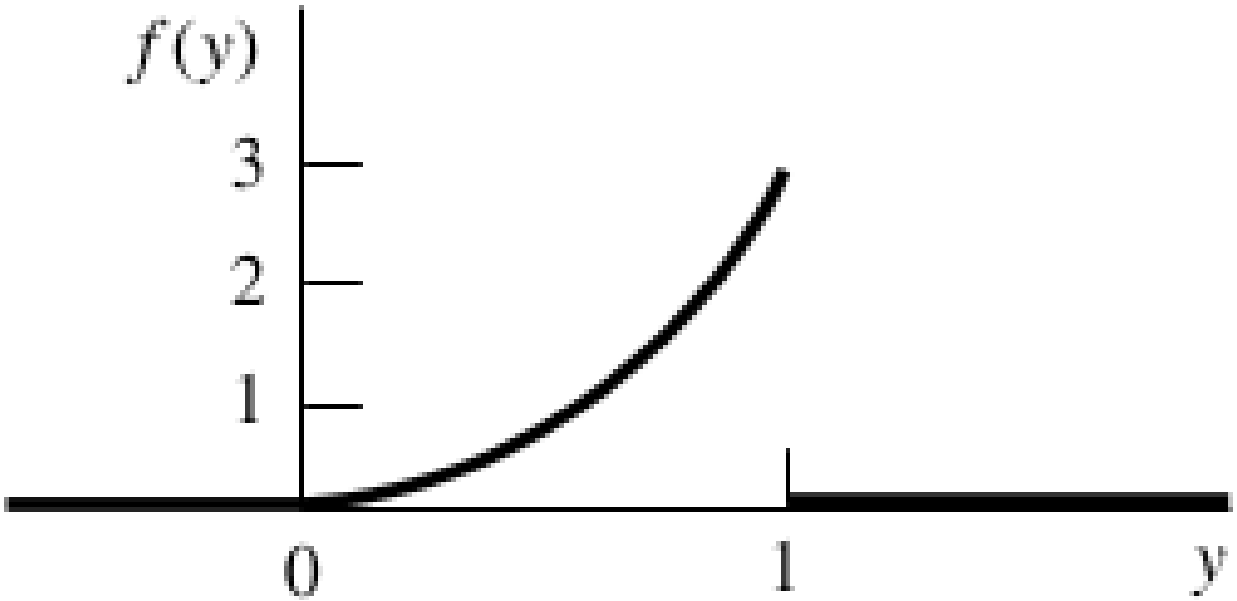
## Example

$$f(y) = \begin{cases} 3y^2 & \text{for } 0 \leq y \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find $F(y)$. Graph both $f(y)$ and $F(y)$

# Continuous random variable

Graph of the probability density function:



# Continuous random variable

Analizing the cases we have when $y < 0$:

$$\int_{-\infty}^{y} 0dt = 0$$

Since we are calculating the area under the curve we need to add the previous area found so when $0 \leq y \leq 1$ we have:

$$\int_{-\infty}^{0} 0dt + \int_{0}^{y} 3t^2 dt = 0 + t^3\Big|_{0}^{y} = y^3$$

And when $y > 1$ we have:

$$\int_{-\infty}^{0} 0dt + \int_{0}^{1} 3t^2 dt + \int_{1}^{y} 0dt = 0 + t^3\Big|_{0}^{1} + 0 = 1$$

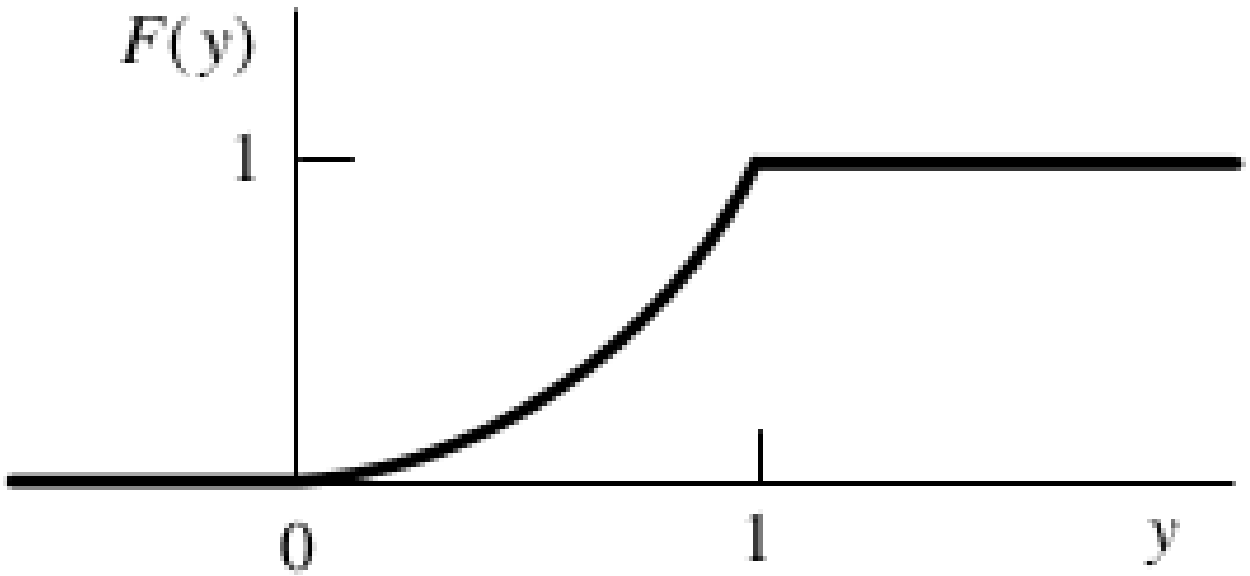
# Continuous random variable

## Example

$$F(y) = \begin{cases} 0 & \text{for } y < 0 \\ y^3 & \text{for } 0 \leq y \leq 1 \\ 1 & \text{for } y > 1 \end{cases}$$

# Continuous random variable

Graph of the probability distribution function:



# Continuous random variable

## Probability of a continuous random variable

If the random variable Y has density function $f(y)$ and $a < b$, then the probability that $Y$ falls in the interval $[a, b]$ is:

$$P(a \leq Y \leq b) = \int_a^b f(y)dy$$

# Continuous random variable

## Expected value

The expected value of a continuous random variable $Y$ is:

$$E(Y) = \int_{-\infty}^{\infty} yf(y)dy$$

provided that the integral exists.

# Continuous random variable

## Expected value

Let $g(Y)$ be a function of $Y$; then the expected value of $g(Y)$ is giben by:

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)dy$$

provided that the integral exists.

# Continuous random variable

Let $c$ be a constant and let $g(Y)$, $g_1(Y)$, $g_2(Y)$, $\ldots$, $g_k(Y)$ be functions of a continuous random variable $Y$. Then the following results hold:

1. $E(c) = c$
2. $E[cg(Y)] = cE[g(Y)]$
3. $E[g_1(Y) + g_2(Y) + \cdots + g_k(Y)] = E[g_1(Y)] + E[g_2(Y)] + \cdots + E[g_k(Y)]$

# Continuous random variable

## Example

The density function for a random vatiable $Y$ is given by $f(y) = \frac{3}{8}y^2$ for $0 \leq y \leq 2$, $f(y) = 0$ elsewhere.
Find $\mu = E(Y)$ and $\sigma^2 = V(Y)$

# Continuous random variable

$$E(Y) = \int_{-\infty}^{\infty} yf(y)dy$$

$$= \int_{0}^{2} y\frac{3}{8}y^2 dy$$

$$\left(\frac{3}{8}\right)\left(\frac{1}{4}\right)y^4\Big|_0^2 = \frac{3}{32}(2^4 - 0^4) = \frac{16(3)}{32} - 0 = \frac{3}{2} = 1.5$$

# Continuous random variable

The variance can be found if we determine the $E(Y^2)$, so:

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 f(y) dy$$

$$= \int_0^2 y^2 \left(\frac{3}{8}\right) y^2 dy$$

$$= \int_0^2 \left(\frac{3}{8}\right) y^4 dy$$

$$\left(\frac{3}{8}\right)\left(\frac{1}{5}\right) y^5 \Big|_0^2 = \frac{3}{40}(2^5 - 0^5) = \frac{32(3)}{40} - 0 = \frac{24}{10} = 2.4$$

Thus, $\sigma^2 = V(Y) = E(Y^2) - [E(Y)]^2 = 2.4 - (1.5)^2 = 0.15$

# Let's code

# References

[1] Kotu V., Deshpande B.: Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner (2009). ISBN: 978-0128016503

[2] Zhang A.: Algorithms, Data Science, Data Mining, Statistics, Big Data, and Predictive Analysis to Improve Business, Work, and Life ISBN: 978-0-596-51649-9

[3] Kumar A. Learning Predictive Analytics With Python Edition (2016). ISBN: 978-1783983261

[4] Siegel E. Predictive Analytics: The Power to Predict Who Whill Click, Buy, Lie, or Die (2013). ISBN: 978-1118416853

[5] Wackerly D., Mendenhall W.,Scheaffer R. Mathematical Statistics with Applications (2008) ISBN: 978-0-495-38508-0

[6] Mendenhall W., Beaver R., Beaver B. Introduction to Probability and Statistics (2009) ISBN: 978-0-495-38953-8