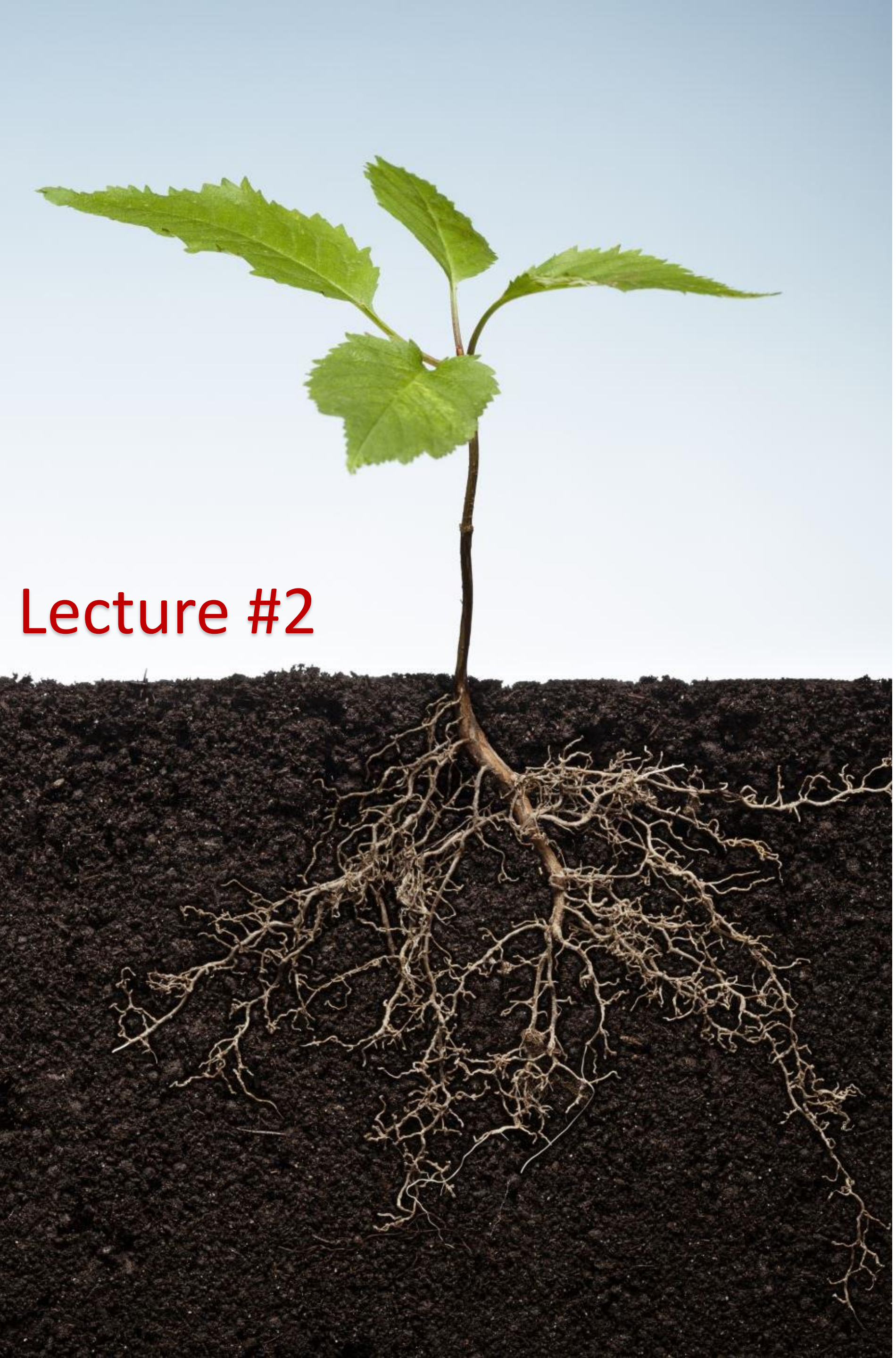


Pattern Recognition
ECSE 4410/6410 CAPA
Fall 2021

Background Material

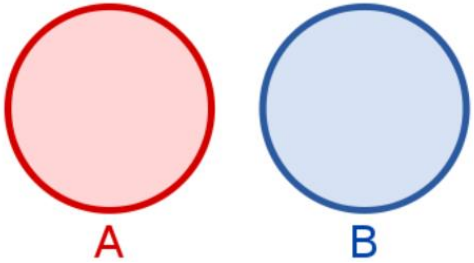
Course Instructor - Thirimachos Bourlai

The Axioms of Probability

*Let **S** be the sample space of a random phenomenon. Suppose that to each event **A** of **S**. a number denoted by **P(A)**, is associated with **A**. If **P** satisfies the following axioms, then it is called a probability and the number **P(A)** is said to be the probability of A.*

- $P(A) \geq 0$ for any event A.
- $P(S) = 1$ where S is the sample space.
- If $\{A_i\}$, i=1,2,..., is a sequence of mutually exclusive events (that is, $A_iA_j=\phi$ for all $i\neq j$), then

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$



Useful Rules & Properties

Monotonicity

$$\text{if} \quad A \subseteq B \quad \text{then} \quad P(A) \leq P(B).$$

The probability of the empty set

$$P(\emptyset) = 0.$$

The numeric bound

It immediately follows from the monotonicity property that

$$0 \leq P(E) \leq 1 \qquad \text{for all } E \in F.$$

Another important property is:

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

This is called the addition law of probability, or the sum rule

Overview

- **Discrete Random Variables**
- Expected Value
- Pairs of Discrete Random Variables
  - Conditional Probability
  - Bayes Rule
- Continuous Random Variables

Discrete Random Variables

- A Random Variable is a **measurement** on an **outcome** of a <u>random experiment</u>

  — denoted by r.v. *x*

- **Discrete** versus **Continuous** random variable:

❖ An r.v. *x* is **discrete** if it may take on only a countable number of distinct values

❖ An r.v. *x* is **continuous** when it takes an infinite number of possible values.

Questions

Which of the following random variables are discrete and which are continuous?

1. X = Number of houses sold by real estate developer per week?
2. X = Number of heads in ten tosses of a coin?
3. X = Weight of a child at birth?
4. X = Time required to run 100 yards?

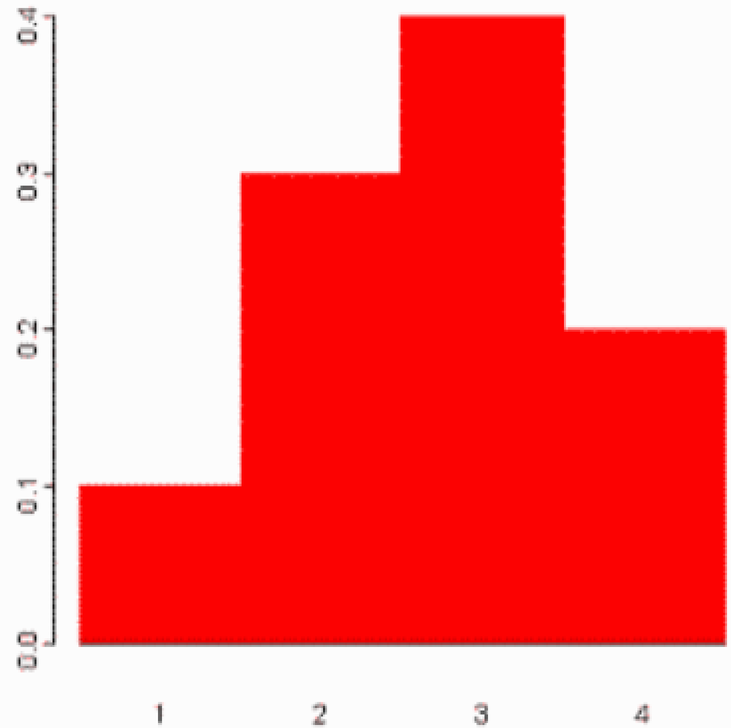
Example

Suppose a variable X can take the values 1, 2, 3, or 4. The probabilities associated with each outcome are described by the following table:

Outcome	1	2	3	4
Probability	0.1	0.3	0.4	0.2

The probability that X is equal to 2 or 3 is the sum of the two probabilities: $P(X = 2 \text{ or } X = 3) = P(X = 2) + P(X = 3) = 0.3 + 0.4 = 0.7$. Similarly, the probability that X is greater than 1 is equal to $1 - P(X = 1) = 1 - 0.1 = 0.9$, by the complement rule.

This distribution may also be described by the ***probability histogram*** shown to the right:



www.stat.yale.edu/Courses/1997-98/101/ranvar.htm

Definition CDF

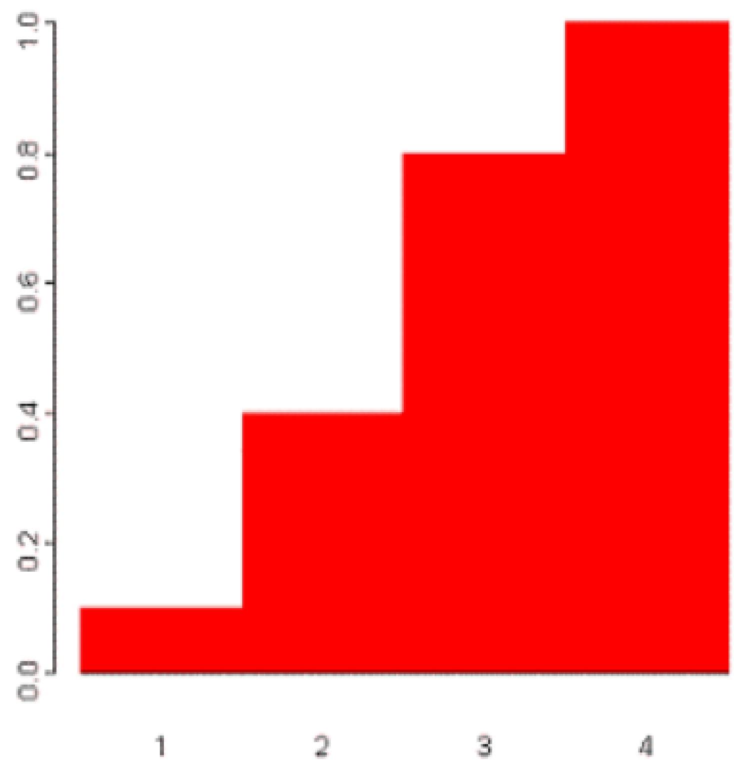
- All random variables (discrete and continuous) have a ***cumulative distribution function***. It is a function **giving the probability** that the random variable *X* is *less than or equal* to *x*, for every value *x*.
- For a **discrete** r. v., the CDF is found by summing up the probabilities

(*Definition taken from Valerie J. Easton and John H. McColl's Statistics Glossary v1.1*)

Example - Calculating the CDF

The cumulative distribution function for the above probability distribution is calculated as follows:
The probability that X is less than or equal to 1 is 0.1,
the probability that X is less than or equal to 2 is
0.1+0.3 = 0.4,
the probability that X is less than or equal to 3 is
0.1+0.3+0.4 = 0.8, and
the probability that X is less than or equal to 4 is
0.1+0.3+0.4+0.2 = 1.

The probability histogram for the cumulative distribution of this random variable is shown to the right:



www.stat.yale.edu/Courses/1997-98/101/ranvar.htm

Probability Distribution Example:
X is the Sum of Two Dice

red	1	2	3	4	5	6

- **This sequence provides an example of a discrete random variable. Suppose that you have a red die which, when thrown, takes the numbers from 1 to 6 with equal probability.**

Probability Distribution Example: X is the Sum of Two Dice

red green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

- **Suppose that you also have a green die that can take the numbers from 1 to 6 with equal probability.**

Probability Distribution Example: X is the Sum of Two Dice

red green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

- **We will define a <u>random variable X</u> as the sum of the numbers when the dice are thrown.**

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6				10		

- **For example, if the red die is 4 and the green one is 6, *X is equal to 10.***

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1						
2						
3						
4						
5		7				
6						

- **Similarly, if the red die is 2 and the green one is 5, *X is equal to 7.***

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

- **The table shows all the possible outcomes.**

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X
2
3
4
5
6
7
8
9
10
11
12

- **If you look at the table, you can see that X can be any of the numbers from 2 to 12.**

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

- **We will now define *f, the frequencies associated with the possible values of X.***

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f
2	
3	
4	
5	4
6	
7	
8	
9	
10	
11	
12	

- **For example, there are four outcomes which make X equal to 5.**

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1

- **Similarly you can work out the frequencies for all the other values of *X*.**

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f	p
2	1	
3	2	
4	3	
5	4	
6	5	
7	6	
8	5	
9	4	
10	3	
11	2	
12	1	

- **Finally we will derive the probability of obtaining each value of *X*.**

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f	p
2	1	
3	2	
4	3	
5	4	
6	5	
7	6	
8	5	
9	4	
10	3	
11	2	
12	1	

- **If there is 1/6 probability of obtaining each number on the red die, and the same on the green die, each outcome in the table will occur with 1/36 probability.**

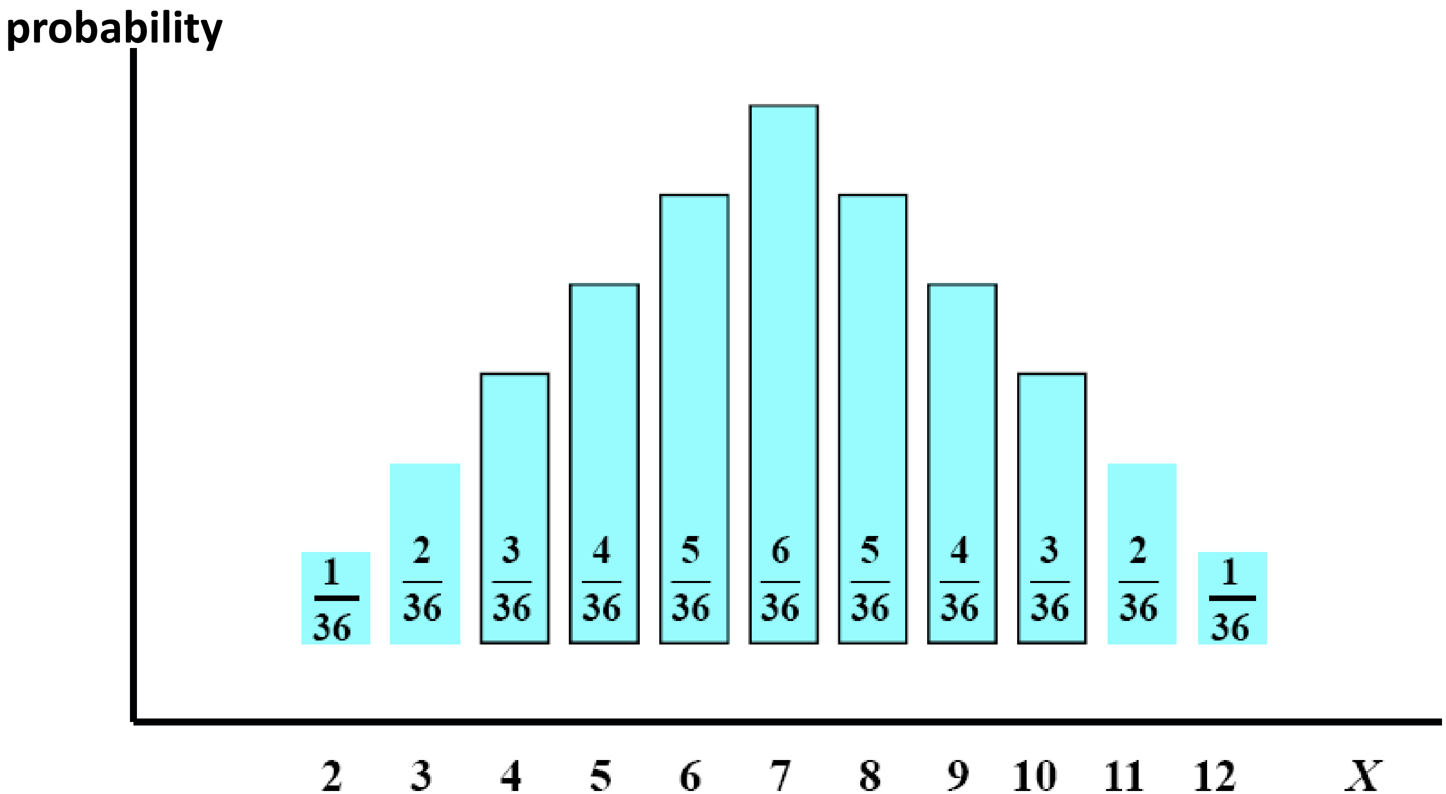
Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f	p
2	1	1/36
3	2	2/36
4	3	3/36
5	4	4/36
6	5	5/36
7	6	6/36
8	5	5/36
9	4	4/36
10	3	3/36
11	2	2/36
12	1	1/36

- **Hence to obtain the probabilities associated with the different values of X, we divide the frequencies by 36.**

Probability Distribution Example: X is the Sum of Two Dice



- **The distribution is shown graphically. In this example it is <u>symmetrical</u>, highest for *X equal to 7 and* declining on either side.**

Overview

- Discrete Random Variables
- **Expected Value**
- Pairs of Discrete Random Variables
  - Conditional Probability
  - Bayes Rule
- Continuous Random Variables

Expected Value

The **expected value** of a **random variable**, also known as its **population mean**,
=> is the *weighted average* of its *possible values*, the ***weights being*** *the probabilities attached to the values*

- Definition of E(X), the expected value of X:

$$E(X) = x_1 p_1 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

Expected Value Example

x_i	p_i	$x_i p_i$
x_1	p_1	$x_1 p_1$
x_2	p_2	$x_2 p_2$
x_3	p_3	$x_3 p_3$
x_4	p_4	$x_4 p_4$
x_5	p_5	$x_5 p_5$
x_6	p_6	$x_6 p_6$
x_7	p_7	$x_7 p_7$
x_8	p_8	$x_8 p_8$
x_9	p_9	$x_9 p_9$
x_{10}	p_{10}	$x_{10} p_{10}$
x_{11}	p_{11}	$x_{11} p_{11}$

$$\Sigma\, x_i p_i = E(X)$$

x_i	p_i	$x_i p_i$
2	1/36	2/36
3	2/36	6/36
4	3/36	12/36
5	4/36	20/36
6	5/36	30/36
7	6/36	42/36
8	5/36	40/36
9	4/36	36/36
10	3/36	30/36
11	2/36	22/36
12	1/36	12/36
		252/36 = 7

Expected Value Properties

- Linear

$$E(X + Y) = E(X) + E(Y)$$
$$E(bX) = bE(X)$$
$$E(b) = b$$

$$\begin{aligned} Y &= b_1 + b_2X \\ E(Y) &= E(b_1 + b_2X) \\ &= E(b_1) + E(b_2X) \\ &= b_1 + b_2 E(X) \end{aligned}$$

- Also denoted by μ

Variance

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \sum (x_i - \mu)^2 P(X = x_i)$$

$$\mathrm{Var}(X) = \sigma^2$$

$$\mathrm{Var}(X) = E[(X-\mu)^2] = E[X^2] - (E[X])^2$$

Overview

- Discrete Random Variables
- Expected Value
- **Pairs of Discrete Random Variables**
  - Conditional Probability
  - Bayes Rule
- Continuous Random Variables

Pairs of Discrete Random Variables

- Let *x* and *y* be two discrete r.v.
- For each possible pair of values, we can define a *joint probability* $p_{ij}=\Pr[x=x_i, y=y_j]$
- We can also define a *joint probability mass function* P(x,y) which offers a complete characterization of the pair of r.v.

$$P_x(x) = \sum_{y \in Y} P(x, y)$$

$$P_y(y) = \sum_{x \in X} P(x, y)$$

Marginal distributions

Note that P_x and P_y are different functions

Statistical Independence

Two random variables *x* and *y* are said to be independent, if and only if

$$P(x,y)=P_x(x)\ P_y(y)$$

that is, when knowing the value of *x* does not give us additional information for the value of *y*.

Or, equivalently

$$E[f(x)g(y)] = E[f(x)]\ E[g(y)]$$

for any functions *f*(*x*) and *g*(*y*).

Conditional Probability

- When two r.v. are not independent, knowing one allows better estimate of the other (e.g. outside temperature, season)

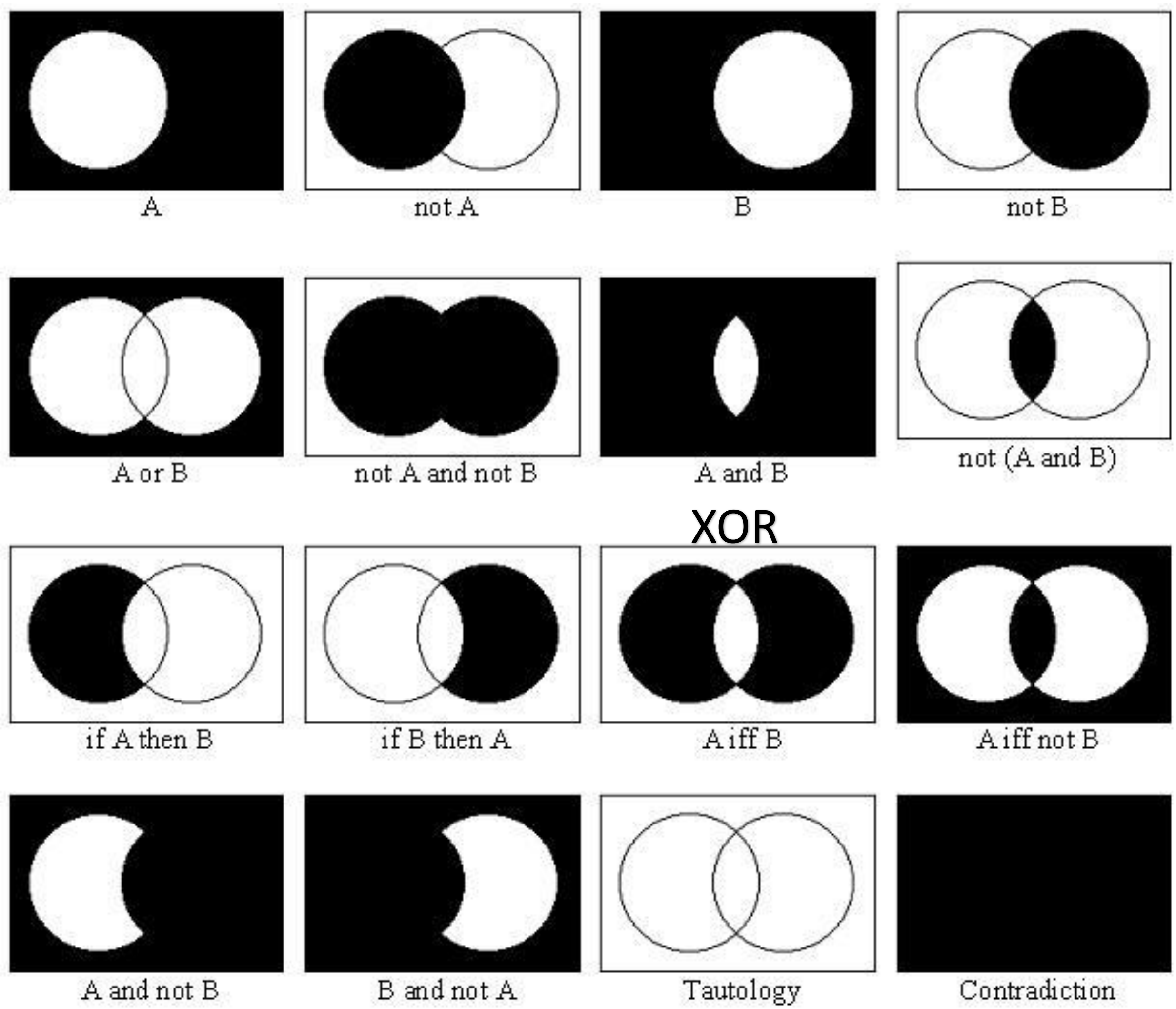
$$\Pr[x = x_i \mid y = y_j] = \frac{\Pr[x = x_i \,,\, y = y_j]}{\Pr[y = y_j]}$$

- If independent P(x|y)=P(x)

Conditional Probability Examples
Solve In-Class Example 1

- A jar contains black and white marbles.
- Two marbles are chosen without replacement.
- The probability of selecting a black marble and then a white marble is 0.34
- The probability of selecting a black marble on the first draw is 0.47
  - What is the probability of selecting a white marble on the second draw, given that the first marble drawn was black?

Reminder - Venn Diagrams



XNOR

www.cat4mba.com

Conditional Probability Examples
Solve In-Class Example 2

Example 2

A machine produces parts that are either good (90%), slightly defective (2%), or obviously defective (8%).

Produced parts get passed through an automatic inspection machine, which is able to detect any part that is obviously defective and discard it.

What is the probability that good quality parts make it through the inspection machine and get shipped?

Conditional Probability Examples
Solve In-Class Example 3

- **Example 3**

Your neighbor has 2 children.

You learn that he has a son, Joe.

What is the probability that Joe's sibling is a brother?

Joe's sibling is equally likely to have been born male or female suggests that the probability the other child is a boy is 1/2. Is this correct?

Multiplication Rule

- **The "Multiplication Rule" (also known as the "Law of Multiplication") states that, assuming P(F) > 0,** $P(E \cap F) = P(F) \cdot P(E|F)$

  > which is (trivially) just a *rewriting of the definition of conditional probability*

- The more general form is equally easy to prove from the definition:

  $P(E1 \cap E2 \cap \cdot \cdot \cdot \cap En) =$

  $P(E1)P(E2|E1)P(E3|E1 \cap E2) \cdot \cdot \cdot$

  $P(En|E1 \cap E2 \cap \cdot \cdot \cdot \cap En-1)$

Conditional Probability Examples
Solve In-Class Example 4

Example 4

- *Suppose that five good fuses and two defective ones have been mixed up*

- *To find the defective fuses, we test them one-by-one, at random and without replacement*

- *What is the probability that we are lucky and find both* of the defective fuses *in the first two tests*?

Conditional Probability Examples
Solve In-Class Example 5

Example 5

- *Six cards are selected at random (without replacement) from a standard deck of 52 cards*

- *What is the probability there will be no pairs? (two cards of the same denomination)*

Law of Total Probability

- The "Law of Total Probability" (also known as the "Method of Conditioning")
- It allows one to **compute** the probability of an event E by **conditioning on cases**, according to a partition of the sample space

For example, one way to partition S is to break into sets F and Fc, for any event F. This gives us the simplest form of the law of total probability:

$$P(E) = P(E \cap F) + P(E \cap F^c) = P(E|F)P(F) + P(E|F^c)P(F^c)$$

Law of Total Probability

More generally for any partition of S into sets

F1, . . . , Fn:

$$P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

Conditional Probability Examples

Solve In-Class Example 6

Example 6

- *Consider the parts problem again, but now assume that a one-year warranty is given for the parts that are shipped to customers. Suppose that a good part fails within the first year with probability 0.01, while a slightly defective part fails within the first year with probability 0.10*

- *What is the probability that a customer receives a part that fails within the first year and therefore is entitled to a warranty replacement?*

Bayes Rule

$$P(x \mid y) = \frac{P(x, y)}{P(y)} = \frac{P(y \mid x) P(x)}{\sum_{x \in X} P(x, y)}$$

$$\text{posterior} = \frac{\text{likelihood} * \text{prior}}{\text{evidence}}$$

- *x* is the unknown cause
- y is the observed evidence
- Denominator often omitted (maximum a posteriori solution)
- Bayes rule shows how probability of *x* changes after we have observed *y*

Bayes Formula 1/3

- Often, **for a given partition** of S into sets F1, . . . , Fn, we want to know the **probability** that **some particular case**, Fj occurs given that some event E occurs.

- We can **compute** this easily using the definition:

$$P(F_j|E) = \frac{P(F_j \cap E)}{P(E)}$$

Bayes Formula 2/3

- Now, using the Multiplication Rule (i.e., the definition of conditional probability), we can rewrite the numerator:

$$P(F_j \cap E) = P(E|F_j)P(F_j)$$

- Using the Law of Total Probability, we can rewrite the denominator: $P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$

Bayes Formula 3/3

- Thus, we get:

$$P(F_j|E) = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

This ***expression is often called*** ***Bayes Formula****, though it can almost always just be derive from scratch with each problem we solve*

Bayes Formula Examples
Solve In-Class Example 7

Example 7

- *Urn 1 contains 5 white balls and 7 black balls.*
- *Urn 2 contains 3 white and 12 black balls.*

A fair coin is flipped; if it is Heads, a ball is drawn from Urn 1, and if it is Tails, a ball is drawn from Urn 2. Suppose that this experiment is done and you learn that a white ball was selected.

- *What is the probability that this ball was in fact taken from Urn 2? (i.e., that the coin flip was Tails)*

Overview

- Discrete Random Variables
- Expected Value
- Pairs of Discrete Random Variables
  - Conditional Probability
  - Bayes Rule
- **Continuous Random Variables**

Continuous Random Variables

- Examples: room temperature, time to run 100m, weight of child at birth…
- Cannot talk about probability of that x has a particular value
- Instead, *probability* that x falls in an interval => ***probability density function***

$$\Pr[x \in (a,b)] = \int_a^b p(x)dx$$

$$p(x) \geq 0 \text{ and } \int_{-\infty}^{\infty} p(x)dx = 1$$

Expected Value

$$E[x] = \mu = \int_{-\infty}^{\infty} x p(x) dx$$

$$E[f(x)] = \int_{-\infty}^{\infty} f(x) p(x) dx$$

$$Var[x] = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx$$

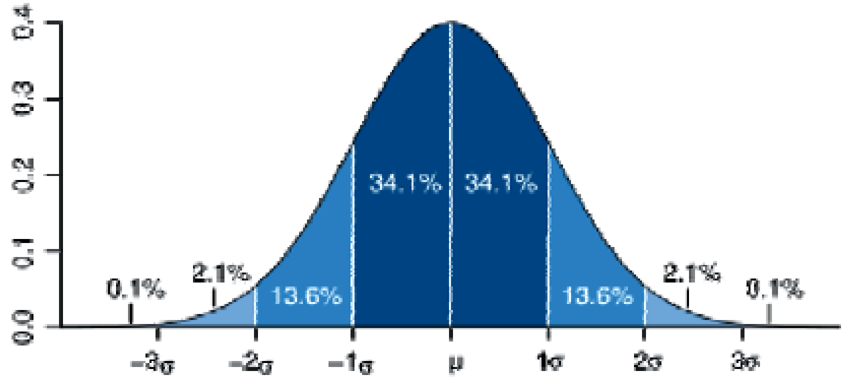
- Bayes rule $$p(x \mid y) = \frac{p(y \mid x) p(x)}{\int_{-\infty}^{\infty} p(y \mid x) p(x) dx}$$

$$\text{posterior} = \frac{\text{likelihood} * \text{prior}}{\text{evidence}}$$

Normal (Gaussian) Distribution

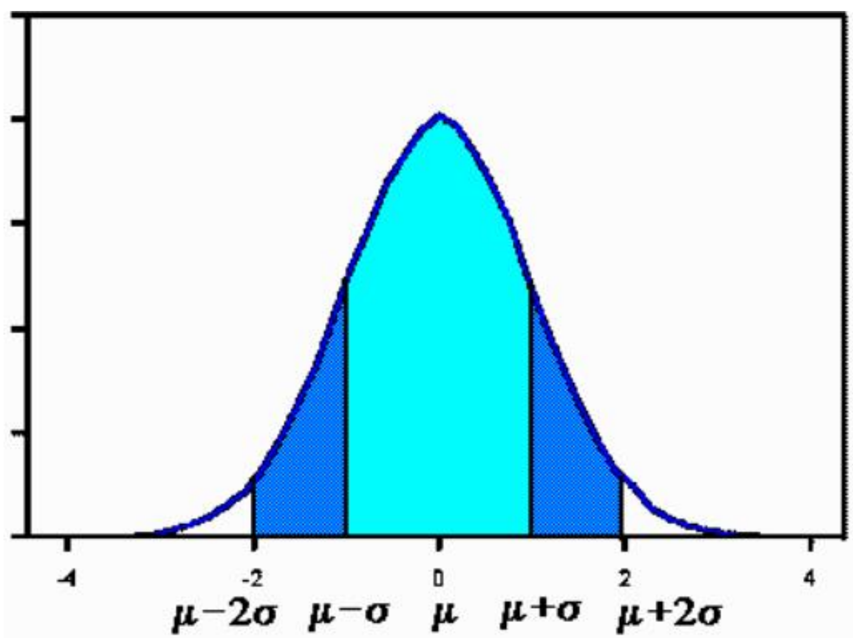
Central Limit Theorem: under various conditions, the *distribution of the sum of d independent random variables* approaches a limiting form *known as* the **normal distribution**

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

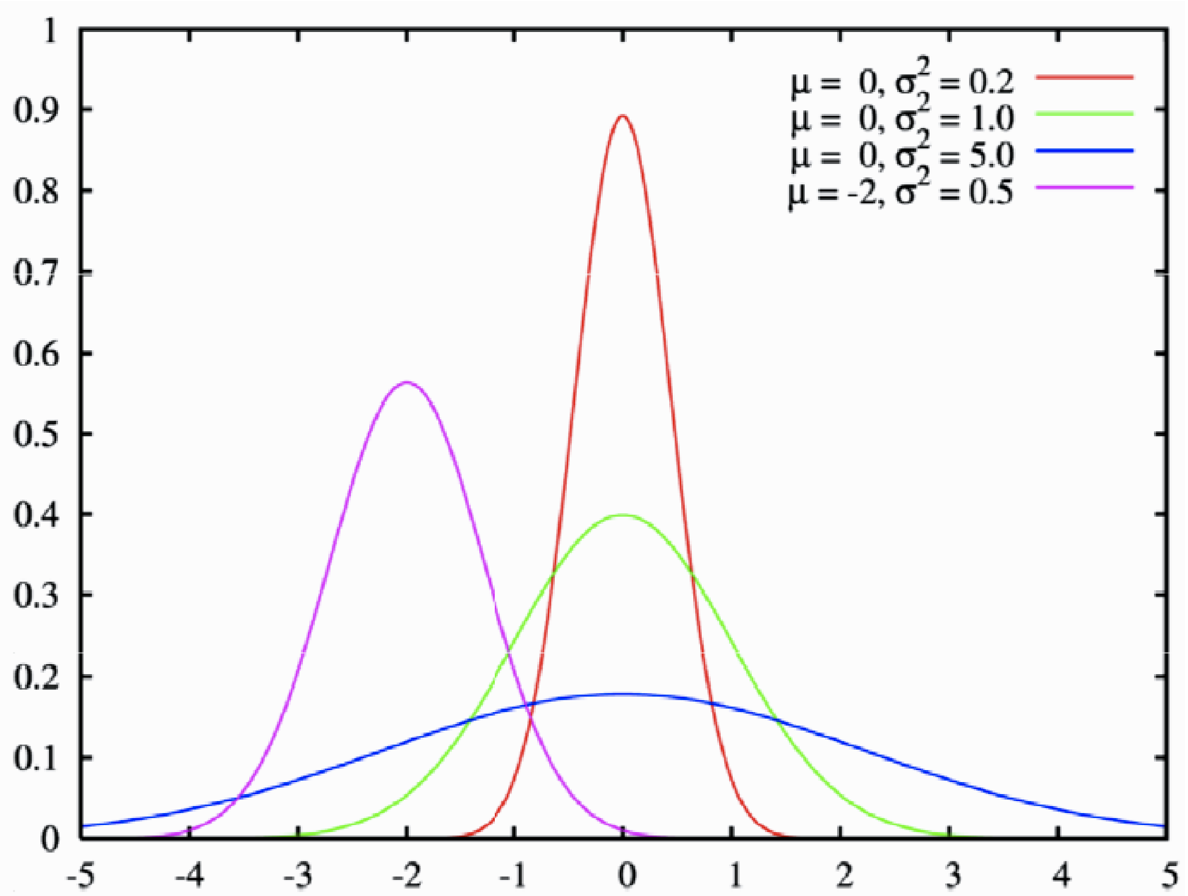


Normal (Gaussian) Distribution

- The **Standard Normal curve**, below, has **mean 0** and **standard deviation 1**.
- If a dataset follows a normal distribution, then **about 68%** of the observations will fall within of the **mean** , which in this case is with the interval (-1,1).
- **About 95%** of the observations will fall within 2 standard deviations of the mean, which is the **interval (-2,2)** for the standard normal
- **About 99.7%** of the observations will fall within 3 standard deviations of the mean, which corresponds to the **interval (-3,3)**



Normal (Gaussian) Distribution



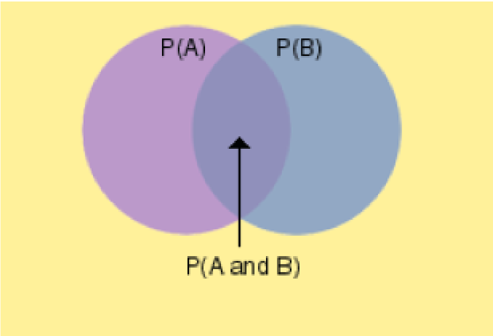
Solutions to Class Questions 1 - 7

Solution #1: Conditional Probability

- What is the probability of selecting a white marble on the second draw, given that the first marble drawn was black?

$$P(White \mid Black) = \frac{P(Black \wedge White)}{P(Black)} = \frac{0.34}{0.47} = 0.72$$

A is black in first draw, B is white in second draw



Solution #2

- G, SD, OD: **events** that a **randomly chosen part** is *good*, *slightly defective* or *obviously defective*
- P(G) = .90, P(SD) = 0.02, and P(OD) = 0.08
- We want to compute the probability that a part is good given that it passed the inspection machine (i.e., it is not obviously defective)

$$P(G|OD^c) = \frac{P(G \cap OD^c)}{P(OD^c)} = \frac{P(G)}{1 - P(OD)} = \frac{.90}{1 - .08} = \frac{90}{92} = .978$$

Solution #3

- **Sample space** is S = {BB,BG,GB,GG}, where, e.g., **outcome** “BG” means that the first-born child is a boy the second-born is a girl
- Assume that boys and girls are equally likely to be born, the 4 elements of S are equally likely.
- The **<u>event, E</u>,** that the neighbor has a son **is the set**

  E = {BB,BG,GB}
- The **<u>event, F</u>**, that the neighbor has two boys (i.e., Joe has a brother) **is the set**

  F = {BB}

$$P(F|E) = \frac{P(F \cap E)}{P(E)} = \frac{P(\{BB\})}{P(\{BB, BG, GB\})} = \frac{1/4}{3/4} = \frac{1}{3}$$

Solution #4

- Let D1 be the event that we find a defective fuse in the first test

- We want to compute $P(D1 \cap D2)$

- $P(D1 \cap D2) = P(D1)P(D2|D1) = 2/7 \,.\, 1/6 = 1/21$

Solution #5

- Let E*i* be the **event** that the first *i* cards have no pair among them
- Then we want to compute P(E6), which is actually the same as P(E1 ∩ E2 ∩ · · · ∩ E6),

  since E6 ⊂ E5 ⊂ · · · ⊂ E1, implying that E1 ∩ E2 ∩ · · · ∩ E6 = E6

$$P(E_1 \cap E_2 \cap \cdots \cap E_6) = P(E_1)P(E_2|E_1)\cdots = \frac{52}{52}\frac{48}{51}\frac{44}{50}\frac{40}{49}\frac{36}{48}\frac{32}{47}$$

0.3452

Solution #6

- From before, we know that:

  P(G) = 90/92 and P(SD) = 2/92

- Also: P(E|G) = .01 and P(E|SD) = 0.10
- Let **E** be the **event** that a randomly selected customer's part fails in the first year
- We want to compute:

$$P(E) = P(E|G)P(G) + P(E|SD)P(SD) =$$

$$= (.01)\frac{90}{92} + (.10)\frac{2}{92} = \frac{11}{920} = 0.012$$

Solution #7 (1/2)

- Let **T** be the **event** that the <u>coin flip was Tails</u>.
- Let **W** be the **event** that a <u>white ball is selected</u>.
- From the given data, we know that:

  > P(W|T) = 3/15

  > P(W|Tc) = 5/12

Since the coin is fair, we also know that

> P(T) = P(T*c*) = ½ (c = complement)

Solution #7 (2/2)

- We want to compute P(T|W), which we do using the definition (and the same simple manipulation that results in Bayes Formula):

$$P(T|W) = \frac{P(T \cap W)}{P(W)} = \frac{P(W|T)P(T)}{P(W|T)P(T) + P(W|T^c)P(T^c)} =$$

$$= \frac{(3/15)(1/2)}{(3/15)(1/2) + (5/12)(1/2)} = \frac{12}{37}$$

0.3243

Let **T** be the **event** that the coin flip was Tails.
Let **W** be the **event** that a white ball is selected.
From the given data, we know that:

> P(W|T) = 3/15
> P(W|Tc) = 5/12

Since the coin is fair, we also know that

> P(T) = P(T*c*) = ½ (c = complement)

We want to compute P(T|W), which we do using the definition (and the same simple manipulation that results in Bayes Formula):

$$P(T|W) = \frac{P(T \cap W)}{P(W)} = \frac{P(W|T)P(T)}{P(W|T)P(T) + P(W|T^c)P(T^c)} =$$

$$= \frac{(3/15)(1/2)}{(3/15)(1/2) + (5/12)(1/2)} = \frac{12}{37}$$

0.3243