

# **Discrete Mathematics**

## **PROBABILITY-II**

**Adil M. Khan**

**Professor of Computer Science**

**Innopolis University**

“The Consequences of an Act Affect the Probability of  
its Occurring Again!”

- B. F. Skinner -

# Recap

1. Why should we learn Probability?
  1. Formulating questions in terms of probability
  1. Building the probability model
    - ❖ Four-step Method
  2. Uniform sample spaces
    1. Counting

# Today's Objectives

1. Counting subsets of a set
1. Conditional Probability
1. Independence
1. Total Probability Theorem
1. Baye's theorem
1. Random variables

# Counting Subsets of a Set

$\binom{n}{k}$  = The number of  $k$ -element subsets of an  $n$  element set.

Is read as “ $n$  choose  $k$ ”

# Why Count Subsets of Set?

## □ Example:

Suppose we select 5 cards at random from a deck of 52 cards.

What is the probability that we will end up having a full house?

Doing this using the possibility tree will take some effort.

# Counting Subsets of a Set

$$\binom{n}{k} = ?$$

# Counting Subsets of a Set

$$\binom{n}{k} = ?$$

- To solve for this, let's start by looking at two ways of constructing an **ordered** sequence of  $k$  distinct items!

# Counting Subsets of a Set

- To solve for this, let's start by looking at two ways of constructing an **ordered** sequence of  $k$  distinct items!
  1. Pick an element one at a time



# Counting Subsets of a Set

- To solve for this, let's start by looking at two ways of constructing an **ordered** sequence of  $k$  distinct items!

1. Pick an element one at a time

$$n(n - 1)(n - 2) \dots (n - k + 1) = \frac{n!}{(n - k)!}$$

# Counting Subsets of a Set

- To solve for this, let's start by looking at two ways of constructing an **ordered** sequence of  $k$  distinct items!

1. Pick an element one at a time

$$n(n-1)(n-2) \dots (n-k+1) = \frac{n!}{(n-k)!}$$

2. Choose  $k$  elements and order them

# Counting Subsets of a Set

- To solve for this, let's start by looking at two ways of constructing an **ordered** sequence of  $k$  distinct items!

1. Pick an element one at a time

$$n(n-1)(n-2) \dots (n-k+1) = \frac{n!}{(n-k)!}$$

2. Choose  $k$  elements and order them

$$\binom{n}{k} k!$$

# Counting Subsets of a Set

- To solve for this, let's start by looking at two ways of constructing an **ordered** sequence of  $k$  distinct items!

❖ Thus

$$\frac{n!}{(n-k)!} = \binom{n}{k} k!$$

# Counting Subsets of a Set

- To solve for this, let's start by looking at two ways of constructing an **ordered** sequence of  $k$  distinct items!

❖ Thus

$$\frac{n!}{(n-k)!} = \binom{n}{k} k!$$

Hence

$$\binom{n}{k} = \frac{n!}{k! (n-k)!}$$

# Conditional Probability

- **An Interesting Kind of Probability Question**
- “After this lecture, when I go to UI canteen for lunch, what is the probability that today they will be serving **biryani** (my favorite food)?

# Biryani 😊



# Conditional Probability

- Of course, the vast majority of the food that the cafeteria prepares is **NEITHER** delicious **NOR** is it ever biryani (low probability).
- But they do cook dishes that contain rice, so now the question is “what’s the probability that food from UI is delicious given that it contains rice?”
- This is called “**Conditional Probability**”



# Conditional Probability

- What is the probability that it will rain this afternoon, given that it is cloudy this morning?
- What is the probability that two rolled dice sum to 10, given that both are odd?

Written as

- **$P(A|B)$**  – denotes the probability of event A, given that event B happens.

# Conditional Probability

- **So, how to answer the “Food Court” question?**

# Conditional Probability

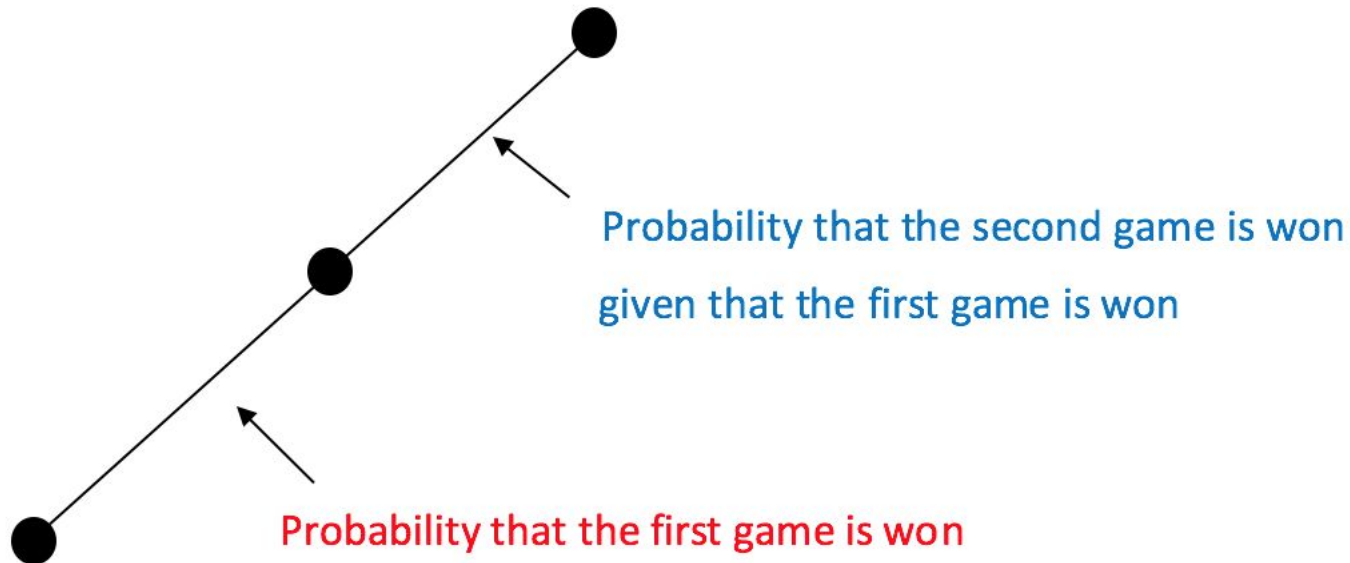
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

# Why Do Tree Diagrams Work?

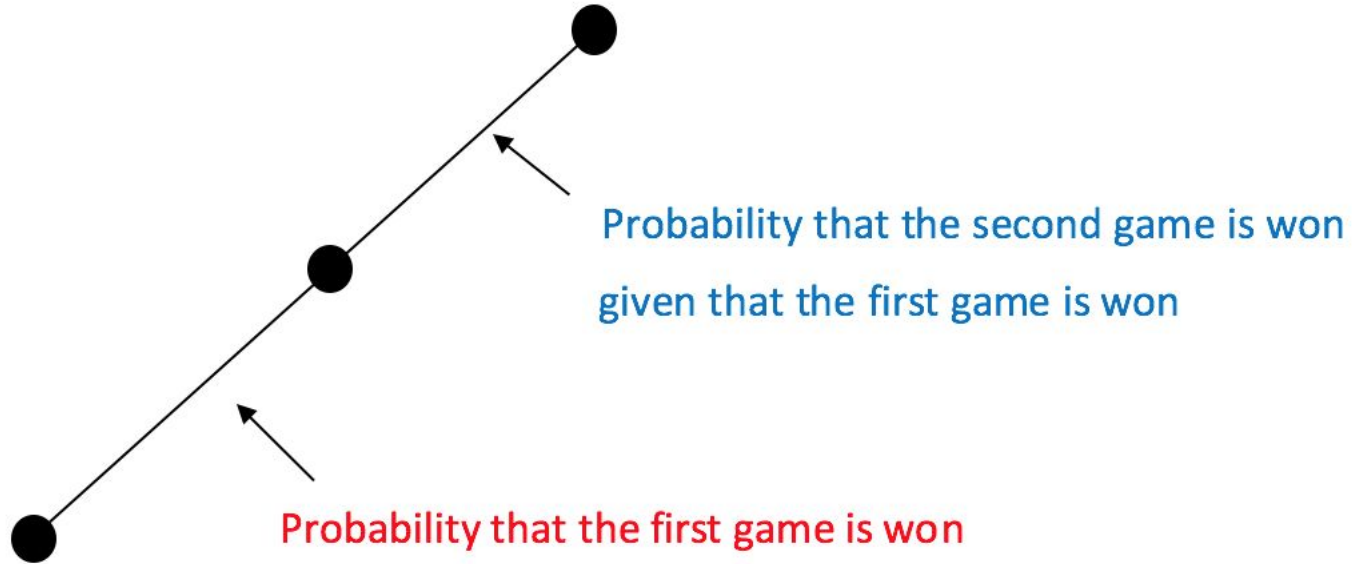
- We have solved multiple probability problems using tree diagrams
- Let's think for a moment about “why do tree diagrams work?”
- The answer involves conditional probabilities
- In fact, the probabilities that we have been recording on the edges of a tree diagram are conditional probabilities
- More generally, on each edge of a tree diagram, we record that the probability that the experiment proceeds along that part, given that it reaches the parent vertex

# Why Do Tree Diagrams Work?

Let's look the upper most edges of the probability tree for the previous example!

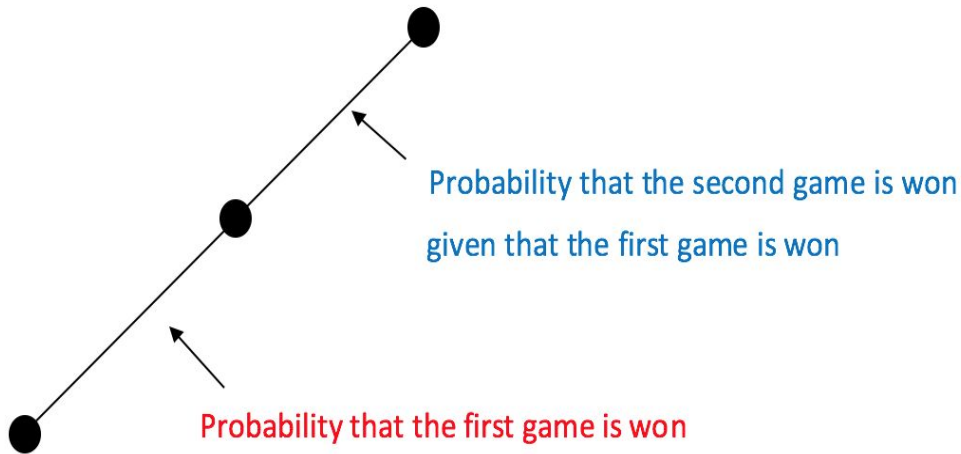


# Why Do Tree Diagrams Work?



$$P(W_1W_2) = P(W_1 \cap W_2) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$$

# Why Do Tree Diagrams Work?



$$P(WW) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$$

$P(\text{win first game} \cap \text{win second game})$

$= P(\text{win first game}) \cdot P(\text{win second game} | \text{win first game})$

# Why Do Tree Diagrams Work?

**Rule** (Product Rule for 2 Events). *If  $\Pr[E_1] \neq 0$ , then:*

$$\Pr[E_1 \cap E_2] = \Pr[E_1] \cdot \Pr[E_2 \mid E_1].$$

**Rule** (Product Rule for  $n$  Events).

$$\Pr[E_1 \cap E_2 \cap \dots \cap E_n] = \Pr[E_1] \cdot \Pr[E_2 \mid E_1] \cdot \Pr[E_3 \mid E_1 \cap E_2] \cdots \\ \cdot \Pr[E_n \mid E_1 \cap E_2 \cap \dots \cap E_{n-1}]$$

*provided that*

$$\Pr[E_1 \cap E_2 \cap \dots \cap E_{n-1}] \neq 0.$$

**“So the *Product Rule* is the formal justification for multiplying edge probabilities in a probability tree to get outcome probabilities”**



# Independence

- Intuitively, two events  $A$  and  $B$  are independent if knowing that  $A$  happens does not affect the probability that  $B$  happens
- Thus

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

# Independence

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

- Now, we already know that

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

# Independence

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

- Now, we already know that

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

- Putting two together

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

# Independence

- Why use this definition instead of the intuitive one?

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

# Independence

- Why use this definition instead of the intuitive one?

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

Because it is symmetric in the roles of  $A$  and  $B$

# Independence

- Why use this definition instead of the intuitive one?

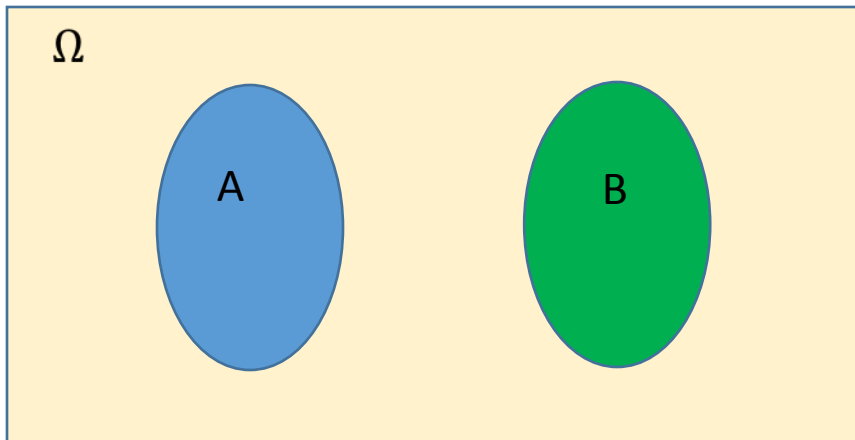
$$P(B \cap A) = P(A)P(B)$$

Because it is symmetric in the roles of  $A$  and  $B$

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

# What Independence Really Means?

- Are these events independent?



$$P(A) > 0 \text{ and } P(B) > 0$$

# What Independence Really Means?

- Thus being dependent is completely different from being disjoint!



# What Independence Really Means?

- Thus being dependent is completely different from being disjoint!
- Two events are independent, **if the occurrence of one does not change our belief about the occurrence of the other.**

# What Independence Really Means?

- Thus being dependent is completely different from being disjoint!
- Two events are independent, if the occurrence of one does not change our belief about the occurrence of the other.
- Typically the case when the two events are determined by two physically distinct and non-interacting processes.
  - **Getting heads in a coin toss** and **snowing outside**

# Independence---Cont.

- Generally, independence is an assumption that we assume when modeling a phenomenon.
- The reason we so-often assume statistical independence is not because of its real-world accuracy
- It is because of its armchair appeal: It makes the math easy

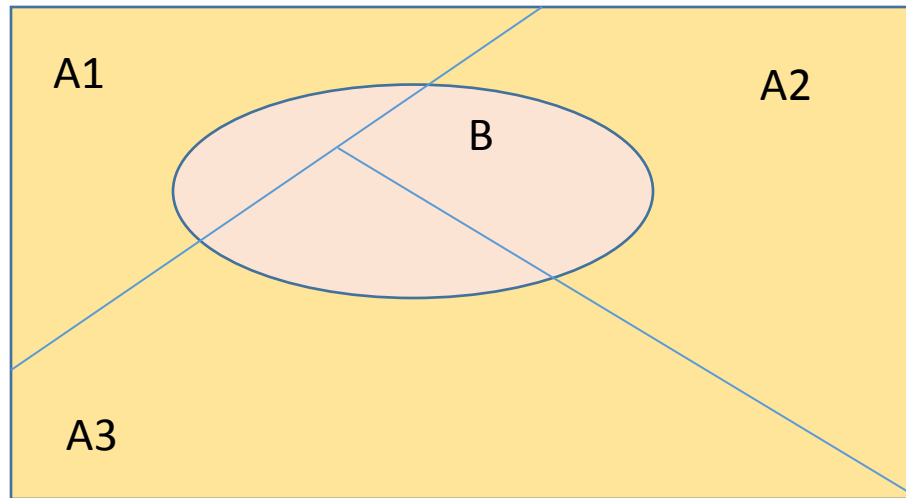
How does it do that?

- By splitting a compound probability into a product of individual probabilities.

(Note for TAs: Include example of Independence assumption in tutorials)

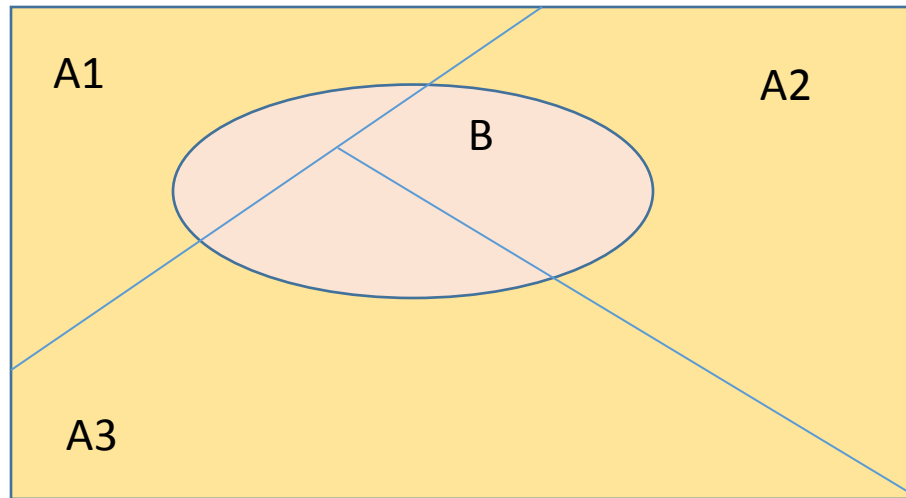
# Total Probability Theorem

- Take a look at the figure below



# Total Probability Theorem

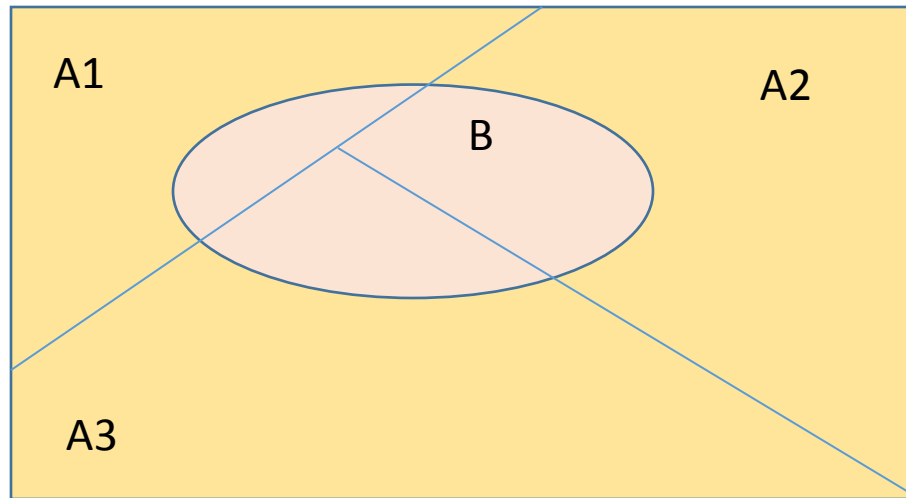
- Take a look at the figure below



$P(B)$  ?

# Total Probability Theorem

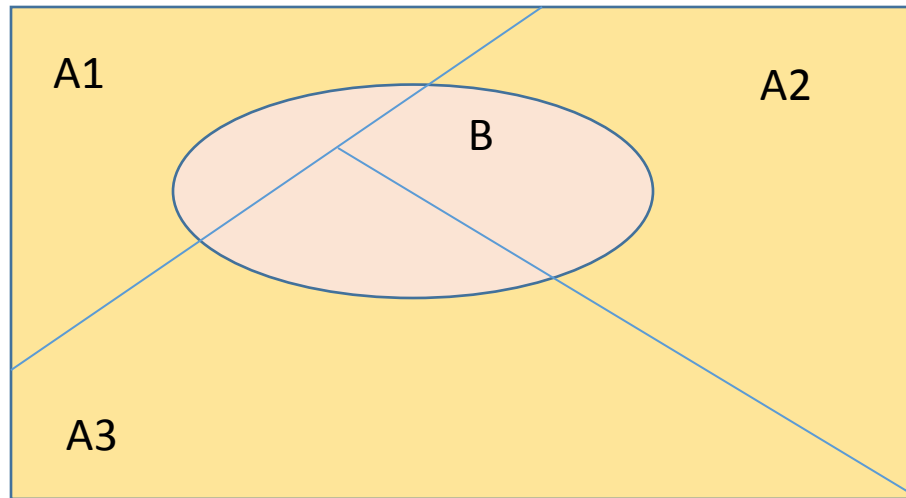
- Take a look at the figure below



$$P(B) = P(B \cap A1) + P(B \cap A2) + P(B \cap A3)$$

# Total Probability Theorem

- Take a look at the figure below

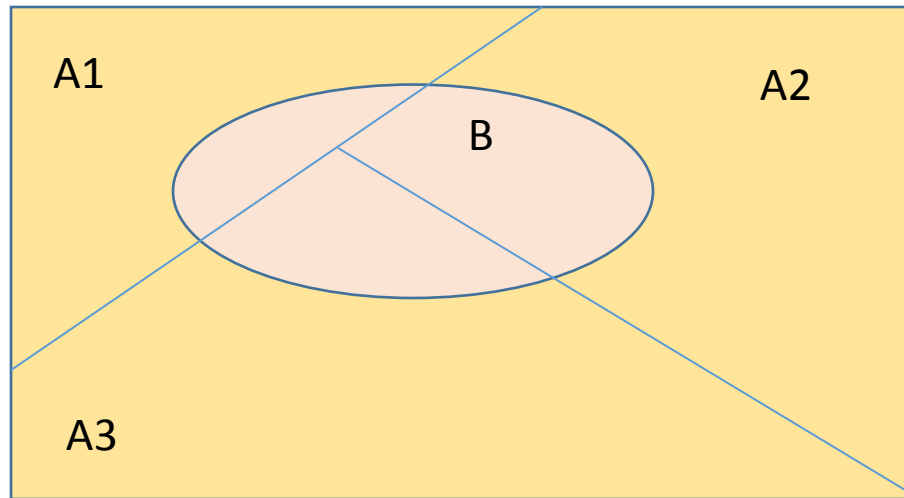


$$P(B) = P(B \cap A1) + P(B \cap A2) + P(B \cap A3)$$

$$= P(A1)P(B|A1) + P(A2)P(B|A2) + P(A3)P(B|A3)$$

# Total Probability Theorem

- Take a look at the figure below



$$P(B) = P(B \cap A1) + P(B \cap A2) + P(B \cap A3)$$

$$= P(A1)P(B|A1) + P(A2)P(B|A2) + P(A3)P(B|A3)$$

$$= \sum_i P(A_i)P(B|A_i)$$



# Total Probability Theorem

- Where do we use it?
  - ❖ Baye's Theorem!

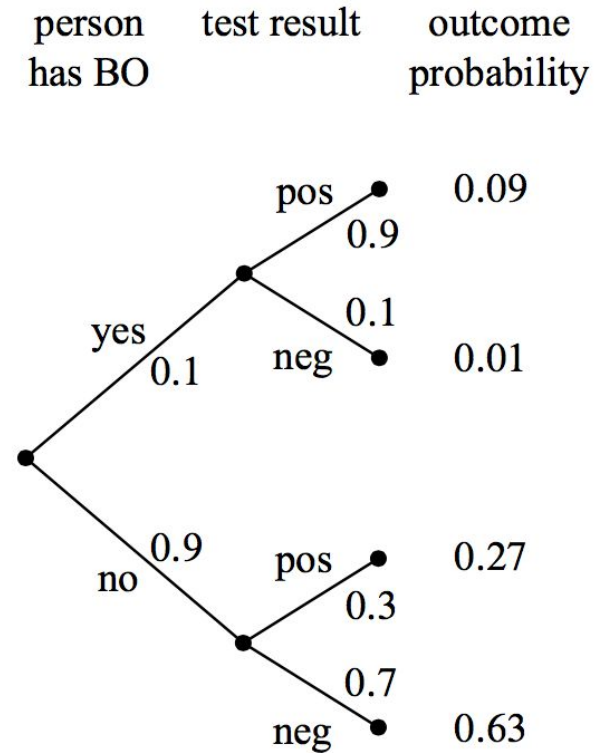
# Medical Testing Problem

- Let's assume a “not-so-perfect” test for a medical condition called BO suffered by 10% of the population
- The test is not-so-perfect because
  - 90% of the tests come positive if you have BO
  - 70% of the tests come negative if you don't have BO
- If we randomly test a person for BO, and if the test comes positive, what is the probability that the person has BO.

# Probability Tree

**A:** The test came positive

**B:** The person has BO



BO is suffered by **10%** of the population

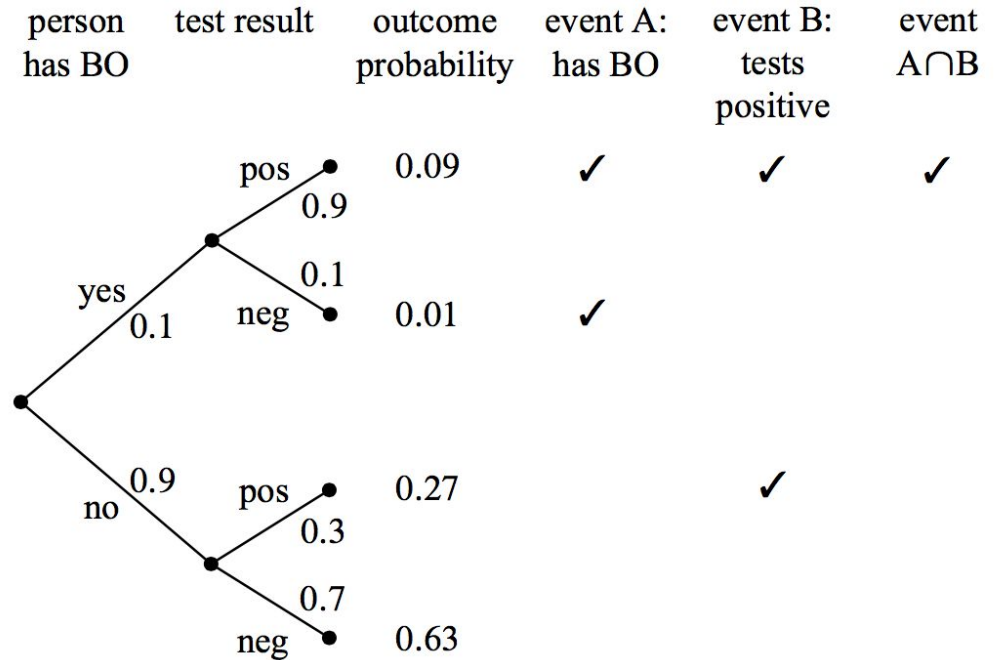
If someone has BO, there is a **90%** chance that the test will be positive

If someone does not have the condition, there is a **70%** chance that the test will be negative.

# Probability Tree

**A:** The test is positive

**B:** The person has BO



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

$$P(B|A) = \frac{0.09}{0.09 + 0.27} = \frac{1}{4}$$

# Conditional Probability Tree---Cont.

- **Surprising, Right!**
- So if the test comes out positive, the person has only 25% chance of having the diseases
- **Conclusion:**
  - Tests are flawed
  - Tests give test probabilities not the real probabilities

# Bayes Theorem

- **How to correct for such Flawed Tests**
- **Bayes Theorem**
  - It lets you relate the test probabilities with the real probabilities.
  - More specifically, it lets you relate  $P(A|B)$  with  $P(B|A)$ .
  - **What is  $P(B|A)$ ?**

# Bayes Theorem---Cont.

- **A Posteriori Probabilities**
  - A conditional probability in reverse  $P(B|A)$  is called a **posteriori probability**.
  - You can understand this by considering that event B precedes event A in time.

# Bayes Theorem---Cont.

- **A Posteriori Probabilities**
- **For example:**
  - The probability that it was cloudy this morning, **given that it rained in the afternoon.**
- Mathematically speaking, there is no difference between a posteriori probability and a conditional probability.



# Flawed Test

- **Coming Back to Flawed Test**

- *Let*

**A:** The test came positive

**B:** Person has BO

# Flawed Test

- **Then**
  - $P(A|B)$  means the chance that indicator **A** (a person's test came positive) happened given that the event **B** occurred (the person has the disease).

# Flawed Test

- $P(A|B)$  means the chance that indicator **A** (a person's test came positive) happened given that the event **B** occurred (the person has the disease).
- $P(B|A)$  means the probability that event **B** (a person having disease) happened given the indicator **A** (the person's test came positive)

# Flawed Test

- $P(A|B)$  means the chance that indicator **A** (a person's test came positive) happened given that the event **B** occurred (the person has the disease).
- $P(B|A)$  means the probability that event **B** (a person having disease) happened given the indicator **A** (the person's test came positive)

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

# Flawed Test

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

$$P(\text{Has BO} | \text{Pos Test}) = \frac{P(\text{Pos Test} | \text{Has BO})}{P(\text{Pos test})}$$

# Random Variables

- So far, we focused on probabilities of events.
- For example,
  - The probability that someone wins the Monty Hall Game
  - The probability that someone has a rare medical condition given that he/she tests positive

# Random Variables

- But most often, we are interested in knowing more than this.
- For example,
  - ❖ How many players must play Monty Hall Game before one of them finally wins?
  - ❖ How long will a weather certain condition last?
  - ❖ How long will I loose gambling with a strange coin all night?
- To be able to answer such questions, we have to learn about “Random Variables”

# Random Variables---Cont.

- “Random Variables” are nothing but “functions”
- A *random variable*  $R$  on a probability space is a function whose domain is the sample space and whose range is a set of Real numbers.



# Random Variables---Cont.

- “Random Variables” are nothing but “functions”
- A *random variable*  $R$  on a probability space is a function whose domain is the sample space and whose range is a set of Real numbers.
- Let’s look at this example!
  - Tossing three independent coins and noting
    - **C: the number of heads that appear**
    - **M: 1 if all are heads or tails, 0 otherwise**
- If we look closely, we will see that C and M are in fact functions that map every outcome of the experiment to a number.

# Random Variables---Cont.

- Example ---Cont.

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

$$C(HHH) = 3$$

$$C(HHT) = 2$$

$$C(HTH) = 2$$

$$C(HTT) = 1$$

$$M(HHH) = 1$$

$$M(HHT) = 0$$

$$M(HTH) = 0$$

$$M(HTT) = 0$$

$$C(THH) = 2$$

$$C(THT) = 1$$

$$C(TTH) = 1$$

$$C(TTT) = 0$$

$$M(THH) = 0$$

$$M(THT) = 0$$

$$M(TTH) = 0$$

$$M(TTT) = 1$$

- Thus **C** and **M** are random variables!

# Indicator Random Variables

- Maps every outcome to either 0 or 1 – its range is  $\{0,1\}$  – *indicates* that a sample point has/hasn't a certain property
- **M** from our example: Partitions the sample space into two blocks
- Such random variables are called indicator random variables

HHH TTT  
 $M=1$

HHT HTH HTT THH THT TTH  
 $M=0$

# Random Variables and Events

- General random variables partition the sample space into several blocks).
- Note that  $C$  from our previous example is a general random variable

$$\underbrace{TTT}_{C=0} \quad \underbrace{TTH \ THT \ HTT}_{C=1} \quad \underbrace{TTH \ HTH \ HHT}_{C=2} \quad \underbrace{HHH}_{C=3}$$

- Notice that each sample in the block has the same value for the random variable
- An equation or an inequality involving a random variable can be regarded as an event.

# Random Variables and Events---Cont.

- For example

$$\begin{aligned}P(C = 2) &= P(HHT, HTH, HHT) \\ &= P(THH) + P(HTH) + P(HHT)\end{aligned}$$

# Random Variables.

- More generally, an event can be defined as

$\{w | R(w) = x\}$  is the event that  $R = x$

- And its probability can be defined as

$$P(R = x) = \sum_{w | R(w) = x} P(w)$$

- A random variable could be continuous or discrete
- When dealing with continuous random variables, use “integrals” instead of “summations”

# Expected Value

- Weighted average of the values of a random variable
- Provides a central point for the distribution of the values of a random variable
- We can solve many problems using the notion of expected values
  - ❖ How many heads are expected to appear if a coin is tossed 100 times?
  - ❖ What is the expected number of comparisons used to find an element in a list using the linear search?

# Expected Value---Cont.

$$Ex[R] ::= \sum_{w \in S} R(w) Pr[w]$$



# Expected Value---Cont.

$$Ex[R] ::= \sum_{w \in S} R(w) Pr[w]$$

- For example, the expected value of a random variable with **uniform distribution** on  $\{1, 2, \dots, n\}$  is

$$Ex[R_n] = \sum_{i=1}^n i \cdot \frac{1}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

# Variance

Consider the following two gambling games:

- ❖ **Game A:** You win \$2 with probability  $\frac{2}{3}$  and lose \$1 with probability  $\frac{1}{3}$ .
  - ❖ **Game B:** You win \$1002 with probability  $\frac{2}{3}$  and lose \$2001 with probability  $\frac{1}{3}$ .
- Which game would you play?

# Variance

Let's compute the **expected** return for both games:

# Variance

- ❖ **Game A:** You win \$2 with probability  $2/3$  and lose \$1 with probability  $1/3$ .

$$Ex[A] = 2 \cdot \frac{2}{3} + (-1) \cdot \frac{1}{3} = 1$$

# Variance

- ❖ **Game B:** You win \$1002 with probability  $\frac{2}{3}$  and lose \$2001 with probability  $\frac{1}{3}$ .

$$Ex[B] = 1002 \cdot \frac{2}{3} + (-2001) \cdot \frac{1}{3} = 1$$

**Expected return is the same.** Thus expected value is not enough to make the decision

# Variance

The variance  $\text{Var}[R]$  of a random variable  $R$  is

$$\text{Var}[R] = \text{Ex}[(R - \text{Ex}[R])^2]$$

# Variance

- ❖ **Game A:** You win \$2 with probability  $\frac{2}{3}$  and lose \$1 with probability  $\frac{1}{3}$ .

$$A - \text{Ex}[A] = \begin{cases} 1 & \text{with probability } \frac{2}{3} \\ -2 & \text{with probability } \frac{1}{3} \end{cases}$$

$$(A - \text{Ex}[A])^2 = \begin{cases} 1 & \text{with probability } \frac{2}{3} \\ 4 & \text{with probability } \frac{1}{3} \end{cases}$$

$$\text{Ex}[(A - \text{Ex}[A])^2] = 1 \cdot \frac{2}{3} + 4 \cdot \frac{1}{3}$$

$$\text{Var}[A] = 2.$$

# Variance

For game B

$$B - \text{Ex}[B] = \begin{cases} 1001 & \text{with probability } \frac{2}{3} \\ -2002 & \text{with probability } \frac{1}{3} \end{cases}$$

$$(B - \text{Ex}[B])^2 = \begin{cases} 1,002,001 & \text{with probability } \frac{2}{3} \\ 4,008,004 & \text{with probability } \frac{1}{3} \end{cases}$$

$$\text{Ex}[(B - \text{Ex}[B])^2] = 1,002,001 \cdot \frac{2}{3} + 4,008,004 \cdot \frac{1}{3}$$

$$\text{Var}[B] = 2,004,002.$$

- Intuitively, this means that the payoff in Game A is usually close to the expected value of \$1, but the payoff in Game B can deviate very far from this expected value – **high variance means high risk.**



# Standard Deviation

- Because of its definition in terms of the **square** of a random variable, the **variance** of a random variable may be very **far from a typical deviation from the mean**.

# Standard Deviation

- For example, in Game B above, the deviation from the mean is **1001** in one outcome and **-2002** in the other. But the variance is a whopping **2,004,002**
- The problem is with the “***units***” of variance.
  - **If a random variable is in dollars**, then the expected value is also in dollars, **but the variance is in *square dollars***

# Standard Deviation

- For this reason, *standard deviation* is often used to describe the deviation of a random variable from its expected value

$$\sigma_R = \sqrt{Var[R]} = \sqrt{Ex[(R - Ex[R])^2]}$$

- For example, the standard deviation for games A and B are

$$\sigma_A = \sqrt{Var[A]} = \sqrt{2} \approx 1.41$$

$$\sigma_B = \sqrt{Var[B]} = \sqrt{2,004,002} \approx 1416$$

Why bother squaring in the first place?