Topic 4

Probability

- How Probability Quantifies Randomness
- Three Approaches in Finding Probabilities
- Visualizing Events & Their Probabilities
- Finding Probabilities
  (The Addition Rule & The Multiplication Rule)

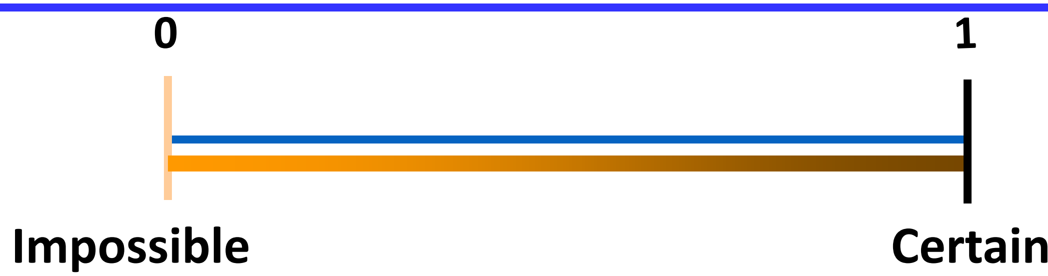
Random Phenomena

- For random phenomena, the outcome is uncertain.
- In the short-run, the proportion of times that something happens is unpredictable.
- In the long-run, the proportion of times that something happens becomes very predictable.
- As we make more observations, the proportion of times that a particular outcome occurs gets closer and closer to a certain number we would expect.
- Probability *quantifies* long-run *randomness.*

Law of Large Numbers and Probability

- As the number of trials increase, the proportion of occurrences of any given outcome approaches a particular number “in the long run”.
- We will interpret the probability of an outcome to represent long-run results.
- With random phenomena, the *probability* of a particular outcome is the proportion of times that the outcome would occur in a long run of observations.

$$\boxed{\text{Probability = Long Run Proportion}}$$



Example: Probability

When rolling a die, the outcome of “6” has probability = 1/6.

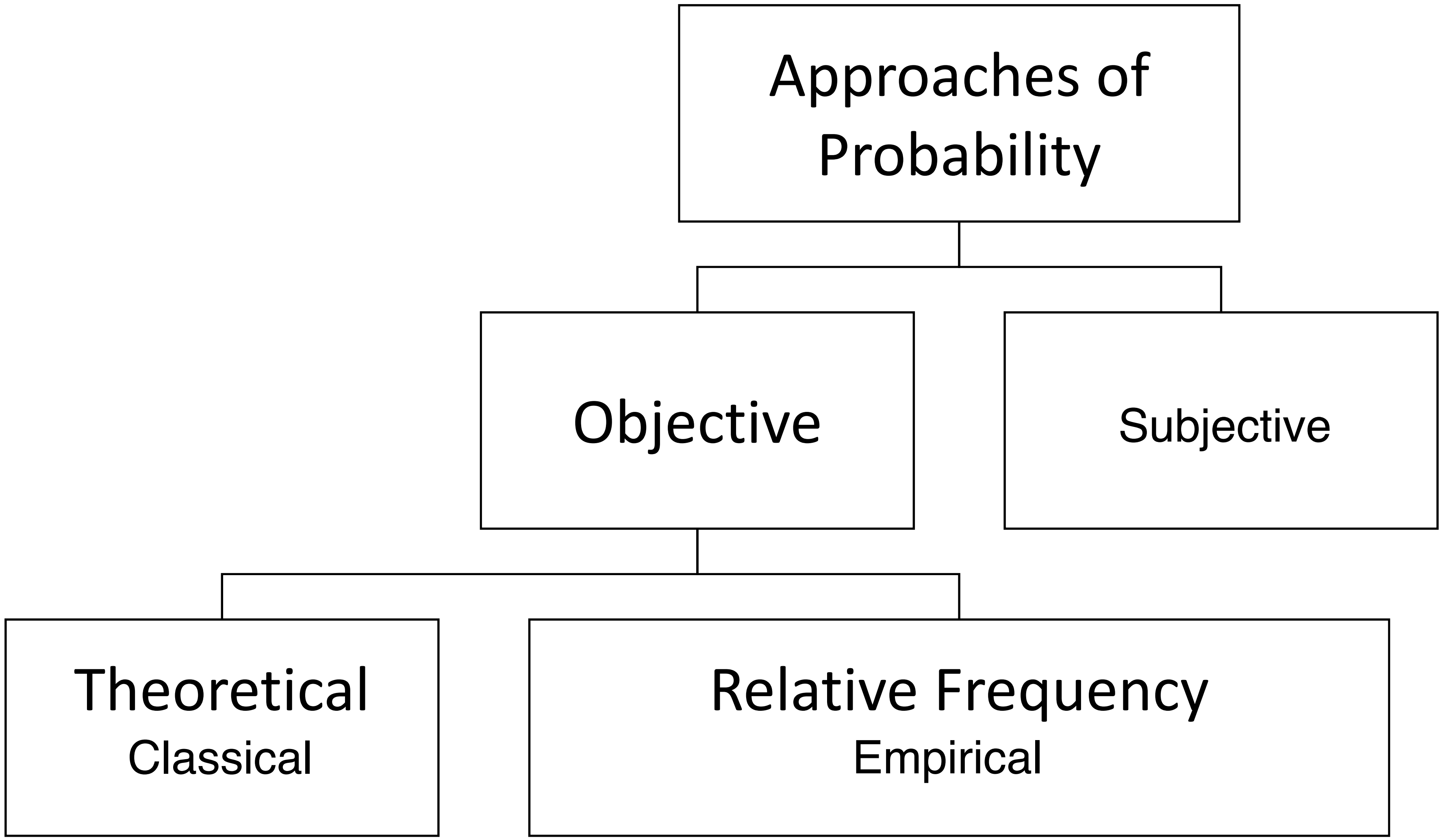
In other words, the proportion of times that a 6 would occur in a long run of observations is 1/6.

Finding Probabilities

- We calculate theoretical probabilities based on assumptions about the random phenomena. For example, it is often reasonable to assume that outcomes are equally likely such as when flipping a coin, or a rolling a die.

- We observe many trials of the random phenomenon and use the sample proportion of the number of times the outcome occurs as its probability. This is merely an estimate of the actual probability.
- The *relative frequency definition of probability* is the long run proportion of times that the outcome occurs in a very large number of trials

- When a long run of trials is not feasible, you must rely on subjective information. In this case, the *subjective definition of the probability* of an outcome is your degree of belief that the outcome will occur based on the information available.


Approaches of Probability
Objective
Subjective
Theoretical
Classical
Relative Frequency
Empirical

Examples:

1. We wish to determine what proportion of students at a certain school have type AB blood

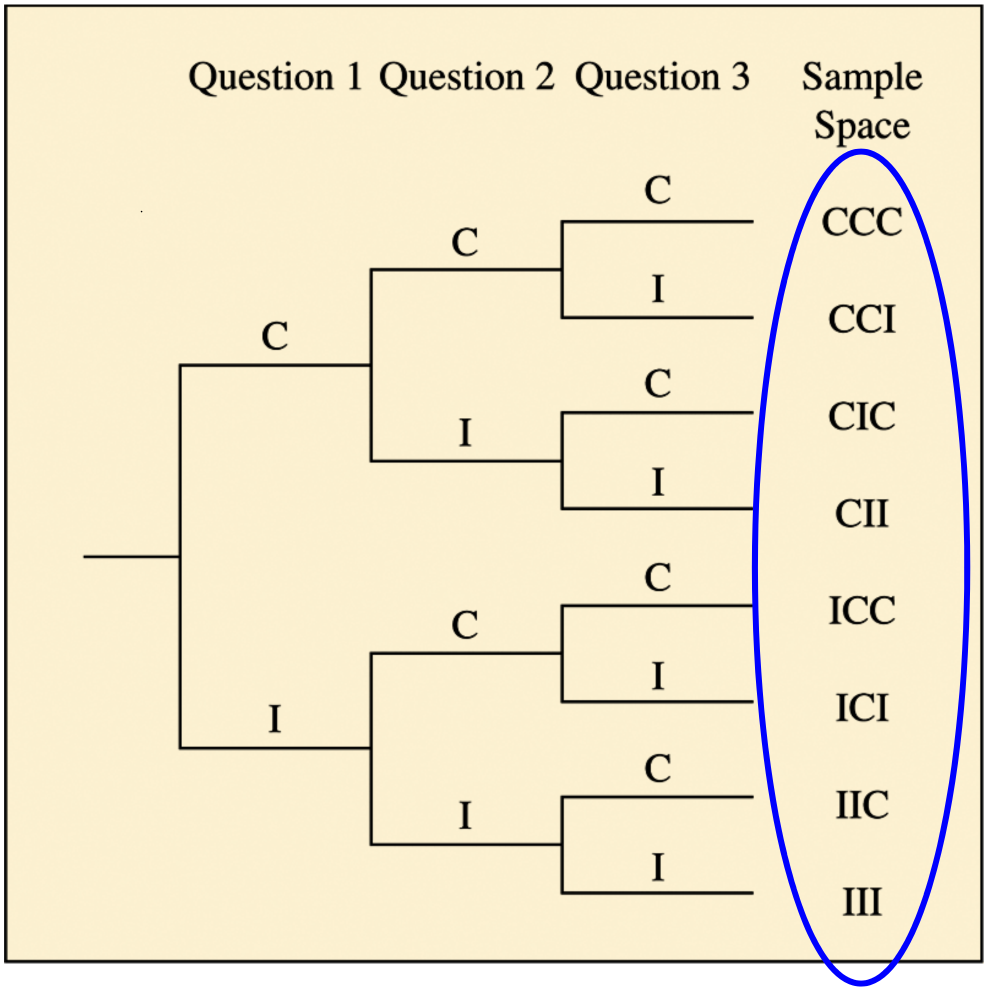
   We take *a simple random sample* with 100 students

   If 5 of those students have type AB blood, then we would estimate that the proportion of students at this school with type AB blood is 5%

2. Throughout his teaching career, Professor Gan has awarded 3000 A's out of 15,000 students. What is the probability that a student in his section this semester will receive an A?

Sample Space

- For a random phenomenon, the **sample space** is the set of all possible outcomes.



From the **tree diagram**, a student's performance has eight possible outcomes:

{CCC, CCI, CIC, CII, ICC, ICI, IIC, III}

Question: How many possible outcomes would there be if the quiz had four questions?

Event

- An **event** is a subset of the **sample space**.
  An event corresponds to a *particular* outcome or a group of possible outcomes.

Example:

For a student taking the three-question pop quiz, some possible events are:

- Event of the student answers all 3 questions correctly = (CCC)
- Event of the student answers at least 2 questions correctly = (CCI, CIC, ICC, CCC)

Finding Probabilities of Events

- Each outcome in a sample space has a probability.
- So does each event.

- The probability of each individual outcome is between 0 and 1.

- The total of all the individual probabilities equals 1.

Example: Tax Audit

Sample Space

Income Level	Audited Yes	Audited No	Total
Under $200,000	1,260	132,147	**33,407**
$200,000–$1,000,000	131	4,311	**4,442**
More than $1,000,000	22	371	**393**
Total	**1,413**	**136,829**	**138,242**

The frequencies in the **contingency table** are reported in thousands. For example, 1260 represents 1,260,000 tax forms that reported income under $200,000 and were audited.

What is the sample space for selecting a taxpayer?

1.{(under $200,000, Yes)

2. (under $200,000, No)

3. ($200,000 - $1,000,000, Yes)

.

.

6. (More than $1,000,000, No)}

Example: Tax Audit

For a randomly selected taxpayer in 2002:

1. What is the probability of an audit?

   1413/ 138242

2. What is the probability of an income of more than $1,000,000?

   393/ 138242

3. What income level has the greatest probability of being audited?

   22/ 393 ? 131/ 4442? 1260 / 33407?

Basic rules for finding probabilities about a pair of events

Some events are expressed as the outcomes that:

- are **not** in some other event
  (complement of the event).

- are in one event **and** in another event
  (intersection of two events).

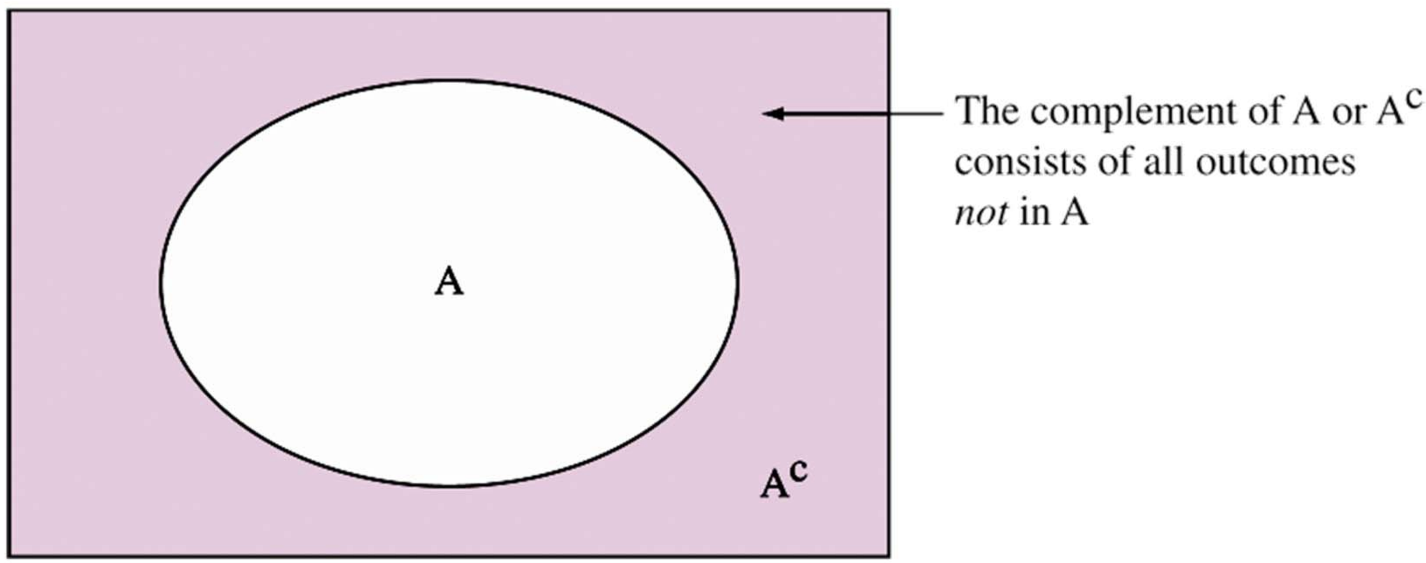
- are in one event **or** in another event **or** in both events
  (union of two events).

The Complement of an Event

•The complement of an event A consists of all outcomes in the sample space that are not in A.

•The probabilities of A and of A^c add to 1,

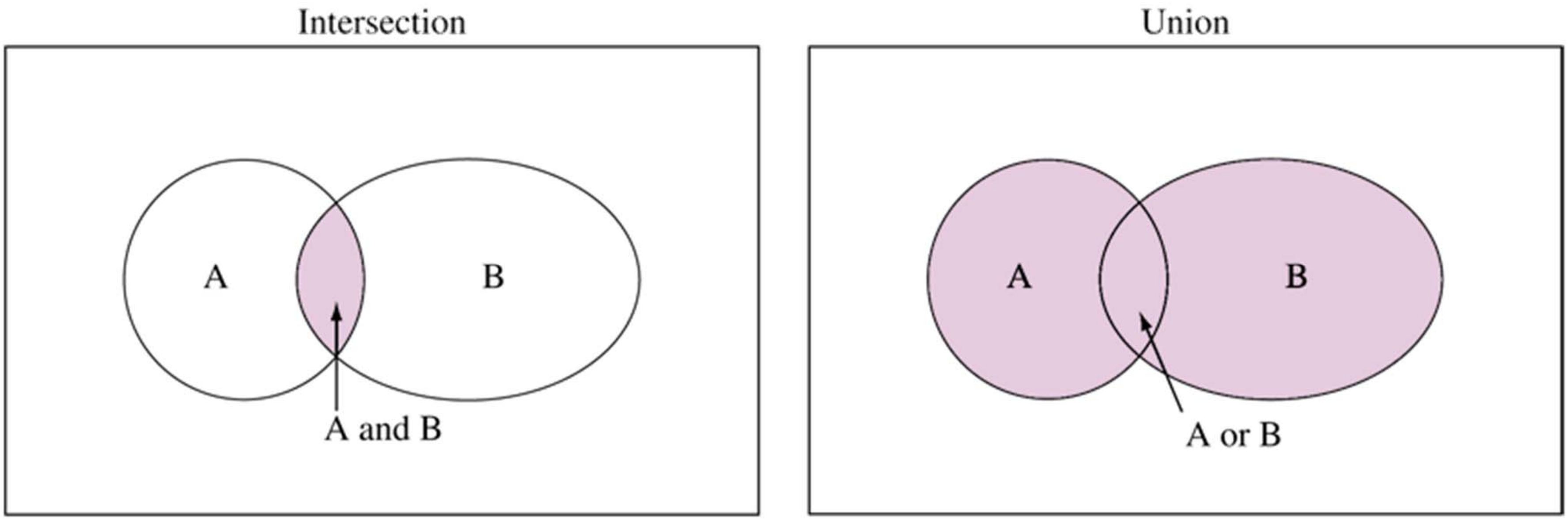
•so $P(A^c) = 1 – P(A)$.



Venn Diagram Illustrating an Event A and Its Complement A^c.

Intersection of Two Events

- The **intersection** of A and B consists of outcomes that are in both A and B, denoted "A and B."

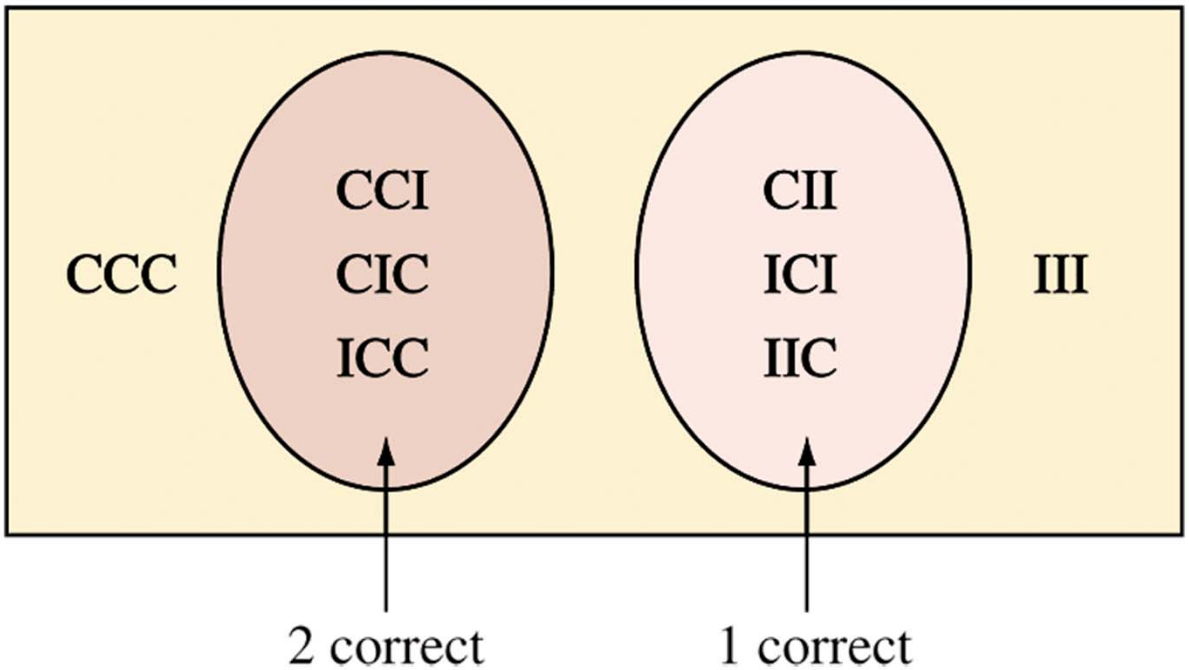


The **union** of A and B consists of outcomes that are in A or in B or in both A and B, denoted "A or B."

Question: How could you find P(A or B) if you know P(A), P(B), and P(A and B)?

Disjoint Events

•Two events, A and B, are disjoint if they do not have any common outcomes. mutually exclusive

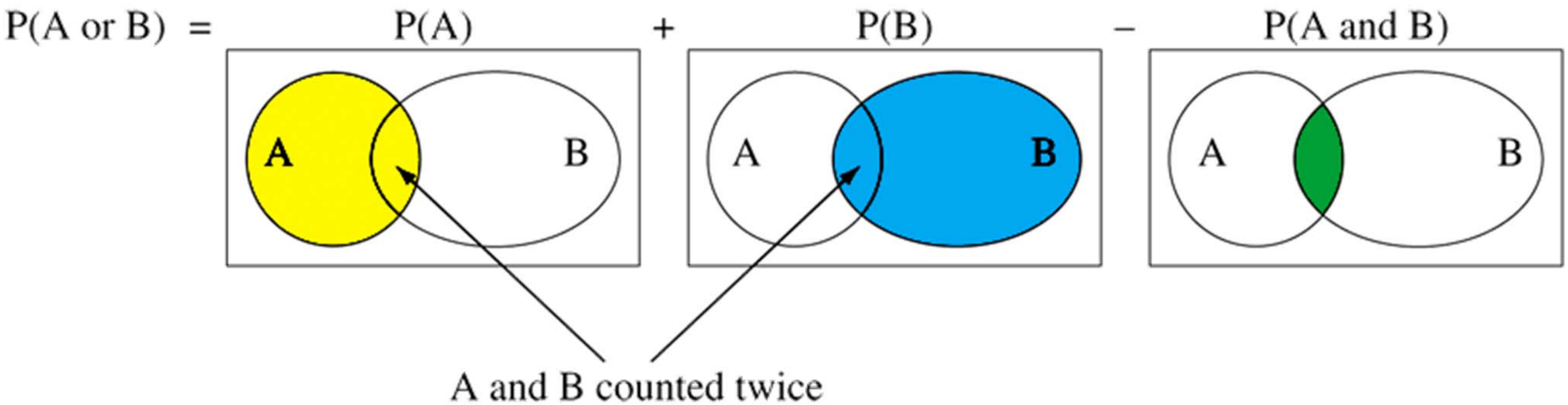


Question:
Identify on this figure the event that the student answers the first question correctly.
Is this event disjoint from either of the two events identified in the Venn diagram?

Probability of the Union of Two Events

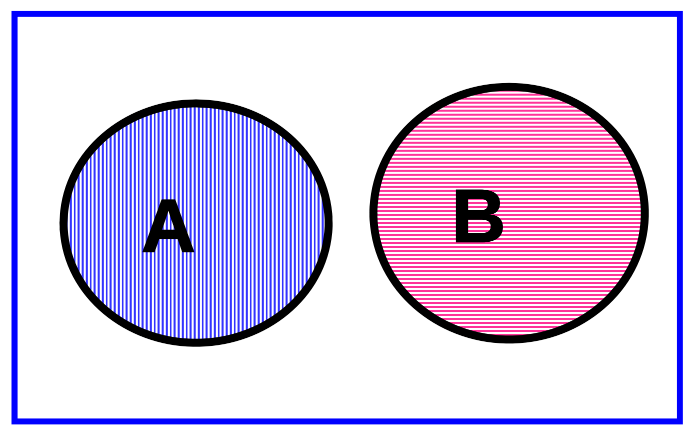
Addition Rule for the ***union*** of two events,

P(A **or** B) = P(A) + P(B) – P(A and B)



If the events are ***disjoint***

P(A **or** B) = P(A) + P(B)



Example: Probability of the Union of Two Events

- Consider a family with two children.
- The sample space possibilities for the genders of the two children are {FF, FM, MF, MM}.
- Assume the four outcomes in the sample space are equally likely.

- Let A = {first child a girl}
- B = {second child a girl}
- P(A) = P({FF, FM}) = 0.50
- P(B) = P({FF, MF}) = 0.50
- P(A and B) = P({FF}) = 0.25

- The event A or B is the event that the first child is a girl, or the second child is a girl, or both are girls. In other words, that at least one child is a girl. Its probability is:

- P(A or B) = P(A) + P(B) - P(A and B) =

Independent Trials & Independent Events

Different trials of a random phenomenon are *independent* if the outcome of any one trial is not affected by the outcome of any other trial.

Two events, A and B are independent if the occurrence of A does not affect the probability of event B.

Other ways of saying the same thing:

- Knowing event A does not give any additional information about B
- A and B are totally unrelated

Not independent = dependent

Example:

If you have 20 flips of a coin in a row that are "heads", you are not "due" a "tail" - the probability of a tail on your next flip is still ½.

Probability of the Intersection of Two Events

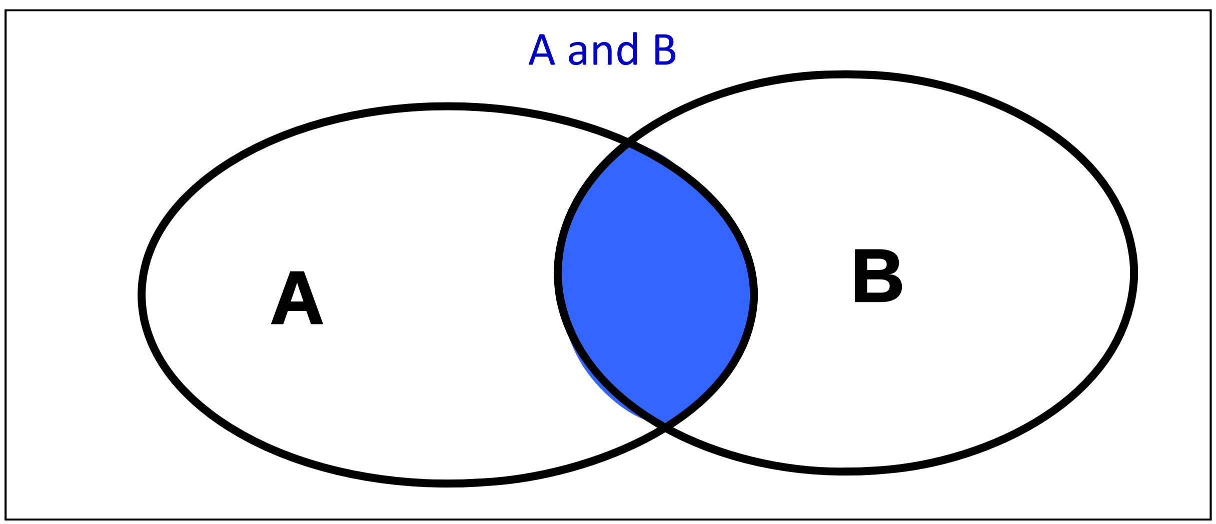
- **Multiplication Rule** for the ***intersection*** of two events

  P(A **and** B) = P(A) x P(B|A)

  P(A **and** B) = P(B) x P(A|B)

  If the events are ***independent***

  P(A **and** B) = P(A) x P(B)



Are the two events independent?

Examples:
Probability of the Intersection of Two Events

Two rolls of a die.

P(6 on roll 1 *and* 6 on roll 2) = P(6 on roll 1) x P(6 on roll 2)

=

Color blindness.

For genetic reasons, color blindness is more common in men than women: 5 in 100 men and 25 in 10,000 women suffer from color blindness. If the population is half male and half female, what is the probability that a randomly chosen person is male and is color blind?

P(male) = 0.50

P(color blind | male) =

P(male and color blind) =

Example: Guessing yet Passing a Pop Quiz

- What is the probability of getting 3 questions correct by guessing?
- If each question has five options, then the probability of selecting the correct answer for any given question is 1/5, or 0.20.

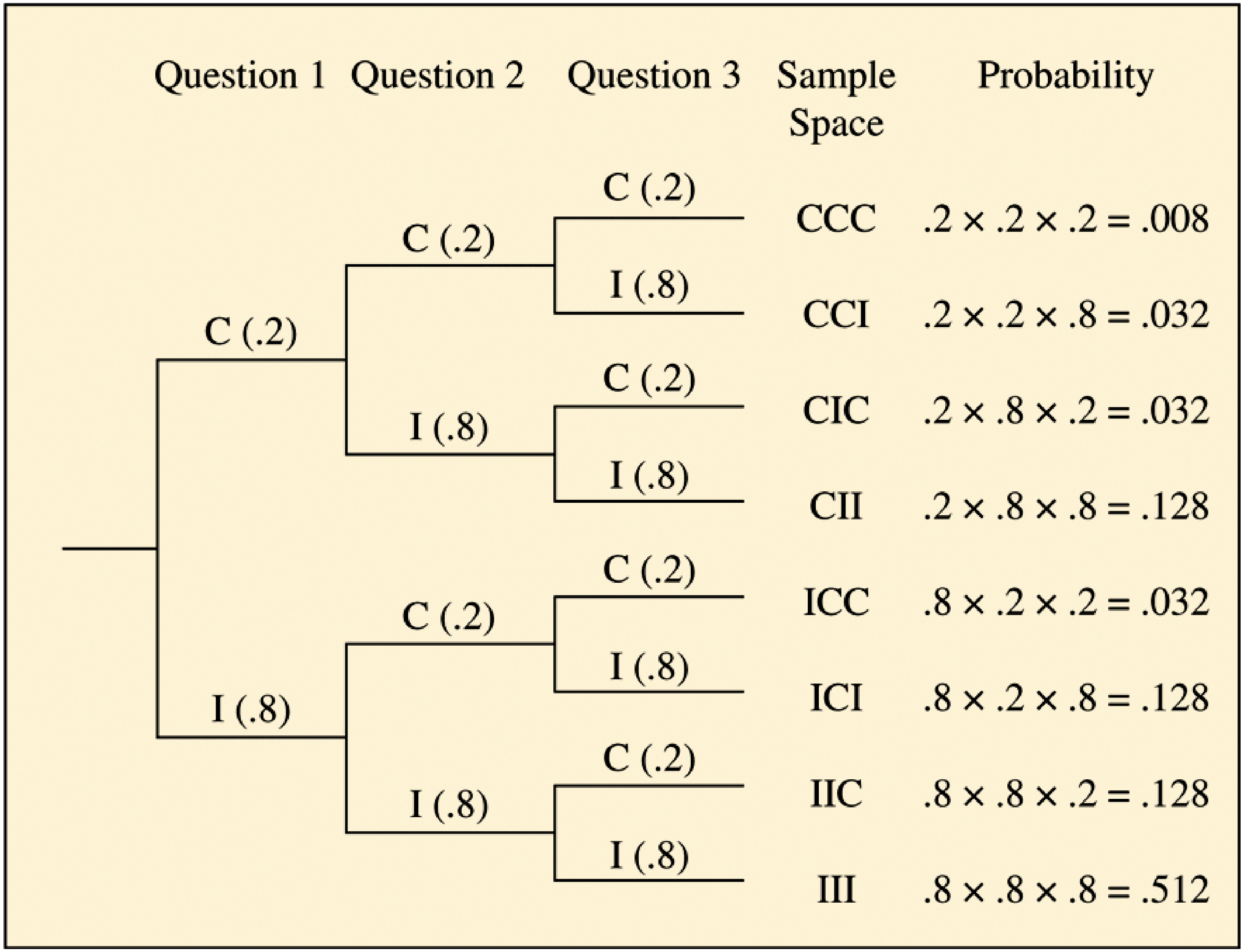
- The probability that the student answers all three questions correctly is:

- P(CCC) = P(C) x P(C) x P(C)

  =

Example: Guessing yet Passing a Pop Quiz

- Multiplication of the probabilities along that path gives its probability.



Example: Guessing yet Passing a Pop Quiz

What is the probability that a student passes the test (answers *at least* 2 questions correctly)?

P(CCC) + P(CCI) + P(CIC) + P(ICC)

=

Example: Events Often Are Not Independent

- **A Pop Quiz with 2 Multiple Choice Questions**
  - Data giving the proportions for the actual responses of students in a class.

    (I = incorrect, C = correct) :

Outcome	:	II	IC	CI	CC
Probability	:	0.26	0.11	0.05	0.58

	2nd Question	
1st Question	**C**	**I**
C	0.58	0.05
I	0.11	0.26

A and B

Example: Events Often Are Not Independent

Define the events A and B as follows:

- A: {first question is answered correctly}
- B: {second question is answered correctly}

P(A) = P{(CI), (CC)} = 0.05 + 0.58 = 0.63

P(B) = P{(IC), (CC)} = 0.11 + 0.58 = 0.69

P(A and B) = P{(CC)} = 0.58

If A and B were independent,

P(A and B) = P(A) x P(B) =

Thus, in this case, A and B are not independent!

Example:
Left handed or Right handed

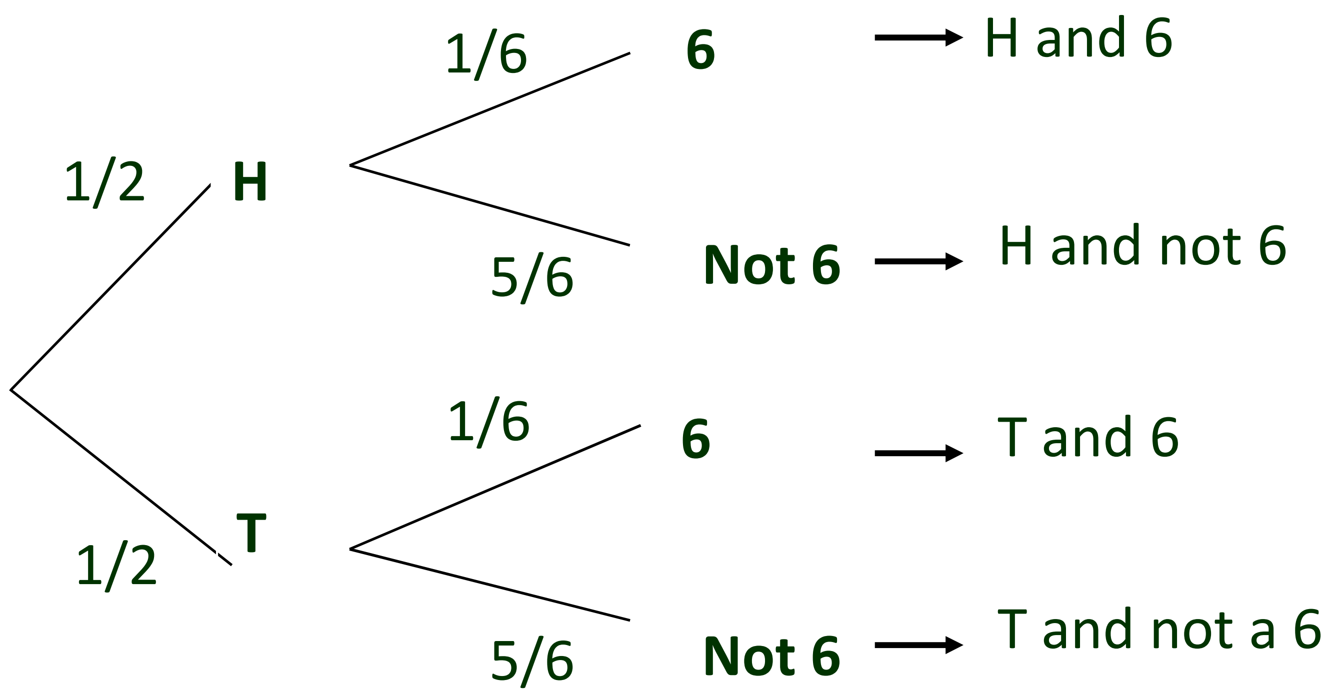
- There are 50 males and 50 females in a sample. 76% of the males are right handed while 16% of females are left handed. To illustrate,
- Construct a contingency table
- Draw a tree diagram
- What is the probability of choosing a subject from this sample and find that the subject is a:
  - Male?
  - Left handed?
  - Right handed male?
  - Left handed male?
  - Right handed female?

Example:
Left handed or Right handed

	Left handed	Right handed	**Total**
Male			**50**
Female			**50**
Total			**100**

Example:
Left handed or Right handed

Example: Independent trials
(Roll a die and flip a coin)



Conditional Probability

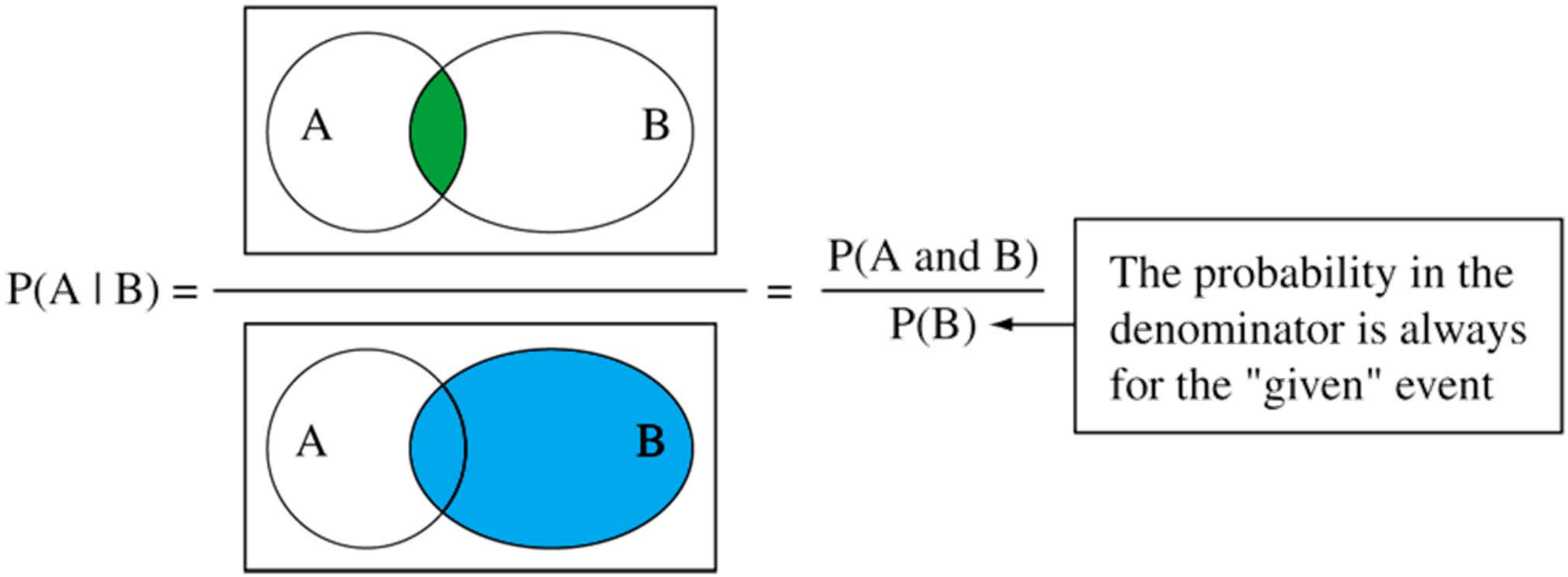
- For events A and B, the **conditional probability** of event A, given that event B has occurred, is:

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

- $P(A|B)$ is read as "the probability of event A, given event B." The vertical slash represents the word "given".

- Of the times that B occurs, $P(A|B)$ is the proportion of times that A also occurs.

Conditional Probability



Venn Diagram of Conditional Probability of Event A Given Event B. Of the cases in which B occurred, P(A|B) is the proportion in which A also occurred.

Question: Sketch a representation of P(B|A). Is P(A|B) necessarily equal to P(B|A)?

Example 1: Conditional Probability

Income Level	Audited		Total
	Yes	No	
	These 6 probabilities sum to 1.0		
Under $200,000	0.0091	0.9559	**0.9650**
$200,000–$1,000,000	0.0009	0.0312	**0.0321**
More than $1,000,000	0.0002	0.0027	**0.0029**
Total	**0.0102**	**0.9898**	**1.0000**

What was the probability of being audited, given that the income was $\geq$ $1,000,000?

- Event A: Taxpayer is audited
- Event B: Taxpayer's income $\geq$ $1,000,000

Example 2: Conditional Probability

•A study of 5282 women aged 35 or over analyzed the Triple Blood Test to test its accuracy.

	Blood Test		
Down Syndrome Status	**POS**	**NEG**	**Total**
D (Down)	48	6	**54**
D^c (unaffected)	1307	3921	**5228**
Total	**1355**	**3927**	**5282**

- A **positive test result** states that the condition is present.
- A **negative test result** states that the condition is not present.
- **False Positive:**
  Test states the condition is present, but it is actually absent.
- **False Negative:**
  Test states the condition is absent, but it is actually present.

Example 2: Conditional Probability

	Blood Test		
Down Syndrome Status	**POS**	**NEG**	**Total**
D (Down)	48	6	**54**
D^c (unaffected)	1307	3921	**5228**
Total	**1355**	**3927**	**5282**

- Assuming the sample is representative of the population, find the estimated probability of a positive test for a randomly chosen pregnant woman 35 years or older

  P(POS) = 1355/5282=0.257

- Given that the diagnostic test result is positive, find the estimated probability that Down syndrome truly is present.

  48/1355=0.0354

Checking for Independence

- To determine whether events A and B are independent:

- Is $P(A \mid B) = P(A)$?

- Is $P(B \mid A) = P(B)$?

- Is $P(A \; and \;\; B) = P(A) \times P(B)$?

- If any one of these is true, the other two are also true, and the events A and B are independent.

Example: Down Syndrome Again

- The diagnostic blood test for Down syndrome:
  - **POS** = positive result **NEG** = negative result
  - **D** = Down Syndrome **D^C** = Unaffected

Status	POS	NEG	Total
D	0.009	0.001	0.010
D^c	0.247	0.742	0.990
Total	0.257	0.743	1.000

- Are the events POS and D independent or dependent?

Is $P(POS \mid D) = P(POS)$?

P(POS) = 0.257

The events POS and D are dependent

Probability Model

- We’ve dealt with finding probabilities in many idealized situations.

- In practice, it’s difficult to tell when outcomes are equally likely or events are independent.

- In most cases, we must specify a *probability model* that approximates reality.

Probability Model

- A **probability model** specifies the possible outcomes for a sample space and provides assumptions on which the probability calculations for events composed of these outcomes are based.

- Probability models merely *approximate* reality. They are rarely *exactly* satisfied.

Example: Safety of the Space Shuttle

- Let S = event that the space shuttle mission is successful.
- A risk assessment study by the Air Force estimated P(S) = 0.971.
- What is the probability of at least one failure in a total of 100 missions?
- P(at least 1 failure) =1-P(0 failures)

$$= 1 - P(S1\ and\ \ S2\ and\ S3 \ldots and\ S100)$$

$$= 1 - P(S1) \times P(S2) \times \ldots \times P(S100)$$

joint event, use multiplication, conditional

=1-[P(S)]^100=1-[0.971]^100=0.9473

Example: Safety of the Space Shuttle

- This answer relies on the assumptions of:
  - Same success probability on each flight.
  - Independence.
  - These assumptions are suspect since other variables (temperature at launch, experience of crew, age of craft used, quality of O-ring seals, etc.) could affect the probability.

Probabilities and Diagnostic Testing

	Blood Test		
Down Syndrome Status	**POS**	**NEG**	**Total**
D (Down)	48	6	**54**
D^c (unaffected)	1307	3921	**5228**
Total	**1355**	**3927**	**5282**

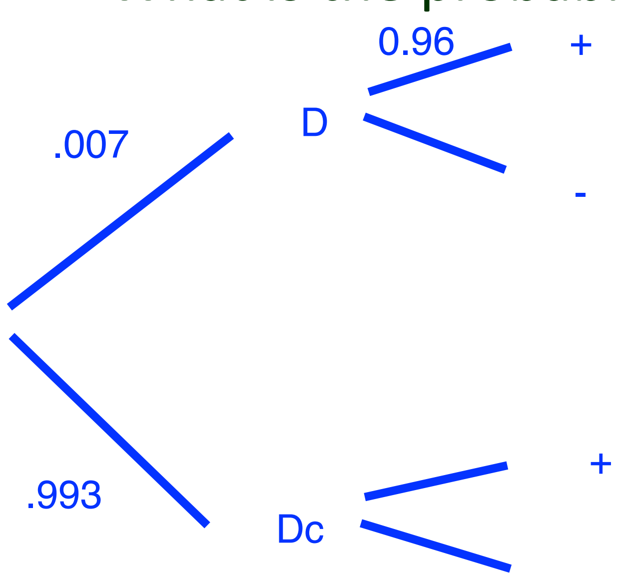
	Diagnostic Test Result		
State Present?	**Positive (POS)**	**Negative (NEG)**	**Total Probability**
Yes (S)	Sensitivity $P(POS \mid S)$	False negative rate $P(NEG \mid S)$	**1.0**
No (S^c)	False positive rate $P(POS \mid S^c)$	Specificity $P(NEG \mid S^c)$	**1.0**

Example:
Random Drug Testing of Air Traffic Controllers

Sensitivity = $P(POS \mid S)$

Specificity = $P(NEG \mid S^c)$

- Probability of drug use at a given time ≈ 0.007 (prevalence of drug use)
- ▪ Sensitivity of test = 0.96
- ▪ Specificity of test = 0.93
- What is the probability of a positive test result?



Example:
Random Drug Testing of Air Traffic Controllers

- What is the probability of a positive test result?

$$P(POS) = P(S \ and \ POS) + P(S^c \ and \ POS)$$

$$P(S \ and \ POS) = P(S)P(POS \mid S)$$
$$= 0.007 \times 0.96 = 0.0067$$

$$P(S^c \ and \ POS) = P(S^c)P(POS \mid S^c)$$
$$= 0.993 \times 0.07 = 0.0695$$

$$P(POS) = 0.0067 + 0.0695 = 0.0762$$

- Even though the prevalence is < 1%, there is an almost 8% chance of the test suggesting drug use!