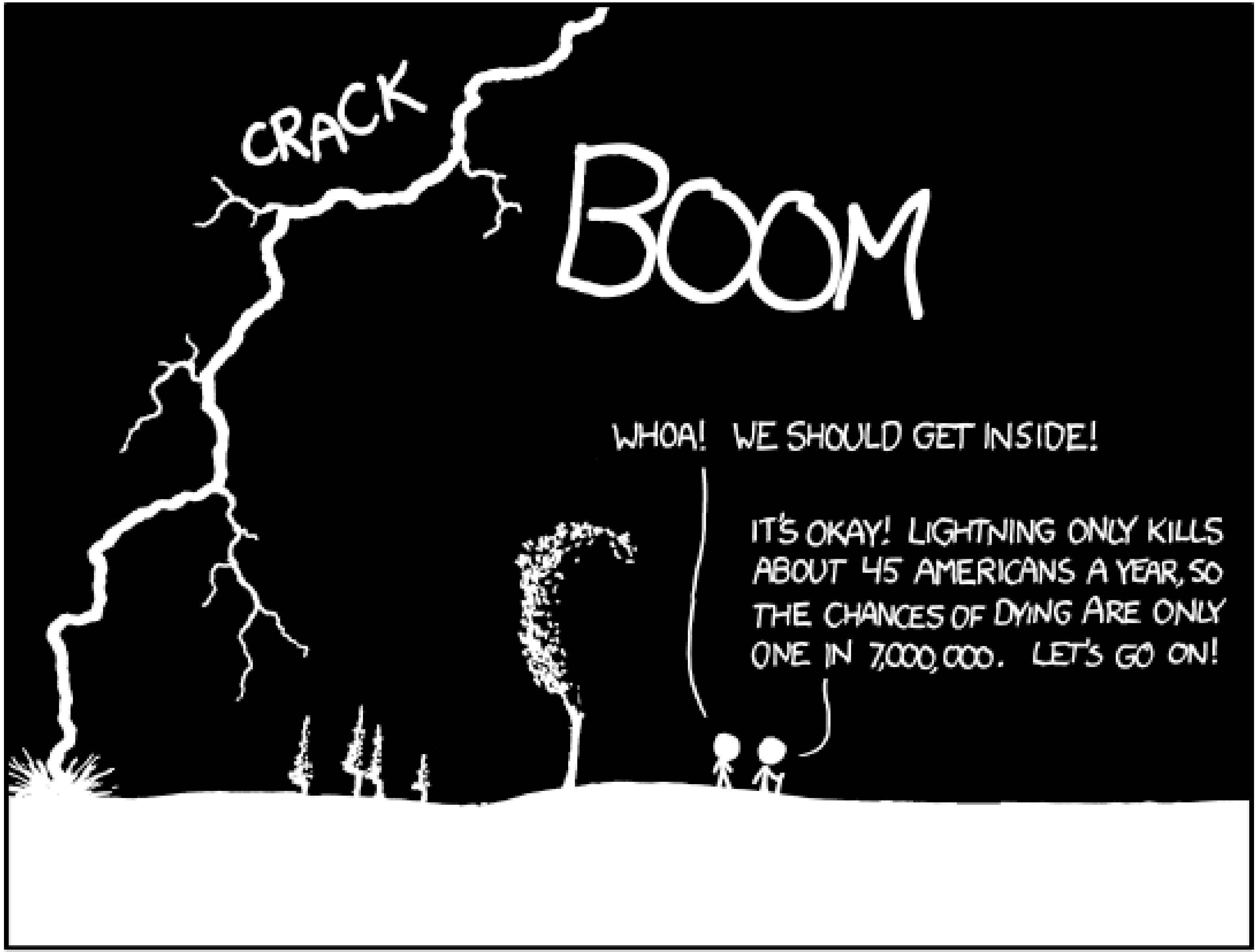
STAT 201 Chapter 5

Probability



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

Introduction to Probability

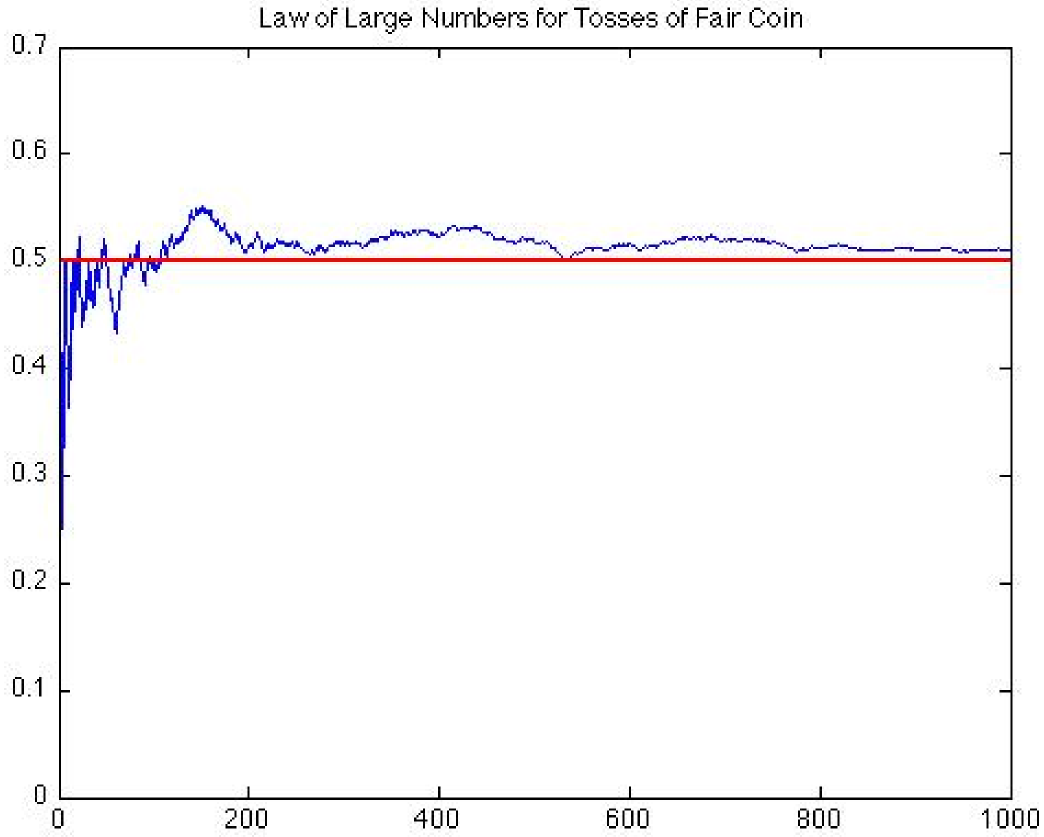
- **Probability –** The way we quantify uncertainty.
- **Subjective Probability –** A probability derived from an individual's personal judgment about whether a specific outcome is likely to occur.
  - **Examples:** Let’s assume Shiwen is a big fan of Gamecock. He claims that Gamecock will win all the games in next season. You and Shiwen have bet on his statement. Shiwen will put the same amount of money you put on the table. How much do you offer?
  - **Note:** Subjective probability is **NOT** what we will discuss in this course!

Introduction to Probability

- **Randomness –** Possible outcomes are known but it is uncertain which will occur for any observation.
  - **Examples:** flipping a coin, rolling a dice, etc
  - **Note:** Random phenomena is highly uncertain so we look at patterns in the long run (law of large numbers)

Law of Large Numbers (LLN)

- (**LLN**) – As the number of observations/trials increase, the proportion of occurrences of any given outcome approaches a particular number in the long run.
- **Long-run Probability -** The probability of a particular outcome is the proportion of times that the outcome would occur in the long-run, as our sample size goes to infinity.



Probability

- Think about a football game. The captains of both teams go to the center of the field to call heads or tails in hopes that they'll be able to get the ball first. What is the probability that the **home team** wins the toss?

- Clearly, we should all answer they have a 50/50 shot – 50% chance – probability of 0.5.

Independence

- **Independence –** when the outcome of one trial is not affected by the outcome of any other trial(s).

Example: Independent or Not?

- Does the fact that it is extremely cold in Moscow, Russia today gives us any information about whether children in Columbia will have recess today?

- Does the fact that it is extremely cold in Columbia today gives us any information about whether children in Columbia will have recess today?

Probability - Definitions

- **Sample space –** the set of all possible outcomes in an experiment
  - **Examples**
    - **Rolling a dice:** {1,2,3,4,5,6}
    - **Flipping a coin:** {Heads, Tails}
- **Event** – We refer to a particular outcome occurring as an event
  - Examples: we rolled a 4, the coin landed heads

Probability – How to Calculate

- Probability of event A

$$P(A) = \frac{N(A)}{N(S)} = \frac{(Number\ of\ ways\ A\ can\ happen)}{(Total\ number\ of\ outcomes\ in\ sample\ space)}$$

- Example: Rolling a dice
  - Sample Space = S = {1,2,3,4,5,6}
  - Let event A = roll a even number = {2,4,6}
  - Probability we roll a multiple of three = P(A) = P(2,4,6)=3/6 = 1/2

Probability Rules

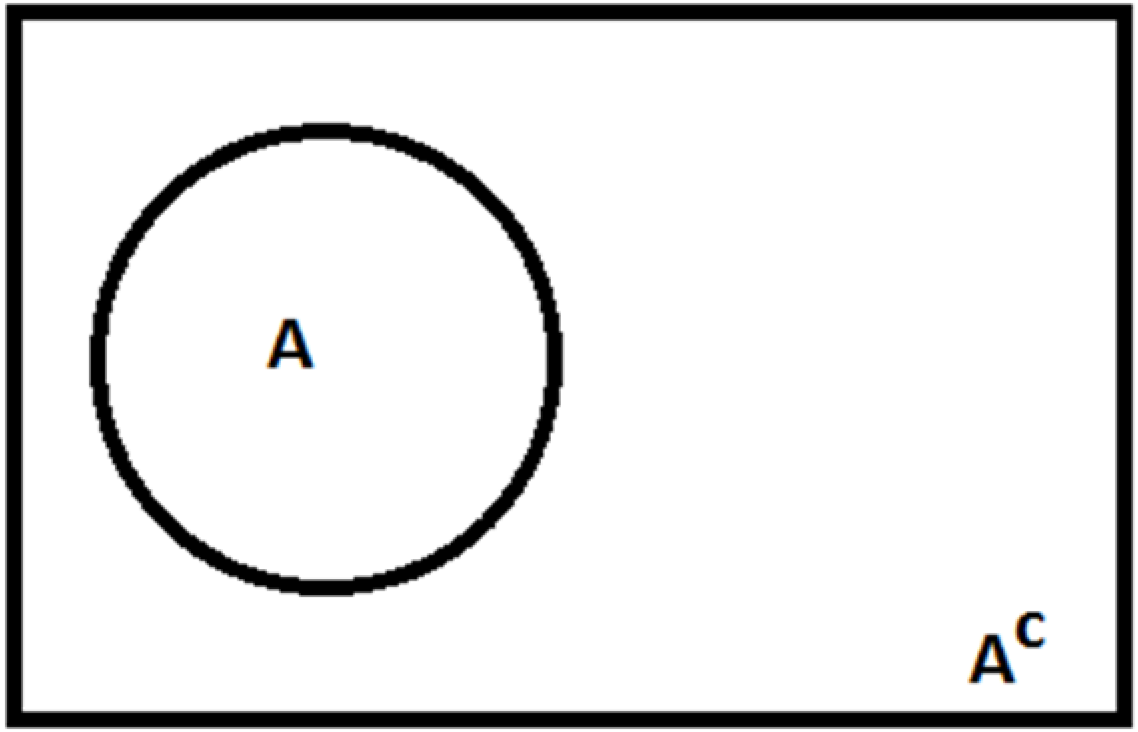
- 1) Probability is always between 0 and 1. (Probabilities are all positive !!! If you give me negative probability in exam, I will give you ZERO!!)

- 2) The sum of the probability of all outcomes must be 1.

- A **probability model** lists all of the possible outcomes and their corresponding probabilities

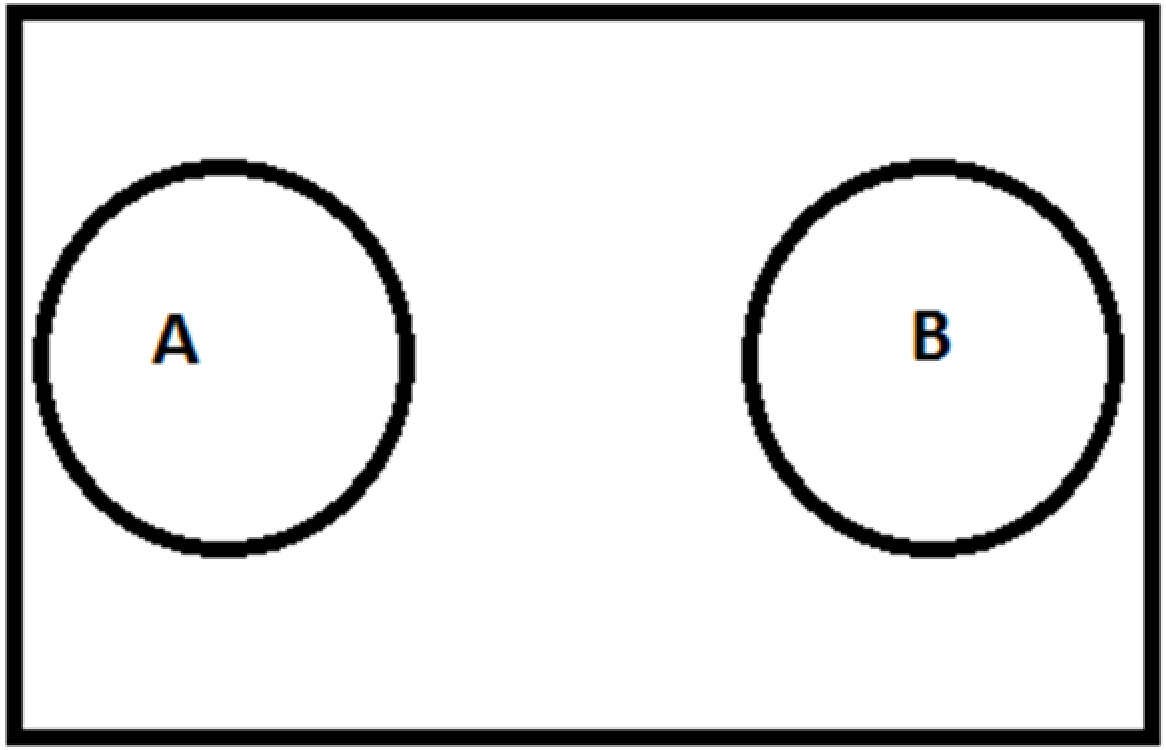
Adjectives for Events: Complement

- **The complement** of an event A is the set of all outcomes not in A



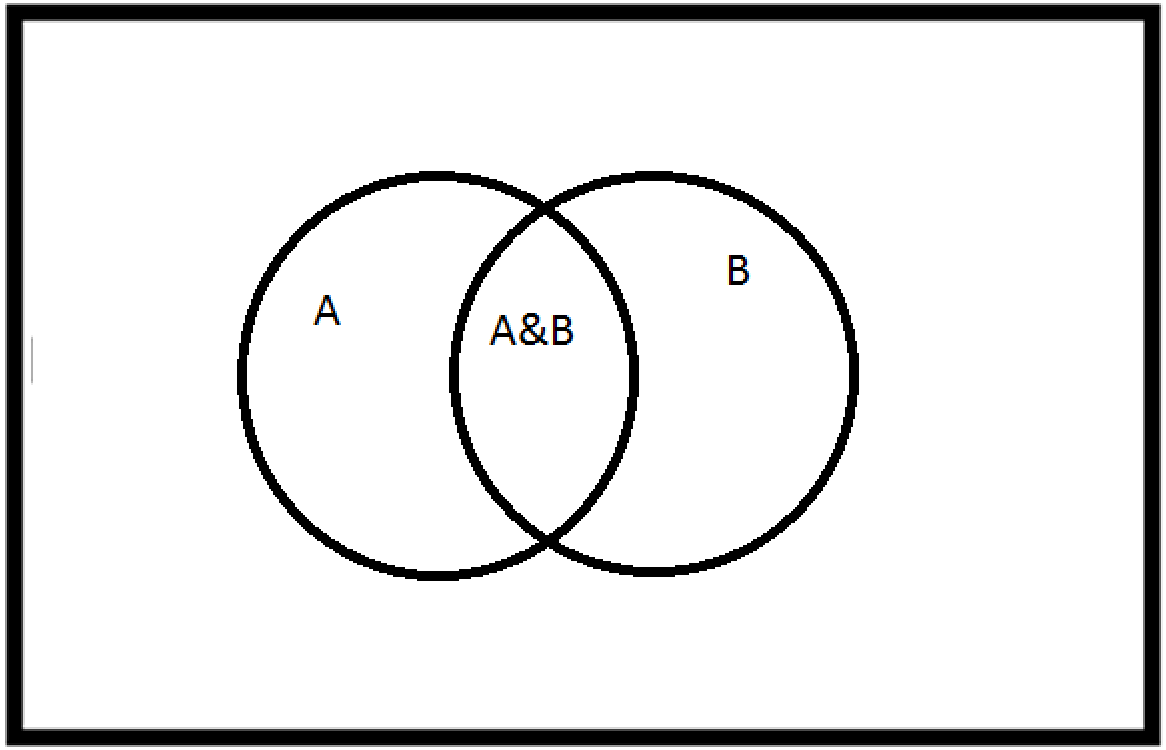
Adjectives for Events: Disjoint

- Two events are **disjoint** if they do not have any common outcomes



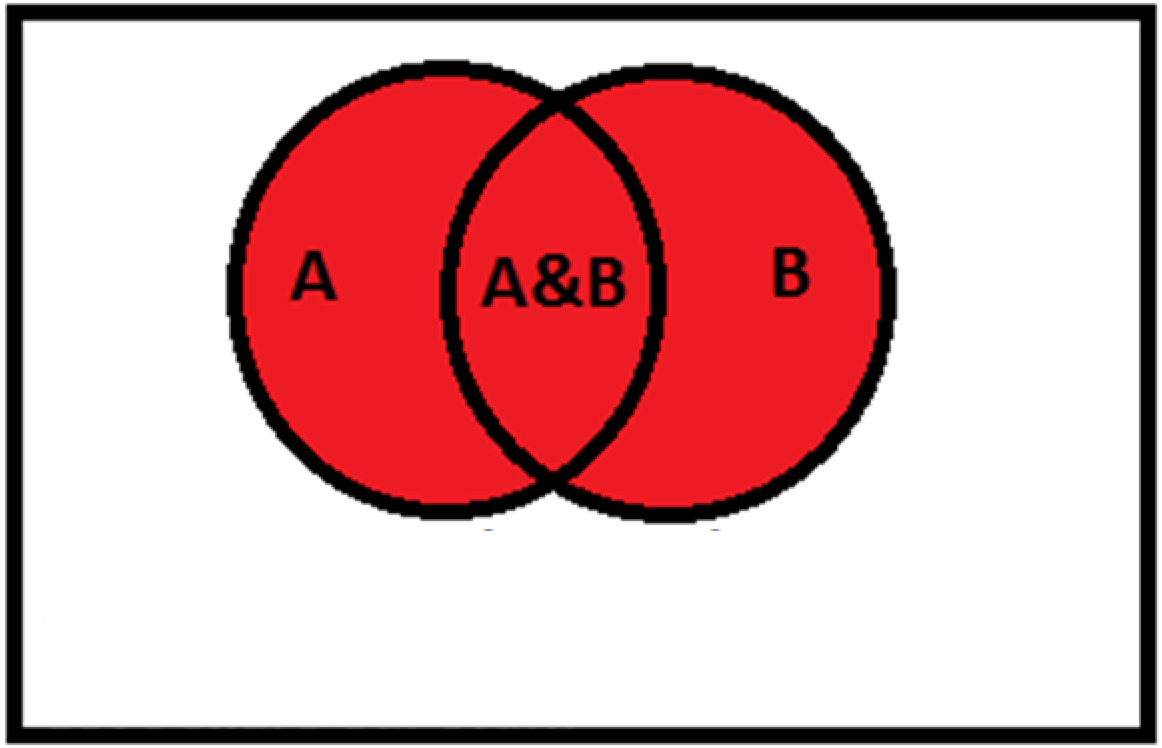
Adjectives for Events: Intersection

- The **intersection** of A and B (A&B) consists of outcomes that are both in A and B



Adjectives for Events: Union

- The **union** of A and B consists of outcomes that are in A or B

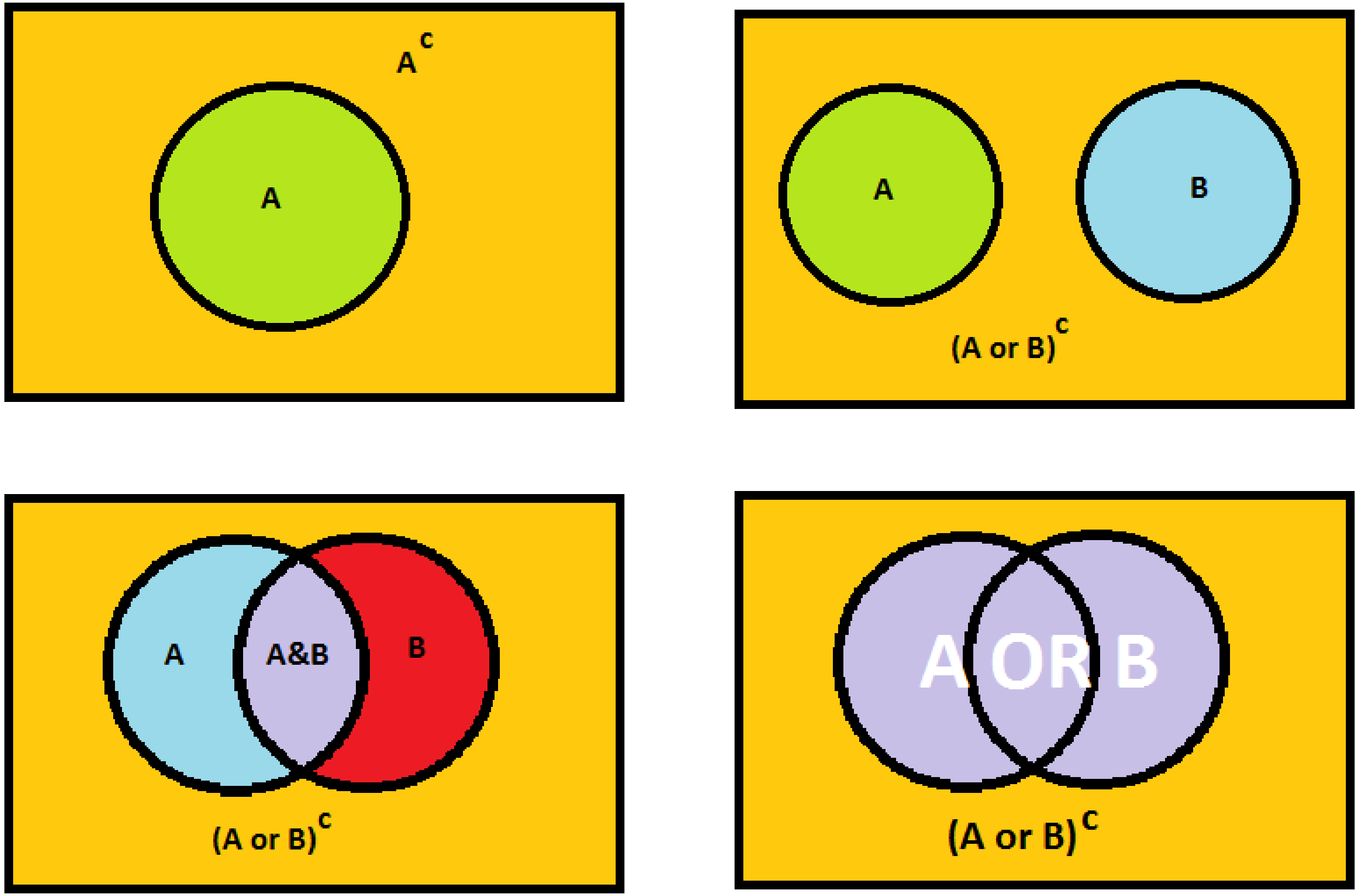


Remember

- **Outside of statistics:** "You can have soup or salad."
  - You can only have one (either soup or salad)

- **In statistics:** "You can have soup or salad."
  - You can have both!

Adjectives for Events: Summary



Probability Rules

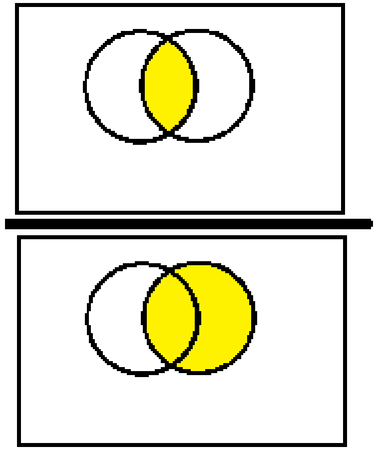
- **Complement Rule:**
  - The probability of something not happening is 1 minus the probability of it happening

$$P(A^c) = 1 - P(A)$$

Probability Rules

- **Conditional Probabilities**
  - The probability of event A given event B (probability of A conditional on B):

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

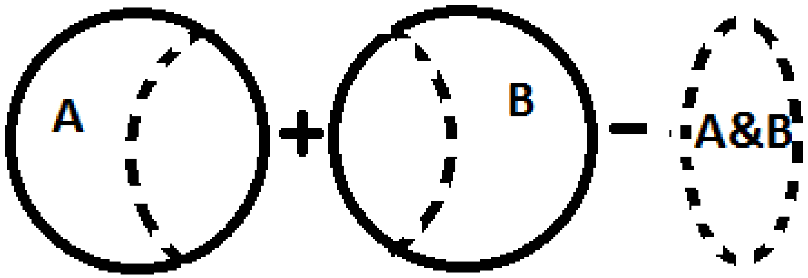


Transformation of Conditional Probabilities

- $\mathrm{P(A\ and\ B)} = \mathrm{P(A)} * \mathrm{P(B|A)}$
  - The probability of A and B happening is the probability of A times the probability of B given A
  - For **independent** events
    - $P(A\ and\ B) = P(A) * P(B)$
    - This is because $\mathrm{P(B|A)}$=P(B) when A and B are independent

Probability Rules

- $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$
  - The probability of A or B happening is the probability of A, plus the probability of B, minus the probability of them A & B because we're double counting the probability that they both happen
  - For disjoint events $P(A\ or\ B) = P(A) + P(B)$



“At Least” Probability

- Question: what is the probability that we roll at least a 5 when using a six sided dice.
  - X, a random variable, is the number rolled
  - Well, what is at least 5?
    - 5 and greater: 5 and 6
- The probability that we roll at least a five:

$$P(At\ least\ 5) = P(X \geq 5) = P(5) + P(6)$$
$$= \left(\frac{1}{6}\right) + \left(\frac{1}{6}\right) = \frac{1}{3}$$

To Check for Independence

- Two events, A and B, are independent if:
    1. If P(A|B) = P(A) or
    2. If P(B|A) = P(B) or
    3. If P(A and B) = P(A)*P(B)

  -**Note:** If any of these are true, the others are also true and the events A and B are independent

Disjoint and Independent

- Question: If events A and B are **independent**, does that mean they're **disjoint** as well?

- No! Disjoint means event A and event B share no common outcomes; while independent means the probability of A cannot be influenced by B.

- Example: A={1,3,5}, B={2,4,6} are disjoint but not independent

Other Types of Events

- **Impossible:** when the probability of the event occurring is zero
- **Certain:** when the probability of the event occurring is one
- **Unusual:** when the probability of the event occurring is low. We consider low to be less than 0.05.

Example: Probability

↓Wear Seat Belt?	Survived (S)	Died (D)	Total
Yes (Y)	412,368	510	412,878
No (N)	16,001	162,527	178,528
Total	428,369	163,037	591,406

- The probability a randomly selected participant wore seat belt:
- $P(Y) = \frac{number\ of\ Y\ observations}{total\ number\ of\ observations} = \frac{412{,}878}{591{,}406} = .69812954$

Example: Probability

↓Wear Seat Belt?	Survived (S)	Died (D)	Total
Yes (Y)	412,368	510	412,878
No (N)	16,001	162,527	178,528
Total	428,369	163,037	591,406

- The probability a randomly selected participant survived
- $P(S) = \frac{number\ of\ S\ observations}{total\ number\ of\ observations} = \frac{428{,}369}{591{,}406} = .72432305$

Example: Probability

↓Wear Seat Belt?	Survived (S)	Died (D)	Total
Yes (Y)	412,368	510	412,878
No (N)	16,001	162,527	178,528
Total	428,369	163,037	591,406

- The probability a randomly selected participant did not survive:
- $P(S^c) = \frac{number\ of\ D\ observations}{total\ number\ of\ observations} = \frac{163{,}037}{591{,}406} =$ .27567695

Example: Probability

↓Wear Seat Belt?	Survived (S)	Died (D)	Total
Yes (Y)	412,368	510	412,878
No (N)	16,001	162,527	178,528
Total	428,369	163,037	591,406

- The probability a randomly selected participant did not survive:
- Let's try the complement rule:
- $P(S^c) = 1 - P(S) = 1 - .72432305 = .27567695$

Example: Probability

↓Wear Seat Belt?	Survived (S)	Died (D)	Total
Yes (Y)	412,368	510	412,878
No (N)	16,001	162,527	178,528
Total	428,369	163,037	591,406

- The probability a randomly selected participant survived and he wore seat belt:
- $P(S\&Y) = \frac{number\ of\ S\&Y\ observations}{total\ number\ of\ observations} = \frac{412{,}368}{591{,}406} = .69726719$

Example: Probability

↓Wear Seat Belt?	Survived (S)	Died (D)	Total
Yes (Y)	412,368	510	412,878
No (N)	16,001	162,527	178,528
Total	428,369	163,037	591,406

- The probability a randomly selected participant survived given he wore seat belt:
- $P(S|Y) = \frac{number\ of\ S\&Y\ observations}{total\ number\ of\ Y\ observations} = \frac{412{,}368}{412{,}878} = .99876477$

Example: Probability

↓Wear Seat Belt?	Survived (S)	Died (D)	Total
Yes (Y)	412,368	510	412,878
No (N)	16,001	162,527	178,528
Total	428,369	163,037	591,406

- The probability a randomly selected participant survived given he wore seat belt:
- Let's try the conditional probability formula:
- $P(S|Y) = \frac{P(S\&Y)}{P(Y)} = \frac{.69726719}{.69812954} = .99876477$

Example: Probability

↓Wear Seat Belt?	Survived (S)	Died (D)	Total
Yes (Y)	412,368	510	412,878
No (N)	16,001	162,527	178,528
Total	428,369	163,037	591,406

- The probability a randomly selected participant survived and he wore seat belt:
- Let's try the formula:
- $P(Y\&S) = P(Y) * P(S|Y) = .69812954 * .998764777 = .69726719$

Example: Probability

↓Wear Seat Belt?	Survived (S)	Died (D)	Total
Yes (Y)	412,368	510	412,878
No (N)	16,001	162,527	178,528
Total	428,369	163,037	591,406

- The probability a randomly selected participant survived given he didn't wear seat belt:
- $P(S|N) = \frac{number\ of\ S\&N\ observations}{total\ number\ of\ N\ observations} = \frac{16{,}001}{178{,}528} = .0896274$

Example: Probability

↓Wear Seat Belt?	Survived (S)	Died (D)	Total
Yes (Y)	412,368	510	412,878
No (N)	16,001	162,527	178,528
Total	428,369	163,037	591,406

- The probability a randomly selected participant wore seat belt or survived:
- Let's try the formula:
- $P(S \text{ or } Y) = P(S) + P(Y) - P(S\&Y)$

$$= .72432305 + .69812954 - .69726719$$
$$= .7251854$$

Example: Probability

↓Wear Seat Belt?	Survived (S)	Died (D)	Total
Yes (Y)	412,368	510	412,878
No (N)	16,001	162,527	178,528
Total	428,369	163,037	591,406

- Question: Would it be worth wearing seat belt?
- $P(S|Y) = .99876477$
- $P(S|N) = .0896274$
- Your answer?

Example: Probability

↓Wear Seat Belt?	Survived (S)	Died (D)	Total
Yes (Y)	412,368	510	412,878
No (N)	16,001	162,527	178,528
Total	428,369	163,037	591,406

- Question: Would it be worth wearing seat belt?
- $P(D|Y) = 1 - .99876477 = 0.001235232$
- $P(D|N) = 1 - .0896274 = 0.9103726$
- Your answer?

Another Example: Should I trust my doctor?

- If Shiwen doubts that he has diabetes, he would probably visit doctor asking for diagnostics.
- In this scenario, there are two possibilities for the truth: he has diabetes or not; also, there are two possible answers from doctor: doctor believes he has diabetes or not.
- If he has diabetes and doctor believes diabetes, or if he has no diabetes and doctor believes no diabetes, doctor makes a correct conclusion. If not, doctor is wrong.

Example 2: Probability

	Positive Test (Pos)	Negative Test (Neg)	Total
Diabetes (D)	48	6	54
No Diabetes (D^c)	1307	3921	5228
Total	1355	3927	5282

- Let Event…
  - D = Has Diabetes
  - D^c= Has No Diabetes
  - Pos = Positive Test (Doctor believes diabetes)
  - Neg = Negative Test (Doctor believes no diabetes)

Example 2: Probability

	Positive Test (Pos)	Negative Test (Neg)	Total
Diabetes (D)	48	6	54
No Diabetes (D^c)	1307	3921	5228
Total	1355	3927	5282

- A **true positive** is when a participant tests positive for diabetes and does have it
- Here there are 48 true positives

Example 2: Probability

	Positive Test (Pos)	Negative Test (Neg)	Total
Diabetes (D)	48	6	54
No Diabetes (D^c)	1307	3921	5228
Total	1355	3927	5282

- A **true negative** is when a participant tests negative for the diabetes and doesn't have it
- Here there are 3,921 true negatives

Example 2: Probability

	Positive Test (Pos)	Negative Test (Neg)	Total
Diabetes (D)	48	6	54
No Diabetes (D^c)	1307	3921	5228
Total	1355	3927	5282

- A **false positive** is when a participant tests positive for the diabetes but doesn't have it
- Here there are 1,307 false positives

Example 2: Probability

	Positive Test (Pos)	Negative Test (Neg)	Total
Diabetes (D)	48	6	54
No Diabetes (D^c)	1307	3921	5228
Total	1355	3927	5282

- A **false negative** is when a participant tests negative for the diabetes but does have it
- Here there are 6 false negatives

Example 2: Probability

	Positive Test (Pos)	Negative Test (Neg)	Total
Diabetes (D)	48	6	54
No Diabetes (D^c)	1307	3921	5228
Total	1355	3927	5282

- The probability a randomly selected participant has a positive test
- $\mathrm{P}(Pos) = \frac{Number\ of\ positive\ observations}{Total\ number\ of\ observations} = \frac{1355}{5282} = .25653162$

Example 2: Probability

	Positive Test (Pos)	Negative Test (Neg)	Total
Diabetes (D)	48	6	54
No Diabetes (D^c)	1307	3921	5228
Total	1355	3927	5282

- The probability a randomly selected participant has diabetes:
- $P(D) = \frac{Number\ of\ D\ observations}{Total\ number\ of\ observations} = \frac{54}{5282} = .0102234$

Example 2: Probability

	Positive Test (Pos)	Negative Test (Neg)	Total
Diabetes (D)	48	6	54
No Diabetes (D^c)	1307	3921	5228
Total	1355	3927	5282

- The probability a randomly selected participant has diabetes given they tested positive:
- $P(D|Pos) = \frac{P(D\&Pos)}{P(Pos)} = \frac{(\frac{48}{5282})}{.25653162} = .03542435$

Example 2: Probability

- $P(D|Pos) = .03542435$
- $P(D) = .0102234$
- Because $P(D|Pos) \neq P(D)$ events D and Positive Test are not independent events

Definition for Tests

- **Sensitivity** – Probability that a test detects a disease correctly by giving the condition that the disease exists
  - $Sensitivity = P(Pos|Has\ disease) = P(Pos|D)$

- **Specificity** – Probability that a test correctly detect no disease by giving the condition that the disease does not exist
  - $Specificity = P(Neg|Has\ no\ disease) = P(Neg|D^c)$

Example 2 (cont.): Probability

	Positive Test (Pos)	Negative Test (Neg)	Total
Diabetes (D)	48	6	54
No Diabetes (D^c)	1307	3921	5228
Total	1355	3927	5282

- $Sensitivity = P(Pos|D) = \frac{P(Pos\ \&\ D)}{P(D)} = \frac{\frac{48}{5282}}{\left(\frac{54}{5282}\right)} = \frac{48}{54} = .8889$
- Probability that a participant tests positive given he has diabetes

Example 2 (cont.): Probability

	Positive Test (Pos)	Negative Test (Neg)	Total
Diabetes (D)	48	6	54
No Diabetes (D^c)	1307	3921	5228
Total	1355	3927	5282

- $Specificity = P(Neg|D^c) = \frac{P(Neg \,\&\, D^C)}{P(D^C)} = \frac{\frac{3921}{5282}}{\left(\frac{5228}{5282}\right)} = \frac{3921}{5228}$ = .75
- Probability that a participant tests negative given he has no diabetes

Should I believe my doctor?

- $Sensitivity = P(Pos|D) = \ .8889 = 88.89\%$
- $Specificity = P(Neg|D^c) = \ .75$ = 75%

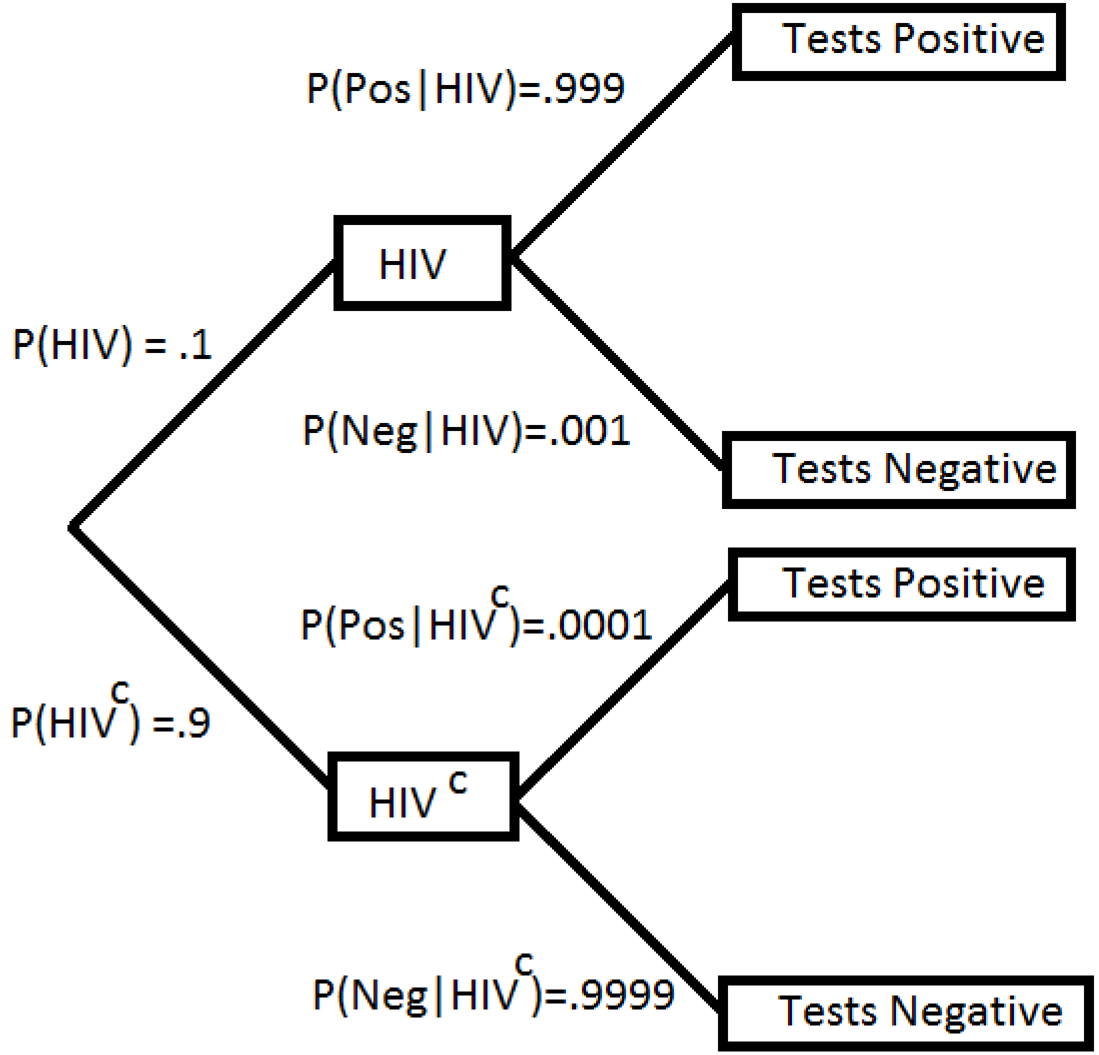
- What's your answer?

Tree Diagram

- For an HIV blood test, the sensitivity is about .999 and the specificity is about .9999

- Consider a high risk group here 10% truly have HIV

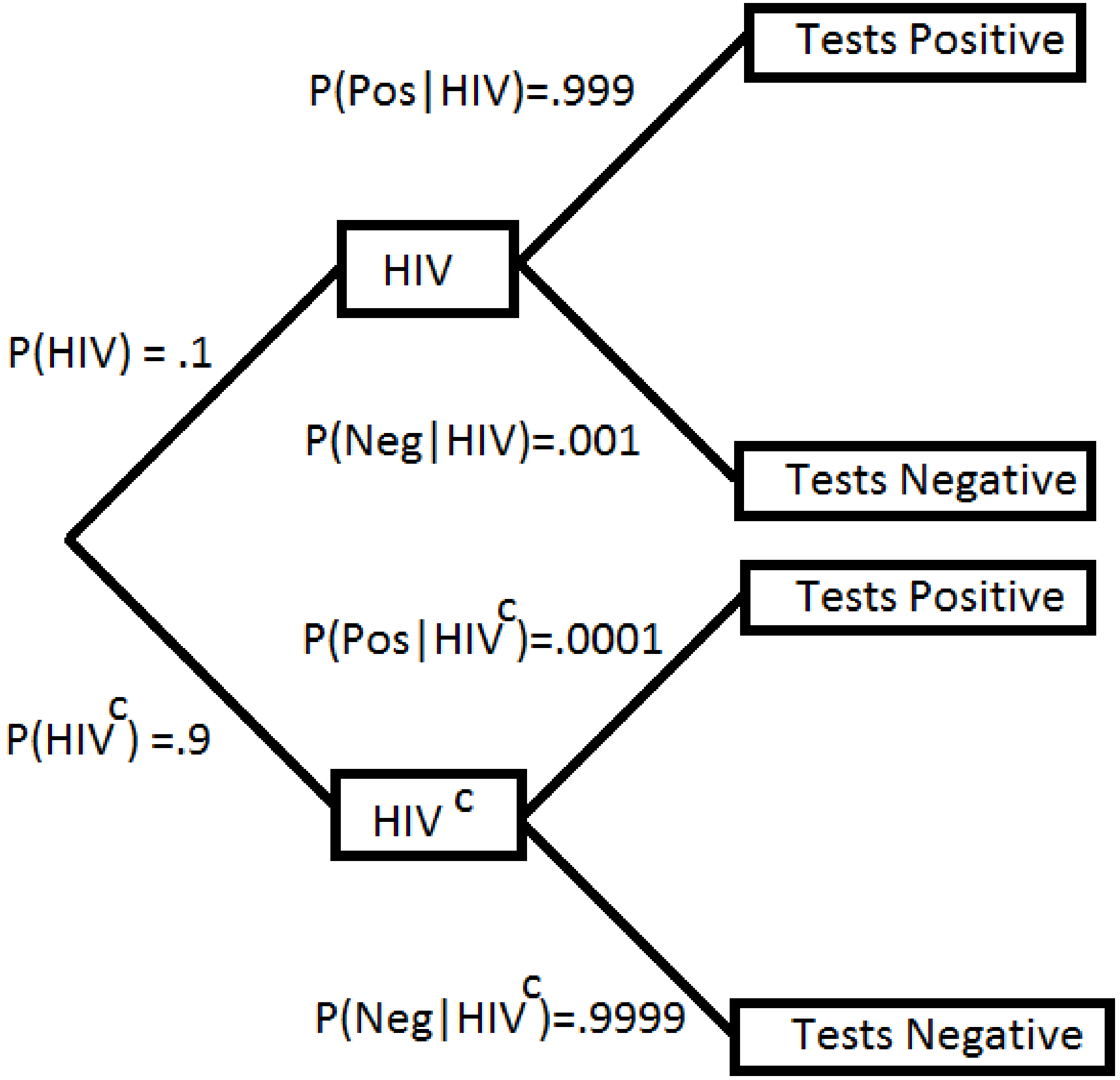
Tree Diagram



Tree Diagram

- The probability that a randomly selected participant tests positive and has HIV:

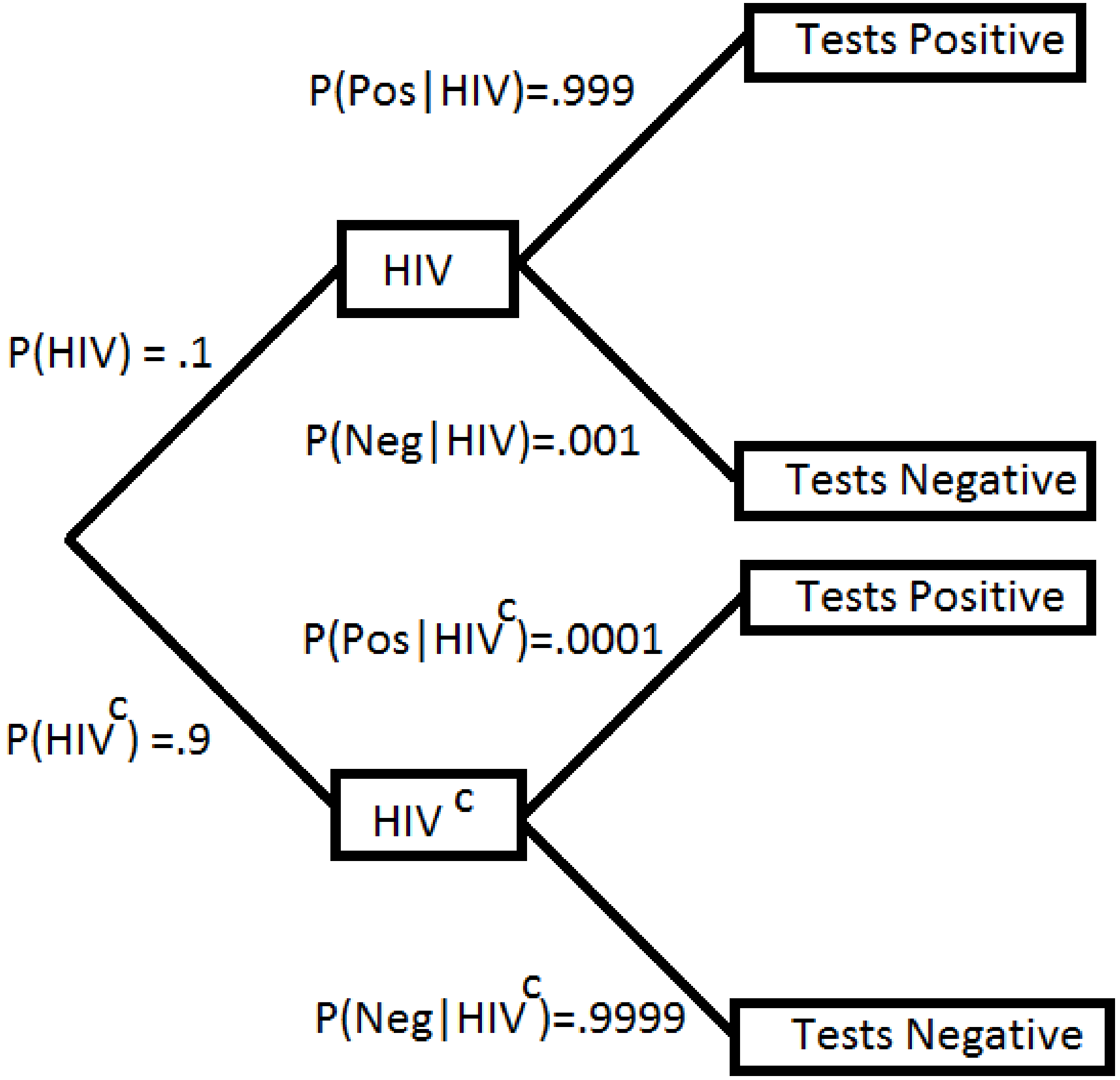
$$P(HIV\ and\ Pos) \\ = P(HIV) * P(Pos|HIV) \\ = .1 * .999 = .0999$$



Tree Diagram

- The probability that a randomly selected participant tests negative and has HIV:

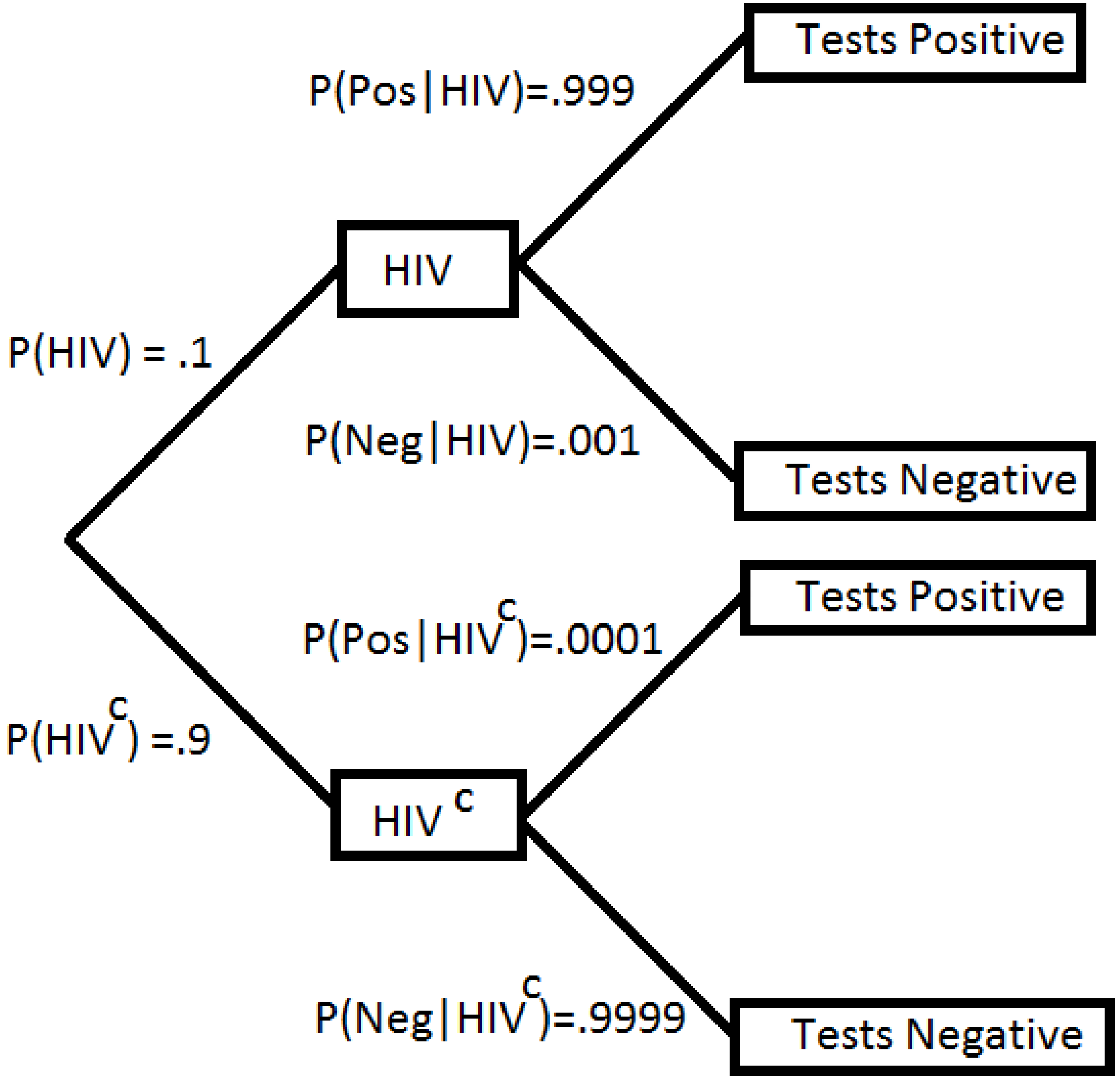
$$P(HIV and\ Neg)$$
$$= P(HIV) * P(Neg|HIV)$$
$$= .1 * .001 = .0001$$



Tree Diagram

- The probability that a randomly selected participant tests positive and doesn't have HIV:

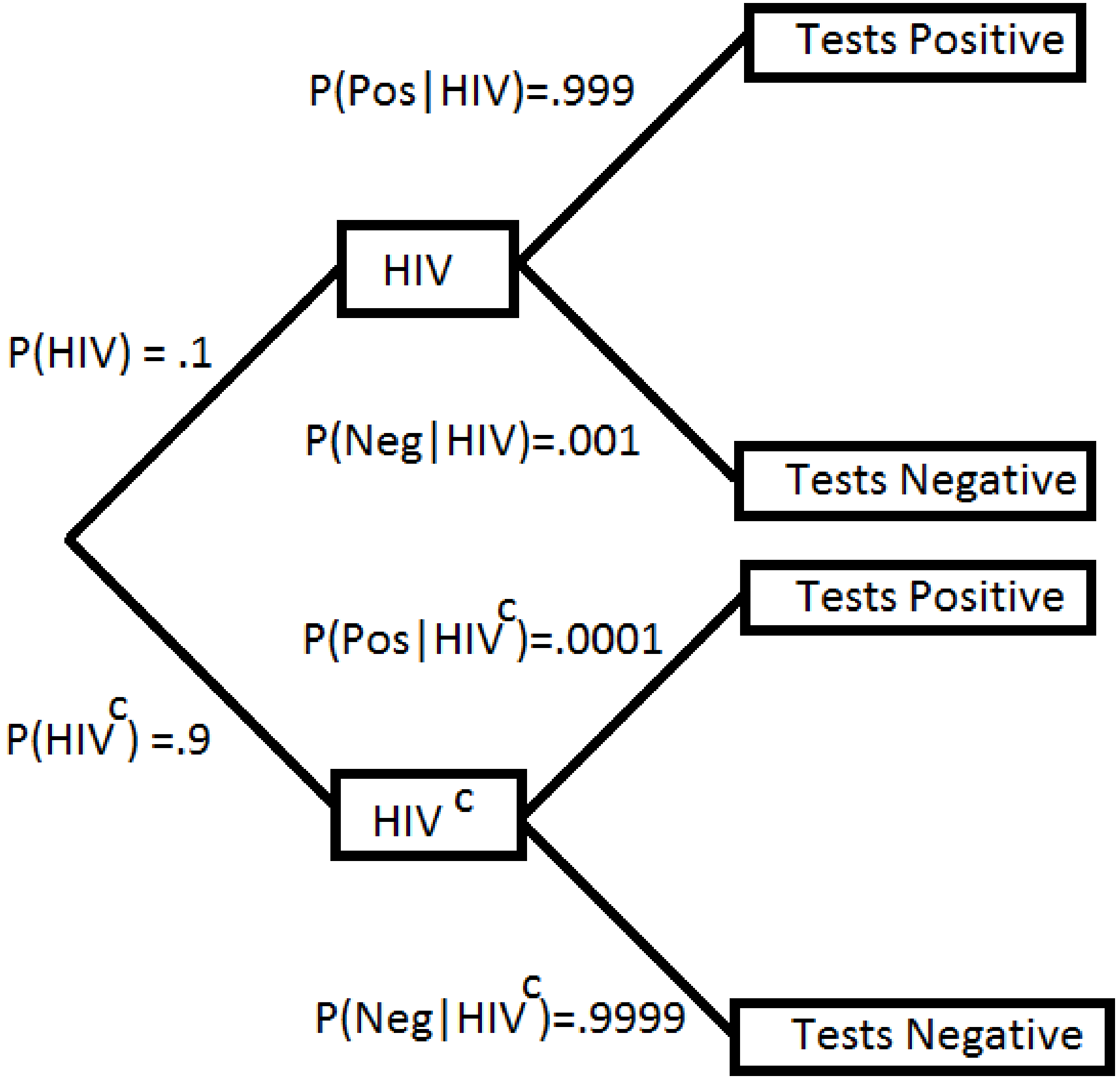
$$P(HIV^c and\ Pos)$$
$$= P(HIV^c) * \mathrm{P(Pos|HIV^c)}$$
$$= .9 * .0001 = .00009$$



Tree Diagram

- The probability that a randomly selected participant tests negative and doesn't have HIV:

$$P(HIV^c and\ Neg)$$
$$= P(HIV^c) * \mathrm{P(Neg|HIV^c)}$$
$$= .9 * .9999 = .89991$$



What Does the Table Look like?

	Pos	**Neg**	**Totals**
HIV	P(Pos&HIV)	P(Neg&HIV)	P(HIV)
HIV^c	P(Pos&HIV^c)	P(Neg&HIV^c)	P(HIV^c)
Totals	P(Pos)	P(Neg)	1

What Does the Table Look like?

	Pos	Neg	Totals
HIV	0.0999	0.0001	0.1
HIV^c	0.00009	0.89991	0.9
Totals	0.09999	0.90001	1

Should I believe my doctor?

- The probability a randomly selected person has HIV given they tested positive
  - $P(HIV|Pos) = \frac{P(HIV\&Pos)}{P(Pos)} = \frac{.0999}{.09999} = .9991$

- Since this is very close to one, if you test positive, it is very likely that you have HIV

Example: Flowers

- 9 flower seeds: 4 are red and 5 are white
- Choose 2 seeds at random:
  - One at a time **without replacement**
  - **Without replacement** means we don’t put our first choice back
  - In other words, you choose one seed, plant it and it’s gone forever, and then you choose another from the sack

Example: Flowers

- The first choice is from nine seeds
  - 4 are red
  - 5 are white

Example: Flowers

- The second choice is from eight seeds because we chose one **without replacement**

Example: Flowers

- The second choice is from eight seeds because we chose one **without replacement**
- If the first was red:
  - 3 are red
  - 5 are white

Example: Flowers

- The second choice is from eight seeds because we chose one **without replacement**
- If the first was white:
  - 4 are red
  - 4 are white

Example: Flowers

- The probability of selecting a red on our first try $P(Red_1) = \frac{4}{9}$

Example: Flowers

- The probability of selecting a white on our first try $P(White_1) = \frac{5}{9}$

Example: Flowers

- The probability of selecting a red on our second try given we got a red on our first try
  - We started with nine seeds and we selected one without replacement, so now we have eight seeds
  - We started with four red seeds and we selected a red on our first try, so now we have three red seeds

$$P(Red_2|Red_1) = \frac{3}{8}$$

Example: Flowers

- The probability of selecting a white on our second try given we got a red on our first try
  - We started with nine seeds and we selected one without replacement, so now we have eight seeds
  - We started with five white seeds and we selected a red on our first try, so we still have five white seeds

$$P(White_2|Red_1) = \frac{5}{8}$$

Example: Flowers

- The probability of selecting a red on our second try given we got a white on our first try
  - We started with nine seeds and we selected one without replacement, so now we have eight seeds
  - We started with four red seeds and we selected a white on our first try, so we still have four red seeds

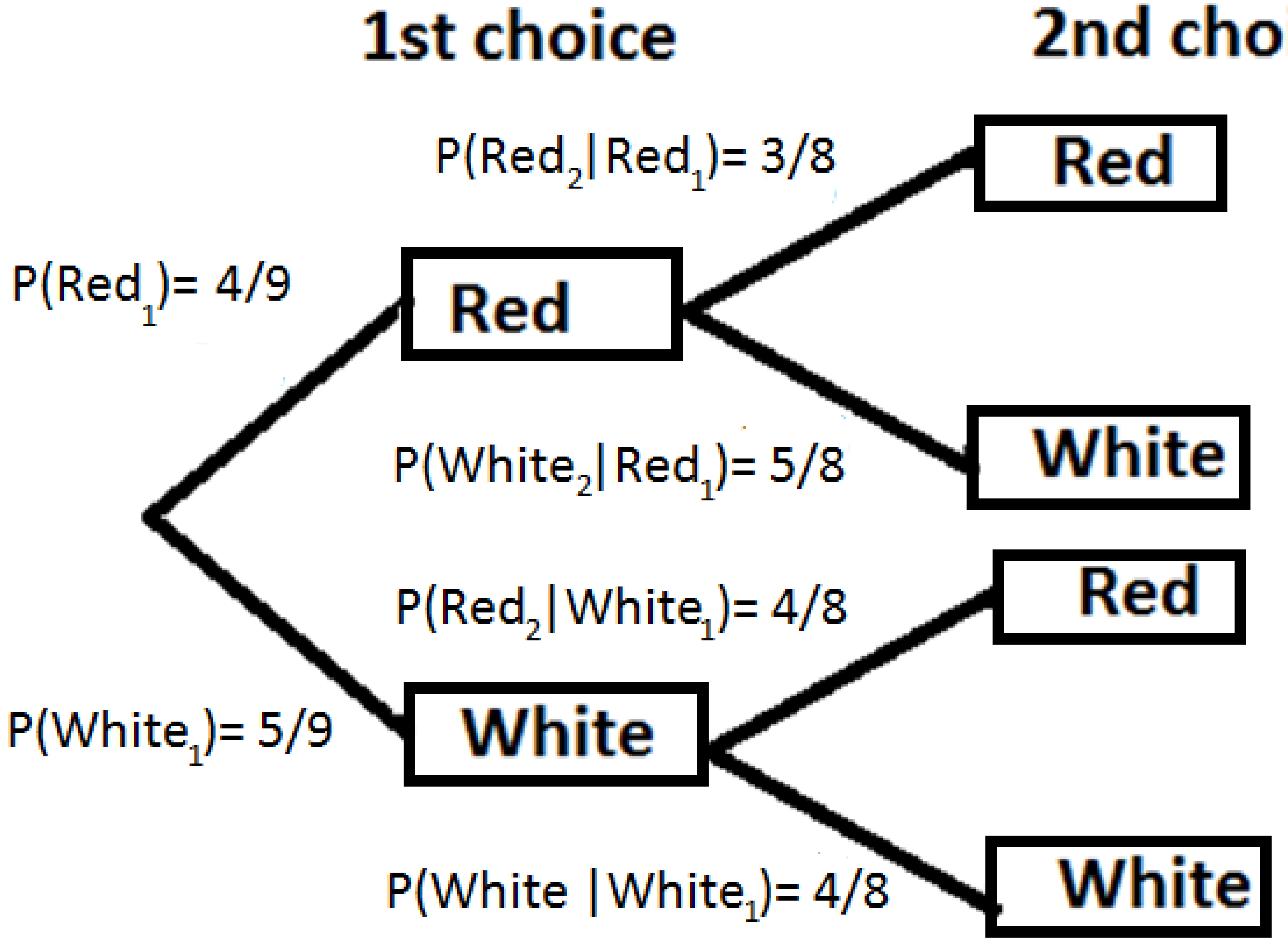
$$P(Red_2|White_1) = \frac{4}{8}$$

Example: Flowers

- The probability of selecting a white on our second try given we got a white on our first try
  - We started with nine seeds and we selected one without replacement, so now we have eight seeds
  - We started with five white seeds and we selected a white on our first try, so now we have four white seeds

$$P(White_2|White_1) = \frac{4}{8}$$

Example: Flowers



Example: Flowers

- To find 'and' probabilities we just multiply across the branches
- Remember $\mathrm{P(A\ and\ B)} = \mathrm{P(A)} * \mathrm{P(B|A)}$
- The probability of choosing two red seeds in a row

$$P(R_1\ and\ R_2) = P(R_1) * P(R_2|R_1)$$
$$= \left(\frac{4}{9}\right) * \left(\frac{3}{8}\right) = \left(\frac{1}{6}\right)$$

Example: Flowers

- To find 'and' probabilities we just multiply across the branches
- Remember $\mathrm{P(A\ and\ B) = P(A) * P(B|A)}$
- The probability of choosing a red seed first, and then a white seed

$$P(R_1\ and\ W_2) = P(R_1) * P(W_2|R_1)$$
$$= \left(\frac{4}{9}\right) * \left(\frac{5}{8}\right) = \left(\frac{5}{18}\right)$$

Example: Flowers

- To find 'and' probabilities we just multiply across the branches
- Remember $\mathrm{P(A\ and\ B)} = \mathrm{P(A)} * \mathrm{P(B|A)}$
- The probability of choosing a white seed first, and then a red seed

$$P(W_1\ and\ R_2) = P(W_1) * P(R_2|W_1)$$
$$= \left(\frac{5}{9}\right) * \left(\frac{4}{8}\right) = \left(\frac{5}{18}\right)$$

Example: Flowers

- To find ‘and’ probabilities we just multiply across the branches
- Remember $\mathrm{P(A\ and\ B)} = \mathrm{P(A)} * \mathrm{P(B|A)}$
- The probability of choosing two white seeds in a row

$$P(W1\ and\ W2) = P(W_1) * P(W_2|W_1)$$
$$= \left(\frac{5}{9}\right) * \left(\frac{4}{8}\right) = \left(\frac{5}{18}\right)$$

Example: Flowers

- To find ‘and’ probabilities we just multiply across the branches
- Remember $\mathrm{P(A\ and\ B)} = \mathrm{P(A)} * \mathrm{P(B|A)}$
- The probability of choosing a white and a red, regardless of order

$$P(1\ red\ \&\ 1\ white) = P(R_1\&W_2\ or\ W_1\&R_2)$$
$$= P(R_1\&W_2) + P(W_1\&R_2)$$
$$= \left(\frac{5}{18}\right) + \left(\frac{5}{18}\right) = \left(\frac{10}{18}\right) = \left(\frac{5}{9}\right)$$

- Can we use complement rule to calculate?

Example: Flowers

- What if we choose two flower seeds **with replacement**?

- Two trials are independent!

Multiplication Rule of Counting

- If a task consists of a sequence of choices in which there are p selections for the first choice, q selections for the second choice, r selections for the third choice and so forth
- Then, together, the task can be done in (p*q*r*…) different ways

Multiplication Rule of Counting w/ Replacement

- A South Carolina license plate is three letters followed by three numbers. How many unique plates can they make in this format?
- 26 letters in the alphabet
- 10 single digit numbers
- 26*26*26*10*10*10 = 17,576,000 plates

Multiplication Rule of Counting w/o Replacement

- We do not allow repeats for letters and numbers. How many unique plates can they make in this format?
- 26 letters in the alphabet
- 10 single digit numbers
- 26*25*24*10*9*8 = 11,232,000 plates. Notice this is a lot less than before!

Factorial

- A **factorial** of a number $n \geq 0$ is defined as
- n! = n*(n-1)*(n-2)*…*3*2*1

- n! = n* (n-1)!

- 0! = 1

Permutation

- A **permutation** is an ordered arrangement in which r objects are chosen from n distinct objects.
  - $r \leq n$
  - No repetition
- How many permutation exists?

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

Permutation

- For example, we have a set of three letters: A, B, and C.
- We might ask how many ways we can arrange 2 letters from that set. Each possible arrangement would be an example of a permutation.
- The complete list of possible permutations would be
- {AB, BA, AC, CA, BC, CB}
- ${}_3P_2 = \frac{3!}{(3-2)!} = \frac{3!}{1!} = \frac{3*2*1}{1} = \frac{6}{1} = 6$

Combination

- A **combination** is an unordered arrangement in which r objects are chosen from n distinct objects.
  - $r \leq n$
  - No repetition
- How many permutation exists?

$$ {}_{n}C_{r} = \frac{n!}{r!\,(n-r)!} $$

Combination

- For example, suppose we have a set of three letters: A, B, and C.
- We might ask how many ways we can select 2 letters from that set. Each possible selection would be an example of a combination.
- The complete list of possible selections is
- {AB, AC, and BC}
- ${}_3C_2 = \frac{3!}{2!(3-2)!} = \frac{3*2*1}{(2*1)*1} = \frac{6}{2} = 3$