APPM 4570/5570

Unit #2: Probability Theory

(Ch 3.1, 3.2, 3.3)

Why Probability Theory?

- One main objective of statistics/data science is to help make good decisions under conditions of **uncertainty** or **chance**.
  - *Example:* In trying to determine how prevalent a certain disease is in the population, we examine a sample of the population for the disease. The inference from sample to population is uncertain.

- **Probability Theory** is one way to quantify outcomes that cannot be predicted with certainty.

Sample Space

- Definition: A **probabilistic process** is a system or experiment whose outcome is uncertain, i.e. what is the outcome of flipping a coin?

- Definition: An **outcome** is a possible result of a probabilistic process

- Definition: A **sample space** of a probabilistic process is the **set** of *all possible outcomes* of that process, typically denoted by $\mathcal{S}$ or Ω.

Set Theory

- A set is a collection of objection or elements, denoted A={set description}
- Roster Notation for a set: Simply list the elements of the set

$$E = \{0, 2, 4, 6, \ldots\} = \text{set of non-negative even integers}$$

- Set builder notation: Give a description of the elements that make up the set:

$$E = \{\text{all integers } n \geq 0 \mid n \text{ is even}\}$$

Sample Space

- Definition: A **probabilistic process** is a system or experiment whose outcome is uncertain, i.e. what is the outcome of flipping a coin?

- Definition: An **outcome** is a possible result of a probabilistic process

- Definition: A **sample space** of a probabilistic process is the **set** of *all possible outcomes* of that process, typically denoted by $\mathcal{S}$ or Ω.

Set Theory

- A **set** is a collection of objection or elements, denoted A={set description}
- Roster Notation for a set: Simply list the elements of the set

$$E = \{0, 2, 4, 6, \ldots\} = \text{set of non-negative even integers}$$

- Set builder notation: Give a description of the elements that make up the set:

$$E = \{\text{all integers } n \geq 0 \mid n \text{ is even}\}$$

Sample Space - Example

- Sample spaces for:
  - Tossing a coin: $S = \{ H, T \}$, **cardinality** is $|S| = 2$

Sample Space - Example

- Sample spaces for:
  - Tossing a coin: $S = \{ H, T \}$, **cardinality** is $|S| = 2$
  - Selecting a card from a standard deck:

$$S = \{2D, 2C, 2H, 2S, \ldots, AD, AC, AH, AS\},\ |S| = 52$$

Sample Space - Example

- Sample spaces for:
  - Tossing a coin: $S = \{ H, T \}$, **cardinality** is $|S| = 2$
  - Selecting a card from a standard deck:

    S = {2D,2C,2H,2S,…,AD,AC,AH,AS}, $|S|=52$

  - Measuring the commuting time on a particular morning:

    $S = \{ t \geq 0 \mid t \text{ is a real number}\}$

  (Here S is **uncountable**.)

Events

•Definition: An **event** is any collection (subset) of outcomes from the sample space S.

•An event is **simple** if it consists of exactly one outcome and **compound** if it consists of more than one outcome.

•When an experiment is performed, a particular event (let's call that event *A)* is said to occur if the resulting experimental outcome is contained in *A.*

Combining Events

- Given events *A* and *B* we can create new events like the event that "*A* <u>or</u> *B* happens" or "*A* <u>and</u> *B* happens" or "*A* does not happen"
- <u>*Definition:*</u>

1. The <u>**union**</u> of two events *A* and *B*, denoted by $A \cup B$ and read "*A or B*," is the event consisting of all outcomes that are ***either in A or in B or <u>in both</u>*** (called an "inclusive or") that is, all outcomes in at least one of the events.

Combining Events

- Given events *A* and *B* we can create new events like the event that "*A* <u>or</u> *B* happens" or "*A* <u>and</u> *B* happens" or "*A* does not happen"
- <u>*Definition:*</u>

1. The <u>**union**</u> of two events *A* and *B*, denoted by $A \cup B$ and read "*A or B*," is the event consisting of all outcomes that are ***either in A or in B or <u>in both</u>*** (called an "inclusive or") that is, all outcomes in at least one of the events.

2. The <u>**intersection**</u> of two events *A* and *B*, denoted by $A \cap B$ and read "*A and B*," is the event consisting of all outcomes that are in ***both A and B***.

Combining Events

- Given events *A* and *B* we can create new events like the event that "*A* or *B* happens" or "*A* and *B* happens" or "*A* does not happen"
- *Definition:*

1. The **union** of two events *A* and *B*, denoted by $A \cup B$ and read "*A or B*," is the event consisting of all outcomes that are ***either in A or in B or in both*** (called an "inclusive or") that is, all outcomes in at least one of the events.

2. The **intersection** of two events *A* and *B*, denoted by $A \cap B$ and read "*A and B*," is the event consisting of all outcomes that are in ***both A and B***.

3. The **complement** of an event *A*, denoted by A' (or A^c), is the set of all outcomes in $\mathcal{S}$ that are not contained in *A*.

The Empty Set

- Sometimes sets *A* and *B* have no outcomes in common, so that the intersection of sets *A* and *B* is empty.

Definition:

- Let Ø ={ } denote the empty set or ***null event*** (the event consisting of no outcomes whatsoever).

The Empty Set

- Sometimes sets *A* and *B* have no outcomes in common, so that the intersection of sets *A* and *B* is <u>empty</u>.

<u>*Definition:*</u>

- Let $\emptyset$ ={ } denote the <u>empty set</u> or <u>*null event*</u> (the event consisting of no outcomes whatsoever).

- When $A \cap B = \emptyset$, then *A* and *B* are said to be **mutually exclusive events** or **disjoint events** and there is <u>no chance</u> of the event "*A* and *B*" occurring.

The Empty Set

- Sometimes sets *A* and *B* have no outcomes in common, so that the intersection of sets *A* and *B* is <u>empty</u>.

<u>Definition:</u>

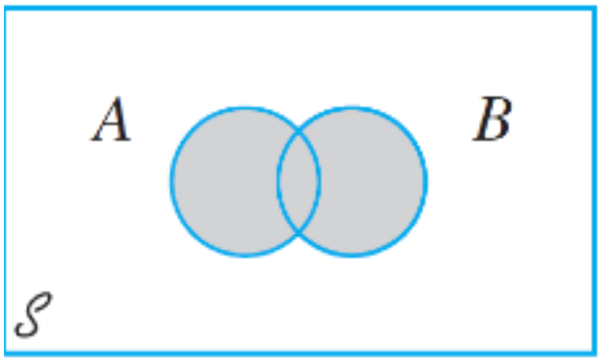
- Let $\emptyset$ ={ } denote the <u>empty set</u> or *<u>null event</u>* (the event consisting of no outcomes whatsoever).

- When $A \cap B = \emptyset$, then *A* and *B* are said to be **mutually exclusive events** or **disjoint events** and there is <u>no chance</u> of the event "*A* and *B*" occurring.

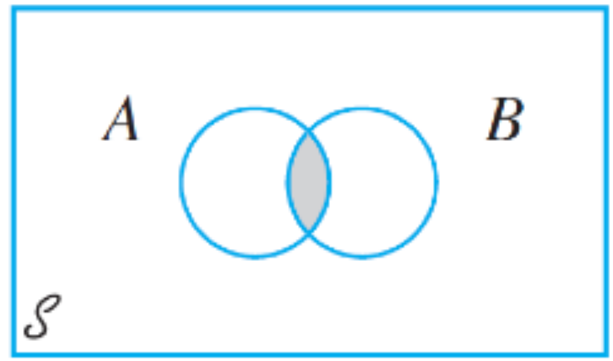
- ***Example:*** Roll a dice once, what's the chance of getting an even number and an odd number?

  *This is a **null event** so there is <u>no chance,</u> i.e. 0 probability!*

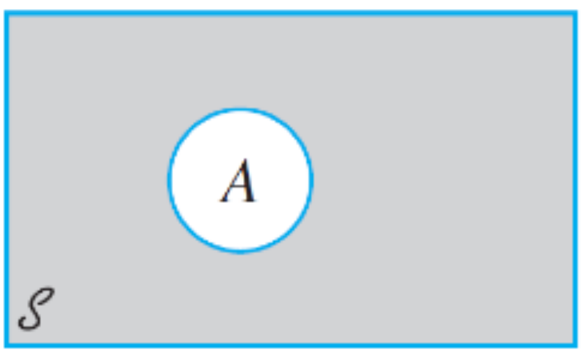
Set Operations



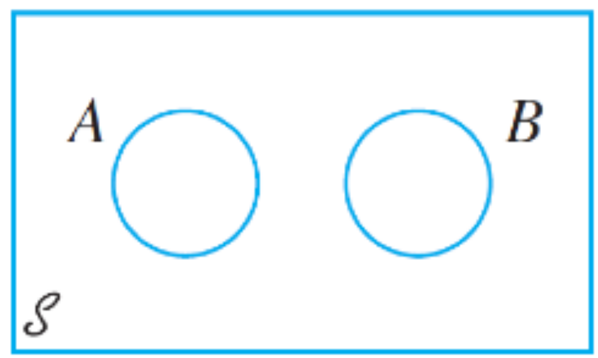
$A \cup B$



$A \cap B$



A' or A^c



$A \cap B = \emptyset$

Events Example

- Flip a coin 3 times.

(a) The sample space is

$$S = \{HHH, THH, HTH, HHT, TTH, THT, HTT, TTT\} \text{ and } |S|=8$$

Events Example

- Flip a coin 3 times.

(a) The sample space is

S = {HHH, THH, HTH, HHT, TTH, THT, HTT,TTT} and $|S|=8$

(b) Some possible events include:

A = getting all heads = {HHH} and

B = getting more than one tail = {TTH, THT, HTT, TTT}

Events Example

- Flip a coin 3 times.

(a) The sample space is

$S = \{HHH, THH, HTH, HHT, TTH, THT, HTT, TTT\}$ and $|S|=8$

(b) Some possible events include:

A = getting all heads = $\{HHH\}$ and

B = getting more than one tail = $\{TTH, THT, HTT, TTT\}$

(c) Note that

$A \cup B = \{HHH, TTH, THT, HTT, TTT\}$ = getting all heads **or** more than one tail

Events Example

- Flip a coin 3 times.

(a) The sample space is

S = {HHH, THH, HTH, HHT, TTH, THT, HTT,TTT} and $|S|$=8

(b) Some possible events include:

A = getting all heads = {HHH} and

B = getting more than one tail = {TTH, THT, HTT, TTT}

(c) Note that

$A \cup B$ = {HHH, TTH, THT, HTT, TTT} = getting all heads **or** more than one tail

(d) And $A \cap B$ = getting all heads **and** more than one tail = { } = Ø

More Set Theory

(★) For any three events *A*, *B*, and *C*,

$A \cup B \cup C$ = *"at least one of the events A <u>or</u> B <u>or</u> C happens"*

and,

$A \cap B \cap C$ = *"event that A <u>and</u> B <u>and</u> C happens"*

(★) **DeMorgans Laws**: Note that from the Venn diagram we see that

$$(A \cap B)^c = A^c \cup B^c$$

That is *"events A <u>and</u> B do not happen"* is the same as *"A does not happen <u>or</u> B does not happen or both"*

And, similarly

$$(A \cup B)^c = A^c \cap B^c$$

(★) Finally note that we have $(A^c)^c = A$

More Set Theory

(★) For any three events A, B, and C,

$A \cup B \cup C$ = *"at least one of the events A <u>or</u> B <u>or</u> C happens"*

and,

$A \cap B \cap C$ = *"event that A <u>and</u> B <u>and</u> C happens"*

More Set Theory

(★) For any three events *A*, *B*, and *C*,

$A \cup B \cup C$ = *"at least one of the events A <u>or</u> B <u>or</u> C happens"*

and,

$A \cap B \cap C$ = *"event that A <u>and</u> B <u>and</u> C happens"*

(★) **DeMorgans Laws**: Note that from the Venn diagram we see that

$$(A \cap B)^c = A^c \cup B^c$$

That is *"events A <u>and</u> B do not happen"* is the same as *"A does not happen <u>or</u> B does not happen or both"*

And, similarly

$$(A \cup B)^c = A^c \cap B^c$$

More Set Theory

(★) For any three events *A*, *B*, and *C*,

$A \cup B \cup C$ = *"at least one of the events A <u>or</u> B <u>or</u> C happens"*

and,

$A \cap B \cap C$ = *"event that A <u>and</u> B <u>and</u> C happens"*

(★) **DeMorgans Laws**: Note that from the Venn diagram we see that

$$(A \cap B)^c = A^c \cup B^c$$

That is *"events A <u>and</u> B do not happen"* is the same as *"A does not happen <u>or</u> B does not happen or both"*

And, similarly

$$(A \cup B)^c = A^c \cap B^c$$

(★) Finally note that we have $(A^c)^c = A$

Rules (Axioms) of Probability

Given an experiment and a sample space $\mathcal{S}$, *the objective of probability theory is to assign to each set/event A, a number P(A), called the probability of the event A*, which quantifies how likely it is that A will occur.

The probability must satisfy the following assumptions:

Axiom 1: For any event A, we have $0 \leq P(A) \leq 1$.

Rules (Axioms) of Probability

Given an experiment and a sample space $\mathcal{S}$, *the objective of <u>probability theory</u> is to assign to each set/event A, a number P(A), called <u>the probability of the event A</u>*, which quantifies how likely it is that A will occur.

The probability must satisfy the following assumptions:

Axiom 1: For any event A, we have $0 \leq P(A) \leq 1$.

Axiom 2: $P(\mathcal{S}) = 1$ (the probability of the entire sample space is 1)

Rules (Axioms) of Probability

Given an experiment and a sample space $\mathcal{S}$, *the objective of* *<u>probability theory</u>* *is to assign to each set/event A, a number P(A), called* *<u>the probability of the event A</u>*, which quantifies how likely it is that A will occur.

The probability must satisfy the following assumptions:

Axiom 1: For any event A, we have $0 \leq P(A) \leq 1$.

Axiom 2: $P(\mathcal{S}) = 1$ (the probability of the entire sample space is 1)

Axiom 3: If $A_1, A_2, \ldots, A_n$ is any collection of <u>disjoint</u> events then:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

Some Theorems of Probability

Law of Complements:

For any event A,

$$P(A) + P(A^c) = 1, \text{ so } P(A) = 1 - P(A^c)$$

Some Theorems of Probability

Law of Complements:

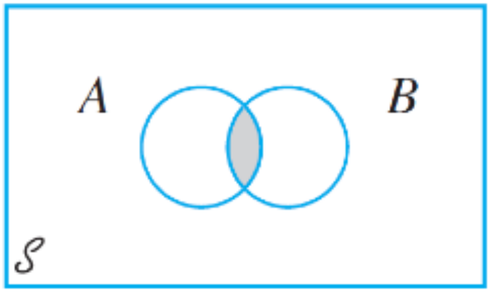
For any event *A*,

$$P(A) + P(A^c) = 1, \text{ so } P(A) = 1 - P(A^c)$$

Inclusion-Exclusion Principle:

For any sets A and *B*,

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



Some Theorems of Probability

Law of Complements:

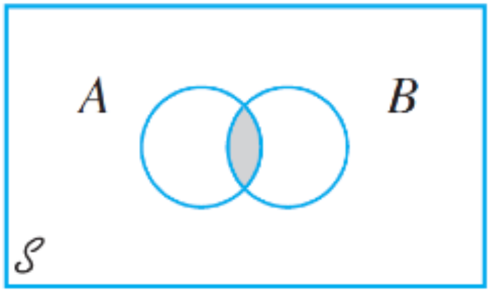
For any event *A*,

$$P(A) + P(A^c) = 1, \text{ so } P(A) = 1 - P(A^c)$$

Inclusion-Exclusion Principle:

For any sets A and *B*,

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



And, similarly, $P(A \cup B \cup C) = P(A) + P(B) + P(C)$

$$- P(A \cap B) - P(A \cap C) - P(B \cap C)$$

$$+ P(A \cap B \cap C)$$

Independent Events

Events A and B are said to be independent if the probability of event A occurring has no effect on event B occurring.

Independent Events

Events A and B are said to be <u>independent</u> if the probability of event A occurring has no effect on event B occurring.

We can mathematically describe two events being independent in the following way:

<u>Proposition:</u> Events *A* and *B* are **independent events** if and only if

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

Independent Events

Events A and B are said to be independent if the probability of event A occurring has no effect on event B occurring.

We can mathematically describe two events being independent in the following way:

Proposition: Events A and B are **independent events** if and only if

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

For example, if we flip a fair coin twice, the probability that the first flip is a head and the second flip is a head, using independence, is

$$P(H_1 \cap H_2) = P(H_1) \cdot P(H_2) = \frac{1}{2} \cdot \frac{1}{2} = 1/4$$

Since the outcome of the first flip does not effect the outcome of the second flip these events are independent.

Independent Events

Events A and B are said to be <u>independent</u> if the probability of event A occurring has no effect on event B occurring.

We can mathematically describe two events being independent in the following way:

<u>Proposition:</u> Events *A* and *B* are **independent events** if and only if

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

For example, if we flip a fair coin twice, the probability that the first flip is a head <u>and</u> the second flip is a head, <u>using independence</u>, is

$$P(H_1 \cap H_2) = P(H_1) \cdot P(H_2) = \frac{1}{2} \cdot \frac{1}{2} = 1/4$$

Since the outcome of the first flip does not effect the outcome of the second flip these events are independent.

(**Caution!** **Mutually exclusive** events and **independent events** are <u>not</u> the same!)

Independent Events vs. Mutually Exclusive

(★) Events A and B are **independent events** if and only if the probabilities P(A), P(B) and $P(A \cap B)$ satisfy the equation

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

Independent Events vs. Mutually Exclusive

(★) Events A and B are **independent events** if and only if the probabilities P(A), P(B) and P($A \cap B$) satisfy the equation

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

(★) Recall that events A and B are said to be **mutually exclusive** or **disjoint** if the intersection of A and B is empty, that is

$$A \cap B = \emptyset$$

And in this case there is no chance of the event "A and B" occurring, that is $P(A \cap B) = 0$ and $P(A \cup B) = P(A) + P(B)$.

Independent Events vs. Mutually Exclusive

(★) Events A and B are **independent events** if and only if the probabilities P(A), P(B) and $P(A \cap B)$ satisfy the equation

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

(★) Recall that events A and B are said to be **mutually exclusive** or **disjoint** if the intersection of A and B is empty, that is

$$A \cap B = \emptyset$$

And in this case there is no chance of the event "*A* and *B*" occurring, that is $P(A \cap B) = 0$ and $P(A \cup B) = P(A) + P(B)$.

NOTE: If $P(A)>0$ and $P(B)>0$ and events A and B are mutually exclusive then these events cannot also be independent $P(A \cap B) = 0 \neq P(A) \cdot P(B)$.

Random Variables

DEFINITION: A ***random variable*** , r.v., is a function, *X*, that ***measures the outcome of an experiment*** and ***always has a probability associated to it***.

Random Variables

DEFINITION: A ***random variable*** , r.v., is a function, *X*, that ***measures the outcome of an experiment*** and ***always has a probability associated to it***.

Examples:

(a) Suppose we flip a fair coin once and suppose we *let X denote the number of heads then this is a **random variable***.

(b) Suppose we have a *bias coin* that comes up heads 70% of the time and comes up tails 30% of the time. Suppose we flip this coin 3 times. *If the variable X counts the total number of heads then X is a **random variable***.

Random Variables - Example

If we have a bias coin that comes up heads 70% of the time and comes up tails 30% of the time and if we flip this coin 3 times then:

(a) The sample space is

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

Note that if the outcomes were **equally likely** then each event above in *S* would have equal probability, i.e. 1/8, but since the coin is not fair, these events are **<u>not</u>** equally likely and so the probability of each event has to be calculated explicitly.

Random Variables - Example

If we have a bias coin that comes up heads 70% of the time and comes up tails 30% of the time and if we flip this coin 3 times then:

(a) The sample space is

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

Note that if the outcomes were **equally likely** then each event above in S would have equal probability, i.e. 1/8, but since the coin is not fair, these events are **not** equally likely and so the probability of each event has to be calculated explicitly.

(b) Since each flip of the coin is independent, we can calculate, for example,

$$P(TTH) = P(T) \cdot P(T) \cdot P(H) = (0.3)(0.3)(0.7) = (0.3)^2(0.7) = 0.063$$

Random Variables - Example

If we have a bias coin that comes up heads 70% of the time and comes up tails 30% of the time and if we flip this coin 3 times then:

(a) The sample space is

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

Note that if the outcomes were **equally likely** then each event above in *S* would have equal probability, i.e. 1/8, but since the coin is not fair, these events are **<u>not</u>** equally likely and so the probability of each event has to be calculated explicitly.

(b) Since each flip of the coin is <u>independent</u>, we can calculate, for example,

$$P(TTH) = P(T) \cdot P(T) \cdot P(H) = (0.3)(0.3)(0.7) = (0.3)^2(0.7) = 0.063$$

Caution! Since the events are *<u>not</u>* equally likely $P(TTH) \neq 1/8$.

Random Variables - Example

(c) If the **random variable** *X* counts the total number of heads then the probability of each event or **probability distribution of X** is

$$P(X = 0) = P(TTT) = P(T) \cdot P(T) \cdot P(T) = (0.3)^3 = 0.027,$$

Random Variables - Example

(c) If the **random variable** *X* counts the total number of heads then the probability of each event or **probability distribution of X** is

$$P(X = 0) = P(TTT) = P(T) \cdot P(T) \cdot P(T) = (0.3)^3 = 0.027,$$

$$P(X = 1) = P(HTT \cup THT \cup TTH)$$

$$= P(HTT) + P(THT) + P(TTH) = 3(0.7)(0.3)^2 = 0.189$$

Random Variables - Example

(c) If the **random variable** X counts the total number of heads then the probability of each event or **probability distribution of X** is

$$P(X = 0) = P(TTT) = P(T) \cdot P(T) \cdot P(T) = (0.3)^3 = 0.027,$$

$$P(X = 1) = P(HTT \cup THT \cup TTH)$$
$$= P(HTT) + P(THT) + P(TTH) = 3(0.7)(0.3)^2 = 0.189$$

$$P(X = 2) = P(THH \cup HTH \cup HHT) = 3(0.7)^2\,(0.3) = 0.441$$

Random Variables - Example

(c) If the **random variable** X counts the total number of heads then the probability of each event or **probability distribution of X** is

$$P(X = 0) = P(TTT) = P(T) \cdot P(T) \cdot P(T) = (0.3)^3 = 0.027,$$

$$P(X = 1) = P(HTT \cup THT \cup TTH)$$

$$= P(HTT) + P(THT) + P(TTH) = 3(0.7)(0.3)^2 = 0.189$$

$$P(X = 2) = P(THH \cup HTH \cup HHT) = 3(0.7)^2 (0.3) = 0.441$$

$$P(X = 3) = P(HHH) = (0.7)^3 = 0.343$$

Example: Equally Likely Outcomes

Think of drawing a card at random from a deck of *52* cards.

Each of the *52* cards has an <u>equal chance of being selected</u>, and $|S|= 52$ so

$$P(\text{selecting any specific card}) = \frac{1}{|S|} = \frac{1}{52}$$

Example: Equally Likely Outcomes

Think of drawing a card at random from a deck of *52* cards.

Each of the *52* cards has an equal chance of being selected, and $|S|$= 52 so

$$P(\text{selecting any specific card}) = \frac{1}{|S|} = \frac{1}{52}$$

So if we have an event, like picking an ace, that is if

A = { an ace } = event that we select an ace from a 52 card deck

Then we have to count the number of ways that event A can occur, $|A|$, and divide it by the total number of outcomes in the sample space, $|S|$:

Example: Equally Likely Outcomes

Think of drawing a card at random from a deck of *52* cards.

Each of the *52* cards has an equal chance of being selected, and $|S|$= 52 so

$$P(\text{selecting any specific card}) = \frac{1}{|S|} = \frac{1}{52}$$

So if we have an event, like picking an ace, that is if

A = { an ace } = event that we select an ace from a 52 card deck

Then we have to count the number of ways that event *A* can occur, $|A|$, and divide it by the total number of outcomes in the sample space, $|S|$:

$$P(A) = \frac{|A|}{|S|} = \frac{4}{52} = \text{ probability of picking an ace}$$

Example: Equally Likely Outcomes

We can use the properties and theorems of Probability to determine the chance of more complicated outcomes.

Draw a card from a standard 52 card deck, what's the probability of picking an ace <u>or</u> a spade? By Inclusion-Exclusion, we have

$$P(A \cup S) = P(A) + P(S) - P(A \cap S) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = 16/52 \approx 0.31$$

Example: Equally Likely Outcomes

We can use the properties and theorems of Probability to determine the chance of more complicated outcomes.

Draw a card from a standard 52 card deck, what's the probability of picking an ace <u>or</u> a spade? By Inclusion-Exclusion, we have

$$P(A \cup S) = P(A) + P(S) - P(A \cap S) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = 16/52 \approx 0.31$$

And, for example, using the Law of Complements

$$P(A^c \cap S^c) = P((A \cup S)^c) = 1 - P(A \cup S) = 1 - 16/52 = 36/52$$

Example: Equally Likely Outcomes

We can use the properties and theorems of Probability to determine the chance of more complicated outcomes.

Draw a card from a standard 52 card deck, what's the probability of picking an ace <u>or</u> a spade? By Inclusion-Exclusion, we have

$$P(A \cup S) = P(A) + P(S) - P(A \cap S) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = 16/52 \approx 0.31$$

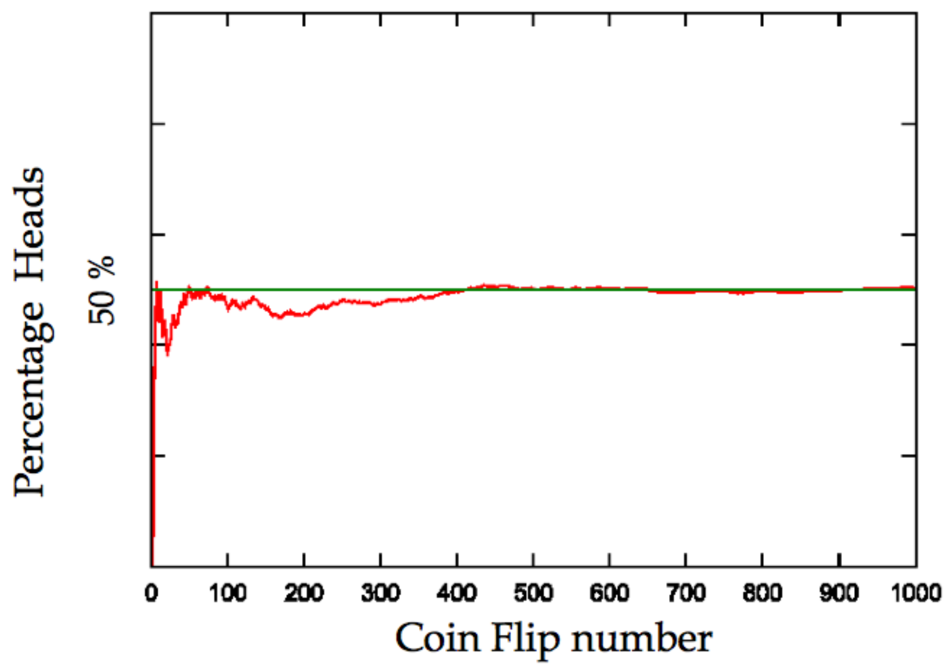
And, for example, using the Law of Complements

$$P(A^c \cap S^c) = P((A \cup S)^c) = 1 - P(A \cup S) = 1 - 16/52 = 36/52$$

This is the probability of picking a card that is <u>neither an ace nor a spade</u>.

Interpretations of Probability

- Although the probability P is well-defined mathematically, how we *interpret* the probability P in real world situations is not always clear. E.g., coin flips vs. P(rain).
- The **relative frequency interpretation** of probability
    - This interpretation says that P is just long run relative frequency of events.



Interpretations of Probability

- The **subjective interpretation** of probability is also accepted by many.
  - This interpretation says that P represents one's "subjective degree of belief" in a claim about a random process, i.e. *faith*

There are other interpretations, and reasonable people disagree about the best interpretation. This has real consequences for statistical practice!

(For more on this, take a philosophy of statistics course)

Conditional Probability

In this section, we examine how the information that "an event *B* has occurred" affects the probability assigned to event *A*.

Conditional Probability

In this section, we examine how the information that "an event *B* has occurred" affects the probability assigned to event *A.*

For example, event *A* might refer to an individual having a particular disease.

If a certain blood test is performed on the individual and the result is negative, then the **updated probability of disease** will be different than if the blood test result was positive.

Conditional Probability

In this section, we examine how the information that "an event *B* has occurred" affects the probability assigned to event *A.*

For example, event *A* might refer to an individual having a particular disease.

If a certain blood test is performed on the individual and the result is negative, then the **updated probability of disease** will be different than if the blood test result was positive.

We will use the notation $P(A \mid B)$ to represent the **conditional probability of event *A* <u>given</u> that the event *B* has occurred.** *B* is the "conditioning event."

Conditional Probability

<u>**Definition:**</u> The **conditional probability of A given B**, $P(A/B)$, is defined as:

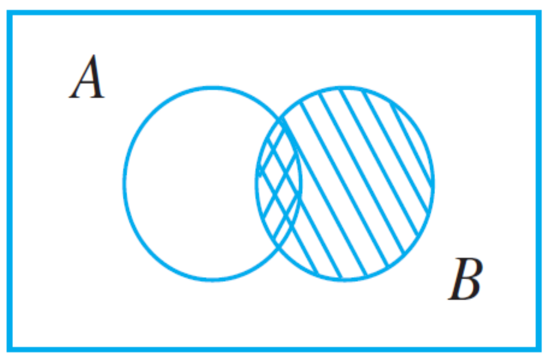
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Conditional Probability

<u>Definition:</u> The **conditional probability of A given B**, $P(A/B)$, is defined as:

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

Given that *B* has occurred, the relevant sample space is no longer all of *S* but it boils down to only the outcomes in *B.*



Conditional Probability - Example

Specific computer parts are assembled in a plant that uses two different assembly lines, line A and line A'.

Line A uses older equipment than A', so it is somewhat slower and less reliable.

Suppose on a given day line A has assembled 8 parts, whereas A' has produced 10.

Conditional Probability - Example

Specific computer parts are assembled in a plant that uses two different assembly lines, line A and line A'.

Line A uses older equipment than A', so it is somewhat slower and less reliable.

Suppose on a given day line A has assembled 8 parts, whereas A' has produced 10.

Let B denote **the event that a defective part has occurred**.

Conditional Probability - Example

Specific computer parts are assembled in a plant that uses two different assembly lines, line A and line A'.

Line A uses older equipment than A', so it is somewhat slower and less reliable.

Suppose on a given day line A has assembled 8 parts, whereas A' has produced 10.

Let B denote **the event that a defective part has occurred**.

Suppose from the 8 parts from line A, 2 are defective and 6 are non-defective and from the 10 parts from line A', 1 was defective and 9 non-defective.

Conditional Probability - Example

This information is summarized in the accompanying table.

		Condition	
		B	B'
Line	A	2	6
	A'	1	9

Unaware of this information, the sales manager randomly selects 1 of these 18 parts for a test. Note **before** the test:

$$P(\text{part from line A selected}) = P(A) = \frac{|A|}{|S|} = \frac{8}{18} = 0.44$$

Conditional Probability - Example

Now, say, for example, that the manager chose a part that turned out to be defective – i.e., event *B = {defective part}* <u>has occurred</u>.

Conditional Probability - Example

Now, say, for example, that the manager chose a part that turned out to be defective – i.e., event *B = {defective part}* has occurred.

Then, the selected part must have been one of the 3 total defective parts made and the probability that it was made by the line *A* would be 2/3 or using **conditioning** we can calculate:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/18}{3/18} = \frac{2}{3}$$

The Multiplication Rule for $P(A \cap B)$

The definition of conditional probability yields the following result:

The Multiplication Rule:

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

This rule is important because it is often the case that $P(A \cap B)$ is desired, whereas only $P(B)$ and $P(A|B)$ can be found from the information available.

By definition of $P(B|A)$ we also have $P(A \cap B) = P(B|A) \cdot P(A)$

(★) **Alternate Definition of Independence**

Two events A and B are **independent** iff $P(A \mid B) = P(A)$ and are **dependent** otherwise.

The Multiplication Rule for $P(A \cap B)$

The definition of conditional probability yields the following result:

The Multiplication Rule:

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

This rule is important because it is often the case that $P(A \cap B)$ is desired, whereas only $P(B)$ and $P(A|B)$ can be found from the information available.

By definition of $P(B|A)$ we also have $P(A \cap B) = P(B|A) \cdot P(A)$

(★) **<u>Alternate Definition of Independence</u>**

Two events A and B are **independent** iff $P(A \mid B) = P(A)$ and are **dependent** otherwise.

Example

Suppose events A and B are independent, we show that A' and B are independent.

Example

Suppose events A and B are independent, we show that A' and B are independent. Note that

$$B = (B \cap A) \cup (B \cap A')$$

And so, by the inclusion exclusion principle, we have

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

Example

Suppose events A and B are independent, we show that A' and B are independent. Note that

$$B = (B \cap A) \cup (B \cap A')$$

And so, by the inclusion exclusion principle, we have

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

And so,

$$P(A'|B) = \frac{P(A' \cap B)}{P(B)} = \frac{P(B) - P(B \cap A)}{P(B)} = 1 - P(A|B) = 1 - P(A) = P(A')$$

Thus, A' and B are independent. Note that we can show that P(•|B) satisfies the axioms of probability for any fixed set B.

Independence of More Than Two Events

Definition

Events $A_1, \ldots, A_n$ are **<u>mutually independent</u>** if for every k ($k = 2, 3, \ldots, n$) and every subset of indices $i_1, i_2, \ldots, i_k$,

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \cdots \cdot P(A_{i_k})$$

Definition

Events $A_1, \ldots, A_n$ are **<u>exhaustive events</u>** if

$$A_1 \cup A_2 \cup \cdots \cup A_n = \mathcal{S}$$

(Recall $A_1, \ldots, A_n$ are <u>mutually exclusive events</u> if $A_i \cap A_j = \emptyset$ for $i \neq j$.)

Law of Total Probability

Let $A_1, \ldots, A_k$ be *mutually exclusive* and *exhaustive events*. Then for any other event B we have,

$$
\begin{aligned}
P(B) &= P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_k) \\
&= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k) \\
&= \sum_{i=1}^{k} P(B|A_i)P(A_i)
\end{aligned}
$$

(Note that if we divide both sides by *P(B)* the probabilities will sum to 1)

Bayes’ Theorem

The multiplication rule is most useful when the experiment consists of **several stages in succession**.

The conditioning event ***B*** **then describes the outcome of the first stage** and ***A*** **the outcome of the second**, so that $P(A|B)$, conditioning on what occurs first, will often be known.

The rule is easily extended to experiments involving **more than two stages**.

Bayes' Theorem

The multiplication rule is most useful when the experiment consists of **several stages in succession**.

The conditioning event ***B*** **then describes the outcome of the first stage** and ***A*** **the outcome of the second**, so that $P(A|B)$, conditioning on what occurs first, will often be known.

The rule is easily extended to experiments involving **more than two stages**.

Bayes Theorem:

The computation of a **posterior probability** $P(A_j|B)$ from given **prior probabilities** $P(A_i)$ and **conditional probabilities** $P(B|A_i)$ occupies a central position in elementary probability.

The general rule, called **Bayes' Theorem**, for such computations goes back to Reverend Thomas Bayes, who lived in the 18th century.

Bayes' Theorem - Example

An individual has 3 different email accounts. Most of her messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3.

Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts
#2 and #3 are 2% and 5% spam, respectively.

(**Q**) What is the probability that a randomly selected message is spam?

Bayes' Theorem - Example

To answer this question, let's first establish some notation:

A_i = {message is from account # *i*} for i = 1, 2, 3,

B = {message is spam}, we wish to find $P(B)$

Bayes' Theorem - Example

To answer this question, let's first establish some notation:

A_i = {message is from account # i} for $i = 1, 2, 3$,

B = {message is spam}, we wish to find $P(B)$

Then the given percentages imply that

$$P(A_1) = .70, \quad P(A_2) = .20, \quad P(A_3) = .10$$

and,

$$P(B|A_1) = .01, P(B|A_2) = .02, P(B|A_3) = .05$$

Bayes' Theorem - Example

Now it is simply a matter of substituting into the equation for the *law of total probability* to find *P(B)*:

$$\begin{aligned} P(B) &= P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) \\ &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \\ &= (.01)(.70) + (.02)(.20) + (.05)(.10) = 0.016 \end{aligned}$$

In the long run, 1.6% of this individual's messages will be spam.

Bayes' Theorem - Example

Now it is simply a matter of substituting into the equation for the *law of total probability* to find *P(B)*:

$$
\begin{aligned}
P(B) &= P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) \\
&= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \\
&= (.01)(.70) + (.02)(.20) + (.05)(.10) = 0.016
\end{aligned}
$$

In the long run, 1.6% of this individual's messages will be spam.

(**Q**) Say she randomly selected a message and it was indeed spam. What is the probability that it came from account #1?

Bayes' Theorem - Example

Now it is simply a matter of substituting into the equation for the *law of total probability* to find *P(B)*:

$$
\begin{aligned}
P(B) &= P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) \\
&= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \\
&= (.01)(.70) + (.02)(.20) + (.05)(.10) = 0.016
\end{aligned}
$$

In the long run, 1.6% of this individual's messages will be spam.

(**Q**) Say she randomly selected a message and it was indeed spam. What is the probability that it came from account #1?

We wish to find $P(A_1|B)$.

Bayes' Theorem

Let $A_1, A_2, \ldots, A_k$ be a collection of k mutually exclusive and exhaustive events with *prior* probabilities $P(A_i)$

Then for any other event B for which $P(B) > 0$, the *posterior* probability of A_j given that B has occurred is

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}, \; j = 1, 2, \ldots, k \qquad (✪)$$

Bayes' Theorem

Let $A_1, A_2, \ldots, A_k$ be a collection of k mutually exclusive and exhaustive events with ***prior*** probabilities $P(A_i)$

Then for any other event B for which $P(B) > 0$, the ***posterior*** probability of A_j given that B has occurred is

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}, \; j = 1, 2, \ldots, k \qquad (✪)$$

So the answer to the spam question is

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B)} = \frac{(0.01)(0.70)}{0.016} = 0.4375$$

Note that Equation (✪) is know as **Bayes' Theorem**.