STA250 Probability and Statistics

Chapter 3 Notes

Probability

Asst. Prof. Abdullah YALÇINKAYA

Ankara University, Faculty of Science, Department of Statistics

ayalcinkaya@ankara.edu.tr

2023

STA250 Probability and Statistics

Reference Book

This lecture notes are prepared according to the contents of

"PROBABILITY & STATISTICS FOR ENGINEERS & SCIENTISTS by Walpole, Myers, Myers and Ye"

Probability Intro

- **The <u>sample space</u> is the set consisting of all possible sample points. A sample space is represented by the symbol S.**
  - Each outcome in a sample space is called a sample point.
  - The sample space S is a set, the domain of the probability function.
  - Each probability value, p, is a real number $0 \leq p \leq 1$.

- **An <u>event</u> is a subset of a sample space ($E \subseteq S$)**
  - Note that both S and $\varnothing$ are events as well.
  - Note that the number of the events are 2^n

- **Sample spaces can be continuous or discrete.**
  - A discrete sample space is one that contains either a finite or a countable number of distinct sample points.

Probability Intro

- **Example: Life in years of a component. S = ?**
  - S = {t | t ≥ 0} => "all values of t such that t ≥ 0"
  - A = component fails before the end of the fifth year.
  - A = {t | t < 5}.
- **Example: Flip a coin one times. S = ?**
  - n(S)=2 (number of elements)
  - S = {H,T}, Head or Tail
- **<u>Events:</u> the number of the events are** $2^2 = 4$.
  - $A_1 = \{H\}$, The flip is Head.
  - $A_2 = \{T\}$, The flip is Tail.
  - $A_3 = \{\emptyset\}$, null set
  - $A_4 = \{H, T\}$, The flip is H or T, it is sample space.

Probability: Tossing a Die

- **The number of points on the top face.**
  - S1={1,2,3,4,5,6}

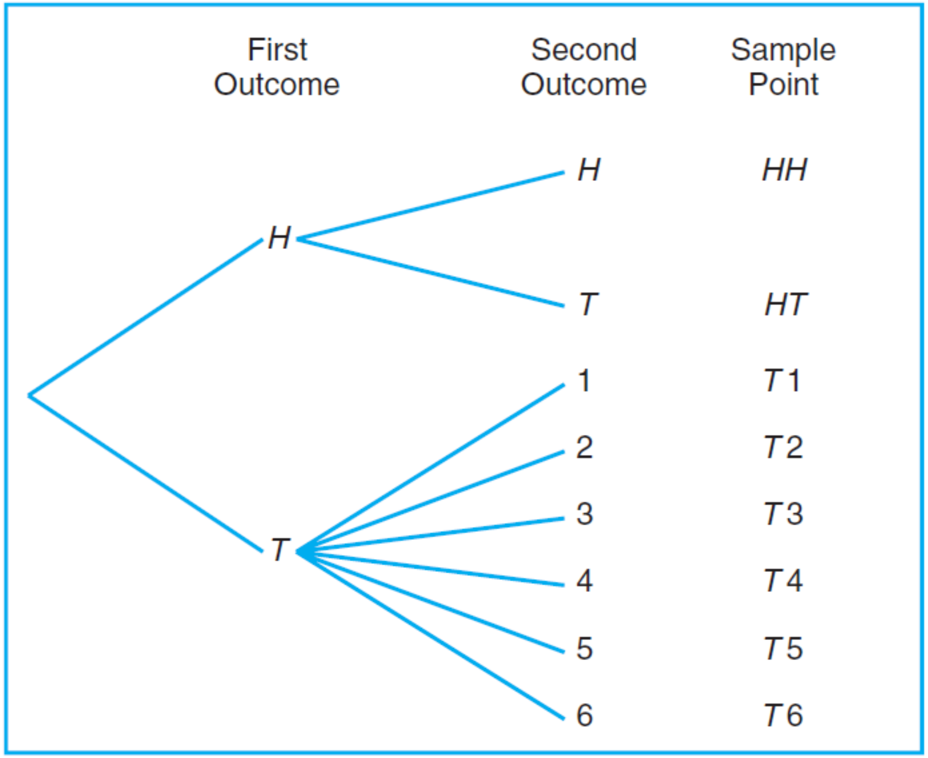
- **The number of points on the top face is even or odd.**
  - S2={even, odd}

- S1 provides more information than S2.

- **More than one sample space can be used to describe the outcomes of an experiment. Which one we use?**

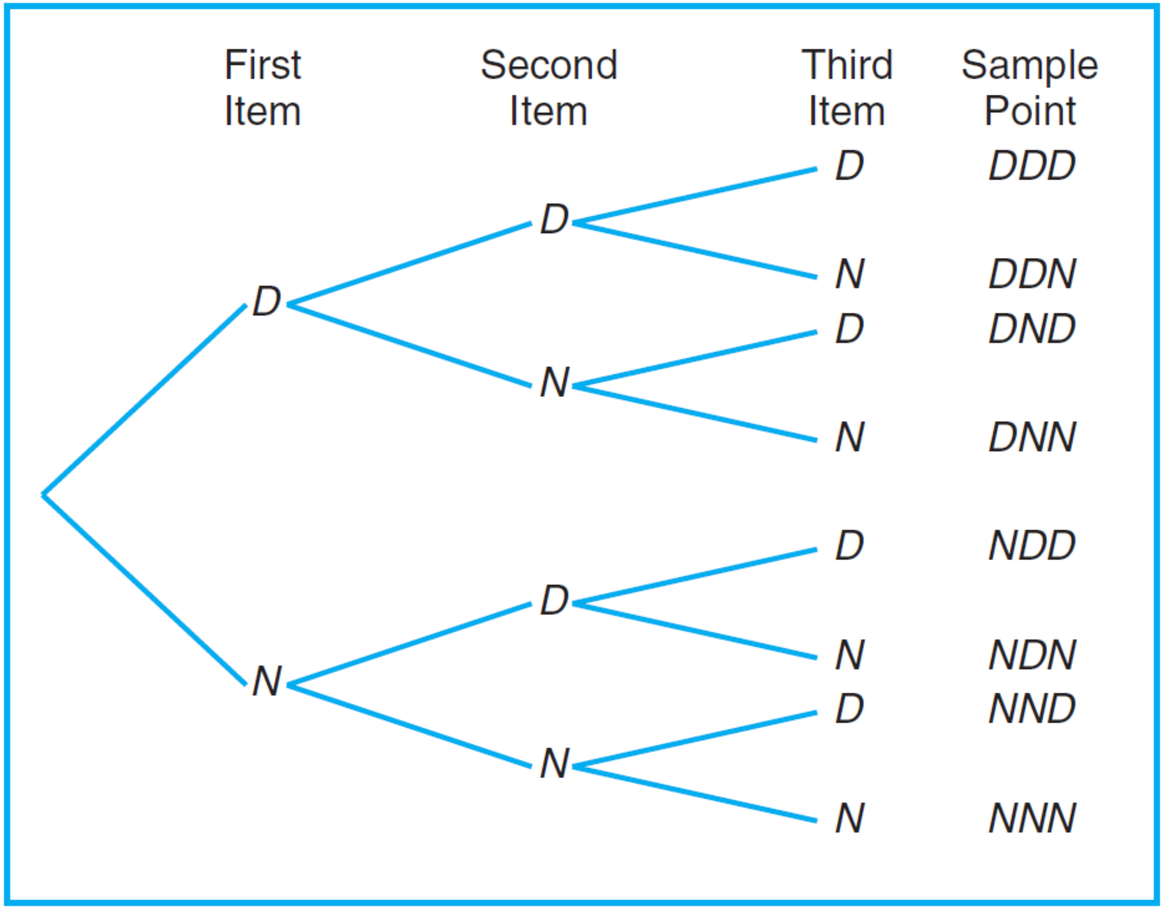
Probability: Tree Diagram

- An experiment consists of flipping a coin and then flipping it a second time if a head (H) occurs.
- If a tail (T) occurs on the first flip, then a die is tossed once.
- T2: Coin Shows T and Die shows 2
- $S = \{HH, HT, T1, T2, T3, T4, T5, T6\}$



Example: a manufacturing process

- 3 items are selected at random.
- Each item is classified defective (D) or nondefective (N).
- DDD: All 3 items inspected are defective.
- *S* = {*DDD*, *DDN*, *DND*, *DNN*, *NDD*, *NDN*, *NND*, *NNN*}.



Probability Intro

- Sample spaces with a large or infinite number of sample points are best described by a **statement** or **rule method.**
- If the possible outcomes of an experiment are the set of cities in the world with a population over one million, our sample space is written
  - S ={ x | x is a city with a population over one million }.
- S is the set of all x **such that** x is a city with a population over one million. The vertical line is read **"such that".**
- If S is the set of all points (x, y) on the boundary or the interior of a circle of radius 2 with center at the origin, we write the **rule**
  - $S = \{ (x, y) \mid x^2 + y^2 \leq 4 \}$.

Probability Intro: Example

- Consider the manufacturing process again.
- Items are either *D*, defective, or *N*, nondefective.
- There are many important statistical procedures called sampling plans that determine whether or not a “lot” of items is considered satisfactory.
- One such plan involves sampling until *k* defectives are observed.
- Suppose the experiment is to sample items randomly until one defective item is observed.
- The sample space for this case is

$S = \{D, ND, NND, NNND, \ldots\}$.

Probability Intro

- **<u>Example:</u> Flip a coin two times. S = ?**
  - n(S)=4 (number of elements)
  - $S = \{TT, TH, HT, HH\}$
- **Events:** **the number of the events are $2^4 = 16$.**
  - $A_1 = \{HH\}$
  - $A_2 = \{TT\}$
  - $A_3 = \{TH\}$
  - $A_4 = \{HT\}$
  - $A_5 = \{TT, TH\}$
  - $A_6 = \{TT, HT\}$
  - ……. goes on
- **<u>Example:</u> Flip a coin three times. S = ?**
  - n(S)=8 (number of elements)
  - S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
  - Number of events: $2^8 = 64$
- **Event A = 1st flip is heads.**
  - A = {HHH, HHT, HTH, HTT}

Probability Intro

- **<u>Example:</u> t is the life in years of a certain electronic component.**
- The sample space is $S = \{t \mid t \geq 0\}$
- Then the event A that the component fails before the end of the fifth year is the subset $A = \{t \mid 0 \leq t < 5\}$.

Example: if we let A be the event of detecting a microscopic organism by the naked eye in a biological experiment, then A is null set.

- If $B = \{ x \mid x$ is an even factor of 7 $\}$, then B must be the null set, since the only possible factors of 7 are the odd numbers 1 and 7.

Probability Intro

The **complement** of an event A is the subset of all elements of S **that are not in A**. We denote the complement of A by the symbol A'.

Consider an experiment where the smoking habits of the employees of a manufacturing firm are recorded.

- A possible sample space might classify an individual as a **nonsmoker**, a **light smoker**, a **moderate smoker**, or a **heavy smoker**.
- Let the subset of smokers be some event. Then **all the nonsmokers** correspond to a different event, also a subset of S, which is called the **complement of the set of smokers**.

Let R be the event that a red card is selected from an ordinary deck of 52 playing cards, and let S be the entire deck.

- Then R′ is the event that the card selected from the deck is **not a red card but a black card**.

Event/Set Operations

- **The <u>complement</u> of an event A?**
  - The set of all elements of S not in A. Denoted A′.
  - A = 1st flip is heads. A′ = first flip is not heads.
- **Example: Consider the sample space;**
  - S={book, cell phone, mp3, paper, stationery, laptop}.
  - Let A={book, stationery, laptop, paper}.
  - Then the complement of A′={cell phone, mp3}.
- **The <u>intersection</u> of two events A and B?**
  - The set of all elements in both A <u>and</u> B. Denoted $A \cap B$.
- **Example: Suppose that *A* and *B* are subsets of the same sample space *S*.**
  - *The tossing of a die: The sample space: S = {1, 2, 3, 4, 5, 6}.*
  - *A* is the event that an even number occurs.
  - $A = \{2, 4, 6\}$
  - *B* shows the event that a number greater than 3.
  - $B = \{4, 5, 6\}$
  - $A \cap B = \{4,6\}$ is the intersection of *A* and *B*.

Event/Set Operations

- Two events *A* and *B* are **mutually exclusive**, or **disjoint**, if *A* and *B* have no elements in common.
  - $A \cap B = \varnothing$

- **Example:** Let $V = \{a, e, i, o, u\}$ and $C = \{l, r, s, t\}$; then it follows that $V \cap C = \varnothing$
  - *V* and *C* have no elements in common and, therefore, cannot both simultaneously occur.

- The **union** of the two events *A* and *B* containing all the elements that belong to *A* or *B* or both.
  - It is represented by the symbol $A \cup B$.

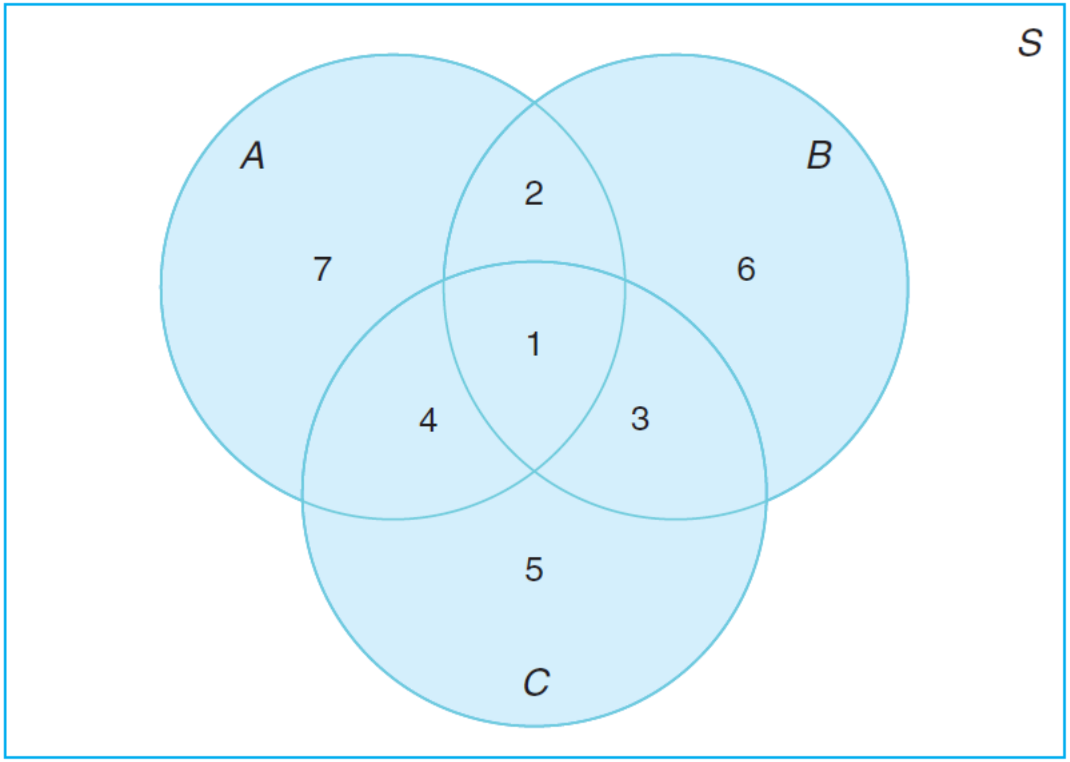
- **Example:** Let $A = \{a, b, c\}$ and $B = \{b, c, d, e\}$
  - then $A \cup B = \{a, b, c, d, e\}$.

- If $M = \{ x \mid 3 < x < 9\}$ and $N = \{ y \mid 5 < y < 12\}$, then $M \cup N = \{ z \mid 3 < z < 12\}$.

Venn Diagrams

- **Venn Diagrams** show various events graphically, and are sometimes helpful in understanding set theory problems.
- **Standard set theory results hold:**
  - $A \cap \varnothing = \varnothing$
  - $A \cup \varnothing = A$
  - $A \cap A' = \varnothing$
  - $A \cup A' = S$
  - $S' = \varnothing$
  - $\varnothing' = S$
  - $(A')' = A$
  - $(A \cap B)' = A' \cup B'$
  - $(A \cup B)' = A' \cap B'$

Venn Diagrams



- $A \cap B$ = regions 1 and 2,
- $B \cap C$ = regions 1 and 3,
- $A \cup C$ = regions 1, 2, 3, 4, 5, and 7,
- $B' \cap A$ = regions 4 and 7,
- $A \cap B \cap C$ = region 1,
- $(A \cup B) \cap C'$ = regions 2, 6, and 7,

Counting Sample Points

If an operation can be performed in n_1 ways, and if for each of these ways a second operation can be performed in n_2 ways, then the two operations can be performed together in $n_1 n_2$ ways.

Example: How many sample points are there in the sample space when a pair of dice is thrown once?
Solution: The first die can land face-up in any one of $n_1 = 6$ ways. For each of these 6 ways, the second die can also land face-up in $n_2 = 6$ ways. Therefore, the pair of dice can land in $n_1 n_2 = (6)(6) = 36$ possible ways.

Example: If a 22-member club needs to elect a chair and a treasurer, how many different ways can these two to be elected?
Solution: For the chair position, there are 22 total possibilities. For each of those 22 possibilities, there are 21 possibilities to elect the treasurer. Using the multiplication rule, we obtain $n_1 \times n_2 = 22 \times 21 = 462$ different ways.

Counting Sample Points

If an operation can be performed in n_1 ways, and if for each of these a second operation can be performed in n_2 ways, and for each of the first two a third operation can be performed in n_3 ways, and so forth, then the sequence of k operations can be performed in $n_1 n_2 \cdots n_k$ ways.

Example: Sam is going to assemble a computer by himself. He has the choice of chips from two brands, a hard drive from four, memory from three, and an accessory bundle from five local stores.
How many different ways can Sam order the parts?
Solution: Since $n_1 = 2$, $n_2 = 4$, $n_3 = 3$, and $n_4 = 5$, there are

$$n_1 \times n_2 \times n_3 \times n_4 = 2 \times 4 \times 3 \times 5 = 120$$

different ways to order the parts.

Counting Sample Points

Example: How many even four-digit numbers can be formed from the digits 0, 1, 2, 5, 6, and 9 if each digit can be used only once?

Solution:

Since the number must be even, we have only $n_1 = 3$ choices (0, 2, 6) for the units position. However, for a four-digit number the thousands position cannot be 0. Hence, we consider the units position in two parts, 0 or not 0.

If the units position is 0 (i.e., $n_1 = 1$), we have $n_2 = 5$ choices for the thousands position, $n_3 = 4$ for the hundreds position, and $n_4 = 3$ for the tens position. Therefore, in this case we have a total of

$$n_1 n_2 n_3 n_4 = (1)(5)(4)(3) = 60$$

even four-digit numbers.

On the other hand, if the units position is not 0 (i.e., $n_1 = 2$), we have $n_2 = 4$ choices for the thousands position, $n_3 = 4$ for the hundreds position, and $n_4 = 3$ for the tens position. In this situation, there are a total of

$$n_1 n_2 n_3 n_4 = (2)(4)(4)(3) = 96$$

even four-digit numbers.

Since the above two cases are mutually exclusive, the total number of even four-digit numbers can be calculated as $60 + 96 = 156$.

Counting Sample Points

A **permutation** is an arrangement of all or part of a set of objects

Example: Consider the three letters *a*, *b*, and *c*.

The possible permutations are *abc*, *acb*, *bac*, *bca*, *cab*, and *cba*.

There are 6 distinct arrangements.

We could arrive at the answer 6 without actually listing the different orders by the following arguments:

There are $n_1 = 3$ choices for the first position. No matter which letter is chosen, there are always $n_2 = 2$ choices for the second position. No matter which two letters are chosen for the first two positions, there is only $n_3 = 1$ choice for the last position, giving a total of

$$n_1 n_2 n_3 = (3)(2)(1) = 6$$

permutations.

Counting Sample Points

In general, n distinct objects can be arranged in $n(n-1)(n-2)\ldots(3)(2)(1)$ ways. There is a notation for such a number.

For any non-negative integer n, $n!$, called "n factorial", is defined as

$$n! = n(n-1)\cdots(2)(1)$$

with special case $0! = 1$.

The number of permutations of n objects is $n!$

The number of permutations of n distinct objects taken r at a time is

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

Example: In one year, three awards (research, teaching, and service) will be given to a class of 25 graduate students in a statistics department. If each student can receive at most one award, how many possible selections are there?

Solution: Since the awards are distinguishable, it is a permutation problem. The total number of sample points is

$${}_{25}P_3 = \frac{25!}{(25-3)!} = \frac{25!}{22!} = (25)(24)(23) = 13800$$

The number of distinct permutations of n things of which n_1 are of one kind, n_2 of a second kind, ..., n_k of a kth kind is

$$\frac{n!}{n_1!n_2!\cdots n_k!}.$$

In a college football training session, the defensive coordinator needs to have 10 players standing in a row. Among these 10 players, there are 1 freshman, 2 sophomores, 4 juniors, and 3 seniors. How many different ways can they be arranged in a row if only their class level will be distinguished?

we find that the total number of arrangements is

$$\frac{10!}{1!\ 2!\ 4!\ 3!} = 12,600.$$

Counting Sample Points

The number of ways of partitioning a set of n objects into r cells with n_1 elements in the first cell, n_2 elements in the second, and so forth, is

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!},$$

where $n_1 + n_2 + \cdots + n_r = n$.

In how many ways can 7 graduate students be assigned to 1 triple and 2 double hotel rooms during a conference?
The total number of possible partitions would be

$$\binom{7}{3, 2, 2} = \frac{7!}{3!\ 2!\ 2!} = 210.$$

The number of combinations of n distinct objects taken r at a time is

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

A young boy asks his mother to get 5 Game-Boy™ cartridges from his collection of 10 arcade and 5 sports games. How many ways are there that his mother can get 3 arcade and 2 sports games?

$$\binom{10}{3} = \frac{10!}{3!\ (10-3)!} = 120. \quad \binom{5}{2} = \frac{5!}{2!\ 3!} = 10. \quad (120)(10) = 1200 \text{ ways.}$$

ANKARA ÜNİVERSİTESİ 1946

Probability of an Event

- For now, we only consider discrete sample spaces (contains a finite number of elements).
- Each point in a sample space is assigned a weight or probability value. The higher the probability, the more likely that outcome is to occur.
- To every point in the sample space, sum of all probabilities is 1.

The **probability** of an event A is the sum of the weights of all sample points in A. Therefore,

$$0 \leq P(A) \leq 1, \quad P(\phi) = 0, \quad \text{and} \quad P(S) = 1.$$

Furthermore, if $A_1, A_2, A_3, \ldots$ is a sequence of mutually exclusive events, then

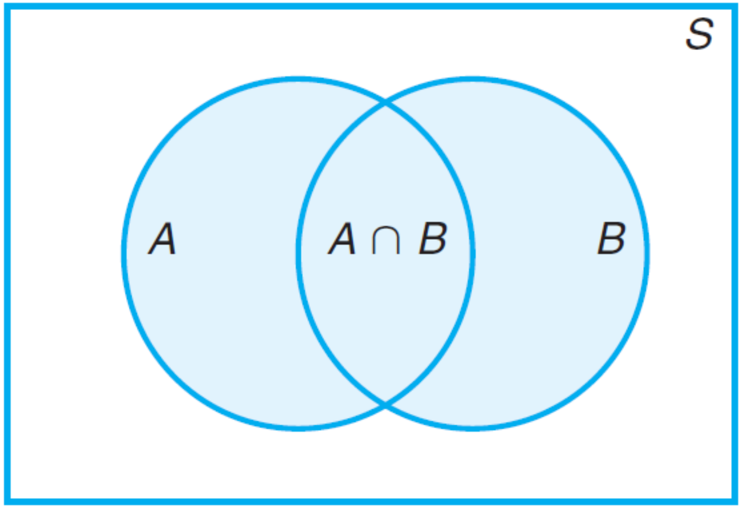
$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots.$$

Probability of an Event

- **Example: A coin is tossed twice. What is the probability that at least 1 head occurs?**
  - **Solution:** The sample space is *S={HH,HT,TH,TT}.*
  - If the coin is balanced, each of these outcomes is equally likely to occur. Therefore, each sample point is assigned a probability of w.
  - *4w=1 or w=1/4.*
  - A=even of at least 1 head occurring.
  - *A={HH,HT,TH} and P(A)=1/4+1/4+1/4=3/4.*
- **If an experiment can result in any one of *N* different <u>equally likely</u> outcomes, and if exactly *n* of these outcomes correspond to event *A*, then the probability of event *A* is**
  - $P(A) = \frac{n}{N}$
- **Example:** Since 25 of 53 students are majoring in industrial engineering, the probability of event A, selecting n industrial engineering major at random, is
  - $P(A) = \frac{25}{53}$

Additive Probability Rules

- **If A and B are two events, then**
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
  - Look at the Venn diagram, the sample points in $P(A \cap B)$ are double counted.
- **If A and B are mutually exclusive, then**
  - $P(A \cup B) = P(A) + P(B)$
  - $A \cap B = \varnothing$ so $P(A \cap B)=P(\varnothing) = 0$
- **If three events A,B,C**
  - $$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$
- **If A and A′ are complementary events, then**
  - $P(A') = 1 - P(A)$ or,
  - $P(A) + P(A')=1$.



Conditional Probability

- **<u>Conditional probability</u>, written P(B|A), is the probability of "<u>B, given A</u>", the probability that B occurs, given that we know that A has occurred.**
  - $P(B|A) = \frac{P(A \cap B)}{P(A)}$, provided $P(A)>0$

- **Example: The data are given in Table.**

	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

P(M|E) = ? $P(M|E) = \frac{P(M \cap E)}{P(E)} = \frac{460/900}{600/900} = \frac{23}{30}$

M: a man is chosen, *E*: the one chosen is employed.

Conditional Probability Example

- **The population is 900 people.**
- **If a person is selected at random from this group,**

- $P(E) = ?$ $\quad P(E)=600/900=2/3$
- $P(M) = ?$ $\quad P(M)=500/900=5/9$
- $P(E \cap M) = ?$ $\quad P(E \cap M) = 460/900=23/45$
- $P(E\,|\,M) = ?$ $\quad P(E\,|\,M) = \dfrac{P(E\cap M)}{P(M)} = \dfrac{460/900}{500/900} = \dfrac{23}{25}$
- $P(M\,|\,E) = ?$ $\quad P(M\,|\,E) = \dfrac{P(M\cap E)}{P(E)} = \dfrac{460/900}{600/900} = \dfrac{23}{30}$

	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

M: a man is chosen, *E*: the one chosen is employed.

Independence and multiplicative rule

- **Suppose P(B | A) = P(B).**
  - Whether A occurs or not, the probability of B occurring doesn't change.
- **If P(B | A) = P(B), then A and B are <u>independent</u>.**
  - Can show that if P(B | A) = P(B) is true, then P(A | B) = P(A) is always also true.
- **From the above, and the definition of conditional probability, if A and B are independent,**
  - $P(A \cap B) = P(A)\,P(B)$
- **Rearranging the conditional probability formula, if both A and B can occur, then**
  - $P(A \cap B) = P(B \mid A)\,P(A)$
  - Or, the probability that both *A* and *B* occur is equal to the probability that *A* occurs multiplied by the conditional probability that *B* occurs, given that *A* occurs.
  - It is <u>**multiplicative rule.**</u>
- **Note that it is also true that**
  - $P(A \cap B) = P(A \mid B)\,P(B)$

Bayes Theorem

- **Bayes Theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event.**

  - If the events $B_1, B_2, \ldots \ldots, B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0\ for\ i = 1,2, \ldots, k$, then for any event A in S such that $P(A) \neq 0$,

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{1=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{1=1}^{k} P(B_i)P(A|B_i)} \qquad \text{for} \qquad r=1,2,\ldots,k$$

Theorem of total probability

Bayes Theorem Example

- **Example:**

  A manufacturing firm employs three analytical plans for the design and development of a particular product. For cost reasons, all three are used at varying times. In fact, plans 1, 2, and 3 are used for 30%, 20%, and 50% of the products, respectively. The defect rate is different for the three procedures as follows:

  $P(D|P_1) = 0.01,\qquad P(D|P_2) = 0.03,\qquad P(D|P_3) = 0.02,$

  where $P(D \mid Pj)$ is the probability of a defective product, given plan j. If a random product was observed and found to be defective, which plan was most likely used and thus responsible?

- **Solution:** From the statement of the problem

  $P(P_1) = 0.30,\qquad P(P_2) = 0.20,\qquad P(P_3) = 0.50,$

we must find $P(Pj \mid D)$ for $j = 1, 2, 3$. Bayes Theorem shows

Bayes Theorem Example

- **Example:**

$$P(P_1|D) = \frac{P(P_1)P(D|P_1)}{P(P_1)P(D|P_1) + P(P_2)P(D|P_2) + P(P_3)P(D|P_3)}$$

$$= \frac{(0.30)(0.01)}{(0.30)(0.01) + (0.20)(0.03) + (0.50)(0.02)} = \frac{0.003}{0.019} = 0.158$$

Similarly,

$$P(P_2|D) = \frac{(0.03)(0.20)}{0.019} = 0.316$$

$$P(P_3|D) = \frac{(0.02)(0.50)}{0.019} = 0.526$$

- The conditional probability of a defect given plan 3 is the largest of the three; thus a defective for a random product is most likely the result of the use of plan 3.

Next Lesson

- **Discrete Random Variables And Their Probability Distributions**

See you☺