Probability

Sample Space

- Statisticians use the word **experiment** to describe any process that generates a set of outcomes.

- For example: a statistical experiment is the tossing of a coin. In this experiment, there are only two possible outcomes, head or tail.

- The set of all possible outcomes of a statistical experiment is called the **sample space** and is represented by the symbol *S*.

- Each outcome in a sample space is called an **element** or a **member** of the sample space, or simply a **sample point**.

- If the sample space has a finite number of elements, we may *list* the members separated by commas and enclosed in braces. Thus, the sample space *S*, of possible outcomes when a coin is flipped, may be written as **S = {H,T}** where *H* and *T* are sample points correspond to heads and tails, respectively.
- Consider the experiment of rolling a die. If we are interested in the number that shows on the top face, the sample space is **S = {1,2,3,4,5,6}**

- Sample spaces with a large or infinite number of sample points are best described by a **statement** or **rule method**.

- For example, if the possible outcomes of an experiment are the set of cities in the India with a population over 1 million, our sample space is written
- $S = \{x \mid x$ is a city in India with a population over 1 million$\}$,

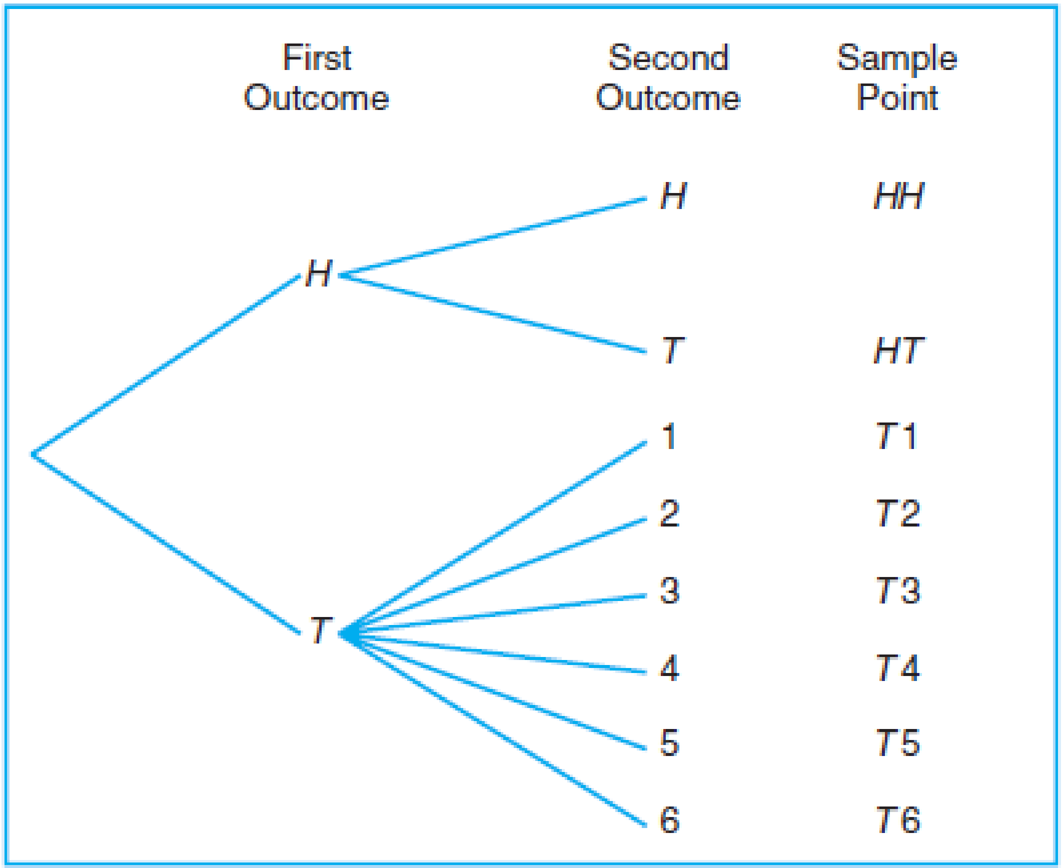
- Similarly, if S is the set of all points (x, y) on the boundary or the interior of a circle of radius 2 with center at the origin, we can write the **rule**

$$S = \{(x, y) \mid x^2 + y^2 \leq 4\}.$$

Sample space by means of tree diagram

- In some experiments, it is helpful to list the elements of the sample space systematically by means of a **tree diagram**.

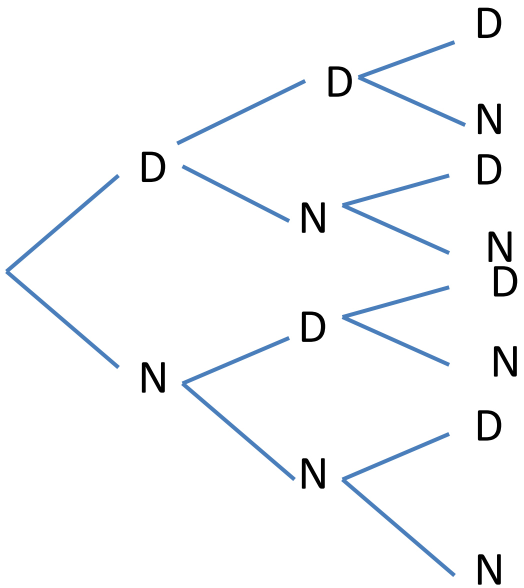
- Example: An experiment consists of flipping a coin and then flipping it a second time if a head occurs. If a tail occurs on the first flip, then a die is rolled once.



S = {*HH*, *HT*, *T*1, *T*2, *T*3, *T*4, *T*5, *T*6}.

Exercise

- Suppose that three items are selected at random from a manufacturing process. Each item is inspected and classified as defective, D, or non-defective, N. What will be the sample space? You may use tree diagram.



$S = \{DDD, DDN, DND, DNN, NDD, NDN, NND, NNN\}.$

- Four students are selected at random from a chemistry class and classified as male or female. List the elements of the sample space $S1$, using the letter M for male and F for female. Define a second sample space $S2$ where the elements represent the number of females selected.
- S1 = {MMMM, MMMF, MMFM, MFMM, FMMM, MMFF, MFMF, MFFM, FMFM, FFMM, FMMF, MFFF, FMFF, FFMF, FFFM, FFFF}.
- S2 = {0, 1, 2, 3, 4}.

Events

- An event is a collection of sample points, which constitute a subset of the sample space where some desired condition(s) is satisfied.

- For any given experiment, we may be interested in the occurrence of certain **events** rather than in the occurrence of a specific element in the sample space. For instance, we may be interested in the event A that the outcome when a die is tossed is even.

- This will occur if the outcome is an element of the subset $A = \{2, 4, 6\}$ of the sample space $S1=\{1,2,3,4,5,6\}$.

- Given the sample space $S = \{t \mid t \geq 0\}$, where t is the life in years of a certain electronic component, then the event A that the component fails before the end of the fifth year is the subset $A = \{t \mid 0 \leq t < 5\}$.

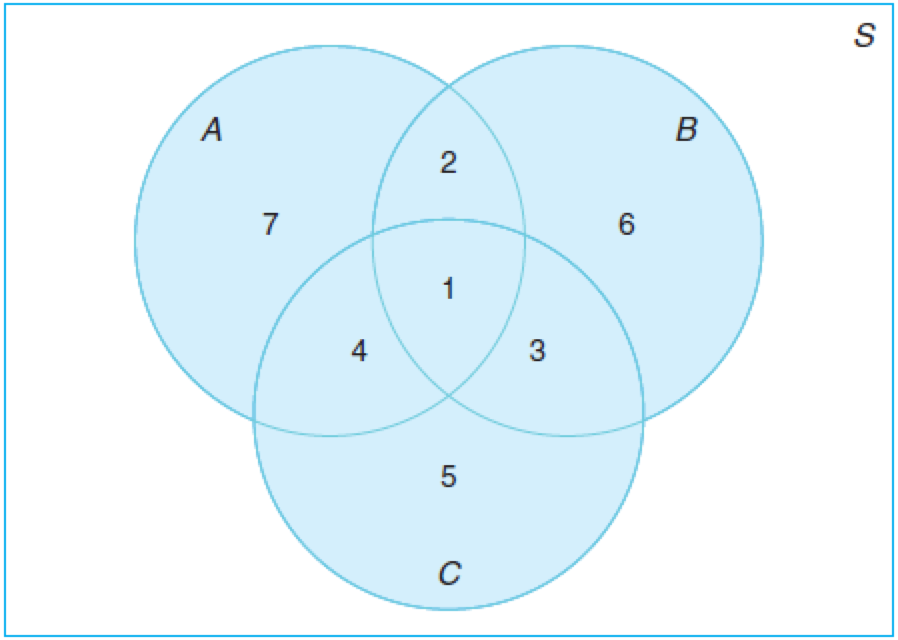
- The **complement** of an event A with respect to S is the subset of all elements of S that are not in A. We denote the complement of A by the symbol A'.

- The **intersection** of two events *A* and *B*, denoted by the symbol $A \cap B$, is the event containing all elements that are common to *A* and *B*.

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

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

The relationship between events and the corresponding sample space can be illustrated graphically by means of **Venn diagrams**.



- $A \cup C$ = 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,

Exercise

- The resumes of two male applicants for a college teaching position in chemistry are placed in the same file as the resumes of two female applicants. Two positions become available, and the first, at the rank of assistant professor, is filled by selecting one of the four applicants at random. The second position, at the rank of instructor, is then filled by selecting at random one of the remaining three applicants. Using the notation $M2F1$, for example, to denote the simple event (sample element) that the first position is filled by the second male applicant and the second position is then filled by the first female applicant,

a) list the elements of a sample space S;

b) list the elements of S corresponding to event A that the position of assistant professor is filled by a male applicant;

c) list the elements of S corresponding to event B that exactly one of the two positions is filled by a male applicant;

a) S = {M1M2,M1F1,M1F2,M2M1,M2F1,M2F2, F1M1, F1M2, F1F2, F2M1, F2M2,F2F1}.

b) A = {M1M2,M1F1,M1F2,M2M1,M2F1,M2F2}.

c) B = {M1F1,M1F2,M2F1,M2F2, F1M1, F1M2, F2M1, F2M2}.

Counting Sample Points

- One of the problems that the statistician must consider and attempt to evaluate is the chance associated with the occurrence of certain events when an experiment is performed. These problems belong in the field of probability.

- In many cases, we shall be able to solve a probability problem by counting the number of points in the sample space without actually listing each element.

Example

- How many sample points are there in the sample space when a pair of dice is thrown once?

- There are 6 possible outcome from first die and for each of these outcomes, there are six possible outcome for second die. So there are total 36 possible outcomes or sample points.

Multiplication rule: 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.

More Example

- If a 22-member club needs to elect a chairman and a treasurer, how many different ways can these two posts be elected?

- Chairman can be chosen in 22 ways. After that there will be 21 choices left so treasurer can be chosen in 21 ways. So one chairman and one treasurer can be chosen in 22*21 ways.

Generalization of multiplication rule

- If an operation can be performed in $n1$ ways, and if for each of these a second operation can be performed in $n2$ ways, and for each of the first two a third operation can be performed in $n3$ ways, and so forth, then the sequence of k operations can be performed in $n1n2 \cdot \cdot \cdot nk$ 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?

Permutation

- 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.*
- *So there are 6 possible arrangements. We can get this number using generalized multiplication rule.*
- There are $n1 = 3$ choices for the first position, no matter which letter is chosen. there are always $n2 = 2$ choices for the second position and only one choice for third position n3=1.
- Total number of possible arrangement = 3*2*1=6

- In general, n distinct objects can be arranged in $n(n-1)(n-2)\cdots(3)(2)(1)$ ways
- For any non-negative integer n, $n!$, called "n factorial," is defined as
- $n! = n(n-1)\cdots(2)(1)$,
- with special case $0! = 1$.

Permuting r letter out of n

- Let us say there are 4 letters a,b,c,d. We are allowed to take any two letters from these 4 letters and arrange them in all possible ways. So what are possible arrangements?
- *ab*, *ac*, *ad*, *ba*, *bc*, *bd*, *ca*, *cb*, *cd*, *da*, *db*, *dc*.
- This can also be derived from multiplication rule.
- For first position we have 4 (n1=4) choices and for second position we have 3 (n2=3) choices. So we have total $n1n2 = (4)(3) = 12$ choices.

- In general, *n* distinct objects taken *r* at a time can be arranged in
- $n(n - 1)(n - 2) \cdot\cdot\cdot (n - r + 1)$=
- [n(n-1)....1]/[(n-r)...1]=n!/(n-r)!
- We represent this product by the symbol

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

- For last example n=4 r=2 So $_{4}P_{2}$=4!/2!=24/2=12

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?
- It is equivalent to choose 3 persons out of 25 and arrange them. n=25, r=3
- ${}_{25}P_3 = 25!/22! = 25*24*23 = 13800$

Combination

- My fruit salad is a combination of apples, grapes and bananas“. We don't care what order the fruits are in, they could also be "bananas, grapes and apples" or "grapes, apples and bananas", its the same fruit salad.
- When the order doesn't matter, it is a Combination.
- Choosing k elements out of n elements can be done in ${}^{n}C_{k}$ ways. We normally read it as n choose k.
- ${}^{n}C_{k} = \frac{n*(n-1)*\cdots*(n-k+1)}{k*(k-1)*\cdots*1} = \frac{n!}{k!*(n-k)!}$

Example

- Number of ways you can choose 2 aces out of 4 aces $= {}^{4}C_{2} = \frac{4!}{2! * 2!} = \frac{24}{4} = 6$
- Number of ways you can choose 3 clubs out of 13 clubs $= {}^{13}C_{3} = \frac{13!}{3! * 10!} = \frac{13 * 12 * 11}{6} = 286$

Exercise

- president and a treasurer are to be chosen from a student club consisting of 50 people. How many different choices of officers are possible if

I. *A* will serve only if he is president

II. *B* and *C* will serve together or not at all

- Since *A* will serve only if he is president, we have two situations here: (i) *A* is selected as the president, which yields 49 possible outcomes for the treasurer's position, or (ii) officers are selected from the remaining 49 people without *A*, which has the number of choices ${}^{49}P_2 = (49)(48) = 2352$. Therefore, the total number of choices is $49 + 2352 = 2401$.

- The number of selections when B and C serve together is 2. The number of selections when both B and C are not chosen is ${}^{48}P_2$ = 2256. Therefore, the total number of choices in this situation is 2 + 2256 = 2258.

- 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 k^{th} kind is $\frac{n!}{n_1! * n_2! * \cdots * n_k!}$
- In a fruit bucket there are 10 fruits. Among these 10 fruits, there are 1 mango, 2 guava, 4 pine apple, and 3 apples. How many different ways fruits can be arranged in a row if there is no way to distinguish among fruits of same type? $\frac{10!}{1! * 2! * 4! * 3!}$

How counting helps in probability?

Probability of an Event

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

$0 \leq P(A) \leq 1$, $P(\varphi) = 0$, and $P(S) = 1$.

- Furthermore, if $A1, A2, A3, \ldots$ is a sequence of mutually exclusive events, then

$$P(A1 \cup A2 \cup A3 \cup \cdots) = P(A1) + P(A2) + P(A3) + \cdots$$

Example

- A coin is tossed twice. What is the probability that at least 1 head occurs?
- The sample space for this experiment is

  $S = \{HH, HT, TH, TT\}$.
- If the coin is balanced, each of these outcomes is **equally likely** to occur. Therefore, we assign a probability/weight ω to each sample point. Then $4\omega = 1$ (as outside this 4 outcomes nothing can happen), or $\omega = 1/4$. If A represents the event of at least 1 head occurring, then Event A comprises of sample point HH, HT and TH
- $A = \{HH, HT, TH\}$ and $P(A) = \omega + \omega + \omega = 1/4+1/4+1/4=3/4$.

- If the outcomes of an experiment are not equally likely to occur, the probabilities/weight must be assigned on the basis of prior knowledge or experimental evidence. For example, if a coin is not balanced, we could estimate the probabilities of heads and tails by tossing the coin a large number of times and recording the outcomes.

Example

- A die is loaded in such a way that an even number is twice as likely to occur as an odd number. If E is the event that a number less than 4 occurs on a single toss of the die, find $P(E)$.
- The sample space is $S = \{1, 2, 3, 4, 5, 6\}$. We assign a probability of w to each odd number and a probability of $2w$ to each even number. Since the sum of the probabilities must be 1, we have $9w = 1$ or $w = 1/9$. Hence, probabilities of 1/9 and 2/9 are assigned to each odd and even number, respectively. Therefore,

$E = \{1, 2, 3\}$ and $P(E) = 1/9 + 2/9 + 1/9 = 4/9$

- let A be the event that an even number turns up and let B be the event that a number divisible by 3 occurs. Find $P(A \cup B)$ and $P(A \cap B)$. Use sample space and probability of last example.
- For the events $A = \{2, 4, 6\}$ and $B = \{3, 6\}$, we have

  $A \cup B = \{2, 3, 4, 6\}$ and $A \cap B = \{6\}$.
- $P(A \cup B) = P(2)+P(3)+P(4)+P(6)$

  $=2/9+1/9+2/9+2/9=7/9$
- $P(A \cap B) = P(6) =2/9$

How counting helps?

- If an experiment can result in any one of N different equally likely outcomes, and if exactly n of these outcomes correspond to event A, then the probability of event A is
- $P(A) = n/N$.

Exercise

- president and a treasurer are to be chosen from a student club consisting of 50 people. How many different choices of officers are possible if

I. *Event E1: A* will serve only if he is president

II. *Event E2: B* and *C* will serve together or not at all

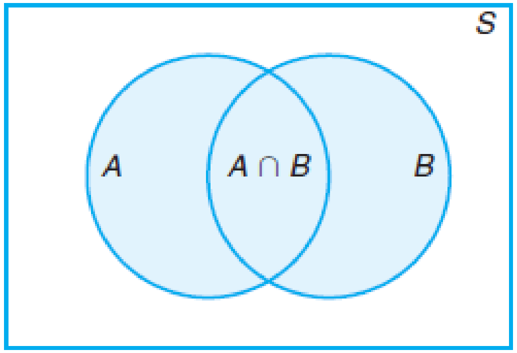
- The total number of choices of officers, without any restrictions, is $^{50}P_2 = 50!/48! = (50)(49) = 2450$.
- Since *A* will serve only if he is president, we have two situations here: (i) *A* is selected as the president, which yields 49 possible outcomes for the treasurer's position, or (ii) officers are selected from the remaining 49 people without *A*, which has the number of choices $49P2 = (49)(48) = 2352$. Therefore, the total number of choices is $49 + 2352 = 2401$.

- Probability of event E1=2401/2450

- The number of selections when *B* and *C* serve together is 2. The number of selections when both *B* and *C* are not chosen is ${}^{48}P_2$ = 2256. Therefore, the total number of choices in this situation is 2 + 2256 = 2258.
- Probability of event E2 = 2258/2450.

Addition Rule

- If A and B are two events, then
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.



- For three events A, B, and 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 $A1, A2, \ldots, An$ is a partition of sample space S, then
- $P(A1 \cup A2 \cup \cdots \cup An) = P(A1) + P(A2) + \cdots + P(An) = P(S) = 1.$

Example

- John is going to graduate from an industrial engineering department in an university by the end of this semester. After being interviewed at two companies he likes, he assesses that his probability of getting an offer from company *A* is 0.8, and his probability of getting an offer from company *B* is 0.6. If he believes that the probability that he will get offers from both companies is 0.5, what is the probability that he will get at least one offer from these two companies?
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.8 + 0.6 - 0.5 = 0.9.$

More Example

- What is the probability of getting a total of 7 or 11 when a pair of fair dice is rolled?
- Let A be the event that 7 occurs and B the event that 11 comes up. Now, a total of 7 occurs for 6 out of the 36 sample points, and a total of 11 occurs for only 2 of the sample points. Since all sample points are equally likely, we have $P(A) = 1/6$ and $P(B) = 1/18$. The events A and B are mutually exclusive, since a total of 7 and 11 cannot both occur on the same toss. Therefore,

$$P(A \cup B) = P(A) + P(B) = 1/6 + 1/18 = 2/9.$$

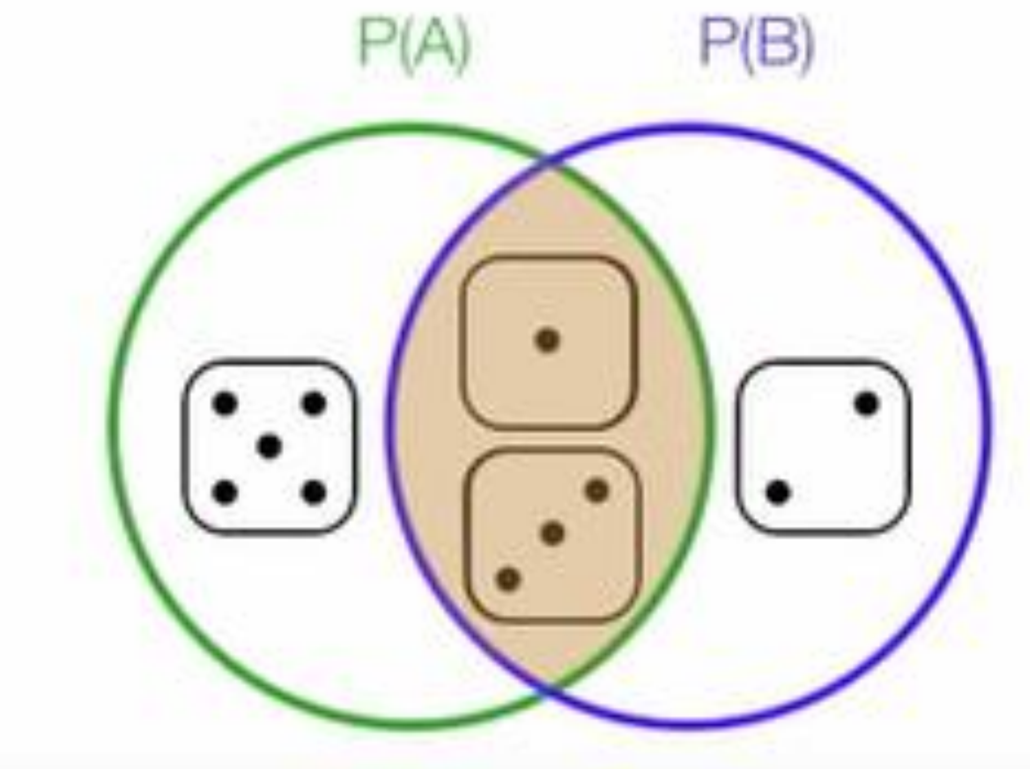
- If A and A' are complementary events, then $P(A) + P(A') = 1$.
- **Example**: If the probabilities that an automobile mechanic will service 3, 4, 5, 6, 7, or 8 or more cars on any given workday are, respectively, 0.12, 0.19, 0.28, 0.24, 0.10, and 0.07, what is the probability that he will service at least 5 cars on his next day at work?
- Let E be the event that at least 5 cars are serviced. Now, $P(E) = 1 - P(E')$, where E' is the event that fewer than 5 cars are serviced. Since
- $P(E') = 0.12 + 0.19 = 0.31$,
- $P(E) = 1 - P(E') = 1 - 0.31 = 0.69$.

Conditional Probability

- Probability of an event say event B occurring, when it is known that some event A is occurred is called a **conditional probability** and is denoted by $P(B|A)$.

- The symbol $P(B|A)$ is usually read "the probability that B occurs given that A occurs" or simply "the probability of B, given A."

Computing P(B|A)



New sample space A= {1,3,5}
Event X = {1,3}
Probability of event X = 2/3

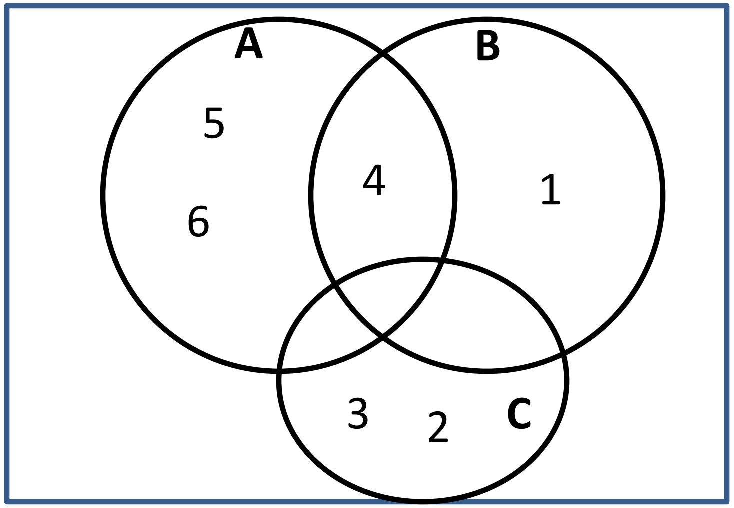
Old Sample space S=AUB= {1,2,3,5}
Probability of A∩B = 2/4
Probability of A = ¾
Probability of (B|A)=P(A∩B)/P(A)=2/3

P(B|A)=$P(A \cap B)/P(A)$ [provided $P(A) > 0$]

- Consider the event *B* of getting a perfect square when a die is rolled. The die is constructed so that the even numbers are twice as likely to occur as the odd numbers. Now suppose that it is known that the toss of the die resulted in a number greater than 3.

- New sample space A={4,5,6}. With the condition that even number appears with twice probability compared to odd numbers,
- probability of 4 and 6=2/5
- and probability of 5 is 1/5.
- P(B|A)=probability of {4(perfect square in new sample space)}=2/5.

- Actual sample space S={1,2,3,4,5,6}
- Probability of ($A \cap B$) = probability of {4} = 2/9
- Probability of A = probability of {4,5,6} = 5/9
- P(B|A)=$P(A \cap B)/P(A)$ = 2/5.



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

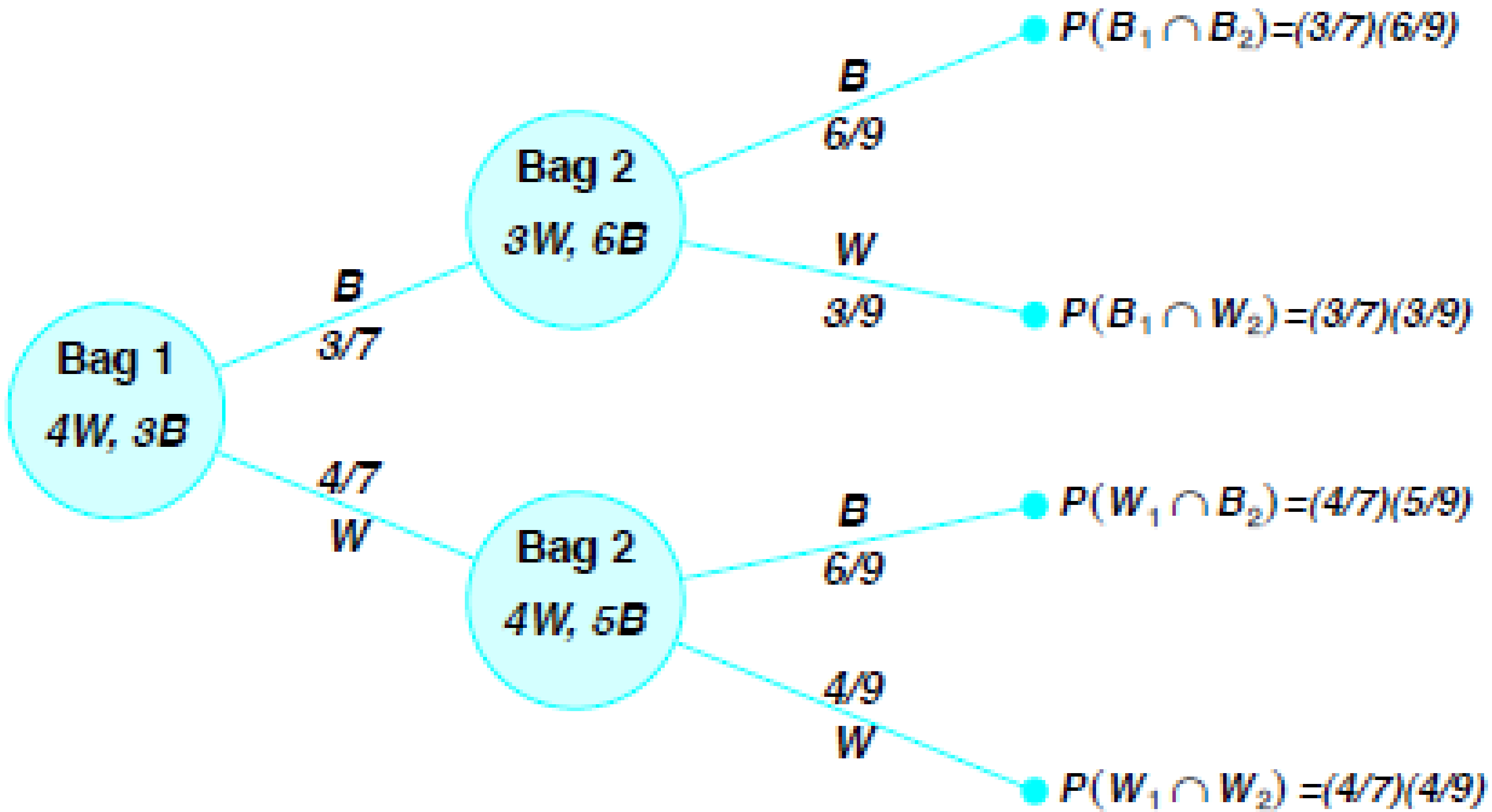
Table : Categorization of the Adults in a Small Town

- M: a man is chosen,
- E: the one chosen is employed.
- Using the reduced sample space E, we find that $P(M|E)$ =460/600=23/30
- *Using original sample space* $P(M|E) = P(E \cap M)/P(E)$
- $P(E \cap M)$=460/900=23/45.
- P(E)=600/900=2/3.
- $P(M|E)$ = 23/30.

- If in an experiment the events *A* and *B* can both occur, then
- $P(A \cap B) = P(A)P(B|A)$, provided $P(A) > 0$.

Example

- One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. What is the probability that a ball now drawn from the second bag is black?
- Let $B1$, $B2$, and $W1$ represent, respectively, the drawing of a black ball from bag 1, a black ball from bag 2, and a white ball from bag 1. We are interested in the union of the mutually exclusive events $B1 \cap B2$ and $W1 \cap B2$.
- $P[(B1 \cap B2) \text{ or } (W1 \cap B2)] = P(B1 \cap B2) + P(W1 \cap B2)$
- $= P(B1)P(B2|B1) + P(W1)P(B2|W1) = (3/7)*(6/9) + (4/7)*(5/9) = 38/63$



Bag 1
4W, 3B
B
3/7
4/7
W
Bag 2
3W, 6B
Bag 2
4W, 5B
B
6/9
W
3/9
B
6/9
4/9
W
P(B1 ∩ B2)=(3/7)(6/9)
P(B1 ∩ W2)=(3/7)(3/9)
P(W1 ∩ B2)=(4/7)(5/9)
P(W1 ∩ W2) =(4/7)(4/9)

Exercise

- The probability that a regularly scheduled flight departs on time is $P(D) = 0.83$; the probability that it arrives on time is $P(A) = 0.82$; and the probability that it departs and arrives on time is $P(D \cap A) = 0.78$. Find the probability that a plane arrives on time, given that it departed on time
- The probability that a plane arrives on time, given that it departed on time, is
- $P(A|D) = P(D \cap A)/P(D) = 0.78/0.83 = 0.94.$

Independent Event

- **Independent Event**: Knowing the probability of one event does not change the probability of the other event.

- Consider tossing a coin for two times S={HH,HT,TH,TT}
- A: We get H at first toss ={HH,HT}
- B: We get T at last toss = {HT,TT}
- The two events are not mutually exclusive as $A \cap B=\{HT\}$
- These two events will be called independent event if $P(A|B)=P(A)$ and $P(B|A)=P(B)$

- $P(A|B) = P(A \cap B)/P(B) = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2} = P(A)$
- $P(B|A) = P(A \cap B)/P(A) = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2} = P(B)$
- So these two events are independent.
- When two events are independent then
- $P(A \cap B) = P(A) * P(B)$

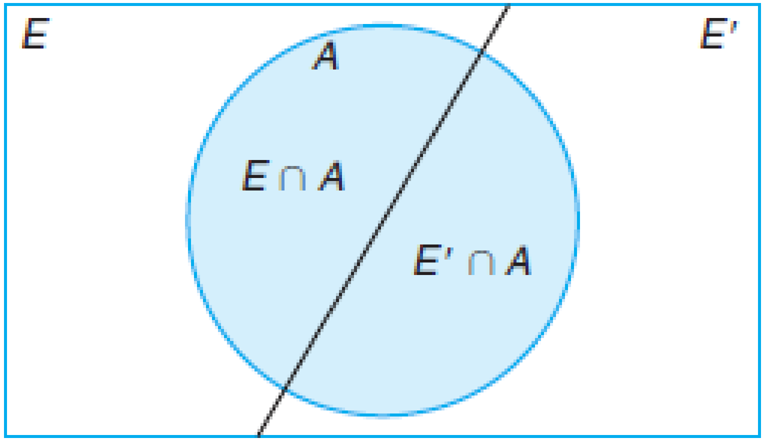
Point to Note

- Let us say we have a dataset containing gender and handedness (left handed vs. right handed). It may seem like a person's gender and whether or not they are left-handed are totally independent events. When we look at probabilities though, we see that about 10% of all people are left-handed, but about 12% of males are left-handed. So these events are not independent, since knowing a random person is a male increases the probability that they are left-handed.

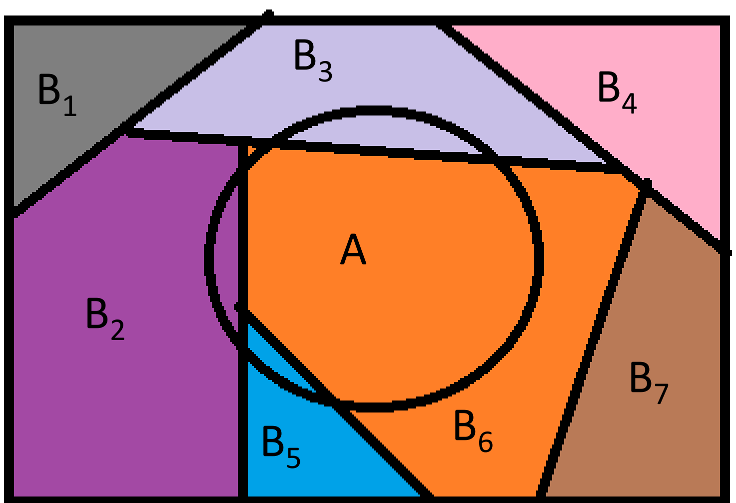
Example

- A small town has one fire engine and one ambulance available for emergencies. The probability that the fire engine is available when needed is 0.98, and the probability that the ambulance is available when called is 0.92. In the event of an injury resulting from a burning building, find the probability that both the ambulance and the fire engine will be available, assuming they operate independently.
- Let A and B represent the respective events that the fire engine and the ambulance are available. Then
- $P(A \cap B) = P(A)P(B) = (0.98)(0.92) = 0.9016.$

Total Probability



$$P(A) = P[(E \cap A) \cup (E' \cap A)]$$
$$= P(E \cap A) + P(E' \cap A)$$
$$= P(E)P(A|E) + P(E')P(A|E').$$



A=(A∩B_1)∪(A∩B_2)∪(A∩B_3)∪(A∩B_4)∪(A∩B_5)∪(A∩B_6)∪(A∩B_7)=$\sum_{i=1}^{7}$ (A∩B_i)

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

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$$

Example

- In a certain assembly plant, three machines, $B1$, $B2$, and $B3$, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective?
- A: the product is defective,
- $B1$: the product is made by machine $B1$,
- $B2$: the product is made by machine $B2$,
- $B3$: the product is made by machine $B3$.
- $P(A) = P(B1)P(A|B1) + P(B2)P(A|B2) + P(B3)P(A|B3)$.

- $P(B1)P(A|B1) = (0.3)(0.02) = 0.006,$
- $P(B2)P(A|B2) = (0.45)(0.03) = 0.0135,$
- $P(B3)P(A|B3) = (0.25)(0.02) = 0.005,$
- and hence
- $P(A) = 0.006 + 0.0135 + 0.005 = 0.0245.$
- Instead of asking for $P(A)$, suppose that we now consider the problem of finding the conditional probability $P(Bi|A)$. In other words, suppose that a product was randomly selected and it is defective. What is the probability that this product was made by machine Bi?

Bay’s Theorem

(Bayes’ Rule) If the events $B_1, B_2, \ldots, B_k$ 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 $P(A) \neq 0$,

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

- With reference from previous example, if a product was chosen randomly and found to be defective, what is the probability that it was made by machine *B*3?
- Using Bayes’ rule to write

$$P(B_3|A) = \frac{P(B_3)P(A|B_3)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)}$$

$$P(B_3|A) = \frac{0.005}{0.006 + 0.0135 + 0.005} = \frac{0.005}{0.0245} = \frac{10}{49}$$