Chapter 10: Probability of Events

1. (10.1) Sample Spaces & Events
2. (10.2) Probability of an Event
    a) Examples from Genetics
3. (10.3) Combinations & Permutations

Example 10.1 (Coin Flipping)

We start with an example (Example 10.1):

What are the possible outcomes of flipping a coin once? twice?

Solution:

- Obviously, when we flip a coin once, there are only two possible outcomes- heads or tails.
- If we flip a coin twice, we have four possible outcomes- heads then heads, heads then tails, tails then heads, and tails then tails.

Experiment, Sample Space, Elementary Event

- The act of flipping a coin once and the act of flipping a coin twice are examples of an **experiment**.
- The set of all possible outcomes of an experiment is called the **sample space** of the experiment.
- The (individual) elements of the sample space are called **elementary events**.
- Thus if we let 'H' stand for 'heads' and 'T' for 'tails,' then the sample space for the experiment of flipping a coin once is given by $S_1 = \{H, T\}$, where H and T are elementary events.
- The sample space for the experiment of flipping a coin twice is given by $S_2 = \{HH, HT, TH, TT\}$, where HH, HT, TH and TT are elementary events.

Events

- Any subset of a sample space is an **event**. For example, consider the generic sample space, $S = \{e_1, e_2, e_3, e_4\}$, where the e_i are elementary events for some experiment.
- Here are all of the events associated with this sample space:

$$\varnothing,$$
$$\{e_1\},\{e_2\},\{e_3\},\{e_4\},$$
$$\{e_1,e_2\},\{e_1,e_3\},\{e_1,e_4\},\{e_2,e_3\},\{e_2,e_4\},\{e_3,e_4\}$$
$$\{e_1,e_2,e_3\},\{e_1,e_2,e_4\},\{e_1,e_3,e_4\},\{e_2,e_3,e_4\}$$
$$\{e_1,e_2,e_3,e_4\}$$

- Elementary events are events
- There are 2^n events associated with a sample space of n elementary events; in this case, $2^4 = 16$

Example 10.2

- Suppose a coin is flipped twice. Identify the event "E = 1st and 2nd flips match"?
- Solution: Recall the sample space for this experiment:

$$S = \{HH, HT, TH, TT\}$$

The event E will be the subset of S containing the elementary events with matching flips. That is,

$$E = \{HH, TT\}$$

Example 10.3

- Suppose a single die is rolled. What is the sample space? Identify the event, "Rolled an odd number greater than two."
- Solution: The sample space is the set of all possible outcomes of rolling a die:

$$S = \{1,2,3,4,5,6\}$$

- The event is the subset of the sample space consisting of all the elementary events satisfying "odd number greater than two":

$$E = \{3,5\}$$

Example 10.3 (continued)

- Suppose two dice are rolled. What is the sample space? Identify the event, "dice sum to seven."
- Solution: The sample space is the set of all possible outcomes of rolling two dice:

$$S = \{ (1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),$$
$$(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),$$
$$(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$$

- The event is the subset of the sample space consisting of all the elementary events satisfying "dice sum to seven":

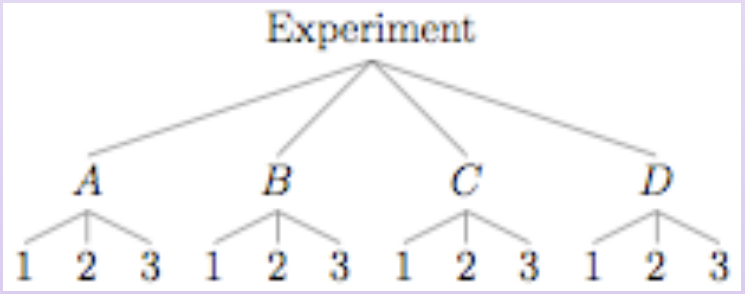
$$E = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

Multiplication Principle

- In the previous example, there were two actions. Action 1 was rolling the first die and action 2 was rolling the second die. For action 1 there were six possible outcomes, $S = \{1,2,3,4,5,6\}$. For action 2 there were the same possible outcomes. Notice this resulted in 6 × 6 = 36 total elementary events for the experiment of rolling two dice. This is an example of the **multiplication principle**:
- If action 1 in an experiment has m possible outcomes and action 2 in the experiment has n possible outcomes, then action 1 followed by action 2 has mn possible outcomes.

Multiplication Principle

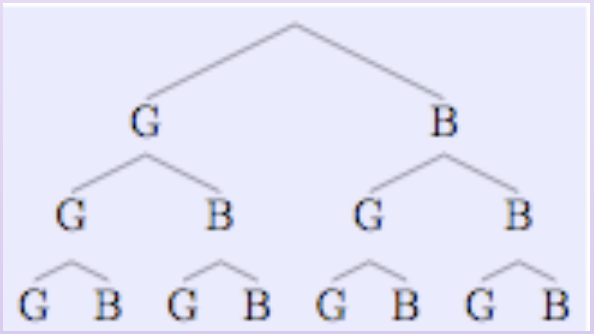
- For example, consider the generic experiment consisting of two actions. The 1st action has 4 possible outcomes and the 2nd action has 3 possible outcomes.
- Then, by the multiplication principle, the sample space of this experiment will consist of 4×3 = 12 elementary events.
- To illustrate, suppose the possible outcomes of action 1 are A, B, C, and D; and the possible outcomes of action 2 are 1, 2, and 3:



$$S = \{A1, A2, A3, B1, B2, B3, C1, C2, C3, D1, D2, D3\}$$

Example

- A couple plans to have 3 children. How many possible orders with respect to gender are there for three children?
- Solution: There are 3 actions in this experiment- having their 1st child, then their 2nd, and then their 3rd. Each action has 2 possible outcomes- boy or girl. By the multiplication principle, then, there are 2×2×2=8 elementary events in the sample space of this experiment.



$$S = \{GGG, GGB, GBG, GBB, BGG, BGB, BBG, BBB\}$$

Example (continued)

- For the same experiment, identify the event "the couple has 3 children, exactly two of which are girls"
- Solution: The event is the subspace of the sample space,

$$S = \{GGG, GGB, GBG, GBB, BGG, BGB, BBG, BBB\},$$

containing each of the elements with exactly two G's:

$$E = \{GGB, GBG, BGG\}$$

- Now identify the event "have at least two girls"

$$E = \{GGG, GGB, GBG, BGG\}$$

Probability Function

- In order to determine the probability of an event for some experiment, we need to first *assign* probabilities to the elementary events comprising the sample space
- Given a sample space $S = \{e_1, e_2, \ldots, e_n\}$, a **probability function** ω is a rule that assigns to each elementary event e_i a real number $p_i = \omega(e_i)$ satisfying:

1. $0 \le p_i \le 1$
2. $\sum_{i=1}^{n} p_i = p_1 + p_2 + \cdots + p_n = 1$

- We are now in a position to define the probability of an event

Probability of an Event

- Let $E = \{s_1, s_2, \ldots, s_k\}$ be an event where 0 ≤ k ≤ n and each s_i is an elementary event in the sample space S. Then we define the probability of E, $P(E) = \omega(s_1) + \omega(s_2) + \cdots + \omega(s_k)$, the sum of the probabilities of the elementary events constituting E.
- We use the symbol ω because we can think of the probability function as a weighting function, i.e. each elementary event is weighted.
- We call the sample space **uniform** or **equiprobable** if equal weights are assigned to each elementary event $e_1, \ldots, e_n$. That is, $\omega(e_1) = \omega(e_2) = \cdots = \omega(e_n) = \frac{1}{n}$.
- It follows that: $P(E) = \frac{|E|}{n}$ ← the # of elements in the set E

Example

- Recall the previous example: Sample space for gender of 3 (ordered) kids:

$$S = \{GGG, GGB, GBG, GBB, BGG, BGB, BBG, BBB\}$$

- Event that 2 are girls:

$$E_1 = \{GGB, GBG, BGG\}$$

- Event that at least 2 are girls:

$$E_2 = \{GGG, GGB, GBG, BGG\}$$

- What is the probability that exactly two out of the three children are girls? at least two out of the three children are girls?
- Solution: The elementary events in this case are equiprobable; i.e. $P(e_i) = 1/8$

Example (continued)

- Thus we have:

$$P(E_1) = \frac{\left|\{GGB, GBG, BGG\}\right|}{8} = \frac{3}{8}$$

and

$$P(E_2) = \frac{\left|\{GGG, GGB, GBG, BGG\}\right|}{8} = \frac{1}{2}$$

Example 10.6 (Rolling Three Dice)

- If three dice are thrown, find the probability that the sum of three dice is six.
- Solution: Each roll of the three dice has six possible outcomes: {1, 2, 3, 4, 5, 6}. By the multiplication principle, there are a total of $6 \times 6 \times 6 = 216$ elementary events.
- Since we are rolling dice we will assume that each elementary event is equally likely.
- Now, let E be the event that the three dice sum to six. We can use a tree diagram to determine how many elementary events are in E. Each path from the top to a bottom node represents one way in which the three dice can sum to six.

Example 10.6 (continued)

We see that there are 10 paths from the top to a bottom node (we determine this by counting the number of bottom nodes). Thus,

$$P(E) = \frac{10}{216} \approx 0.046$$

Terminology

- **Genes** are genetic material on a chromosome that code for a trait. For example, you have a gene for eye color.
- An **allele** is one member of a pair that is located at a specific position on a specific chromosome.
  - Each gene contains one or more pairs of alleles.
  - Some alleles are dominant over others, these are known as dominant alleles.
  - The alleles that are not dominant are known as recessive alleles.
  - For notation, we will use capital letters to denote dominant alleles and lower case letters to denote recessive alleles.
  - For example, for the eye color gene the allele for brown eyes, B, is dominant over the allele for blue eyes, b.
  - For a pair of alleles on a gene, one allele is inherited from the father and the other is inherited from the mother. If a person inherits both the dominant and the recessive alleles, the dominant allele will be the one expressed.

Terminology

- A **genotype** is the actual set of alleles an organism carries.
  - Since each gene contains one or more pairs of alleles, the genotype is expressed as a pair (or pairs) of letters that represent the pair (or pairs) of alleles for that particular gene.
  - For example, for the eye color gene the allele for brown eyes, B, is dominant over the allele for blue eyes, b. Thus, the possible genotypes are BB, Bb, and bb.
  - Genes that have two dominant alleles or two recessive alleles are known as **homozygous**. Genes that have one dominant allele and one recessive allele are known as **heterozygous**.
- A **phenotype** is the physical expression of a gene.
  - If you have the genotype BB you will have brown eyes since both alleles are for brown eyes. Likewise, if you have the genotype bb you will have blue eyes since both alleles are for blue eyes. However, if you have the genotype Bb with one allele for brown and one allele for blue, the dominant allele (B) will mask the recessive allele (b) and you will have the phenotype for brown eyes.

Terminology

- A **Punnett square** is a diagram used to show the potential genotypes resulting from a mating where the genotype of each of the parents is known.
  - For example, if a woman with brown eyes (genotype Bb) and a man with blue eye (genotype bb) mated, the Punnett square below shows the possible genotypes of their offspring:

		♀ B	b
♂	b	bB	bb
	b	bB	bb

- There are many genetic diseases, like sickle-cell anemia and albinism, which are only expressed in the phenotype if an individual has two recessive alleles. In these cases, an individual is referred to as a "**carrier**" of the genetic disease if their genotype for the disease contains one dominant and one recessive allele.
  - For example, the hemoglobin gene has dominant allele S and recessive allele s. A person is a carrier for sickle-cell anemia if they have genotype Ss.

Example 10.7 (Albinism)

- The gene determining albinism can have dominant allele A or recessive allele a. A set of parents both have genotype Aa, so each is a carrier of the defective allele a. Find the probability that their child will be (a) an albino, (b) a carrier.
- Solution: First draw a Punnett square:

		♀	
		A	a
♂	A	AA	Aa
	a	aA	aa

  - We assume that each of the four outcomes shown in the Punnett square is equally likely. Thus, P(AA) = ¼, P(Aa) = ¼, P(aA) = ¼, and P(aa) = ¼
  - The event that their child will be albino is E_1={aa}. Thus, P(E_1)= ¼
  - The event that their child will be a carrier is E_2={Aa,aA}. Thus, P(E_2)= ½

Blood Types

- The next example has to do with the genetics of blood types. Common blood type in humans is determined by three alleles: A, B, and O.
  - The O allele is recessive.
  - If alleles A and B are paired, neither dominates the other, and an additional blood type (type AB) is formed.
  - Possible genotypes are AA, BB, OO, AB, AO, and, BO. Alleles A and B each result in the production of their own antigens, while allele O is inactive.
  - The phenotypes are determined by the antigens produced:

Phenotype	Genotypes
A	AA AO
B	BB BO
AB	AB
O	OO

Example 10.8 (Blood Type)

- A husband and wife have blood types AO and AB, respectively. What is the probability that their child will have blood (a) type A, (b) type B, (c) type AB, and (d) O?
  - Solution: First, let us make a Punnett square of the possible offspring genotypes produced by this husband and wife:

♂ \ ♀	*A*	*B*
A	*AA*	*AB*
O	*OA*	*OB*

a) The phenotype of type A blood corresponds to genotypes AA and AO. E_1 = {AA, AO} thus $P(E_1)$ = ½

b) E_2 = {BB, BO} thus $P(E_2)$ = P(BB) + P(BO) = 0 + ¼ = ¼

c) E_3 = {AB} and $P(E_3)$ = P(AB) = ¼

d) E_4 = {OO} and $P(E_4)$ = P(OO) = 0

We can add the probabilities of elementary events; see example 10.7 (b)

Permutations- Order Important

- As we have seen, *counting* is an important component to solving the problems presented thus far
  - The multiplication principle is one tool that helps us count
- In the sample spaces considered so far, the *order* of actions constituting an experiment is important:
  - HT is **not** the same as TH in the coin flipping example
  - That is, HT & TH are two different permutations
  - A **permutation** is an ordered tuple
- However, the outcome of the first action does not change the possible outcome of subsequent actions:
  - If the first coin flipped is H, there are still the same two options for the second coin, H or T
  - That is, whatever the outcome, we *replace* that possible outcome before the next action
  - The multiplication principle *in this case* counts the number of permutations **with replacement**

Permutations

- It is not always that case that the outcome of the first action does not effect the outcome of subsequent actions:
  - Suppose five cards are dealt from a standard deck of 52 cards, and the order in which they were dealt is noted
  - The first card dealt could be any one of the 52 cards in the deck
  - The second card, however, there is already one card missing from the deck, so there are only 51 possible cards that could be dealt second
  - For the third card, there are now 2 cards missing from the deck, so there are only 50 possible cards that could be dealt third; etc.
  - By the Multiplication Principle, this deal can be made $52\times51\times50\times49\times48=311,875,200$ ways!
  - The multiplication principle *in this case* counts the number of permutations **without replacement**

Permutations: Factorial Notation

- It is useful to introduce factorials to help count. Recall:

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

- For example, 5! = 5×4×3×2×1 = 120
- Thus, the previous problem could be written:

$$52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 = \frac{52!}{47!} = \frac{52!}{(52-5)!}$$

- In general, the number of permutations length k selected from n objects is given by:

$$P[n,k] = \frac{n!}{(n-k)!}$$

Example 10.9 (Genome)

- Gene orders refer to the permutation of genome arrangement. Vesicular stomatitis virus (VSV) is a prototype RNA virus that encodes five genes (N-P-M-G-L). These five genes can occur in any gene order. How many different gene orders (i.e. permutations) are there?
- Solution: There are a total of n = 5 genes to choose from. We must place these genes in k = 5 slots where the order matters. Thus, there are

$$P[5,5] = \frac{5!}{(5-5)!} = 5! = 120$$

possible gene orders.

Note: 0! = 1

Combinations: Order not Important

- Sometimes the *order* of actions constituting an experiment is not important:
  - In a typical card game, the hand with, say, 3D, AS, 8C is the same as the hand with AS, 3D, 8C
  - That is, the two hands are the *same* combination
  - A **combination** is a tuple (not ordered)
- In general, the number of combinations length k selected from n objects is given by:

$$C[n,k] = \frac{n!}{k!(n-k)!}$$

Combinations: Order not Important

- To see where this formula comes from, recall that the number of 5-card hands *if we note the order* is given by:

$$P[52,5] = \frac{52!}{(52-5)!} = \frac{52!}{47!} = 311{,}875{,}200$$

- But in the context of a typical card game, the *order* in which the cards were dealt is irrelevant- we only regard *which* cards we were dealt.
- So, we need to divide out the repetitions that are counted using a permutation.

Combinations: Order not Important

- How many ways (orders) could a particular 5 card hand be dealt? That is, given 5 particular cards, how many ways can one order them? $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5!$
- That is, each different hand appears 120 times in our previous count. If we only want to count each distinct hand once, we divide by 120.
- That is, the number of different 5-card hands is given by:

$$C[52,5] = \frac{52!}{5!(52-5)!} = 2598960 = \frac{P[52,5]}{5!} = \frac{52!}{(52-5)!}$$

Notice that this is the formula for the number of permutations length 5 taken from 52 objects divided by 5!

Example

- Recall from a previous example we counted how many ways could occur the event, “have exactly two girls” in a family with three children:

$$E = \{GGB, GBG, BGG\}$$

- We now use combinations to answer this question. It is helpful to think of three slots and 2 letter G’s:
  - How many ways can we place the letters in the slots?
  - Or, there are 3 slots, how many ways can I choose 2 of them?

$$C[3,2] = \frac{3!}{2!(3-2)!} = \frac{3 \cdot 2 \cdot 1}{2 \cdot 1} = 3$$

Exercise 10.7

Suppose you want to plant 4 trees in a plot and you can choose from 10 different species. How many ways can the trees to be planted in the plot be chosen?

If we nitpick, we can make this question ambiguous enough to consider each of the three scenarios:

1. Permutations (Order important)
   a) With replacement
   b) Without replacement
2. Combinations (Order not important)

Exercise 10.7

1. Perhaps this is a very specific plot we have in mind and we are not only choosing trees, but we are choosing exactly how we will place them; i.e. order matters. There are 2 sub-cases to consider:

   a) Suppose there are plenty of each kind of tree and we are able to choose more than one of a particular species. The question in this case is, "How many permutations with replacement are there?"

   Solution: For each of my (ordered) slots, I have 10 choices. Thus, there are $10\times10\times10\times10=10^4=10000$ ways to choose the trees to be planted.

Exercise 10.7

1. Perhaps this is a very specific plot we have in mind and we are not only choosing trees, but we are choosing exactly how we will place them; i.e. order matters. There are 2 sub-cases to consider:

   b) Suppose there is only one of each kind of tree. The question in this case is, "How many permutations without replacement are there?"

   Solution: For my first (ordered) slot, I have 10 choices. But for my second slot, I have only 9 choices; etc. Thus, there are 10×9×8×7=5040 ways to choose the trees to be planted. Using the notation for permutations:

$$P[10,4] = \frac{10!}{(10-4)!} = \frac{10!}{6!} = 10 \times 9 \times 8 \times 7 = 5040$$

Exercise 10.7

2. Perhaps the way/order we are planting the trees is irrelevant. We simply want to know how many ways we can choose 4 trees from 10 species. The question in this case is, "How many combinations are there of 4 trees selected from 10 species?" (This is probably what the author of the question had in mind.)

   Solution:

$$C[10,4] = \frac{10!}{4!(10-4)!} = \frac{10!}{4!6!} = \frac{10 \times 9 \times 8 \times 7}{4 \times 3 \times 2 \times 1} = 210$$

Homework

- Chapter 10: 1-13