EEE 549- Probability and Stochastic Processes

Prof.Dr. Bayram ÜNAL

Istanbul Gelisim University

Faculty of Engineering and Architecture,

Electrical and Electronics Engineering

2024-2025 Fall Semester

Probability and Stochastic Processes

Chapter 2- Sequential Experiments

2.1 Tree Diagrams
2.2 Counting Methods
2.3 Independent Trials
2.4 Reliability Analysis

Chapter 2. Sequential Experiments

Many applications of probability refer to sequential experiments in which the procedure consists of many actions performed in sequence, with an observation taken after each action.

- Each action in the procedure together with the outcome associated with it can be viewed as a separate experiment with its own probability model.
- In analyzing sequential experiments we refer to the separate experiments in the sequence as *sub-experiments*.

2.1 Tree Diagrams

Tree diagrams display the outcomes of the sub-experiments in a sequential experiment.

- The labels of the branches are probabilities and conditional probabilities.
- The probability of an outcome of the entire experiment is the product of the probabilities of branches going from the root of the tree to a leaf.
- Many experiments consist of a sequence of *sub-experiments*.
- The procedure followed for each sub-experiment may depend on the results of the previous sub-experiments.

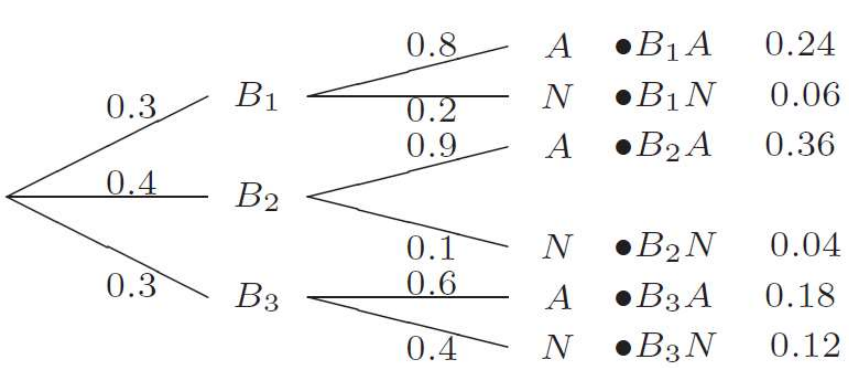
2.1 Tree Diagrams

Useful to use a type of graph referred to as a *tree diagram* to represent the sequence of sub-experiments.

- To do so, we assemble the outcomes of each sub-experiment into sets in a partition.
- Starting at the root of the tree,1 we represent each event in the partition of the first sub-experiment as a branch and we label the branch with the probability of the event.
- Each branch leads to a node.
- The events in the partition of the second sub-experiment appear as branches growing from every node at the end of the first sub-experiment.
- The labels of the branches of the second sub-experiment are the *conditional* probabilities of the events in the second sub-experiment. We continue the procedure taking the remaining sub-experiments in order.
- The nodes at the end of the final sub-experiment are the leaves of the tree. Each leaf corresponds to an outcome of the entire sequential experiment.
- The probability of each outcome is the product of the probabilities and conditional probabilities on the path from the root to the leaf.
- We usually label each leaf with a name for the event and the probability of the event.
- This is a complicated description of a simple procedure as we see in the following five examples.

Example 2.1

For the resistors of Example 1.19, we used A to denote the event that a randomly chosen resistor is "within 50 Ω of the nominal value." This could mean "acceptable." We use the notation N ("not acceptable") for the complement of A. The experiment of testing a resistor can be viewed as a two-step procedure. First we identify which machine (B_1, B_2, or B_3) produced the resistor. Second, we find out if the resistor is acceptable. Draw a tree for this sequential experiment. What is the probability of choosing a resistor from machine B_2 that is not acceptable?

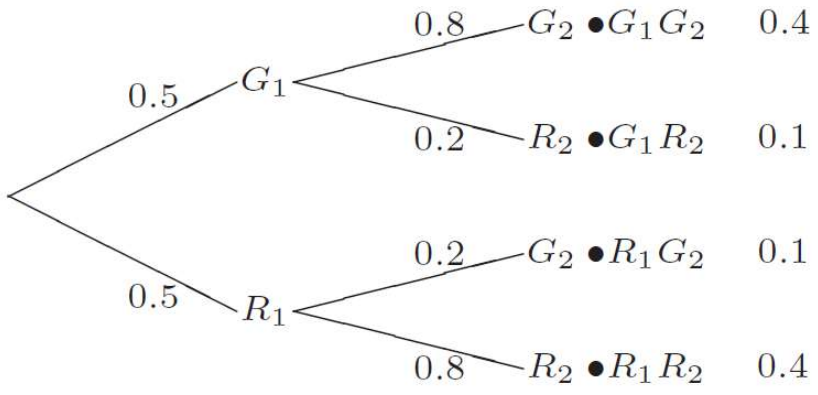


This two-step procedure is shown in the tree on the left. To use the tree to find the probability of the event B_2N, a nonacceptable resistor from machine B_2, we start at the left and find that the probability of reaching B_2 is $\mathrm{P}[B_2] = 0.4$. We then move to the right to B_2N and multiply $\mathrm{P}[B_2]$ by $\mathrm{P}[N|B_2] = 0.1$ to obtain $\mathrm{P}[B_2N] = (0.4)(0.1) = 0.04$.

We observe in this example a general property of all tree diagrams that represent sequential experiments. The probabilities on the branches leaving any node add up to 1. This is a consequence of the law of total probability and the property of conditional probabilities that corresponds to Axiom 3 (Theorem 1.7). Moreover, Axiom 2 implies that the probabilities of all of the leaves add up to 1.

Example 2.2

Traffic engineers have coordinated the timing of two traffic lights to encourage a run of green lights. In particular, the timing was designed so that with probability 0.8 a driver will find the second light to have the same color as the first. Assuming the first light is equally likely to be red or green, what is the probability $\mathrm{P}[G_2]$ that the second light is green? Also, what is $\mathrm{P}[W]$, the probability that you wait for at least one of the first two lights? Lastly, what is $\mathrm{P}[G_1|R_2]$, the conditional probability of a green first light given a red second light?



The tree for the two-light experiment is shown on the left. The probability that the second light is green is

$$\begin{aligned}\mathrm{P}[G_2] &= \mathrm{P}[G_1G_2] + \mathrm{P}[R_1G_2] \\ &= 0.4 + 0.1 = 0.5.\end{aligned}$$

The event W that you wait for at least one light is the event that at least one light is red.

$$W = \{R_1G_2 \cup G_1R_2 \cup R_1R_2\}.$$

The probability that you wait for at least one light is

$$\mathrm{P}[W] = \mathrm{P}[R_1G_2] + \mathrm{P}[G_1R_2] + \mathrm{P}[R_1R_2] = 0.1 + 0.1 + 0.4 = 0.6.$$

An alternative way to the same answer is to observe that W is also the complement of the event that both lights are green. Thus,

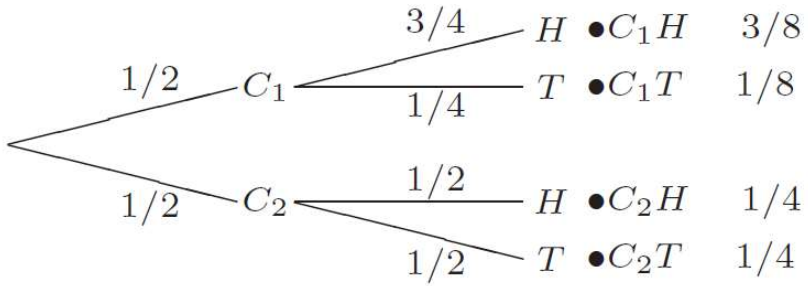
$$\mathrm{P}[W] = \mathrm{P}[(G_1G_2)^c] = 1 - \mathrm{P}[G_1G_2] = 0.6.$$

To find $\mathrm{P}[G_1|R_2]$, we need $\mathrm{P}[R_2] = 1 - \mathrm{P}[G_2] = 0.5$. Since $\mathrm{P}[G_1R_2] = 0.1$, the conditional probability that you have a green first light given a red second light is

$$\mathrm{P}[G_1|R_2] = \frac{\mathrm{P}[G_1R_2]}{\mathrm{P}[R_2]} = \frac{0.1}{0.5} = 0.2.$$

Example 2.3

Suppose you have two coins, one biased, one fair, but you don't know which coin is which. Coin 1 is biased. It comes up heads with probability $3/4$, while coin 2 comes up heads with probability $1/2$. Suppose you pick a coin at random and flip it. Let C_i denote the event that coin i is picked. Let H and T denote the possible outcomes of the flip. Given that the outcome of the flip is a head, what is $\mathrm{P}[C_1|H]$, the probability that you picked the biased coin? Given that the outcome is a tail, what is the probability $\mathrm{P}[C_1|T]$ that you picked the biased coin?



First, we construct the sample tree on the left. To find the conditional probabilities, we see

$$\begin{aligned}\mathrm{P}[C_1|H] &= \frac{\mathrm{P}[C_1H]}{\mathrm{P}[H]} \\ &= \frac{\mathrm{P}[C_1H]}{\mathrm{P}[C_1H]+\mathrm{P}[C_2H]}\end{aligned}$$

From the leaf probabilities in the sample tree,

$$\mathrm{P}[C_1|H] = \frac{3/8}{3/8+1/4} = \frac{3}{5}.$$

Similarly,

$$\mathrm{P}\left[C_1|T\right] = \frac{\mathrm{P}\left[C_1T\right]}{\mathrm{P}\left[T\right]} = \frac{\mathrm{P}\left[C_1T\right]}{\mathrm{P}\left[C_1T\right] + \mathrm{P}\left[C_2T\right]} = \frac{1/8}{1/8 + 1/4} = \frac{1}{3}. \tag{2.6}$$

As we would expect, we are more likely to have chosen coin 1 when the first flip is heads, but we are more likely to have chosen coin 2 when the first flip is tails.

The next example is the ”Monty Hall” game, a famous problem with a solution that many regard as counterintuitive. Tree diagrams provide a clear explanation of the answer.

Example 2.4 Monty Hall

In the Monty Hall game, a new car is hidden behind one of three closed doors while a goat is hidden behind each of the other two doors. Your goal is to select the door that hides the car. You make a preliminary selection and then a final selection. The game proceeds as follows:

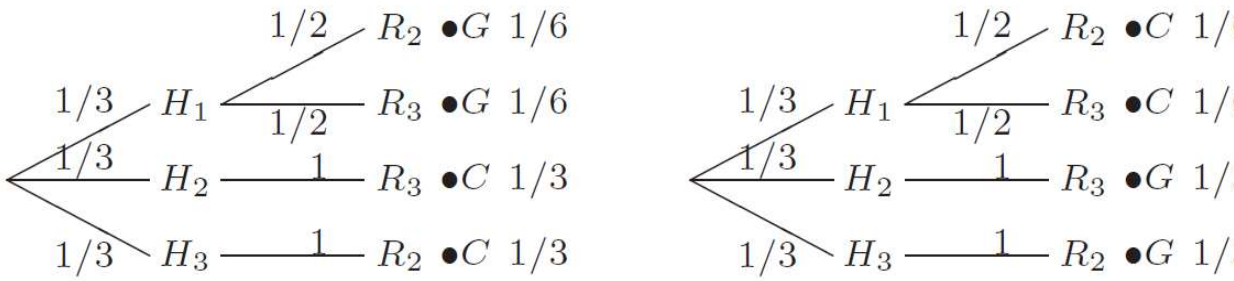
- You select a door.
- The host, Monty Hall (who knows where the car is hidden), opens one of the two doors you didn't select to reveal a goat.
- Monty then asks you if you would like to switch your selection to the other unopened door.
- After you make your choice (either staying with your original door, or switching doors), Monty reveals the prize behind your chosen door.

To maximize your probability $\mathrm{P}[C]$ of winning the car, *is switching to the other door either (a) a good idea, (b) a bad idea or (c) makes no difference?*

To solve this problem, we will consider the "switch" and "do not switch" policies separately. That is, we will construct two different tree diagrams: The first describes what happens if you switch doors while the second describes what happens if you do not switch.

First we describe what is the same no matter what policy you follow. Suppose the doors are numbered 1, 2, and 3. Let H_i denote the event that the car is hidden behind door i. Also, let's assume you first choose door 1. (Whatever door you do choose, that door can be labeled door 1 and it would not change your probability of winning.) Now let R_i denote the event that Monty opens door i that hides a goat. If the car is behind door 1 Monty can choose to open door 2 or door 3 because both hide goats. He chooses door 2 or door 3 by flipping a fair coin. If the car is behind door 2, Monty opens door 3 and if the car is behind door 3, Monty opens door 2. Let C denote the event that you win the car and G the event that you win a goat. After Monty opens one of the doors, you decide whether to change your choice or stay with your choice of door 1. Finally, Monty opens the door of your final choice, either door 1 or the door you switched to.

The tree diagram in Figure 2.1(a) applies to the situation in which you change your choice. From this tree we learn that when the car is behind door 1 (event H_1) and



(a) Switch (b) Do Not Switch

Figure 2.1 Tree Diagrams for Monty Hall

Monty opens door 2 (event R_2), you switch to door 3 and then Monty opens door 3 to reveal a goat (event G). On the other hand, if the car is behind door 2, Monty reveals the goat behind door 3 and you switch to door 2 and win the car. Similarly, if the car is behind door 3, Monty reveals the goat behind door 2 and you switch to door 3 and win the car. For always switch, we see that

$$\mathrm{P}\left[C\right] = \mathrm{P}\left[H_2 R_3 C\right] + \mathrm{P}\left[H_3 R_2 C\right] = 2/3.$$

If you do not switch, the tree is shown in Figure 2.1(b). In this tree, when the car is behind door 1 (event H_1) and Monty opens door 2 (event R_2), you stay with door 1 and then Monty opens door 1 to reveal the car. On the other hand, if the car is behind door 2, Monty will open door 3 to reveal the goat. Since your final choice was door 1, Monty opens door 1 to reveal the goat. For do not switch,

$$\mathrm{P}[C] = \mathrm{P}[H_1 R_2 C] + \mathrm{P}[H_1 R_3 C] = 1/3.$$

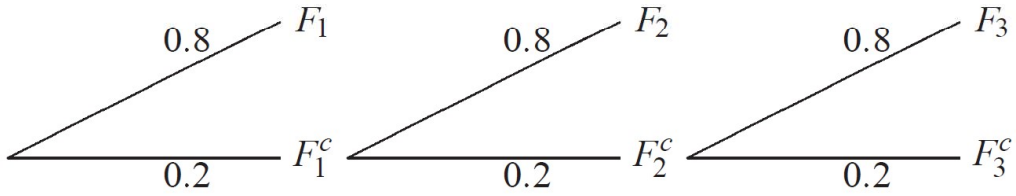
Thus switching is better; if you don't switch, you win the car only if you initially guessed the location of the car correctly, an event that occurs with probability $1/3$. If you switch, you win the car when your initial guess was wrong, an event with probability $2/3$.

Note that the two trees look largely the same because the key step where you make a choice is somewhat hidden because it is implied by the first two branches followed in the tree.

Quiz 2.1

In a cellular phone system, a mobile phone must be paged to receive a phone call. However, paging attempts don't always succeed because the mobile phone may not receive the paging signal clearly. Consequently, the system will page a phone up to three times before giving up. If the results of all paging attempts are independent and a single paging attempt succeeds with probability 0.8, sketch a probability tree for this experiment and find the probability P[F] that the phone receives the paging signal clearly.

Let F_i denote the event that that the user is found on page i. The tree for the experiment is



The user is found unless all three paging attempts fail. Thus the probability the user is found is

$$P[F] = 1 - P[F_1^c F_2^c F_3^c] = 1 - (0.2)^3 = 0.992$$

2.2 Counting Methods

In all applications of probability theory it is important to understand the sample space of an experiment. The methods in this section determine the number of outcomes in the sample space of a sequential experiment

Understanding the sample space is a key step in formulating and solving a probability problem. To begin, it is often useful to know the number of outcomes in the sample space. This number can be enormous as in the following simple example.

Example 2.5

Choose 7 cards at random from a deck of 52 different cards. Display the cards in the order in which you choose them. How many different sequences of cards are possible?

The procedure consists of seven subexperiments. In each subexperiment, the observation is the identity of one card. The first subexperiment has 52 possible outcomes corresponding to the 52 cards that could be drawn. For each outcome of the first subexperiment, the second subexperiment has 51 possible outcomes corresponding to the 51 remaining cards. Therefore there are 52×51 outcomes of the first two subexperiments. The total number of outcomes of the seven subexperiments is

$$52 \times 51 \times \cdots \times 46 = 674{,}274{,}182{,}400\,.$$

Although many practical experiments are more complicated, the techniques for determining the size of a sample space all follow from the fundamental principle of counting in Theorem 2.1:

Theorem 2.1

An experiment consists of two subexperiments. If one subexperiment has k outcomes and the other subexperiment has n outcomes, then the experiment has nk outcomes.

Example 2.6

There are two subexperiments. The first subexperiment is "Flip a coin and observe either heads H or tails T." The second subexperiment is "Roll a six-sided die and observe the number of spots." It has six outcomes, $1, 2, \ldots, 6$. The experiment, "Flip a coin and roll a die," has $2 \times 6 = 12$ outcomes:

$$\begin{array}{cccccc} (H,1), & (H,2), & (H,3), & (H,4), & (H,5), & (H,6), \\ (T,1), & (T,2), & (T,3), & (T,4), & (T,5), & (T,6). \end{array}$$

Generally, if an experiment E has k sub-experiments $E_1, \ldots, E_k$ where E_i has n_i outcomes, then E has $\prod_{i=1}^{k} n_i$ outcomes. In Example 2.5, we chose an ordered sequence of seven objects out of a set of 52 *distinguishable objects*. In general, an ordered sequence of k distinguishable objects is called a *k-permutation*. We will use the notation $(n)k$ to denote the number of possible k-permutations of n distinguishable objects.

To find $(n)k$, suppose we have n distinguishable objects, and the experiment is to choose a sequence of k of these objects. There are n choices for the first object, $n-1$ choices for the second object, etc. Therefore, the total number of possibilities is

$$(n)_k = n(n-1)(n-2)\cdots(n-k+1).$$

Multiplying the right side by $(n-k)!/(n-k)!$ yields our next theorem.

Theorem 2.2

The number of k-permutations of n distinguishable objects is

$$(n)_k = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

Sampling without Replacement

Sampling without replacement corresponds to a sequential experiment in which the sample space of each sub-experiment depends on the outcomes of previous sub-experiments.

- Choosing objects randomly from a collection is called ***sampling***, and the chosen objects are known as a ***sample***.
- A *k*-permutation is a type of sample obtained by specific rules for selecting objects from the collection.
- In particular, once we choose an object for a *k*-permutation, we remove the object from the collection and we cannot choose it again.
- Consequently, this procedure is called ***sampling without replacement***.
- Different outcomes in a *k*-permutation are distinguished by the order in which objects arrive in a sample.
- By contrast, in many practical problems, we are concerned only with the identity of the objects in a sample, not their order.
- For example, in many card games, only the set of cards received by a player is of interest. The order in which they arrive is irrelevant.

Example 2.7

Suppose there are four objects, *A*, *B*, *C*, and *D*, and we define an experiment in which the procedure is to choose two objects without replacement, arrange them in alphabetical order, and observe the result. In this case, to observe *AD* we could choose *A* first or *D* first or both *A* and *D* simultaneously. The possible outcomes of the experiment are *AB*, *AC*, *AD*, *BC*, *BD*, and *CD*.

In contrast to this example with six outcomes, the next example shows that the *k*-permutation corresponding to an experiment in which the observation is the sequence of two letters has 4!/2! = 12 outcomes.

Example 2.8

Suppose there are four objects, *A*, *B*, *C*, and *D*, and we define an experiment in which the procedure is to choose two objects without replacement and observe the result. The 12 possible outcomes of the experiment are *AB*, *AC*, *AD*, *BA*, *BC*, *BD*, *CA*, *CB*, *CD*, *DA*, *DB*, and *DC.*

In Example 2.7, each outcome is a subset of the outcomes of a *k*-permutation. Each subset is called a *k-combination*. We want to find the number of *k*-combinations.

We use the notation $\binom{n}{k}$ to denote the number of *k*-combinations.

The words for this number are "*n* choose *k*," the number of *k*-combinations of *n* objects.

To find $\binom{n}{k}$, we perform the following two sub-experiments to assemble a *k*-permutation of *n* distinguishable objects:

1. Choose a *k*-combination out of the *n* objects.
2. Choose a *k*-permutation of the *k* objects in the *k*-combination.

Theorem 2.2 tells us that the number of outcomes of the combined experiment is $(n)_k$.

The first sub-experiment has $\binom{n}{k}$ possible outcomes, the number we have to derive.

By Theorem 2.2, the second experiment has $(k)_k = k!$ possible outcomes.

Since there are $(n)k$ possible outcomes of the combined experiment,

$$(n)_k = \binom{n}{k} \cdot k!$$

Theorem 2.3

The number of ways to choose k objects out of n distinguishable objects is

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

We encounter $\binom{n}{k}$ in other mathematical studies. Sometimes it is called a *binomial coefficient* because it appears (as the coefficient of $x^k y^{n-k}$) in the expansion of the binomial $(x+y)^n$. In addition, we observe that

$$\binom{n}{k} = \binom{n}{n-k}.$$

The logic behind this identity is that choosing *k* out of *n* elements to be part of a subset is equivalent to choosing *n* − *k* elements to be excluded from the subset.

In most contexts, $\binom{n}{k}$ is undefined except for integers *n* and *k* with $0 \le k \le n$. Here, we adopt the following definition that applies to all nonnegative integers *n* and all real numbers *k*:

Definition 2.1 — **n choose k**

For an integer $n \geq 0$, we define

$$\binom{n}{k} = \begin{cases} \dfrac{n!}{k!(n-k)!} & k = 0, 1, \ldots, n, \\ 0 & \textit{otherwise.} \end{cases}$$

This definition captures the intuition that given, say, $n = 33$ objects, there are zero ways of choosing $k = -5$ objects, zero ways of choosing $k = 8.7$ objects, and zero ways of choosing $k = 87$ objects.

Although this extended definition may seem unnecessary, and perhaps even silly, it will make many formulas in later chapters more concise and easier for students to grasp.

Example 2.9

- The number of combinations of seven cards chosen from a deck of 52 cards is

$$\binom{52}{7} = \frac{52 \times 51 \times \cdots \times 46}{2 \times 3 \times \cdots \times 7} = 133{,}784{,}560,$$

  which is the number of 7-combinations of 52 objects. By contrast, we found in Example 2.5 674,274,182,400 7-permutations of 52 objects. (The ratio is $7! = 5040$).
- There are 11 players on a basketball team. The starting lineup consists of five players. There are $\binom{11}{5} = 462$ possible starting lineups.
- There are $\binom{120}{60} \approx 10^{36}$. ways of dividing 120 students enrolled in a probability course into two sections with 60 students in each section.
- A baseball team has 15 field players and ten pitchers. Each field player can take any of the eight nonpitching positions. The starting lineup consists of one pitcher and eight field players. Therefore, the number of possible starting lineups is $N = \binom{10}{1}\binom{15}{8} = 64{,}350$. For each choice of starting lineup, the manager must submit to the umpire a batting order for the 9 starters. The number of possible batting orders is $N \times 9! = 23{,}351{,}328{,}000$ since there are N ways to choose the 9 starters, and for each choice of 9 starters, there are $9! = 362{,}880$ possible batting orders.

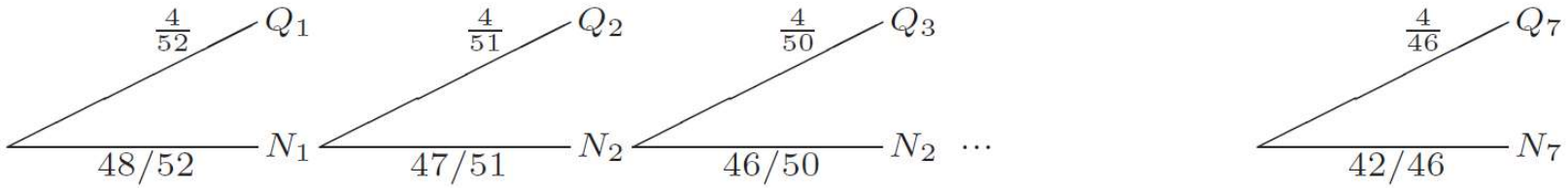
Example 2.10

There are four queens in a deck of 52 cards. You are given seven cards at random from the deck. What is the probability that you have no queens?

Consider an experiment in which the procedure is to select seven cards at random from a set of 52 cards and the observation is to determine if there are one or more queens in the selection. The sample space contains $H = \binom{52}{7}$ possible combinations of seven cards, each with probability $1/H$. There are $H_{NQ} = \binom{52-4}{7}$ combinations with no queens. The probability of receiving no queens is the ratio of the number of outcomes with no queens to the number of outcomes in the sample space. $H_{NQ}/H = 0.5504$.

Another way of analyzing this experiment is to consider it as a sequence of seven subexperiments. The first subexperiment consists of selecting a card at random and observing whether it is a queen. If it is a queen, an outcome with probability $4/52$ (because there are 52 outcomes in the sample space and four of them are in the event $\{\text{queen}\}$), stop looking for queens. Otherwise, with probability $48/52$, select another card from the remaining 51 cards and observe whether it is a queen. This outcome of this subexperiment has probability $4/51$. If the second card is not a queen, an outcome

with probability $47/51$, continue until you select a queen or you have seven cards with no queen. Using Q_i and N_i to indicate a "Queen" or "No queen" on subexperiment i, the tree for this experiment is



The probability of the event N_7 that no queen is received in your seven cards is the product of the probabilities of the branches leading to N_7:

$$(48/52) \times (47/51) \cdots \times (42/46) = 0.5504.$$

Sampling with Replacement

Consider selecting an object from a collection of objects, replacing the selected object, and repeating the process several times, each time replacing the selected object before making another selection.

We refer to this situation as ***sampling with replacement***. Each selection is the procedure of a sub-experiment. The sub-experiments are referred to as ***independent trials***.

In this section we consider the number of possible outcomes that result from sampling with replacement. In the next section we derive probability models for for experiments that specify sampling with replacement.

Example 2.11

There are four queens in a deck of 52 cards. You are given seven cards at random from the deck. After receiving each card you return it to the deck and receive another card at random. Observe whether you have not received any queens among the seven cards you were given. What is the probability that you have received no queens?

The sample space contains 52^7 outcomes. There are 48^7 outcomes with no queens. The ratio is $(48/52)^7 = 0.5710$, the probability of receiving no queens. If this experiment is considered as a sequence of seven subexperiments, the tree looks the same as the tree in Example 2.10, except that all the horizontal branches have probability $48/52$ and all the diagonal branches have probability $4/52$.

Example 2.12

A laptop computer has USB slots A and B. Each slot can be used for connecting a memory card (m), a camera (c) or a printer (p). It is possible to connect two memory cards, two cameras, or two printers to the laptop. How many ways can we use the two USB slots?

This example corresponds to sampling two times with replacement from the set $\{m, c, p\}$. Let xy denote the outcome that device type x is used in slot A and device type y is used in slot B. The possible outcomes are $S = \{mm, mc, mp, cm, cc, cp, pm, pc, pp\}$. The sample space S contains nine outcomes.

The fact that Example 2.12 has nine possible outcomes should not be surprising. Since we were sampling with replacement, there were always three possible outcomes for each of the subexperiments to attach a device to a USB slot. Hence, by the fundamental theorem of counting, Example 2.12 must have $3 \times 3 = 9$ possible outcomes.

In Example 2.12, *mc* and *cm* are distinct outcomes. This result generalizes naturally when we want to choose with replacement a sample of *n* objects out of a collection of *m* distinguishable objects. The experiment consists of a sequence of *n* identical subexperiments with *m* outcomes in the sample space of each subexperiment. Hence there are *mn* ways to choose with replacement a sample of *n* objects.

Theorem 2.4

Given m distinguishable objects, there are m^n ways to choose with replacement an ordered sample of n objects.

Example 2.13

There are $2^{10} = 1024$ binary sequences of length 10.

Example 2.14

The letters A through Z can produce $26^4 = 456{,}976$ four-letter words.

Sampling with replacement corresponds to performing n repetitions of an identical sub-experiment. Using x_i to denote the outcome of the ith sub-experiment, the result for n repetitions of the sub-experiment is a sequence $x_1, \ldots, x_n$.

Example 2.15

A chip fabrication facility produces microprocessors. Each microprocessor is tested to determine whether it runs reliably at an acceptable clock speed. A subexperiment to test a microprocessor has sample space $S_{\text{sub}} = \{0, 1\}$ to indicate whether the test was a failure (0) or a success (1). For test i, we record $x_i = 0$ or $x_i = 1$ to indicate the result. In testing four microprocessors, the observation sequence, $x_1x_2x_3x_4$, is one of 16 possible outcomes:

$$S = \begin{Bmatrix} 0000, & 0001, & 0010, & 0011, & 0100, & 0101, & 0110, & 0111, \\ 1000, & 1001, & 1010, & 1011, & 1100, & 1101, & 1110, & 1111 \end{Bmatrix}.$$

Note that we can think of the observation sequence $x_1, \ldots, x_n$ as the result of sampling with replacement n times from a sample space S_{sub}. For sequences of identical subexperiments, we can express Theorem 2.4 as

Theorem 2.5

For n repetitions of a subexperiment with sample space $S_{sub} = \{s_0, \ldots, s_{m-1}\}$, the sample space S of the sequential experiment has m^n outcomes.

Example 2.16

There are ten students in a probability class. Each earns a grade $s \in S_{\text{sub}} = \{A, B, C, F\}$. We use the notation x_i to denote the grade of the ith student. For example, the grades for the class could be

$$x_1 x_2 \cdots x_{10} = CBBACFBACF \tag{2.14}$$

The sample space S of possible sequences contains $4^{10} = 1{,}048{,}576$ outcomes.

In Example 2.12 and Example 2.16, repeating a subexperiment n times and recording the observation consists of constructing a word with n letters. In general, n repetitions of the same subexperiment consists of choosing symbols from the alphabet $\{s_0, \ldots, s_{m-1}\}$. In Example 2.15, $m = 2$ and we have a binary alphabet with symbols $s_0 = 0$ and $s_1 = 1$.

A more challenging problem than finding the number of possible combinations of m objects sampled with replacement from a set of n objects is to calculate the number of observation sequences such that each object appears a specified number of times. We start with the case in which each subexperiment is a trial with sample space $S_{\text{sub}} = \{0, 1\}$ indicating failure or success.

Example 2.17

For five subexperiments with sample space $S_{\text{sub}} = \{0, 1\}$, what is the number of observation sequences in which 0 appears $n_0 = 2$ times and 1 appears $n_1 = 3$ times?

The 10 five-letter words with 0 appearing twice and 1 appearing three times are:

{00111, 01011, 01101, 01110, 10011, 10101, 10110, 11001, 11010, 11100}

Example 2.17 determines the number of outcomes in the sample space of an experiment with five subexperiments by listing all of the outcomes. Even in this simple example it is not a simple matter to determine all of the outcomes, and in most practical applications of probability there are far more then ten outcomes in the sample space of an experiment and listing them all is out of the question. On the other hand, the counting methods covered in this chapter provide formulas for quickly calculating the number of outcomes in a sample space.

In Example 2.17 each outcome corresponds to the position of three ones in a five-letter binary word. That is, each outcome is completely specified by choosing three positions that contain 1. There are $\binom{5}{3} = 10$ ways to choose three positions in a word. More generally, for length *n* binary words with *n*1 1's, we choose $\binom{n}{n_1}$ slots to hold a 1.

Theorem 2.6

The number of observation sequences for n subexperiments with sample space $S = \{0, 1\}$ with 0 appearing n_0 times and 1 appearing $n_1 = n - n_0$ times is $\binom{n}{n_1}$.

Theorem 2.6 can be generalized to subexperiments with $m > 2$ elements in the sample space. For n trials of a subexperiment with sample space $S_{sub} = \{s_0, \ldots, s_{m-1}\}$, we want to find the number of outcomes in which s_0 appears n_0 times, $s1$ appears $n1$ times, and so on. Of course, there are no such outcomes unless $n_0 + \cdots + n_{m-1} = n$. The notation for the number of outcomes is

$$\binom{n}{n_0, \ldots, n_{m-1}}$$

It is referred to as the *multinomial coefficient*. To derive a formula for the multinomial coefficient, we generalize the logic used in deriving the formula for the binomial coefficient. With n subexperiments, representing the observation sequence by n slots, we first choose n_0 positions in the observation sequence to hold s_0, then n_1 positions to hold s_1, and so on. The details can be found in the proof of the following theorem:

Theorem 2.7

For n repetitions of a subexperiment with sample space $S = \{s_0, \ldots, s_{m-1}\}$, the number of length $n = n_0 + \cdots + n_{m-1}$ observation sequences with s_i appearing n_i times is

$$\binom{n}{n_0, \ldots, n_{m-1}} = \frac{n!}{n_0! n_1! \cdots n_{m-1}!}.$$

Proof Let $M = \binom{n}{n_0,\ldots,n_{m-1}}$. Start with n empty slots and perform the following sequence of subexperiments:

Subexperiment	Procedure
0	Label n_0 slots as s_0.
1	Label n_1 slots as s_1.
$\vdots$	$\vdots$
$m-1$	Label the remaining n_{m-1} slots as s_{m-1}.

There are $\binom{n}{n_0}$ ways to perform subexperiment 0. After n_0 slots have been labeled, there are $\binom{n-n_0}{n_1}$ ways to perform subexperiment 1. After subexperiment $j-1$, $n_0 + \cdots + n_{j-1}$ slots have already been filled, leaving $\binom{n-(n_0+\cdots+n_{j-1})}{n_j}$ ways to perform subexperiment j. From the fundamental counting principle,

$$\begin{aligned} M &= \binom{n}{n_0}\binom{n-n_0}{n_1}\binom{n-n_0-n_1}{n_2}\cdots\binom{n-n_0-\cdots-n_{m-2}}{n_{m-1}} \\ &= \frac{n!}{(n-n_0)!n_0!}\frac{(n-n_0)!}{(n-n_0-n_1)!n_1!}\cdots\frac{(n-n_0-\cdots-n_{m-2})!}{(n-n_0-\cdots-n_{m-1})!n_{m-1}!}. \end{aligned}$$

Canceling the common factors, we obtain the formula of the theorem.

Note that a binomial coefficient is the special case of the multinomial coefficient for an alphabet with $m = 2$ symbols. In particular, for $n = n_0 + n_1$,

$$\binom{n}{n_0, n_1} = \binom{n}{n_0} = \binom{n}{n_1}$$

Lastly, in the same way that we extended the definition of the binomial coefficient, we will employ an extended definition for the multinomial coefficient.

Definition 2.2 Multinomial Coefficient

For an integer $n \geq 0$, we define

$$\binom{n}{n_0, \ldots, n_{m-1}} = \begin{cases} \dfrac{n!}{n_0! n_1! \cdots n_{m-1}!} & n_0 + \cdots + n_{m-1} = n; \\ & n_i \in \{0, 1, \ldots, n\}, i = 0, 1, \ldots, m-1, \\ 0 & \textit{otherwise.} \end{cases}$$

Example 2.18

In Example 2.16, the professor uses a *curve* in determining student grades. When there are ten students in a probability class, the professor always issues two grades of A, three grades of B, three grades of C and two grades of F. How many different ways can the professor assign grades to the ten students?

In Example 2.16, we determine that with four possible grades there are $4^{10} = 1{,}048{,}576$ ways of assigning grades to ten students. However, now we are limited to choosing $n_0 = 2$ students to receive an A, $n_1 = 3$ students to receive a B, $n_2 = 3$ students to receive a C and $n_3 = 4$ students to receive an F. The number of ways that fit the curve is the multinomial coefficient

$$\binom{n}{n_0, n_1, n_2, n_3} = \binom{10}{2, 3, 3, 2} = \frac{10!}{2!3!3!2!} = 25{,}200.$$

Quiz 2.2

Consider a binary code with 4 bits (0 or 1) in each code word. An example of a code word is 0110.

(a) How many different code words are there?

(b) How many code words have exactly two zeroes?

(c) How many code words begin with a zero?

(d) In a constant-ratio binary code, each code word has N bits. In every word, M of the N bits are 1 and the other $N - M$ bits are 0. How many different code words are in the code with $N = 8$ and $M = 3$?

(a) We can view choosing each bit in the code word as a subexperiment. Each subexperiment has two possible outcomes: 0 and 1. Thus by the fundamental principle of counting, there are $2 \times 2 \times 2 \times 2 = 2^4 = 16$ possible code words.

(b) An experiment that can yield all possible code words with two zeroes is to choose which 2 bits (out of 4 bits) will be zero. The other two bits then must be ones. There are $\binom{4}{2} = 6$ ways to do this. Hence, there are six code words with exactly two zeroes. For this problem, it is also possible to simply enumerate the six code words:

$$1100 \quad 1010 \quad 1001 \quad 0101 \quad 0110 \quad 0011$$

(c) When the first bit must be a zero, then the first subexperiment of choosing the first bit has only one outcome. For each of the next three bits, we have two choices. In this case, there are $1 \times 2 \times 2 \times 2 = 8$ ways of choosing a code word.

(d) For the constant ratio code, we can specify a code word by choosing M of the bits to be ones. The other $N - M$ bits will be zeroes. The number of ways of choosing such a code word is $\binom{N}{M}$. For $N = 8$ and $M = 3$, there are $\binom{8}{3} = 56$ code words.

2.3 Independent Trials

Independent trials are identical subexperiments in a sequential experiment.

The probability models of all the subexperiments are identical and independent of the outcomes of previous subexperiments.

Sampling with replacement is one category of experiments with independent trials.

We now apply the counting methods of Section 2.2 to derive probability models for experiments consisting of independent repetitions of a subexperiment.

We start with a simple subexperiment in which there are two outcomes: a success (1) occurs with probability p; otherwise, a failure (0) occurs with probability $1 - p$.

- The results of all trials of the subexperiment are mutually independent.
- An outcome of the complete experiment is a sequence of successes and failures denoted by a sequence of ones and zeroes.

For example, 10101 . . . is an alternating sequence of successes and failures.

Let E_{n_0,n_1} denote the event $n0$ failures and $n1$ successes in $n = n_0 + n_1$ trials.

To find $P[E_{n_0,n_1}]$, we first consider an example.

Example 2.19

What is the probability $\mathrm{P}[E_{2,3}]$ of two failures and three successes in five independent trials with success probability p.

To find $\mathrm{P}[E_{2,3}]$, we observe that the outcomes with three successes in five trials are 11100, 11010, 11001, 10110, 10101, 10011, 01110, 01101, 01011, and 00111. We note that the probability of each outcome is a product of five probabilities, each related to one subexperiment. In outcomes with three successes, three of the probabilities are p and the other two are $1-p$. Therefore each outcome with three successes has probability $(1-p)^2p^3$.

From Theorem 2.6, we know that the number of such sequences is $\binom{5}{3}$. To find $\mathrm{P}[E_{2,3}]$, we add up the probabilities associated with the 10 outcomes with 3 successes, yielding

$$\mathrm{P}\left[E_{2,3}\right] = \binom{5}{3}(1-p)^2p^3.$$

In general, for $n = n_0 + n_1$ independent trials we observe that

- Each outcome with n_0 failures and n_1 successes has probability $(1-p)^{n_0}p^{n_1}$.
- There are $\binom{n}{n_0} = \binom{n}{n_1}$ outcomes that have n_0 failures and n_1 successes.

Therefore the probability of n_1 successes in n independent trials is the sum of $\binom{n}{n_1}$ terms, each with probability $(1-p)^{n_0}p^{n_1} = (1-p)^{n-n_1}p^{n_1}$.

Theorem 2.8

The probability of n_0 failures and n_1 successes in $n = n_0 + n_1$ independent trials is

$$\mathrm{P}\left[E_{n_0,n_1}\right] = \binom{n}{n_1}(1-p)^{n-n_1}p^{n_1} = \binom{n}{n_0}(1-p)^{n_0}p^{n-n_0}.$$

The second formula in this theorem is the result of multiplying the probability of n_0 failures in n trials by the number of outcomes with n_0 failures.

Example 2.20

In Example 1.19, we found that a randomly tested resistor was acceptable with probability $\mathrm{P}[A] = 0.78$. If we randomly test 100 resistors, what is the probability of T_i, the event that i resistors test acceptable?

Testing each resistor is an independent trial with a success occurring when a resistor is acceptable. Thus for $0 \le i \le 100$,

$$\mathrm{P}\left[T_i\right] = \binom{100}{i}(0.78)^i(1-0.78)^{100-i} \tag{2.19}$$

We note that our intuition says that since 78% of the resistors are acceptable, then in testing 100 resistors, the number acceptable should be near 78. However, $\mathrm{P}[T_{78}] \approx 0.096$, which is fairly small. This shows that although we might expect the number acceptable to be close to 78, that does not mean that the probability of exactly 78 acceptable is high.

Example 2.21

To communicate one bit of information reliably, cellular phones transmit the same binary symbol five times. Thus the information "zero" is transmitted as 00000 and "one" is 11111. The receiver detects the correct information if three or more binary symbols are received correctly. What is the information error probability $\mathrm{P}[E]$, if the binary symbol error probability is $q = 0.1$?

In this case, we have five trials corresponding to the five times the binary symbol is sent. On each trial, a success occurs when a binary symbol is received correctly. The probability of a success is $p = 1 - q = 0.9$. The error event E occurs when the number of successes is strictly less than three:

$$\begin{aligned}\mathrm{P}[E] &= \mathrm{P}[E_{0,5}] + \mathrm{P}[E_{1,4}] + \mathrm{P}[E_{2,3}] \\ &= \binom{5}{0} q^5 + \binom{5}{1} pq^4 + \binom{5}{2} p^2 q^3 = 0.00856.\end{aligned}$$

By increasing the number of binary symbols per information bit from 1 to 5, the cellular phone reduces the probability of error by more than one order of magnitude, from 0.1 to 0.0081.

Now suppose we perform n independent repetitions of a subexperiment for which there are m possible outcomes for any subexperiment. That is, the sample space for each subexperiment is $(s_0, \ldots, s_{m-1})$ and every event in one subexperiment is independent of the events in all the other subexperiments. Therefore, in every subexperiment the probabilities of corresponding events are the same and we can use the notation $P[s_k] = p_k$ for all of the subexperiments.

An outcome of the experiment consists of a sequence of n sub-experiment outcomes. In the probability tree of the experiment, each node has m branches and branch i has probability p_i. The probability of an outcome of the sequential experiment is just the product of the n branch probabilities on a path from the root of the tree to the leaf representing the outcome. For example, with $n = 5$, the outcome $s_2s_0s_3s_2s_4$ occurs with probability $p_2p_0p_3p_2p_4$. We want to find the probability of the event

$$E_{n_0,\ldots,n_{m-1}} = \{s_0 \text{ occurs } n_0 \text{ times}, \ldots, s_{m-1} \text{ occurs } n_{m-1} \text{ times}\}$$

Note that the notation $E_{n_0,\ldots,n_{m-1}}$ implies that the experiment consists of a sequence of $n = n_0 + \cdots + n_{m-1}$ trials.

To calculate $\mathrm{P}[E_{n_0,\ldots,n_{m-1}}]$, we observe that the probability of the outcome

$$\underbrace{s_0 \cdots s_0}_{n_0 \text{ times}} \underbrace{s_1 \cdots s_1}_{n_1 \text{ times}} \cdots \underbrace{s_{m-1} \cdots s_{m-1}}_{n_{m-1} \text{ times}}$$

is

$$p_0^{n_0} p_1^{n_1} \cdots p_{m-1}^{n_{m-1}}.$$

Next, we observe that any other experimental outcome that is a reordering of the preceding sequence has the same probability because on each path through the tree to such an outcome there are n_i occurrences of s_i. As a result,

Next, we observe that any other experimental outcome that is a reordering of the preceding sequence has the same probability because on each path through the tree to such an outcome there are n_i occurrences of s_i. As a result,

$$\mathrm{P}\left[E_{n_0,\ldots,n_{m-1}}\right] = M p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

where M, the number of such outcomes, is the multinomial coefficient of $\binom{n}{n_0,\ldots,n_{m-1}}$

Definition 2.2. Applying Theorem 2.7, we have the following theorem:

Theorem 2.9

A subexperiment has sample space $S_{sub} = \{s_0, \ldots, s_{m-1}\}$ *with* $\mathrm{P}[s_i] = p_i$. *For* $n = n_0 + \cdots + n_{m-1}$ *independent trials, the probability of* n_i *occurences of* s_i, $i = 0, 1, \ldots, m-1$, *is*

$$\mathrm{P}\left[E_{n_0,\ldots,n_{m-1}}\right] = \binom{n}{n_0, \ldots, n_{m-1}} p_0^{n_0} \cdots p_{m-1}^{n_{m-1}}.$$

Example 2.22

A packet processed by an Internet router carries either audio information with probability $7/10$, video, with probability $2/10$, or text with probability $1/10$. Let $E_{a,v,t}$ denote the event that the router processes a audio packets, v video packets, and t text packets in a sequence of 100 packets. In this case,

$$\mathrm{P}\left[E_{a,v,t}\right] = \binom{100}{a,v,t}\left(\frac{7}{10}\right)^a\left(\frac{2}{10}\right)^v\left(\frac{1}{10}\right)^t$$

Keep in mind that by the extended definition of the multinomial coefficient, $\mathrm{P}[E_{a,v,t}]$ is nonzero only if $a+v+t=100$ and a, v, and t are nonnegative integers.

Example 2.23

Continuing with Example 2.16, suppose in testing a microprocessor that all four grades have probability 0.25, independent of any other microprocessor. In testing $n = 100$ microprocessors, what is the probability of exactly 25 microprocessors of each grade?

Let $E_{25,25,25,25}$ denote the probability of exactly 25 microprocessors of each grade. From Theorem 2.9,

$$\mathrm{P}\left[E_{25,25,25,25}\right] = \binom{100}{25,25,25,25}(0.25)^{100} = 0.0010.$$

Quiz 2.3

Data packets containing 100 bits are transmitted over a communication link. A transmitted bit is received in error (either a 0 sent is mistaken for a 1, or a 1 sent is mistaken for a 0) with probability $\epsilon = 0.01$, independent of the correctness of any other bit. The packet has been coded in such a way that if three or fewer bits are received in error, then those bits can be corrected. If more than three bits are received in error, then the packet is decoded with errors.

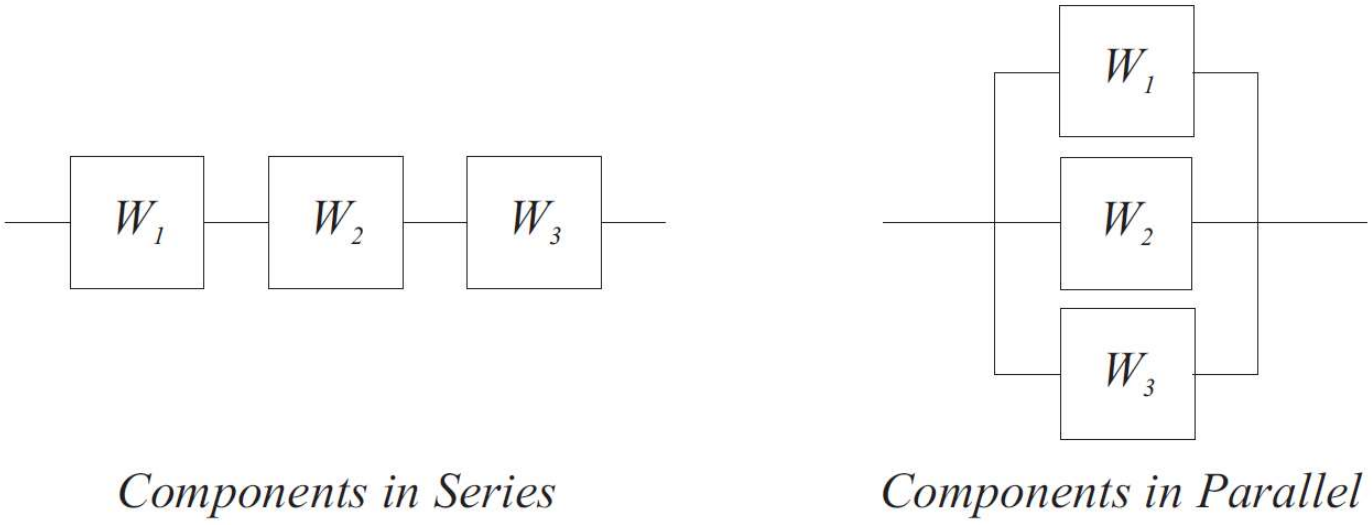
(a) Let $E_{k,100-k}$ denote the event that a received packet has k bits in error and $100-k$ correctly decoded bits. What is $\mathrm{P}[E_{k,100-k}]$ for $k = 0, 1, 2, 3$?

(b) Let C denote the event that a packet is decoded correctly. What is $\mathrm{P}[C]$?

2.4 Reliability Analysis

To find the success probability of a complicated process with components in series and components in parallel, it is helpful to consider a group of components in series as one equivalent component and a group of components in parallel as another equivalent component.

- Sequential experiments are models for practical processes that depend on several operations to succeed.
- Examples are manufacturing processes that go through several stages, and communications systems that relay packets through several routers between a source and destination.
- In some cases, the processes contain redundant components that protect the entire process from the failure of one or more components.
- In this section we describe the simple case in which all operations in a process succeed with probability p independent of the success or failure of other components.



Serial and parallel devices.

Let W_i denote the event that component *i* succeeds. As depicted in Figure 2.2, there are two basic types of operations.

• *Components in series*. The operation succeeds if *all* of its components succeed. One example of such an operation is a sequence of computer programs in which each program after the first one uses the result of the previous program.

Therefore, the complete operation fails if any component program fails. Whenever the operation consists of *k* components in series, we need all *k* components to succeed in order to have a successful operation. The probability that the operation succeeds is

$$\mathrm{P}[W] = \mathrm{P}[W_1 W_2 \cdots W_n] = p \times p \times \cdots \times p = p^n$$

If the independent components in parallel have different success probabilities $p_1, p_2 \ldots p_n$, the operation succeeds with probability

$$\mathrm{P}\left[W\right] = \mathrm{P}\left[W_1 W_2 \cdots W_n\right] = p_1 \times p_2 \times \cdots \times p_n$$

With components in series, the probability of a successful operation is lower than the success probability of the weakest component.

• *Components in parallel.* The operation succeeds if *any* component works. This operation occurs when we introduce redundancy to promote reliability.

In a redundant system, such as a space shuttle, there are n computers on board so that the shuttle can continue to function as long as at least one computer operates successfully. If the components are in parallel, the operation fails when all elements fail, so we have

$$\mathrm{P}\left[W^c\right] = \mathrm{P}\left[W_1^c W_2^c \cdots W_n^c\right] = (1-p)^n.$$

The probability that the parallel operation succeeds is

$$\mathrm{P}\left[W\right] = 1 - \mathrm{P}\left[W^c\right] = 1 - (1-p)^n$$

If the independent components in parallel have different success probabilities $p_1, p_2 \ldots p_n$, the operation fails with probability

The probability that the parallel operation succeeds is

$$\mathrm{P}\left[W\right] = 1 - \mathrm{P}\left[W^c\right] = (1 - p_1) \times (1 - p_2) \times \cdots \times (1 - p_n).$$

With components in parallel, the probability that the operation succeeds is higher than the probability of success of the strongest component.

We can analyze complicated combinations of components in series and in parallel with by reducing several components in parallel or components in series to a single equivalent component.

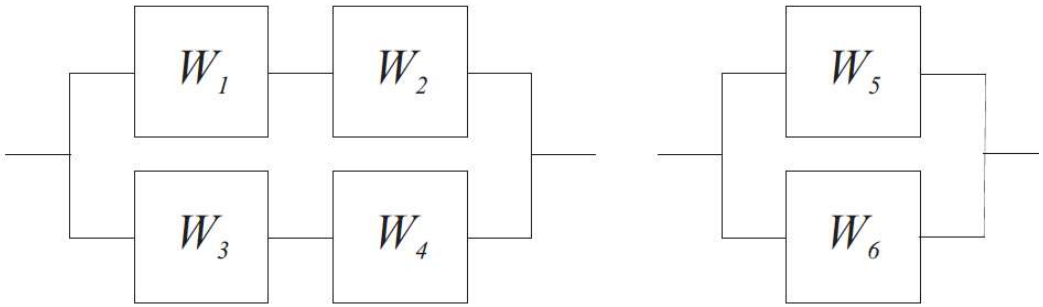


Figure 3 *The operation described in Example 2.24. On the left is the original operation. On the right is the equivalent operation with each pair of series components replaced with an equivalent component.*

Example 2.24

An operation consists of two redundant parts. The first part has two components in series (W_1 and W_2) and the second part has two components in series (W_3 and W_4). All components succeed with probability $p = 0.9$. Draw a diagram of the operation and calculate the probability that the operation succeeds.

A diagram of the operation is shown in Figure 2.3. We can create an equivalent component, W_5, with probability of success p_5 by observing that for the combination of W_1 and W_2,

$$\mathrm{P}\left[W_5\right] = p_5 = P[W_1 W_2] = p^2 = 0.81.$$

Similarly, the combination of W_3 and W_4 in series produces an equivalent component, W_6, with probability of success $p_6 = p_5 = 0.81$. The entire operation then consists of W_5 and W_6 in parallel, which is also shown in Figure 2.3. The success probability of the operation is

$$\mathrm{P}\left[W\right] = 1 - (1 - p_5)^2 = 0.964$$

We could consider the combination of W_5 and W_6 to be an equivalent component W_7 with success probability $p_7 = 0.964$ and then analyze a more complex operation that contains W_7 as a component.

Note that in Equation (2.29) we computed the probability of success of a process with components in series as the product of the success probabilities of the components.

- ❖ The reason is that for the process to be successful, all components must be successful.
- ❖ The event {*all components successful*} is the intersection of the individual success events and the probability of the intersection of two events is the product of the two success probabilities.
- ❑ On the other hand, with components in parallel, the process is successful when one or more components is successful.
- ❑ The event {*one or more components successful*} is the union of individual success probabilities.
- Recall that the probability of the union of two events is the difference between the sum of the individual probabilities and the probability of their intersection.
- The formula for the probability f of more than two events is even more complicated.

On the other hand, with components in parallel, the process fails when *all* of the components fail.

- The event {*all components fail*} is the intersection of the individual failure probabilities.
- Each failure probability is the difference between 1 and the success probability.

In general, *De Morgan's law* (Theorem 1.1) allows us to express a union as the complement of an intersection and vice versa.

- Therefore, in many applications of probability, when it is difficult to calculate directly the probability we need, we can often calculate the probability of the complementary event and then subtract this probability from 1 to find the answer.
- This is how we calculated the probability of success of a process with components in parallel.

Quiz 2.4

A memory module consists of nine chips. The device is designed with redundancy so that it works even if one of its chips is defective. Each chip contains n transistors and functions properly only if all of its transistors work. A transistor works with probability p independent of any other transistor.

(a) What is the probability $\mathrm{P}[C]$ that a chip works?
(b) What is the probability $\mathrm{P}[M]$ that the memory module works?
(c) If $p = 0.999$, what is the maximum number of transistors per chip n that produces $\mathrm{P}[M] \geq 0.9$ (a 90% success probability for the memory module)?
(d) If the memory module can tolerate two defective chips, what is the maximum number of transistors per chip n that produces $\mathrm{P}[M] \geq 0.9$?

(a) Since the chip works only if all n transistors work, the transistors in the chip are like devices in series. The probability that a chip works is $P[C] = p^n$.

(b) The module works if either 8 chips work or 9 chips work. Let C_k denote the event that exactly k chips work. Since transistor failures are independent of each other, chip failures are also independent. Thus the $P[C_k]$ has a binomial probability. Specifically,

$$P[C_8] = \binom{9}{8} (P[C])^8 (1 - P[C])^{9-8} = 9p^{8n}(1 - p^n)$$

$$P[C_9] = (P[C])^9 = p^{9n}$$

The probability a memory module works is

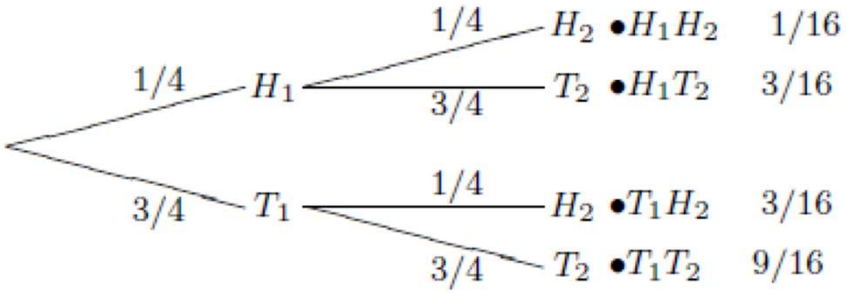
$$P[C] = P[C_8] + P[C_9] = p^{8n}(9 - 8p^n)$$

2.1.1 Suppose you flip a coin twice. On any flip, the coin comes up heads with probability 1/4. Use H_i and T_i to denote the result of flip i.

(a) What is the probability, $\mathrm{P}[H_1|H_2]$, that the first flip is heads given that the second flip is heads?

(b) What is the probability that the first flip is heads and the second flip is tails?

A sequential sample space for this experiment is



(a) From the tree, we observe

$$\mathrm{P}[H_2] = \mathrm{P}[H_1H_2] + \mathrm{P}[T_1H_2] = 1/4.$$

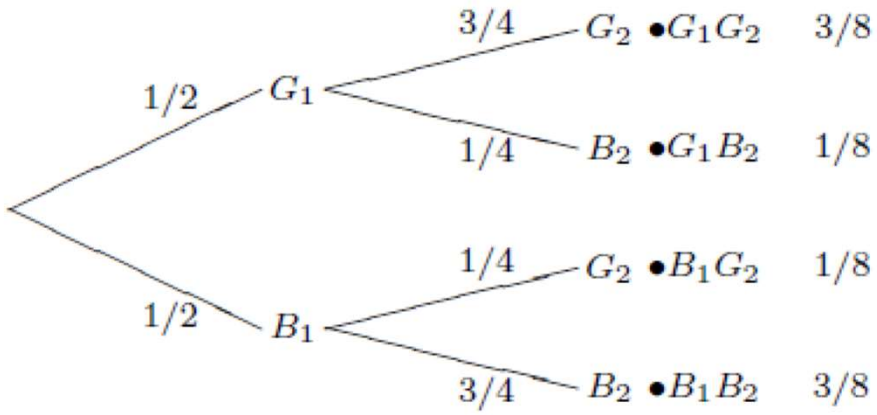
This implies

$$\mathrm{P}[H_1|H_2] = \frac{\mathrm{P}[H_1H_2]}{\mathrm{P}[H_2]} = \frac{1/16}{1/4} = 1/4.$$

(b) The probability that the first flip is heads and the second flip is tails is $\mathrm{P}[H_1T_2] = 3/16$.

2.1.3 At the end of regulation time, a basketball team is trailing by one point and a player goes to the line for two free throws. If the player makes exactly one free throw, the game goes into overtime. The probability that the first free throw is good is 1/2. However, if the first attempt is good, the player relaxes and the second attempt is good with probability 3/4. However, if the player misses the first attempt, the added pressure reduces the success probability to 1/4. What is the probability that the game goes into overtime?

Let G_i and B_i denote events indicating whether free throw i was good (G_i) or bad (B_i). The tree for the free throw experiment is

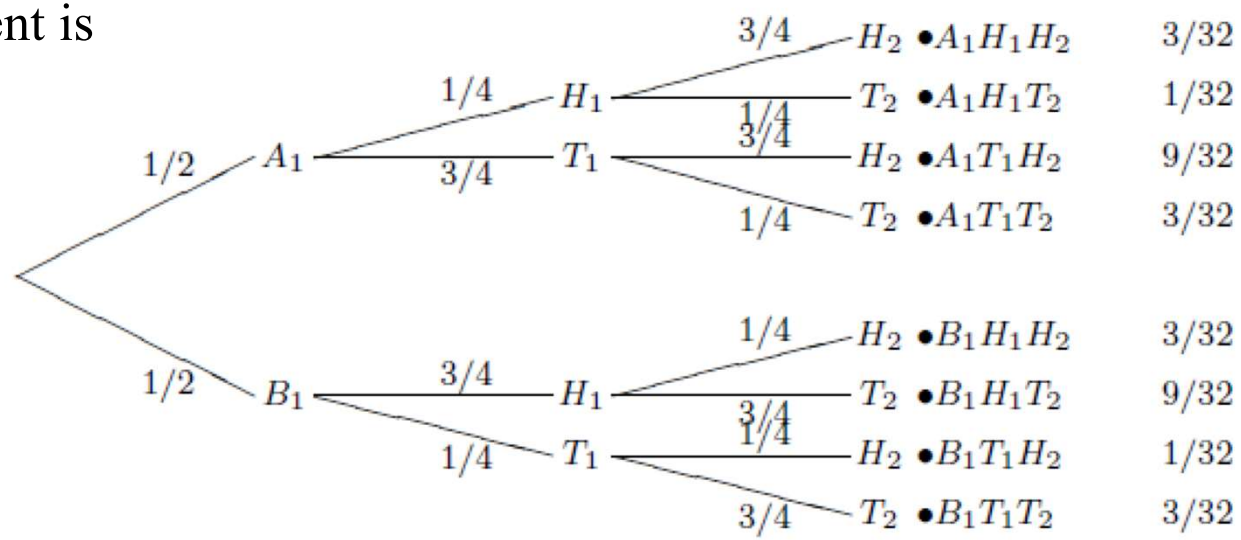


The game goes into overtime if exactly one free throw is made. This event has probability

$$\mathrm{P}[O] = \mathrm{P}[G_1B_2] + \mathrm{P}[B_1G_2] = 1/8 + 1/8 = 1/4.$$

2.1.7 You have two biased coins. Coin A comes up heads with probability 1/4. Coin B comes up heads with probability 3/4. However, you are not sure which is which so you flip each coin once, choosing the first coin randomly. Use Hi and Ti to denote the result of flip i. Let $A1$ be the event that coin A was flipped first. Let $B1$ be the event that coin B was flipped first. What is P[$H1H2$]? Are $H1$ and $H2$ independent? Explain your answer.

The tree for this experiment is



The event H_1H_2 that heads occurs on both flips has probability

$$\mathrm{P}[H_1H_2] = \mathrm{P}[A_1H_1H_2] + \mathrm{P}[B_1H_1H_2] = 6/32.$$

The probability of H_1 is

$$\begin{aligned}\mathrm{P}[H_1] &= \mathrm{P}[A_1H_1H_2] + \mathrm{P}[A_1H_1T_2] + \mathrm{P}[B_1H_1H_2] + \mathrm{P}[B_1H_1T_2] \\ &= 1/2.\end{aligned}$$

Similarly,

$$\begin{aligned} \mathrm{P}[H_2] &= \mathrm{P}[A_1 H_1 H_2] + \mathrm{P}[A_1 T_1 H_2] + \mathrm{P}[B_1 H_1 H_2] + \mathrm{P}[B_1 T_1 H_2] \\ &= 1/2. \end{aligned}$$

Thus $\mathrm{P}[H_1 H_2] \neq \mathrm{P}[H_1]\,\mathrm{P}[H_2]$, implying H_1 and H_2 are not independent. This result should not be surprising since if the first flip is heads, it is likely that coin B was picked first. In this case, the second flip is less likely to be heads since it becomes more likely that the second coin flipped was coin A.

2.1.9 Suppose Dagwood (Blondie's husband) wants to eat a sandwich but needs to go on a diet. Dagwood decides to let the flip of a coin determine whether he eats. Using an unbiased coin, Dagwood will postpone the diet (and go directly to the refrigerator) if either (a) he flips heads on his first flip or (b) he flips tails on the first flip but then proceeds to get two heads out of the next three flips. Note that the first flip is *not* counted in the attempt to win two of three and that Dagwood never performs any unnecessary flips. Let H_i be the event that Dagwood flips heads on try i. Let T_i be the event that tails occurs on flip i.

(a) Draw the tree for this experiment. Label the probabilities of all outcomes.

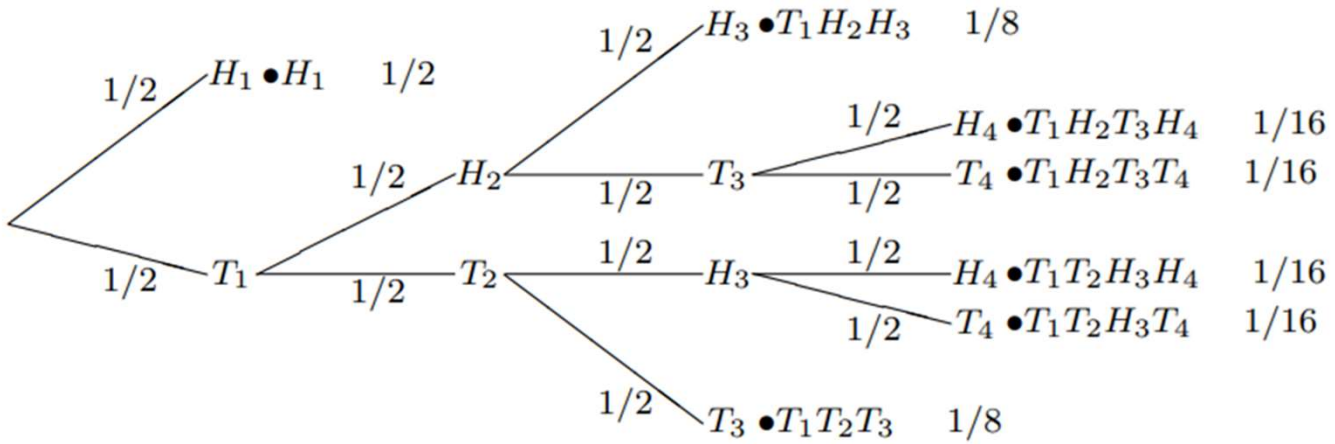
(b) What are $\mathrm{P}[H_3]$ and $\mathrm{P}[T_3]$?

(c) Let D be the event that Dagwood must diet. What is $\mathrm{P}[D]$? What is $\mathrm{P}[H_1|D]$?

(d) Are H_3 and H_2 independent events?

Problem 2.1.9 Solution

(a) The primary difficulty in this problem is translating the words into the correct tree diagram. The tree for this problem is shown below.



(b) From the tree,

$$\begin{aligned} \mathrm{P}[H_3] &= \mathrm{P}[T_1H_2H_3] + \mathrm{P}[T_1T_2H_3H_4] + \mathrm{P}[T_1T_2H_3H_4] \\ &= 1/8 + 1/16 + 1/16 = 1/4. \end{aligned} \tag{1}$$

Similarly,

$$\begin{aligned} \mathrm{P}[T_3] &= \mathrm{P}[T_1H_2T_3H_4] + \mathrm{P}[T_1H_2T_3T_4] + \mathrm{P}[T_1T_2T_3] \\ &= 1/8 + 1/16 + 1/16 = 1/4. \end{aligned} \tag{2}$$

(c) The event that Dagwood must diet is

$$D = (T_1 H_2 T_3 T_4) \cup (T_1 T_2 H_3 T_4) \cup (T_1 T_2 T_3). \tag{3}$$

The probability that Dagwood must diet is

$$\begin{aligned} \mathrm{P}[D] &= \mathrm{P}[T_1 H_2 T_3 T_4] + \mathrm{P}[T_1 T_2 H_3 T_4] + \mathrm{P}[T_1 T_2 T_3] \\ &= 1/16 + 1/16 + 1/8 = 1/4. \end{aligned} \tag{4}$$

The conditional probability of heads on flip 1 given that Dagwood must diet is

$$\mathrm{P}[H_1|D] = \frac{\mathrm{P}[H_1 D]}{\mathrm{P}[D]} = 0. \tag{5}$$

Remember, if there was heads on flip 1, then Dagwood always postpones his diet.

(d) From part (b), we found that $\mathrm{P}[H_3] = 1/4$. To check independence, we calculate

$$\begin{aligned} \mathrm{P}\left[H_2\right] &= \mathrm{P}\left[T_1H_2H_3\right] + \mathrm{P}\left[T_1H_2T_3\right] + \mathrm{P}\left[T_1H_2T_4T_4\right] = 1/4 \\ \mathrm{P}\left[H_2H_3\right] &= \mathrm{P}\left[T_1H_2H_3\right] = 1/8. \end{aligned} \tag{6}$$

Now we find that

$$\mathrm{P}\left[H_2H_3\right] = 1/8 \neq \mathrm{P}\left[H_2\right]\mathrm{P}\left[H_3\right]. \tag{7}$$

Hence, H_2 and H_3 are dependent events. In fact, $\mathrm{P}[H_3|H_2] = 1/2$ while $\mathrm{P}[H_3] = 1/4$. The reason for the dependence is that given H_2 occurred, then we know there will be a third flip which may result in H_3. That is, knowledge of H_2 tells us that the experiment didn't end after the first flip.

2.2.5 In a game of rummy, you are dealt a seven-card hand.
(a) What is the probability P[R_7] that your hand has only red cards?
(b) What is the probability P[F] that your hand has only face cards?
(c) What is the probability P[R_7F] that your hand has only red face cards? (The face cards are jack, queen, and king.)

Since there are $H = \binom{52}{7}$ equiprobable seven-card hands, each probability is just the number of hands of each type divided by H.

(a) Since there are 26 red cards, there are $\binom{26}{7}$ seven-card hands with all red cards. The probability of a seven-card hand of all red cards is

$$\mathrm{P}[R_7] = \frac{\binom{26}{7}}{\binom{52}{7}} = \frac{26!\,45!}{52!\,19!} = 0.0049.$$

(b) There are 12 face cards in a 52 card deck and there are $\binom{12}{7}$ seven card hands with all face cards. The probability of drawing only face cards is

$$\mathrm{P}[F] = \frac{\binom{12}{7}}{\binom{52}{7}} = \frac{12!\,45!}{5!\,52!} = 5.92 \times 10^{-6}.$$

(c) There are 6 red face cards (J, Q, K of diamonds and hearts) in a 52 card deck. Thus it is impossible to get a seven-card hand of red face cards: P[R_7F] = 0.

2.2.7 Consider a binary code with 5 bits (0 or 1) in each code word. An example of a code word is 01010. How many different code words are there? How many code words have exactly three 0's?

There are $2^5 = 32$ different binary codes with 5 bits. The number of codes with exactly 3 zeros equals the number of ways of choosing the bits in which those zeros occur. Therefore there are $\binom{5}{3} = 10$ codes with exactly 3 zeros.

2.3.1 Consider a binary code with 5 bits (0 or 1) in each code word. An example of a code word is 01010. In each code word, a bit is a zero with probability 0.8, independent of any other bit.
(a) What is the probability of the code word 00111?
(b) What is the probability that a code word contains exactly three ones?

(a) Since the probability of a zero is 0.8, we can express the probability of the code word 00111 as 2 occurrences of a 0 and three occurrences of a 1. Therefore

$$\mathrm{P}\,[00111] = (0.8)^2(0.2)^3 = 0.00512.$$

(b) The probability that a code word has exactly three 1's is

$$\mathrm{P}\,[\text{three 1's}] = \binom{5}{3}(0.8)^2(0.2)^3 = 0.0512.$$

2.3.5♦ A collection of field goal kickers are divided into groups 1 and 2. Group i has $3i$ kickers. On any kick, a kicker from group i will kick a field goal with probability $1/(i+1)$, independent of the outcome of any other kicks.

(a) A kicker is selected at random from among all the kickers and attempts one field goal. Let K be the event that a field goal is kicked. Find $\mathrm{P}[K]$.

(b) Two kickers are selected at random; K_j is the event that kicker j kicks a field goal. Are K_1 and K_2 independent?

(c) A kicker is selected at random and attempts 10 field goals. Let M be the number of misses. Find $\mathrm{P}[M = 5]$.

Problem 2.3.5 Solution

(a) There are 3 group 1 kickers and 6 group 2 kickers. Using G_i to denote that a group i kicker was chosen, we have

$$\mathrm{P}[G_1] = 1/3, \qquad \mathrm{P}[G_2] = 2/3. \tag{1}$$

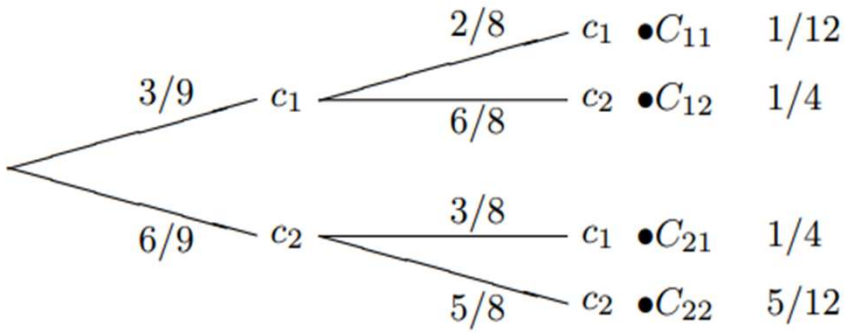
In addition, the problem statement tells us that

$$\mathrm{P}[K|G_1] = 1/2, \qquad \mathrm{P}[K|G_2] = 1/3. \tag{2}$$

Combining these facts using the Law of Total Probability yields

$$\begin{aligned} \mathrm{P}[K] &= \mathrm{P}[K|G_1]\,\mathrm{P}[G_1] + \mathrm{P}[K|G_2]\,\mathrm{P}[G_2] \\ &= (1/2)(1/3) + (1/3)(2/3) = 7/18. \end{aligned} \tag{3}$$

(b) To solve this part, we need to identify the groups from which the first and second kicker were chosen. Let c_i indicate whether a kicker was chosen from group i and let C_{ij} indicate that the first kicker was chosen from group i and the second kicker from group j. The experiment to choose the kickers is described by the sample tree:



Since a kicker from group 1 makes a kick with probability 1/2 while a kicker from group 2 makes a kick with probability 1/3,

$$\mathrm{P}\left[K_1K_2|C_{11}\right] = (1/2)^2, \qquad \mathrm{P}\left[K_1K_2|C_{12}\right] = (1/2)(1/3), \quad (4)$$

$$\mathrm{P}\left[K_1K_2|C_{21}\right] = (1/3)(1/2), \qquad \mathrm{P}\left[K_1K_2|C_{22}\right] = (1/3)^2. \quad (5)$$

By the law of total probability,

$$\begin{aligned}\mathrm{P}\left[K_1K_2\right] &= \mathrm{P}\left[K_1K_2|C_{11}\right]\mathrm{P}\left[C_{11}\right] + \mathrm{P}\left[K_1K_2|C_{12}\right]\mathrm{P}\left[C_{12}\right] \\ &\quad + \mathrm{P}\left[K_1K_2|C_{21}\right]\mathrm{P}\left[C_{21}\right] + \mathrm{P}\left[K_1K_2|C_{22}\right]\mathrm{P}\left[C_{22}\right] \\ &= \frac{1}{4}\frac{1}{12} + \frac{1}{6}\frac{1}{4} + \frac{1}{6}\frac{1}{4} + \frac{1}{9}\frac{5}{12} = 15/96. \qquad (6)\end{aligned}$$

It should be apparent that $\mathrm{P}[K_1] = \mathrm{P}[K]$ from part (a). Symmetry should also make it clear that $\mathrm{P}[K_1] = \mathrm{P}[K_2]$ since for any ordering of two kickers, the reverse ordering is equally likely. If this is not clear, we derive this result by calculating $\mathrm{P}[K_2|C_{ij}]$ and using the law of total probability to calculate $\mathrm{P}[K_2]$.

$$\mathrm{P}\left[K_2|C_{11}\right] = 1/2, \qquad \mathrm{P}\left[K_2|C_{12}\right] = 1/3, \qquad (7)$$

$$\mathrm{P}\left[K_2|C_{21}\right] = 1/2, \qquad \mathrm{P}\left[K_2|C_{22}\right] = 1/3. \qquad (8)$$

By the law of total probability,

$$\begin{aligned}\mathrm{P}\left[K_2\right] &= \mathrm{P}\left[K_2|C_{11}\right]\mathrm{P}\left[C_{11}\right] + \mathrm{P}\left[K_2|C_{12}\right]\mathrm{P}\left[C_{12}\right] \\ &\quad + \mathrm{P}\left[K_2|C_{21}\right]\mathrm{P}\left[C_{21}\right] + \mathrm{P}\left[K_2|C_{22}\right]\mathrm{P}\left[C_{22}\right] \\ &= \frac{1}{2}\frac{1}{12} + \frac{1}{3}\frac{1}{4} + \frac{1}{2}\frac{1}{4} + \frac{1}{3}\frac{5}{12} = \frac{7}{18}. \end{aligned} \tag{9}$$

We observe that K_1 and K_2 are not independent since

$$\mathrm{P}\left[K_1K_2\right] = \frac{15}{96} \neq \left(\frac{7}{18}\right)^2 = \mathrm{P}\left[K_1\right]\mathrm{P}\left[K_2\right]. \tag{10}$$

Note that $15/96$ and $(7/18)^2$ are close but not exactly the same. The reason K_1 and K_2 are dependent is that if the first kicker is successful, then it is more likely that kicker is from group 1. This makes it more likely that the second kicker is from group 2 and is thus more likely to miss.

$$\mathrm{P}\left[M|G_1\right] = \binom{10}{5}(1/2)^5(1/2)^5, \quad \mathrm{P}\left[M|G_2\right] = \binom{10}{5}(1/3)^5(2/3)^5. \quad (11)$$

We use the Law of Total Probability to find

$$\begin{aligned}\mathrm{P}\left[M\right] &= \mathrm{P}\left[M|G_1\right]\mathrm{P}\left[G_1\right] + \mathrm{P}\left[M|G_2\right]\mathrm{P}\left[G_2\right] \\ &= \binom{10}{5}\left((1/3)(1/2)^{10} + (2/3)(1/3)^5(2/3)^5\right). \qquad (12)\end{aligned}$$