

The Pigeonhole Principle

Introduction

Suppose that a flock of 20 pigeons flies into a set of 19 pigeonholes to roost. Because there are 20 pigeons but only 19 pigeonholes, at least one of these 19 pigeonholes must have at least two pigeons in it.

To see why this is true, note that if each pigeonhole had at most one pigeon in it, at most 19 pigeons, one per hole, could be accommodated.

This illustrates a general principle called the **pigeonhole principle**, which states that if there are more pigeons than pigeonholes, then there must be at least one pigeonhole with at least two pigeons in it

THEOREM 1: THE PIGEONHOLE PRINCIPLE / DIRICHLET DRAWER PRINCIPLE

If k is a positive integer and $k + 1$ or more objects are placed into k boxes, then there is at least one box containing two or more of the objects

Proof: We prove the pigeonhole principle using a proof by contraposition. Suppose that none of the k boxes contains more than one object. Then the total number of objects would be at most k . This is a contradiction, because there are at least $k + 1$ objects

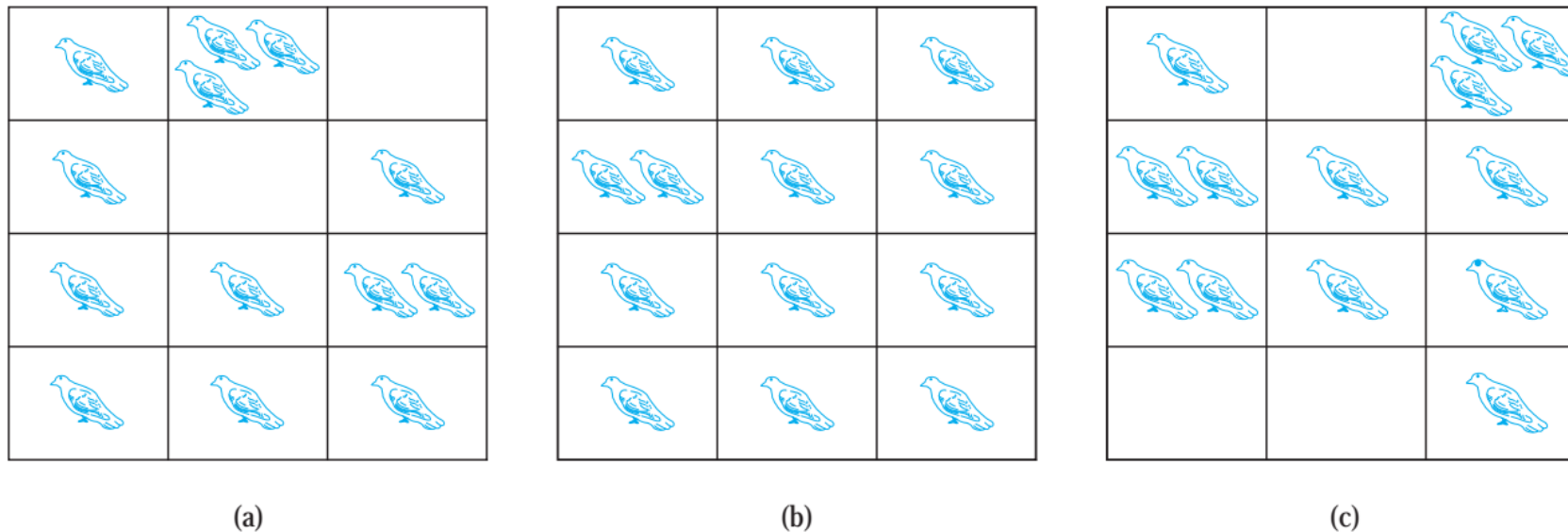


FIGURE 1 There Are More Pigeons Than Pigeonholes.

We now illustrate the usefulness of the pigeonhole principle, first we show that it can be used to prove a useful corollary about functions.

COROLLARY 1 A function f from a set with $k + 1$ or more elements to a set with k elements is not one-to-one

Proof: Suppose that for each element y in the codomain of f we have a box that contains all elements x of the domain of f such that $f(x) = y$. Because the domain contains $k + 1$ or more elements and the codomain contains only k elements, the pigeonhole principle tells us that one of these boxes contains two or more elements x of the domain. This means that f cannot be one-to-one.

The Examples 1–3 given below, how the pigeonhole principle is used

EXAMPLE 1 Among any group of 367 people, there must be at least two with the same birthday, because there are only 366 possible birthdays.

EXAMPLE 2 In any group of 27 English words, there must be at least two that begin with the same letter, because there are 26 letters in the English alphabet.

EXAMPLE 3 How many students must be in a class to guarantee that at least two students receive the same score on the final exam, if the exam is graded on a scale from 0 to 100 points?

Solution: There are 101 possible scores on the final. The pigeonhole principle shows that among any 102 students there must be at least 2 students with the same score. The pigeonhole principle is a useful tool in many proofs, including proofs of surprising results, such as that given in Example 4.

EXAMPLE 4 Show that for every integer n there is a multiple of n that has only 0s and 1s in its decimal expansion.

Solution: Let n be a positive integer. Consider the $n + 1$ integers 1, 11, 111, ..., 11 ... 1 (where the last integer in this list is the integer with $n + 1$ 1s in its decimal expansion). Note that there are n possible remainders when an integer is divided by n .

Because there are $n + 1$ integers in this list, by the pigeonhole principle there must be two with the same remainder when divided by n . The larger of these integers less the smaller one is a multiple of n , which has a decimal expansion consisting entirely of 0s and 1s

The Generalized Pigeonhole Principle

The pigeonhole principle states that there must be at least two objects in the same box when there are more objects than boxes. However, even more can be said when the number of objects exceeds a multiple of the number of boxes.

For instance, among any set of 21 decimal digits there must be 3 that are the same. This follows because when 21 objects are distributed into 10 boxes, one box must have more than 2 objects.

THEOREM 2 THE GENERALIZED PIGEONHOLE PRINCIPLE

If N objects are placed into k boxes, then there is at least one box containing at least $\lceil N/k \rceil$ objects

Proof: We will use a proof by contraposition. Suppose that none of the boxes contains more than $\lceil N/k \rceil - 1$ objects. Then, the total number of objects is at most

$$k \left(\lceil \frac{N}{k} \rceil - 1 \right) < k \left(\left(\lceil \frac{N}{k} \rceil + 1 \right) - 1 \right)$$

where the inequality $\lceil N/k \rceil < (N/k) + 1$ has been used.

This is a contradiction because there are a total of N objects.

A common type of problem asks for the minimum number of objects such that at least r of these objects must be in one of k boxes when these objects are distributed among the boxes. When we have N objects, the generalized pigeonhole principle tells us there must be at least r objects in one of the boxes as long as $\lceil N/k \rceil \geq r$.

Examples 5–8 illustrate how the generalized pigeonhole principle is applied.

EXAMPLE 5 Among 100 people there are at least $\lceil 100/12 \rceil = 9$ who were born in the same month.

EXAMPLE 6 What is the minimum number of students required in a discrete mathematics class to be sure that at least six will receive the same grade, if there are five possible grades, A, B, C, D, and F?

Solution: The minimum number of students needed to ensure that at least six students receive the same grade is the smallest integer N such that $\lceil N/5 \rceil = 6$. The smallest such integer is $N = 5 \cdot 5 + 1 = 26$.

If you have only 25 students, it is possible for there to be five who have received each grade so that no six students have received the same grade. Thus, 26 is the minimum number of students needed to ensure that at least six students will receive the same grade.

EXAMPLE 7

- a) How many cards must be selected from a standard deck of 52 cards to guarantee that at least three cards of the same suit are chosen?
- b) How many must be selected to guarantee that at least three hearts are selected?

A standard deck of 52 cards has 13 kinds of cards, with four cards of each of kind, one in each of the four suits, hearts, diamonds, spades, and clubs

EXAMPLE 8 What is the least number of area codes needed to guarantee that the 25 million phones in a state can be assigned distinct 10-digit telephone numbers? (Assume that telephone numbers are of the form $NXX-NXX-XXXX$, where the first three digits form the area code, N represents a

digit from 2 to 9 inclusive, and X represents any digit.)

Solution: There are eight million different phone numbers of the form $NXX-XXXX$ (as shown in Example 8 of Section 6.1).

Hence, by the generalized pigeonhole principle, among 25 million telephones, at least $\lceil 25,000,000/8,000,000 \rceil = 4$ of them must have identical phone numbers.

Hence, at least four area codes are required to ensure that all 10-digit numbers are different.

Example 9, although not an application of the generalized pigeonhole principle, makes use of similar principles.

EXAMPLE 9 Suppose that a computer science laboratory has 15 workstations and 10 servers. A cable can be used to directly connect a workstation to a server. For each server, only one direct connection to that server can be active at any time. We want to guarantee that at any time any set of 10 or fewer workstations can simultaneously access different servers via direct connections. Although we could do this by connecting every workstation directly to every server (using 150 connections), what is the minimum number of direct connections needed to achieve this goal?

Solution: Suppose that we label the workstations W_1, W_2, \dots, W_{15} and the servers S_1, S_2, \dots, S_{10} . Furthermore, suppose that we connect W_k to S_k for $u = 1, 2, \dots, 10$ and each of $W_{11}, W_{12}, W_{13}, W_{14}$, and W_{15} to all 10 servers. We have a total of 60 direct connections. Clearly any set of 10 or fewer workstations can simultaneously access different servers.

We see this by noting that if workstation W_j is included with $1 \leq j \leq 10$, it can access server S_j , and for each workstation W_k with $k \geq 11$ included, there must be a corresponding workstation W_j with $1 \leq j \leq 10$ not included, so W_k can access server S_j . (This follows because there are at least as many available servers S_j as there are workstations W_j with $1 \leq j \leq 10$ not included.)

Now suppose there are fewer than 60 direct connections between workstations and servers.

Then some server would be connected to at most $\lfloor 59/10 \rfloor = 5$ workstations. (If all servers were connected to at least six workstations, there would be at least $6 \cdot 10 = 60$ direct connections.)

This means that the remaining nine servers are not enough to allow the other 10 workstations to simultaneously access different servers. Consequently, at least 60 direct connections are needed.

It follows that 60 is the answer

Some Elegant Applications of the Pigeonhole Principle

In many interesting applications of the pigeonhole principle, the objects to be placed in boxes must be chosen in a clever way. A few such applications will be described here.

EXAMPLE 10 During a month with 30 days, a baseball team plays at least one game a day, but no more than 45 games. Show that there must be a period of some number of consecutive days during which the team must play exactly 14 games.

Solution: Let a_j be the number of games played on or before the j^{th} day of the month. Then a_1, a_2, \dots, a_{30} is an increasing sequence of distinct positive integers, with $1 \leq a_j \leq 45$. Moreover, $a_1 + 14, a_2 + 14, \dots, a_{30} + 14$ is also an increasing sequence of distinct positive integers, with $15 \leq a_j + 14 \leq 59$.

The 60 positive integers $a_1, a_2, \dots, a_{30}, a_1 + 14, a_2 + 14, \dots, a_{30} + 14$ are all less than or equal to 59. Hence, by the pigeonhole principle two of these integers are equal.

Because the integers $a_j, j = 1, 2, \dots, 30$ are all distinct and the integers $a_j + 14, j = 1, 2, \dots, 30$ are all distinct, there must be indices i and j with $a_i = a_j + 14$. This means that exactly 14 games were played from day $j + 1$ to day i .

EXAMPLE 11 Show that among any $n + 1$ positive integers not exceeding $2n$ there must be an integer that divides one of the other integers.

Solution: Write each of the $n + 1$ integers a_1, a_2, \dots, a_{n+1} as a power of 2 times an odd integer.

In other words, let $a_j = 2^{k_j} q_j$ for $j = 1, 2, \dots, n + 1$, where k_j is a nonnegative integer and q_j is odd.

The integers q_1, q_2, \dots, q_{n+1} are all odd positive integers less than $2n$. Because there are only n odd positive integers less than $2n$, it follows from the pigeonhole principle that two of the integers q_1, q_2, \dots, q_{n+1} must be equal.

Therefore, there are distinct integers i and j such that $q_i = q_j$. Let q be the common value of q_i and q_j . Then, $a_i = 2^{k_i} q$ and $a_j = 2^{k_j} q$. It follows that if $k_i < k_j$, then a_i divides a_j ; while if $k_i > k_j$, then a_j divides a_i .

- Suppose that a_1, a_2, \dots, a_N is a sequence of real numbers. A **subsequence** of this sequence is a sequence of the form $a_{i_1}, a_{i_2}, \dots, a_{i_m}$, where $1 \leq i_1 < i_2 < \dots < i_m \leq N$.
- Hence, a subsequence is a sequence obtained from the original sequence by including some of the terms of the original sequence in their original order, and perhaps not including other terms.
- A sequence is called **strictly increasing** if each term is larger than the one that precedes it, and it is called **strictly decreasing** if each term is smaller than the one that precedes it.

THEOREM 3

Every sequence of $n^2 + 1$ distinct real numbers contains a subsequence of length $n + 1$ that is either strictly increasing or strictly decreasing

Proof: Let $a_1, a_2, \dots, a_{n^2+1}$ be a sequence of $n^2 + 1$ distinct real numbers. Associate an ordered pair with each term of the sequence, namely, associate (i_k, d_k) to the term a_k , where i_k is the length of the longest increasing subsequence starting at a_k , and d_k is the length of the longest decreasing subsequence starting at a_k .

Suppose that there are no increasing or decreasing subsequences of length $n + 1$. Then i_k and d_k are both positive integers less than or equal to n , for $k = 1, 2, \dots, n^2 + 1$.

Hence, by the product rule there are n^2 possible ordered pairs for (i_k, d_k) .

By the pigeonhole principle, two of these $n^2 + 1$ ordered pairs are equal. In other words, there exist terms a_s and a_t , with $s < t$ such that $i_s = i_t$ and $d_s = d_t$.

We will show that this is impossible. Because the terms of the sequence are distinct, either $a_s < a_t$ or $a_s > a_t$.

If $a_s < a_t$, then, because $i_s = i_t$, an increasing subsequence of length $i_t + 1$ can be built starting at a_s , by taking a_s followed by an increasing subsequence of length i_t beginning at a_t .

This is a contradiction. Similarly, if $a_s > a_t$, the same reasoning shows that d_s must be greater than d_t , which is a contradiction.

EXAMPLE 12 The sequence 8, 11, 9, 1, 4, 6, 12, 10, 5, 7 contains 10 terms. Note that $10 = 3^2 + 1$.

There are four strictly increasing subsequences of length four, namely, 1, 4, 6, 12; 1, 4, 6, 7; 1, 4, 6, 10; and 1, 4, 5, 7.

There is also a strictly decreasing subsequence of length four, namely, 11, 9, 6, 5.

The final example shows how the generalized pigeonhole principle can be applied to an important part of combinatorics called **Ramsey theory**, after the English mathematician F. P. Ramsey. In general, Ramsey theory deals with the distribution of subsets of elements of sets.

The **Ramsey number** $R(m, n)$, where m and n are positive integers greater than or equal to 2, denotes the minimum number of people at a party such that there are either m mutual friends or n mutual enemies, assuming that every pair of people at the party are friends or enemies.

Example 13 shows that $R(3, 3) \leq 6$. We conclude that $R(3, 3) = 6$ because in a group of five people where every two people are friends or enemies, there may not be three mutual friends or three mutual enemies (see Exercise 26).

It is possible to prove some useful properties about Ramsey numbers, but for the most part it is difficult to find their exact values. Note that by symmetry it can be shown that $R(m, n) = R(n, m)$ (see Exercise 30). We also have $R(2, n) = n$ for every positive integer $n \geq 2$ (see Exercise 29). The exact values of only nine Ramsey numbers $R(m, n)$ with $3 \leq m \leq n$ are known, including $R(4, 4) = 18$. Only bounds are known for many other Ramsey numbers, including $R(5, 5)$, which is known to satisfy $43 \leq R(5, 5) \leq 49$.

EXAMPLE 13 Assume that in a group of six people, each pair of individuals consists of two friends or two enemies. Show that there are either three mutual friends or three mutual enemies in the group.

Solution: Let A be one of the six people. Of the five other people in the group, there are either three or more who are friends of A , or three or more who are enemies of A . This follows from the generalized pigeonhole principle, because when five objects are divided into two sets, one of the sets has at least $\lceil 5/2 \rceil = 3$ elements. In the former case, suppose that B , C , and D are friends of A . If any two of these three individuals are friends, then these two and A form a group of three mutual friends. Otherwise, B , C , and D form a set of three mutual enemies. The proof in the latter case, when there are three or more enemies of A , proceeds in a similar manner