Chapter 9

Systems of Distinct Representatives

This short chapter serves as an interlude between the basic enumerative Chapters 2 and 4 to 8, and the remaining chapters of the book. We begin by discussing three problems:

Problem 1. Consider an m-by-n chessboard in which certain squares are forbidden and the others are free. What is the largest number of nonattacking rooks that can be placed in free positions on the board?

In previous sections we considered the problem of counting the number of ways to place n nonattacking rooks on an n-by-n board. Our underlying assumption was that this number was positive; that is, it was possible to place n nonattacking rooks on the board. Now we are concerned not only with whether or not it is possible to place n nonattacking rooks on the board but, more generally, with the question of the largest number of nonattacking rooks that can be placed on a rectangular board.

Problem 2. Consider again an m-by-n chessboard where certain squares are forbidden and the others are free. What is the largest number of dominoes that can be placed on the board so that each domino covers two free squares and no two dominoes overlap (cover the same square)?

In Chapter 1 we considered the special case of this problem concerning when a board with forbidden squares has a tiling (perfect cover). For a tiling, we must have, in addition, that every free square is covered by a domino. If p is the total number of free squares, then there is a tiling if and only if p is even, and the answer to Problem 2 is $p/2$. In the general case, some free squares may not be covered by any domino.

Problem 3. A company has n jobs available, with each job requiring certain qualifications. There are m people who apply for the n jobs. What is the largest number of jobs that can be filled from the applicant pool if a job can be filled only by a person who meets its qualifications?

The first two problems are of a recreational nature. The third problem, however, is clearly of a more serious and applied nature. As a matter of fact, Problems 1 and

3 are different formulations of the same abstract problem, and Problem 2 is merely a special case. In this chapter we solve the abstract problem and thereby solve each of Problems 1, 2, and 3. Of course, in Problem 3, we would want to know not only the largest number of jobs that can be filled with qualified applicants, but also a particular assignment of the largest number of applicants to jobs they qualify for. (A similar remark applies to Problems 1 and 2.) We shall discuss this in Chapter 13 in the context of a different model for the problem.

9.1 General Problem Formulation

Each of Problems 1, 2, and 3 has a common abstract formulation which we now discuss.

Let Y be a finite set, and let $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ be a family[1] of n subsets of Y. A family $(e_1, e_2, \ldots, e_n)$ of elements of Y is called a *system of representatives*, abbreviated SR, of $\mathcal{A}$, provided that

$$e_1 \text{ is in } A_1, e_2 \text{ is in } A_2, \ldots, e_n \text{ is in } A_n.$$

In a system of representatives, the element e_i belongs to A_i and thus "represents" the set A_i. If, in a system of representatives, the elements $e_1, e_2, \ldots, e_n$ are all different, then $(e_1, e_2, \ldots, e_n)$ is called a *system of distinct representatives*, abbreviated SDR. Note that even though, for example, A_1 and A_2 may be equal as sets, they must have different representatives in an SDR because they are different terms of the family.

Example. Let (A_1, A_2, A_3, A_4) be the family of subsets of the set $Y = \{a, b, c, d, e\}$, defined by

$$A_1 = \{a, b, c\}, A_2 = \{b, d\}, A_3 = \{a, b, d\}, A_4 = \{b, d\}.$$

Then (a, b, b, d) is an SR, and (c, b, a, d) is an SDR.

□

A family $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ of *nonempty* sets always has an SR. We need only pick any element from each of the sets $A_1, A_2, \ldots, A_n$ to obtain an SR. However, the family $\mathcal{A}$ need not have an SDR even though all the sets in the family are nonempty. For instance, if there are two sets in the family, say, A_1 and A_2, each containing only one element, and the element in A_1 is the same as the element in A_2, that is,

$$A_1 = \{x\}, \quad A_2 = \{x\},$$

then the family $\mathcal{A}$ does not have an SDR. This is because, in any SR, x has to represent both A_1 and A_2, and thus no SDR exists (no matter what $A_3, \ldots, A_n$ are). But this is not the only way in which a family $\mathcal{A}$ can fail to have an SDR.

[1] A family as used here is really the same as a sequence, but not a sequence of numbers. We have here a sequence whose terms are sets. As in sequences of numbers, different terms can be equal; thus, some sets in the family may be equal.

Example. Let the family $\mathcal{A} = (A_1, A_2, A_3, A_4)$ be defined by

$$A_1 = \{a, b\}, A_2 = \{a, b\}, A_3 = \{a, b\}, A_4 = \{a, b, c, d\}.$$

Then $\mathcal{A}$ does not have an SDR because in any system of representatives, A_1 has to be represented by either a or b, A_2 has to be represented by either a or b, and A_3 has to be represented by either a or b. So we have two elements, namely, a and b, from which the representatives of three sets, namely, A_1, A_2, and A_3, have to be drawn. By the pigeonhole principle, two of the three sets A_1, A_2 and A_3 have to be represented by the same element. Hence no SDR is possible. □

Example. Consider the 4-by-5 board with forbidden positions pictured in Figure 9.1 and the problem of placing nonattacking rooks on this board. The rooks have to be placed in the free squares.

	1	2	3	4	5
A_1		×			
A_2			×		×
A_3	×		×		×
A_4	×				

Figure 9.1

In the diagram each row has one of the labels A_1, A_2, A_3, A_4 and each column has one of the labels $1, 2, 3, 4, 5$. These labels indicate that, with this board, we associate the family $\mathcal{A} = (A_1, A_2, A_3, A_4)$ of subsets of $Y = \{1, 2, 3, 4, 5\}$, where A_i is the set of columns in which the free squares in row i lie: thus,

$$A_1 = \{1, 3, 4, 5\}, A_2 = \{1, 2, 4\}, A_3 = \{2, 4\}, A_4 = \{2, 3, 4, 5\}.$$

It is possible to place four nonattacking rooks on this board if and only if the associated family $\mathcal{A}$ has an SDR. For example, the four nonattacking rooks in Figure 9.2

correspond to the SDR $(4, 1, 2, 5)$ of $\mathcal{A}$.[2]

	1	2	3	4	5
A_1		$\times$		$\otimes$	
A_2	$\otimes$		$\times$		$\times$
A_3	$\times$	$\otimes$	$\times$		$\times$
A_4	$\times$				$\otimes$

Figure 9.2

□

The discussion in the previous example applies in general, to any problem of placing nonattacking rooks on a board with forbidden positions. More precisely, with any m-by-n board B with forbidden positions, we associate a family $\mathcal{A} = (A_1, A_2, \ldots, A_m)$ of subsets of the set $Y = \{1, 2, \ldots, n\}$, called the *rook family of the board*, where

$$A_i = \{k : \text{ the } k\text{th square in row } i \text{ is free}\} \quad (i = 1, 2, \ldots, m)$$

is the set of columns having a free square in row i. It is possible to place m nonattacking rooks in free positions on the board if and only if the rook family $\mathcal{A}$ has an SDR. More generally, if k is an integer, then it is possible to place k nonattacking rooks on the board if and only if there is a *subfamily*[3] $\mathcal{A}(i_1, i_2, \ldots, i_k) = (A_{i_1}, A_{i_2}, \ldots, A_{i_k})$ of k sets, where $1 \le i_1 < i_2 < \cdots i_k \le m$, with an SDR. The rooks will go into rows $i_1, i_2, \ldots, i_k$ and the respective columns given by the SDR.

In fact, this is all reversible in that any family $\mathcal{A} = (A_1, A_2, \ldots, A_m)$ of m subsets of $Y = \{1, 2, \ldots, n\}$ of n elements is the rook family of some m-by-n board with forbidden positions, where an SDR corresponding to m nonattacking rooks in free positions on the board. We simply construct the m-by-n board of which the position in row i and column j is free if and only if j belongs to A_i and is forbidden otherwise.

Example. Consider a 4-by-5 board whose squares are alternately colored black and white and where some of the squares are forbidden. For identification we label the free white squares $w_1, w_2, \ldots, w_7$ and the free black squares $b_1, b_2, \ldots, b_7$, as shown in Figure 9.3.

[2]Another was to describe this 4-by-5 board is by a 4-by-5 *bit matrix* or *incidence matrix*. This is the 4-by-5 matrix

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix},$$

which has a 0 in row i and column j if the corresponding position of the board is forbidden and a 1 if the position is free. Placing nonattacking rooks on the board is equivalent to picking a bunch of 1s no two from the same row and no two from the same column. The boldface 1s correspond to the placement of rooks in Figure 9.2.

[3]A family is a sequence of sets; a subfamily is a subsequence of that sequence.

w_1	$\times$	w_2	b_1	w_3
b_2	w_4	$\times$	w_5	b_3
$\times$	b_4	$\times$	b_5	$\times$
$\times$	w_6	b_6	w_7	b_7

Figure 9.3

We associate with this board a family $\mathcal{A} = (A_1, A_2, A_3, A_4, A_5, A_6, A_7)$ of subsets of the set of black squares, one subset for each white square, as follows. We let A_i equal the set of all black squares that share an edge with white square w_i, $(i = 1, 2, 3, 4, 5, 6, 7)$. Thus

$$A_1 = \{b_2\}, A_2 = \{b_1\}, A_3 = \{b_1, b_3\}, A_4 = \{b_2, b_4\}, A_5 = \{b_1, b_3, b_5\},$$

$$A_6 = \{b_4, b_6\}, A_7 = \{b_5, b_6, b_7\}.$$

If a domino is placed on the board and covers square w_i, then it must cover one of the black squares in A_i. Hence A_i consists of all the black squares that can be covered by a domino that also covers white square w_i. We see that the 4-by-5 board has a tiling if and only if $\mathcal{A}$ has an SDR. □

The discussion in the previous example can be carried out for any tiling problem by dominoes. We simply list the free white squares $w_1, w_2, \ldots, w_m$ in some order and list the free black squares $b_1, b_2, \ldots, b_n$ in some order (the number m of white squares must equal the number n of black squares if there is to be a tiling, but we need not restrict ourselves in this way), and form the family $\mathcal{A} = (A_1, A_2, \ldots, A_m)$, one set for each free white square, where A_i is the set of black squares sharing an edge with white square w_i, $(i = 1, 2, \ldots, m)$. The family $\mathcal{A}$ is called the *domino family of the board.* There is a tiling of the board if and only if the domino family $\mathcal{A}$ has an SDR. More generally, if k is an integer, then it is possible to place k nonoverlapping dominoes on the board if and only if there is a subfamily $\mathcal{A}(i_1, i_2, \ldots, i_k) = (A_{i_1}, A_{i_2}, \ldots, A_{i_k})$ of k sets, where $1 \leq i_1 < i_2 < \cdots i_k \leq m$, with an SDR. The dominoes will be placed on white squares $w_{i_1}, w_{i_2}, \ldots, w_{i_k}$ and the respective black squares corresponding to representatives in the SDR.

It should now be clear that Problem 3 in the introduction, of assigning applicants to jobs for which they qualify, is a just a general SDR problem. Let the jobs be labeled $p_1, p_2, \ldots, p_n$. Then to the ith applicant we associate the set A_i of jobs for which he or she qualifies. Assignment of people to jobs to which they qualify is the same as finding an SDR of the family $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ or one of its subfamilies.

We are now ready to formulate our general problem:

Let $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ be a family of subsets of a finite set Y. Determine when $\mathcal{A}$ has an SDR. If $\mathcal{A}$ does not have an SDR, what is the largest number t of sets in a subfamily $\mathcal{A}(i_1, i_2, \ldots, i_t) = (A_{i_1}, A_{i_2}, \ldots, A_{i_t})$ that does have an SDR?

Solving this problem solves each of Problems 1, 2, and 3 in the introduction to this chapter.

9.2 Existence of SDRs

We begin by identifying a general necessary condition for the existence of an SDR.

Let $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ be a family of sets. Let k be an integer with $1 \leq k \leq n$. In order for $\mathcal{A}$ to have an SDR, it is necessary that the union of every k sets of the family $\mathcal{A}$ contain at least k elements. Why is this so? Suppose, to the contrary, that there are k sets, to be explicit, say, $A_1, A_2, \ldots, A_k$, which together contain fewer than k elements; that is,

$$A_1 \cup A_2 \cup \cdots \cup A_k = F, \text{ where } |F| < k.$$

Then the representatives of each of the k sets $A_1, A_2, \ldots, A_k$ have to be drawn from the elements of the set F. Since F has fewer than k elements, it follows from the pigeonhole principle that two of the k sets $A_1, A_2, \ldots, A_k$ have to be represented by the same element. Hence, there can be no SDR. We formulate this necessary condition as the next lemma.

Lemma 9.2.1 *In order for the family $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ of sets to have an SDR, it is necessary that the following condition hold:*

> (MC)*: For each $k = 1, 2, \ldots, n$ and each choice of k distinct indices $i_1, i_2, \ldots, i_k$ from $\{1, 2, \ldots, n\}$,*
>
> $$|A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}| \geq k; \qquad (9.1)$$
>
> *in short, every k sets of the family collectively contain at least k elements.*
>
> □

Condition MC in Lemma 9.2.1 is often called the *marriage condition*. The reason stems from the following amusing and classical formulation of the problem of systems of distinct representatives.

Example (*The Marriage Problem*). There are n men and m women, and all the men are eager to marry. If there were no restrictions on who marries whom, then, in order to marry off all the men, we need only require that the number m of women be at least as large as the number n of men. But we would expect that each man and each woman would insist on some compatibility with a spouse, thereby eliminating some of the women as potential spouses for each man. Thus, each man would arrive at a certain set of compatible women from the set of available women.[4] Let $(A_1, A_2, \ldots, A_n)$ be

[4]This is sounding like the problem of assigning applicants to jobs, isn't it? The women are the "jobs," and the compatible women for a man are the jobs for which he has the qualifications.

the family of subsets of the women, where A_i denotes the set of compatible women for the ith man $(i = 1, \ldots, n)$. Then *marrying off all the men corresponds to an SDR* $(w_1, w_2, \ldots, w_n)$ *of* $(A_1, A_2, \ldots, A_n)$. The correspondence is that the ith man marries the woman w_i, $(i = 1, 2, \ldots, n)$. Since w_i is in A_i, w_i is a woman compatible with the ith man. Since $(w_1, w_2, \ldots, w_n)$ is a system of *distinct* representatives, no two men are claiming the same woman.[5] In the context of this example, the MC asserts that the combined lists of any set of k men have to contain at least k women, and thus this is a necessary condition for all the men to be able to marry a compatible woman. □

The marriage condition (9.1) is not only a necessary condition for the existence of SDR but, surprisngly, a sufficient condition as well. It thus provides a characterization for the existence of an SDR.

Theorem 9.2.2 *The family* $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ *of subsets of a set* Y *has an SDR if and only if the marriage condition* MC *holds.*

Proof. By Lemma 9.2.1 we know that if $\mathcal{A}$ has an SDR, then the marriage condition holds. We now assume that the marriage condition holds and show that $\mathcal{A}$ has an SDR. The proof we give is by induction on the number n of sets of the family $\mathcal{A}$.

To get started, if $n = 1$, that is, $\mathcal{A} = (A_1)$, then MC says that $|A_1| \geq 1$. Hence, choosing any element in A_1, we get an SDR for $\mathcal{A}$ in this case.

Now suppose that $n \geq 2$. There are two cases to be considered, which could be described as the *tight* case and the *room-to-spare* case.

The tight case: There is an integer k with $1 \leq k \leq n-1$ and a subfamily of $\mathcal{A}$ of k sets whose union contains exactly k elements. (By MC the union cannot contain fewer that k elements, so we are tight.) For simplicity of notation, let us assume[6] that the k sets are the first k sets $A_1, A_2, \ldots, A_k$. So letting $E = A_1 \cup A_2 \cup \cdots \cup A_k$, we have

$$|E| = k.$$

Since $\mathcal{A}$ satisfies MC, then so does its subfamily $(A_1, A_2, \ldots, A_k)$. Since $k < n$, it follows by the induction hypothesis that $(A_1, A_2, \ldots, A_k)$ has an SDR $(e_1, e_2, \ldots, e_k)$. Because $E = A_1 \cup A_2 \cup \cdots \cup A_k$, $|E| = k$, and $e_1, e_2, \ldots, e_k$ are distinct, we have $E = \{e_1, e_2, \ldots, e_k\}$. Thus if $\mathcal{A}$ is to have an SDR, none of the remaining sets in the subfamily $(A_{k+1}, A_{k+2}, \ldots, A_n)$ can have their representative from E.

So we consider the family

$$\mathcal{A}^* = (A_{k+1} \setminus E, A_{k+2} \setminus E, \ldots, A_n \setminus E)$$

of $n-k$ sets obtained by removing the elements of E from the sets $A_{k+1}, A_{k+2}, \ldots, A_n$. Since $k \geq 1$, $n-k < n$. So if we can show that $\mathcal{A}^*$ satisfies MC, we can use the induction

[5] We forgot to say that no woman is allowed two spouses.

[6] Actually, listing the sets in the family in a different order affects neither the MC nor the existence of an SDR.

hypothesis again to conclude it has an SDR $(f_{k+1}, f_{k+2}, \dots, f_n)$. None of the f's can equal any of the e's, and it will follow that $(e_1, e_2, \dots, e_k, f_{k+1}, f_{k+2}, \dots, f_n)$ is an SDR for $\mathcal{A}$, completing the induction.

So let's show that $\mathcal{A}^*$ satisfies MC. Take any l sets

$$A_{j_1} \setminus E, A_{j_2} \setminus E, \dots, A_{j_l} \setminus E$$

of $\mathcal{A}^*$, where $k+1 \le j_1 < j_2 < \cdots < j_l \le n$, and consider the $k+l$ sets

$$A_1, A_2, \dots, A_k, A_{j_1}, A_{j_2}, \dots, A_{j_l}$$

of the family $\mathcal{A}$. Since MC holds for $\mathcal{A}$, then, using elementary calculations, we have

$$\begin{aligned}
|A_1 \cup A_2 \cup \cdots \cup A_k \cup A_{j_1} \cup A_{j_2} \cup \cdots \cup A_{j_l}| &\ge k+l \\
|E \cup A_{j_1} \cup A_{j_2} \cup \cdots \cup A_{j_l}| &\ge k+l \\
|E| + |(A_{j_1} \setminus E) \cup (A_{j_2} \setminus E) \cup \cdots \cup (A_{j_l} \setminus E)| &\ge k+l \\
k + |(A_{j_1} \setminus E) \cup (A_{j_2} \setminus E) \cup \cdots \cup (A_{j_l} \setminus E)| &\ge k+l \\
|(A_{j_1} \setminus E) \cup (A_{j_2} \setminus E) \cup \cdots \cup (A_{j_l} \setminus E)| &\ge l.
\end{aligned}$$

Thus $\mathcal{A}^*$ satisfies MC and hence has an SDR, and, as shown, this implies that $\mathcal{A}$ has an SDR.

The room to spare case: For every integer k with $1 \le k \le n-1$ and every subfamily of $\mathcal{A}$ of k sets, the union contains at least $k+1$ elements. (So the union contains more elements than needed for MC, and we have room to spare.) With room to spare, the proof ought to be easier, and it is. Each set of the family $\mathcal{A}$ contains at least one element, indeed two because of room to spare. So take A_n and any element e_n that it contains. Consider the family $\mathcal{A}' = (A'_1, A'_2, \dots, A'_{n-1})$ obtained from $A_1, A_2, \dots, A_{n-1}$ by deleting e_n from each set that contains it. We claim that $\mathcal{A}'$ satisfies MC. Indeed, since we have room to spare and we have only eliminated one element from $A_1, A_2, \dots, A_n$, for each integer k with $1 \le k \le n-1$ and each choice of indices $i_1, i_2, \dots, i_k$ with $1 \le i_1 < i_2 < \cdots < i_k \le n-1$ we have

$$|A'_{i_1} \cup A'_{i_2} \cup \cdots \cup A'_{i_k}| \ge |A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}| - 1 \ge (k+1) - 1 = k.$$

Hence $\mathcal{A}'$ satisfies MC and so by the induction hypothesis has an SDR $(e_1, e_2, \dots, e_{n-1})$. Since none of these elements can equal e_n, $(e_1, e_2, \dots, e_{n-1}, e_n)$ is an SDR of $\mathcal{A}$. Therefore, the theorem holds by induction. $\square$

If the marriage condition fails so that there is no SDR, then we would like to know the largest number of sets in a subfamily with an SDR. To answer this we first prove the following theorem.

Theorem 9.2.3 *Let $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ be a family of subsets of a finite set Y. Let t be an integer with $0 \leq t \leq n$. Then there exists a subfamily of t sets of $\mathcal{A}$ that has an SDR if and only if*

$$|A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}| \geq k - (n - t) \tag{9.2}$$

for all k with $k \geq n-t$ and all choices of k distinct indices $i_1, i_2, \ldots, i_k$ from $\{1, 2, \ldots, n\}$

Proof. Note that Theorem 9.2.2 is the special case obtained by taking $t = n$, and we shall actually derive it from that theorem. Let F be a set of $n - t$ elements completely disjoint from Y: $F \cap Y = \emptyset$. We define a family $\mathcal{A}^* = (A_1^*, A_2^*, \ldots, A_n^*)$ of subsets of $F \cup Y$ by putting all the elements of F into all the sets of $\mathcal{A}$:

$$A_i^* = A_i \cup F \quad (i = 1, 2, \ldots, n).$$

We claim that $\mathcal{A}$ has a family of t sets with an SDR if and only if $\mathcal{A}^*$ has an SDR. First suppose $\mathcal{A}^*$ has an SDR. Then since $|F| = k$, at most k of the elements in this SDR come from F, and hence at least $n - k$ come from Y, thereby forming an SDR of at least $n - k$ sets of the family $\mathcal{A}$. Conversely, suppose $\mathcal{A}$ has a subfamily of t sets having an SDR. For convenience of notation, let these t sets be $A_1, A_2, \ldots, A_t$ and let the SDR be $(y_1, y_2, \ldots, y_t)$. Let the $n - t$ elements of F be $f_{t+1}, f_{t+2}, \ldots, f_n$. Then

$$(y_1, y_2, \ldots, y_t, f_{t+1}, f_{t+2}, \ldots, f_n)$$

is an SDR of $\mathcal{A}^*$. Thus our claim holds.

We now apply Theorem 9.2.2 to $\mathcal{A}^*$. By that theorem, $\mathcal{A}^*$ has an SDR if and only if for each $k = 1, 2, \ldots, n$ and each choice of k distinct indices $i_1, i_2, \ldots, i_k$ from $\{1, 2, \ldots, n\}$,

$$|A_{i_1}^* \cup A_{i_2}^* \cup \cdots \cup A_{i_k}^*| \geq k. \tag{9.3}$$

Since

$$A_{i_1}^* \cup A_{i_2}^* \cup \cdots \cup A_{i_k}^* = (A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}) \cup F,$$

and since

$$\begin{aligned} |A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k} \cup F| &= |(A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k})| + |F| \\ &= |A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}| + n - t, \end{aligned}$$

we see that the conditions of (9.3) are equivalent to the conditions of (9.2). Hence Theorem 9.2.3 follows from Theorem 9.2.2. □

As a corollary, we can obtain an expression for the largest number of sets in a subfamily with an SDR.

[7] If $k < n - t$, then $k - (n - t) < 0$, and (9.2) surely holds, so we need not include it in (9.2).

Corollary 9.2.4 *Let $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ be a family of subsets of a finite set Y. Then the largest number of sets in a subfamily of $\mathcal{A}$ with an SDR equals the smallest value taken by the expression*

$$|A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}| + n - k \tag{9.4}$$

over all choices of $k = 1, 2, \ldots, n$ and all choices of k indices $i_1, i_2, \ldots, i_k$ with $1 \leq i_1 < i_2 < \cdots < i_k \leq n$.

Proof. The largest number of sets in a subfamily equals the largest integer t for which (9.2) holds for all k with $k \geq n - t$ and for all choices of k distinct indices $i_1, i_2, \ldots, i_k$ from $\{1, 2, \ldots, n\}$. Since (9.2) can be rewritten as

$$|A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}| + (n - k) \geq t,$$

the corollary holds; we simply have to choose the smallest value of

$$|A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}| + (n - k)$$

to find the largest t to work. □

Example. We define a family $\mathcal{A} = (A_1, A_2, A_3, A_4, A_5, A_6)$ of subsets of the set $\{a, b, c, d, e, f\}$ by

$$\begin{array}{lll} A_1 = \{a, b, c\}, & A_2 = \{b, c\}, & A_3 = \{b, c\}, \\ A_4 = \{b, c\}, & A_5 = \{c\}, & A_6 = \{a, b, c, d\}. \end{array}$$

We have

$$|A_2 \cup A_3 \cup A_4 \cup A_5| = |\{b, c\}| = 2;$$

hence,

$$|A_2 \cup A_3 \cup A_4 \cup A_5| + 6 - 4 = 2 + 6 - 4 = 4.$$

Thus, with $n = 6$ and $k = 4$, we see by Corollary 9.2.4 that at most four of the sets $\mathcal{A}$ can be chosen so that they have an SDR. Since (A_1, A_2, A_5, A_6) has (a, b, c, d) as an SDR, it follows that 4 is the largest number of sets with an SDR. In terms of marriage, 4 is the largest number of gentlemen that can marry if each gentleman is to marry a compatible woman. □

9.3 Stable Marriages

In this section[8] we consider a variation of the marriage problem discussed in the previous section.

[8]This section is partly based on the article "College Admissions and the Stability of Marriage" by D. Gale and L. S. Shapely, *American Mathematical Monthly*, 69 (1962), 9–15. A comprehensive treatment of the questions considered here can be found in the book *The Stable Marriage Problem: Structure and Algorithms*, by D. Gusfield and R. W. Irving, The MIT Press, Cambridge (1989).

There are n women and n men in a community. Each woman ranks each man in accordance with her preference for that man as a spouse. No ties are allowed, so that if a woman is indifferent between two men, we nonetheless require that she express some preference. The preferences are to be purely ordinal, and thus each woman ranks the men in the order $1, 2, \ldots, n$. Similarly, each man ranks the women in the order $1, 2, \ldots, n$. There are $n!$ ways in which the women and men can be paired so that a *complete marriage* takes place. We say that a complete marriage is *unstable*, provided that there exist two women A and B and two men a and b such that

(1) A and a get married;

(2) B and b get married;

(3) A prefers (i.e., ranks higher) b to a;

(4) b prefers A to B.

Thus, in an unstable complete marriage, A and b could act independently of the others and run off with each other, since both would regard their new partner as more preferable than their current spouse. Thus, the complete marriage is "unstable" in the sense that it can be upset by a man and a woman acting together in a manner that is beneficial to both. A complete marriage is called *stable*, provided it is not unstable. The question that arises first is, *Does there always exist a stable, complete marriage?*

The mathematical model we use for this problem is the *preferential ranking matrix*. This matrix is an n-by-n array of n rows, one for each of the women $w_1, w_2, \ldots, w_n$, and n columns, one for each of the n men $m_1, m_2, \ldots, m_n$. In the position at the intersection of row i and column j, we place the pair p, q of numbers representing, respectively, the ranking of m_j by w_i and the ranking of w_i by m_j. A complete marriage corresponds to a set of n positions of the matrix that includes exactly one position from each row and one position from each column.[9]

Example. Let $n = 2$, and let the preferential ranking matrix be

$$\begin{array}{c c} & \begin{array}{cc} m_1 & m_2 \end{array} \\ \begin{array}{c} w_1 \\ w_2 \end{array} & \left[\begin{array}{cc} 1,2 & 2,2 \\ 2,1 & 1,1 \end{array}\right] \end{array}.$$

Thus, for instance, the entry $1,2$ in the first row and first column means that w_1 has put m_1 first on her list and m_1 has put w_1 second on his list. There are two possible complete marriages:

(1) $w_1 \leftrightarrow m_1,\ w_2 \leftrightarrow m_2$, and

[9]The astute reader has no doubt noticed that a complete marriage corresponds to n nonattacking rooks, where we treat the n-by-n matrix as an n-by-n board.

(2) $w_1 \leftrightarrow m_2$, $w_2 \leftrightarrow m_1$.

The first is readily seen to be stable. The second is unstable since w_2 prefers m_2 to her spouse m_1, and similarly m_2 prefers w_2 to his spouse w_1. □

Example. Let $n = 3$, and let the preferential ranking matrix be

$$\begin{bmatrix} 1,3 & 2,2 & 3,1 \\ 3,1 & 1,3 & 2,2 \\ 2,2 & 3,1 & 1,3 \end{bmatrix}. \tag{9.5}$$

There are $3! = 6$ possible complete marriages. One is

$$w_1 \leftrightarrow m_1,\ w_2 \leftrightarrow m_2,\ w_3 \leftrightarrow m_3.$$

Since each woman gets her first choice, the complete marriage is stable, even though each man gets his last choice. Another stable complete marriage is obtained by giving each man his first choice. But note that, in general, there may not be a complete marriage in which every man (or every woman) gets first choice. For example, this happens when all the women have the same first choice and all the men have the same first choice. □

We now show that a stable complete marriage always exists and, in doing so, obtain an algorithm for determining a stable complete marriage. Thus, complete chaos can be avoided!

Theorem 9.3.1 *For each preferential ranking matrix, there exists a stable complete marriage.*

Proof. We define an algorithm, the *deferred acceptance algorithm*,[10] for determining a complete marriage:

Deferred Acceptance Algorithm

Begin with every woman marked as rejected.

While there exists a rejected woman, do the following:

(1) Each woman marked as rejected chooses the man whom she ranks highest among all those men who have not yet rejected her.

(2) Each man picks out the woman whom he ranks highest among all those women who have chosen him and whom he has not yet rejected, defers decision on her (and removes her rejection status), and now rejects the others.

[10] Also called the Gale-Shapley algorithm.

Thus, during the execution of the algorithm,[11] the women propose to the men, and some men and some women become *engaged*, but the men are able to break engagements if they receive a better offer. Once a man becomes engaged, he remains engaged throughout the execution of the algorithm, but his fiancée may change; in his eyes, a change is always an improvement. A woman, however, may be engaged and disengaged several times during the execution of the algorithm; however, each new engagement results in a less desirable partner for her. It follows from the description of the algorithm that, as soon as there are no rejected women, then each man is engaged to exactly one woman, and since there are as many men as women, each woman is engaged to exactly one man. We now pair each man with the woman to whom he is engaged and obtain a complete marriage. We now show that this marriage is stable.

Consider women A and B and men a and b such that A is paired with a and B is paired with b, but A prefers b to a. We show that b cannot prefer A to B. Since A prefers b to a, during some stage of the algorithm A chose b, but A was rejected by b for some woman he ranked higher. But the woman b eventually gets paired with is at least as high on his list as any woman that he rejected during the course of the algorithm. Since A was rejected by b, b must prefer B to A. Thus, there is no unstable pair, and this complete marriage is stable. □

Example. We apply the deferred acceptance algorithm to the preferential ranking matrix in (9.5), designating the women as A, B, C, respectively, and the men as a, b, c, respectively.[12] In (1), A chooses a, B chooses b, and C chooses c. There are no rejections, the algorithm halts, and A marries a, B marries b, C marries c, and, hopefully, they live happily ever after. □

Example. We apply the deferred acceptance algorithm to the preferential ranking matrix

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} \;a\; & \;b\; & \;c\; & \;d\; \end{array} \\ \left[\begin{array}{cccc} 1,2 & 2,1 & 3,2 & 4,1 \\ 2,4 & 1,2 & 3,1 & 4,2 \\ 2,1 & 3,3 & 4,3 & 1,4 \\ 1,3 & 4,4 & 3,4 & 2,3 \end{array}\right] \end{array}. \tag{9.6}$$

The results of the algorithm are as follows:

(1) A chooses a, B chooses b, C chooses d, D chooses a; a rejects D.

(2) D chooses d; d rejects C.

(3) C chooses a; a rejects A.

(4) A chooses b; b rejects B.

[11] Note that we have reversed the traditional roles of men and women in which men are the suitors.

[12] The BIG guys versus the little guys.

(5) B chooses a; a rejects B.

(6) B chooses c.

In (vi), there are no rejections, and

$$A \leftrightarrow b,\ B \leftrightarrow c,\ C \leftrightarrow a,\ D \leftrightarrow d$$

is a stable complete marriage. □

If, in the deferred acceptance algorithm, we interchange the roles of the women and men and have the men choose women according to their rank preferences, we obtain a stable complete marriage which may, but need not, differ from the one obtained by having the women choose men.

Example. We apply the deferred acceptance algorithm to the preferential ranking matrix in (9.6), where the men choose the women. The results are as follows:

(1) a chooses C, b chooses A, c chooses B, d chooses A; A rejects d.

(2) d chooses B; B rejects d.

(3) d chooses D.

The complete marriage

$$a \leftrightarrow C,\ b \leftrightarrow A,\ c \leftrightarrow B,\ d \leftrightarrow D$$

is stable. This is the same complete marriage obtained by applying the algorithm the other way around. □

Example. We apply the deferred acceptance algorithm to the preferential ranking matrix in (9.5), where the men choose the women. The results are as follows:

(1) a chooses B, b chooses C, c chooses A.

Since there are no rejections, the stable complete marriage obtained is

$$a \leftrightarrow B,\ b \leftrightarrow C,\ c \leftrightarrow A.$$

This is different from the complete marriage obtained by applying the algorithm the other way around. □

A stable complete marriage is called *optimal for a woman*, provided that a woman gets as a spouse a man whom she ranks at least as high as the spouse she obtains in every other stable complete marriage. In other words, there is no stable complete marriage in which the woman gets a spouse who is higher on her list. A stable complete marriage is called *women-optimal* provided that it is optimal for each woman. In a similar way, we define a *men-optimal* stable complete marriage. It is not obvious that there exist women-optimal and men-optimal stable complete marriages. In fact, it is not even obvious that, if each woman is independently given the best partner that she has in all the stable complete marriages, then this results in a pairing of the women and the men (it is conceivable that two women might end up with the same man in this way). Clearly, there can be only one women-optimal complete marriage and only one men-optimal complete marriage.

Theorem 9.3.2 *The stable complete marriage obtained from the deferred acceptance algorithm, with the women choosing the men, is women-optimal. If the men choose the women in the deferred acceptance algorithm, the resulting complete marriage is men-optimal.*

Proof. A man M is called *feasible* for a woman W, provided that there is some stable complete marriage in which M is W's spouse. We shall prove by induction that the complete marriage obtained by applying the deferred acceptance algorithm has the property that the men who reject a particular woman are not feasible for that woman. Because of the nature of the algorithm, this implies that each woman obtains as a spouse the man she ranks highest among all the men that are feasible for her, and hence the complete marriage is women-optimal.

The induction is on the number of rounds of the algorithm. To start the induction, we show that, at the end of the first round, no woman has been rejected by a man that is feasible for her. Suppose that both woman A and woman B choose man a, and a rejects A in favor of B. Then any complete marriage in which A is paired with a is not stable because a prefers B and B prefers a to whichever man she is eventually paired with.

We now proceed by induction and assume that at the end of some round $k \geq 1$, no woman has been rejected by a man who is feasible for her. Suppose that at the end of the $(k + 1)$st round, woman A is rejected by man a in favor of woman B. Then B prefers a over all those men that have not yet rejected her. By the induction assumption, none of the men who have rejected B in the first k rounds is feasible for B, and so there is no stable complete marriage in which B is paired with one of them. Thus, in any stable marriage, B is paired with a man who is no higher on her list than a is.

Now suppose that there is a stable complete marriage in which A is paired with a. Then a prefers B to A and, by the last remark, B prefers a to whomever she is paired with. This contradicts the fact that the complete marriage is stable. The inductive step is now complete, and we conclude that the stable complete marriage obtained from the deferred acceptance algorithm is optimal for the women. □

We now show that in the women-optimal complete marriage, each man has the *worst* partner he can have in any stable complete marriage.

Corollary 9.3.3 *In the women-optimal stable complete marriage, each man is paired with the woman he ranks lowest among all the partners that are possible for him in a stable complete marriage.*

Proof. Let man a be paired with woman A in the women-optimal stable complete marriage. By Theorem 9.3.2, A prefers a to all other men that are possible for her in a stable complete marriage. Suppose there is a stable complete marriage in which a is paired with woman B, where a ranks B lower than A. In this stable marriage, A is paired with some man b different from a whom she therefore ranks lower than a. But then A prefers a, and a prefers A, and this complete marriage is not stable contrary to assumption. Hence, there is no stable complete marriage in which a gets a worse partner than A. □

Suppose the men-optimal and women-optimal stable complete marriages are identical. Then, by Corollary 9.3.3, in the woman-optimal complete marriage, each man gets both his best and worst partner taken over all stable complete marriages. (A similar conclusion holds for the women.) It thus follows in this case that there is exactly one stable complete marriage. Of course, the converse holds as well: If there is only one stable complete marriage, then the men-optimal and women-optimal stable complete marriages are identical.

The deferred acceptance algorithm has been in use since 1952 to match medical residents in the United States to hospitals.[13] We can think of the hospitals as being the women and the residents as being the men. But now, since a hospital generally has places for several residents, polyandrous marriages (in which a woman can have several spouses) are allowed.

We conclude this section with a discussion of a similar problem for which the existence of a stable marriage is no longer guaranteed.

Example. Suppose an even number $2n$ of girls wish to pair up as roommates. Each girl ranks the other girls in the order $1, 2, \ldots, 2n-1$ of preference. A *complete marriage* in this situation is a pairing of the girls into n pairs. A complete marriage is *unstable*, provided there exist two girls who are not roommates such that each of the girls prefers the other to her current roommate. A complete marriage is *stable* provided it is not unstable. Does there always exist a stable complete marriage?

Consider the case of four girls, A, B, C, D, where A ranks B first, B ranks C first, C ranks A first, and each of A, B, and C ranks D last. Then, irrespective of the other rankings, there is no stable complete marriage as the following argument shows. Suppose A and D are roommates. Then B and C are also roommates. But C prefers A to B, and since A ranks D last, A prefers C to D. Thus, this complete marriage is not stable. A similar conclusion holds if B and D are roommates or if C and D are roommates. Since D has a roommate, there is no stable complete marriage. □

[13]It can also be used to match students to colleges, and so on.

9.4 Exercises

1. Consider the chessboard B with forbidden positions shown in Figure 9.4. Construct the rook family $\mathcal{A} = (A_1, A_2, A_3, A_4, A_5, A_6)$ of subsets of $\{1, 2, 3, 4, 5, 6\}$ of this board. Find six positions for six nonattacking rooks on B and the corresponding SDR of $\mathcal{A}$.

			×	×	
×			×		
×					×
×	×	×	×	×	
×	×	×			
		×	×		

Figure 9.4

2. Construct the domino family $\mathcal{A}$ of subsets of the black squares associated with the white squares of the board B in Figure 9.4. (Consider the square in the upper left corner to be white.) Determine a tiling of this board and the associated SDR of $\mathcal{A}$.

3. Give an example of a family $\mathcal{A}$ of sets that is not the domino family of any board.

4. Consider an m-by-n chessboard in which both m and n are odd. The board has one more square of one color, say, black, than of white. Show that, if exactly one black square is forbidden on the board, the resulting board has a tiling with dominoes.

5. Consider an m-by-n chessboard, where at least one of m and n is even. The board has an equal number of white and black squares. Show that if m and n are at least 2 and if exactly one white and exactly one black square are forbidden, the resulting board has a tiling with dominoes.

6. A corporation has seven available positions $y_1, y_2, \ldots, y_7$ and there are ten applicants $x_1, x_2, \ldots, x_{10}$. The set of positions each applicant is qualified for is given, respectively, by $\{y_1, y_2, y_6\}$, $\{y_2, y_6, y_7\}$, $\{y_3, y_4\}$, $\{y_1, y_5\}$, $\{y_6, y_7\}$, $\{y_3\}$, $\{y_2, y_3\}$, $\{y_1, y_3\}$, $\{y_1\}$, $\{y_5\}$. Determine the largest number of positions that can be filled by the qualified applicants and justify your answer.

7. Let $\mathcal{A} = (A_1, A_2, A_3, A_4, A_5, A_6)$, where

$$\begin{aligned} A_1 &= \{a, b, c\},\ A_2 = \{a, b, c, d, e\},\ A_3 = \{a, b\}, \\ A_4 &= \{b, c\},\ A_5 = \{a\},\ A_6 = \{a, c, e\}. \end{aligned}$$

Does the family $\mathcal{A}$ have an SDR? If not, what is the largest number of sets in the family with an SDR?

8. Let $\mathcal{A} = (A_1, A_2, A_3, A_4, A_5, A_6)$, where

$$\begin{aligned} A_1 &= \{1,2\},\ A_2 = \{2,3\},\ A_3 = \{3,4\}, \\ A_4 &= \{4,5\},\ A_5 = \{5,6\},\ A_6 = \{6,1\}. \end{aligned}$$

Determine the number of different SDRs that $\mathcal{A}$ has. Generalize to n sets.

9. Let $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ be a family of sets with an SDR. Let x be an element of A_1. Prove that there is an SDR containing x, but show by example that it may not be possible to find an SDR in which x represents A_1.

10. Suppose $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ is a family of sets that "more than satisfies" the marriage condition. More precisely, suppose that

$$|A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}| \geq k+1$$

for each $k = 1, 2, \ldots, n$ and each choice of k distinct indices $i_1, i_2, \ldots, i_k$. Let x be an element of A_1. Prove that $\mathcal{A}$ has an SDR in which x represents A_1.

11. Let $n > 1$, and let $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ be the family of subsets of $\{1, 2, \ldots, n\}$, where

$$A_i = \{1, 2, \ldots, n\} - \{i\}, \qquad (i = 1, 2, \ldots, n).$$

Prove that $\mathcal{A}$ has an SDR and that the number of SDRs is the nth derangement number D_n.

12. Consider a board with forbidden positions which has the property that, if a square is forbidden, so is every square to its right in its row and every square below it in its column. Prove that the chessboard has a tiling by dominoes if and only if the number of allowable white squares equals the number of allowable black squares.

13. * Let A be a matrix with n columns, with integer entries taken from the set $S = \{1, 2, \ldots, k\}$. Assume that each integer i in S occurs exactly nr_i times in A, where r_i is an integer. Prove that it is possible to permute the entries in each row of A to obtain a matrix B in which each integer i in S appears r_i times in each column.[14]

[14] E. Kramer, S. Magliveras, T. van Trung, and Q. Wu, Some Perpendicular Arrays for Arbitrary Large t, *Discrete Math.*, 96 (1991), 101–110.

14. Let $\mathcal{A} = (A_1, A_2, \ldots, A_m)$ be a family of subsets of a set $Y = \{y_1, y_2, \ldots, y_n\}$. Suppose that there is a positive integer p such that each set of $\mathcal{A}$ contains at least p elements, and each element in Y is contained in at most p sets of $\mathcal{A}$. By counting in two different ways, prove that $n \geq m$.

15. Let p be a positive integer, and let $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ be a family of n subsets of the set $Y = \{y_1, y_2, \ldots, y_n\}$ of n elements. Suppose that each set A_i of $\mathcal{A}$ contains exactly p elements of Y, and each element y_j of Y is contained in exactly p sets of $\mathcal{A}$. Prove that $\mathcal{A}$ has an SDR. Reformulate this problem in terms of nonattacking rooks on a board with forbidden positions.

16. Find a 2-by-2 preferential ranking matrix for which both complete marriages are stable.

17. Consider a preferential ranking matrix in which woman A ranks man a first, and man a ranks A first. Show that, in every stable marriage, A is paired with a.

18. Consider the preferential ranking matrix

$$\begin{bmatrix} 1,n & 2,n-1 & 3,n-2 & \cdots & n,1 \\ n,1 & 1,n & 2,n-1 & \cdots & n-1,2 \\ n-1,2 & n,1 & 1,n & \cdots & n-2,3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 3,n-2 & 4,n-3 & 5,n-4 & \cdots & 2,n-1 \\ 2,n-1 & 3,n-2 & 4,n-3 & \cdots & 1,n \end{bmatrix}.$$

Prove that, for each $k = 1, 2, \ldots, n$, the complete marriage in which each woman gets her kth choice is stable.

19. Use the deferred acceptance algorithm to obtain both the women-optimal and men-optimal stable complete marriages for the preferential ranking matrix

$$\begin{array}{cc} & \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 1,3 & 2,3 & 3,2 & 4,3 \\ 1,4 & 4,1 & 3,3 & 2,2 \\ 2,2 & 1,4 & 3,4 & 4,1 \\ 4,1 & 2,2 & 3,1 & 1,4 \end{bmatrix} \end{array}.$$

Conclude that, for the given preferential ranking matrix, there is only one stable complete marriage.

20. Prove that in every application of the deferred acceptance algorithm with n women and n men, there are at most $n^2 - n + 1$ proposals.

21. $*$ Extend the deferred acceptance algorithm to the case in which there are more men than women. In such a case, not all of the men will get partners.

22. Show, by using Exercise 19, that it is possible that in no stable complete marriage does any person get his or her first choice.

23. Apply the deferred acceptance algorithm to obtain a stable complete marriage for the preferential ranking matrix

$$\begin{array}{c c} & \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \end{array} & \left[\begin{array}{cccc} 1,3 & 2,2 & 3,1 & 4,3 \\ 1,4 & 2,3 & 3,2 & 4,4 \\ 3,1 & 1,4 & 2,3 & 4,2 \\ 2,2 & 3,1 & 1,4 & 4,1 \end{array}\right] \end{array}.$$

24. Consider an n-by-n board in which there is a nonnegative number a_{ij} in the square in row i and column j, $(1 \leq i, j \leq n)$. Assume that the sum of the numbers in each row and in each column equals 1. Prove that it is possible to place n nonattacking rooks on the board at positions occupied by positive numbers.

25. Apply the deferred-acceptance algorithm to obtain a stable marriage for the preferential ranking matrix

$$\left[\begin{array}{cccccc} 1,4 & 2,3 & 3,6 & 4,2 & 5,5 & 6,1 \\ 3,1 & 5,2 & 6,5 & 2,6 & 1,3 & 4,4 \\ 5,5 & 3,6 & 6,1 & 4,4 & 2,2 & 1,3 \\ 6,6 & 5,5 & 4,4 & 3,3 & 2,1 & 1,2 \\ 1,3 & 3,1 & 5,2 & 2,5 & 4,4 & 6,6 \\ 4,2 & 5,4 & 6,3 & 1,1 & 2,6 & 3,4 \end{array}\right]$$

where the rows correspond to A, B, C, D, E, F and the columns correspond to a, b, c, d, d, f.