N.I. Lobachevsky State University
of Nizhni Novgorod

Probability theory and mathematical statistics:

Combinatorics / Art of Counting

Associate Professor
A.V. Zorine

Outline

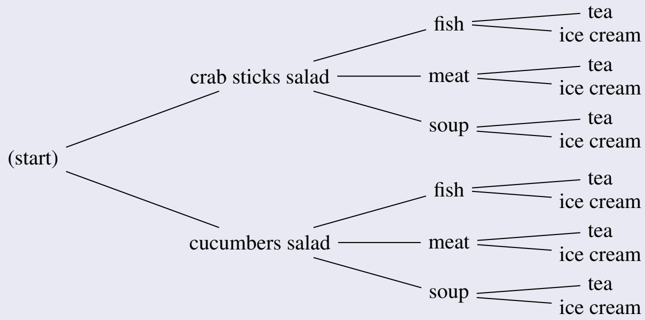
Restaurant problem

You're eating at *'Proschaj, Tovarisch!'* and you have two choices for salads: crab sticks and cucumbers, three choices for the main course: a soup, a meat, or a fish, and two for dessert: ice cream or tea. How many possible choices do you have for your complete meal?

Restaurant problem

You're eating at *'Proschaj, Tovarisch!'* and you have two choices for salads: crab sticks and cucumbers, three choices for the main course: a soup, a meat, or a fish, and two for dessert: ice cream or tea. How many possible choices do you have for your complete meal?

Choice tree



Total number of meals is

$$(2+2+2)+(2+2+2)=3\cdot 2+3\cdot 2=2\cdot 3\cdot 2=12.$$

General rules of combinatorics

Many problems in probability theory require that we count the number of ways that a particular event can occur.

Summation theorem

If you can divide items into classes $K_1, K_2, \ldots, K_m$ of size $N_1, N_2, \ldots, N_m$ respectively, then the total number of items is $N_1 + N_2 + \ldots + N_m$.

Multiplication theorem

A task can be carried out in k stages. There are n_1 ways to carry out the first stage; for each of these n_1 ways there are n_2 ways to carry out the second stage; for each of these ways to carry out the first and the second stage, there are n_3 ways to carry out the third stage and so on. The the total number of ways in which the entire task can be accomplished is given by the product $N = n_1 \cdot n_2 \cdot n_3 \cdots n_k$.

Example. 5 paths lead to the mountain. There are $5 \cdot 5 = 25$ ways to go up and down the mountain. There are $5 \cdot 4 = 20$ ways to go up and down the mountain following different paths.

Example. 1500 people live in a village. At least two of them should have the same initial letters in their first name and last name. Indeed, Russian alphabet consists of 33 letters, so there can be only $33 \cdot 33 = 1089$ different pair of initials.

Example. A secret lock in a safe has two disks with numbers from 0 to 9 and two disks with letters from A to Z. How many trials would it take a burglar to open it in the worst case?

Example. 5 paths lead to the mountain. There are $5 \cdot 5 = 25$ ways to go up and down the mountain. There are $5 \cdot 4 = 20$ ways to go up and down the mountain following different paths.

Example. 1500 people live in a village. At least two of them should have the same initial letters in their first name and last name. Indeed, Russian alphabet consists of 33 letters, so there can be only $33 \cdot 33 = 1089$ different pair of initials.

Example. A secret lock in a safe has two disks with numbers from 0 to 9 and two disks with letters from A to Z. How many trials would it take a burglar to open it in the worst case?

(*Answer:* $10 \cdot 26 \cdot 26 \cdot 10 = 67600$)

Definition

An ordered sample of k items out of n is called *arrangement*. We have an arrangement *without repetitions* if every item in an arrangement occurs only once, otherwise we have an arrangement with repetitions.

Example

We can make twenty-four arrangements (without repetitions) of 3 letters out of the four letters A, B, C, D:

ABC,	BAC,	ACB,	CAB,	CBA,	BCA,
ABD,	BAD,	ADB,	DAB,	DBA,	BDA,
ACD,	CAD,	ADC,	DAC,	DCA,	CDA,
BCD,	CBD,	BDC,	DBC,	DCB,	CDB;

Denote by A_n^k the number of arrangements without repetitions of k items out of n. By multiplication theorem,

$$A_n^k = \overbrace{n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)}^{k \text{ factors}}.$$

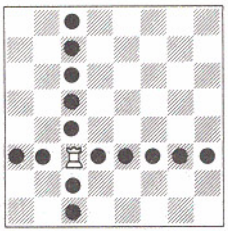
Denote by $\bar{A}_n^k$ the number of arrangements with repetitions of k items out of n items. By multiplication theorem,

$$\bar{A}_n^k = n^k.$$

Example. In how many ways can the first and the second prises be given to a class of ten boys, without giving both to the same boy? Boys to receive prises make an ordered sample without repetitions of 2 persons out of 10, thus there are $A_{10}^2 = 10 \cdot 9 = 90$ ways to do that.

Example. Two persons got into a bus where there are six vacant seats. In how many different ways can they seat themselves? $A_6^2 = 6 \cdot 5 = 30$.

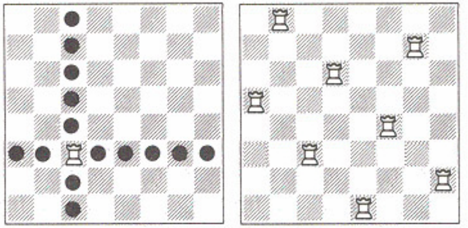
Example. A chess board has 8 vertical and 8 horizontal rows of squares, 64 squares in total. How many ways are there to place 8 rooks so that they don't attack each other?



Example. In how many ways can the first and the second prises be given to a class of ten boys, without giving both to the same boy? Boys to receive prises make an ordered sample without repetitions of 2 persons out of 10, thus there are $A_{10}^2 = 10 \cdot 9 = 90$ ways to do that.

Example. Two persons got into a bus where there are six vacant seats. In how many different ways can they seat themselves? $A_6^2 = 6 \cdot 5 = 30$.

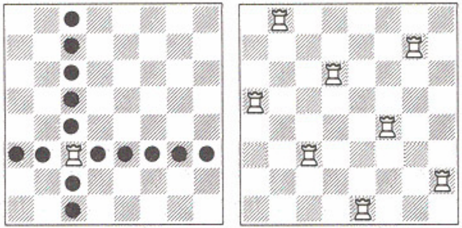
Example. A chess board has 8 vertical and 8 horizontal rows of squares, 64 squares in total. How many ways are there to place 8 rooks so that they don't attack each other?



Example. In how many ways can the first and the second prises be given to a class of ten boys, without giving both to the same boy? Boys to receive prises make an ordered sample without repetitions of 2 persons out of 10, thus there are $A_{10}^2 = 10 \cdot 9 = 90$ ways to do that.

Example. Two persons got into a bus where there are six vacant seats. In how many different ways can they seat themselves? $A_6^2 = 6 \cdot 5 = 30$.

Example. A chess board has 8 vertical and 8 horizontal rows of squares, 64 squares in total. How many ways are there to place 8 rooks so that they don't attack each other?



Each row should hold a single rook. So, a rook in the first horizontal row is placed on one of 8 squares, a rook in the second horizontal row is placed on one of 7 free squares, etc. The total number of possible positions is $8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 40\,320$.

Definition

An arrangement of all items without repetitions is called *a permutation*.

The number of permutations of n items is

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

(pronounces as 'n-factorial'). For example,

$$1! = 1, \quad 2! = 2, \quad 3! = 3 \cdot 2 \cdot 1 = 6,$$
$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 4 \cdot 3! = 24, \quad \ldots, \quad 10! = 3628800, \ldots$$

Observe that $n! = n \cdot (n-1)!$.

$$A_n^k = n(n-1)\cdots(n-k+1) = \frac{n(n-1)\cdots(n-k+1)(n-k)\cdots 2 \cdot 1}{(n-k)\cdots 2 \cdot 1} = \frac{n!}{(n-k)!}$$

Example. ”Placing balls into cells” amount to choosing one cell for each ball. With k balls we have k independent choices, and therefore *k balls can be placed into n cells in n^k different ways*. (Example of arrangement with repetitions)

Example. How many numbers can be expressed using k digits in base $d > 1$? Any such number can be regarded as an arrangement with repetitions of k digits, hence the answer is d^k.

Definition

Suppose we have items $e_1, e_2, \ldots, e_n$. Consider unordered samples of size k, $0 \leqslant k \leqslant n$ without repetitions (no item occurs twice in a sample). These samples are called *combinations*.

Example

Suppose we choose 3 letters out of 5 letters 'a', 'b', 'c', 'd', 'e'. Possible *combinations* (or *selections*) are:

$\{a, b, c\}$, $\{a, b, d\}$, $\{a, b, e\}$, $\{a, c, d\}$,
$\{a, c, e\}$, $\{a, d, e\}$, $\{b, c, d\}$, $\{b, c, e\}$,
$\{b, d, e\}$, $\{c, d, e\}$.

Definition

Suppose we have items $e_1, e_2, \ldots, e_n$. Consider unordered samples of size k, $0 \leqslant k \leqslant n$ without repetitions (no item occurs twice in a sample). These samples are called *combinations*.

Example

Suppose we choose 3 letters out of 5 letters 'a', 'b', 'c', 'd', 'e'. Possible *combinations* (or *selections*) are:

{a, b, c}, {a, b, d}, {a, b, e}, {a, c, d},
{a, c, e}, {a, d, e}, {b, c, d}, {b, c, e},
{b, d, e}, {c, d, e}.

Denote C_n^k the number of all combinations of k objects out of n. In the example above $C_5^2 = 10$. Let's find a recurrent formula for C_n^k. If $k = 0$, only one empty combination is possible, $C_n^0 = 1$. If $k = n$, only one combination containing all elements is possible, $C_n^n = 1$.

Definition

Suppose we have items $e_1, e_2, \ldots, e_n$. Consider unordered samples of size k, $0 \leqslant k \leqslant n$ without repetitions (no item occurs twice in a sample). These samples are called *combinations*.

Example

Suppose we choose 3 letters out of 5 letters 'a', 'b', 'c', 'd', 'e'. Possible *combinations* (or *selections*) are:

{a, b, c}, {a, b, d}, {a, b, e}, {a, c, d},
{a, c, e}, {a, d, e}, {b, c, d}, {b, c, e},
{b, d, e}, {c, d, e}.

Denote C_n^k the number of all combinations of k objects out of n. In the example above $C_5^2 = 10$. Let's find a recurrent formula for C_n^k. If $k = 0$, only one empty combination is possible, $C_n^0 = 1$. If $k = n$, only one combination containing all elements is possible, $C_n^n = 1$. Assume now $k \geqslant 1$. Divide all combinations into two classes. Let the first class contain combinations with e_1. So we have to choose $(k-1)$ elements out of $(n-1)$. Hence there are C_{n-1}^{k-1} items in this class.

Definition

Suppose we have items $e_1, e_2, \ldots, e_n$. Consider unordered samples of size k, $0 \leqslant k \leqslant n$ without repetitions (no item occurs twice in a sample). These samples are called *combinations*.

Example

Suppose we choose 3 letters out of 5 letters 'a', 'b', 'c', 'd', 'e'. Possible *combinations* (or *selections*) are:

$\{a, b, c\}$, $\{a, b, d\}$, $\{a, b, e\}$, $\{a, c, d\}$,
$\{a, c, e\}$, $\{a, d, e\}$, $\{b, c, d\}$, $\{b, c, e\}$,
$\{b, d, e\}$, $\{c, d, e\}$.

Denote C_n^k the number of all combinations of k objects out of n. In the example above $C_5^2 = 10$. Let's find a recurrent formula for C_n^k. If $k = 0$, only one empty combination is possible, $C_n^0 = 1$. If $k = n$, only one combination containing all elements is possible, $C_n^n = 1$. Assume now $k \geqslant 1$. Divide all combinations into two classes. Let the first class contain combinations with e_1. So we have to choose $(k-1)$ elements out of $(n-1)$. Hence there are C_{n-1}^{k-1} items in this class. The second class has the rest. Each combination from the second class includes only some of the elements $e_2, e_3, \ldots, e_n$. There are C_{n-1}^{k} such combinations.

Definition

Suppose we have items $e_1, e_2, \ldots, e_n$. Consider unordered samples of size k, $0 \leqslant k \leqslant n$ without repetitions (no item occurs twice in a sample). These samples are called *combinations*.

Example

Suppose we choose 3 letters out of 5 letters 'a', 'b', 'c', 'd', 'e'. Possible *combinations* (or *selections*) are:

$\{a, b, c\}$, $\{a, b, d\}$, $\{a, b, e\}$, $\{a, c, d\}$,
$\{a, c, e\}$, $\{a, d, e\}$, $\{b, c, d\}$, $\{b, c, e\}$,
$\{b, d, e\}$, $\{c, d, e\}$.

Denote C_n^k the number of all combinations of k objects out of n. In the example above $C_5^2 = 10$. Let's find a recurrent formula for C_n^k. If $k = 0$, only one empty combination is possible, $C_n^0 = 1$. If $k = n$, only one combination containing all elements is possible, $C_n^n = 1$. Assume now $k \geqslant 1$. Divide all combinations into two classes. Let the first class contain combinations with e_1. So we have to choose $(k-1)$ elements out of $(n-1)$. Hence there are C_{n-1}^{k-1} items in this class. The second class has the rest. Each combination from the second class includes only some of the elements $e_2, e_3, \ldots, e_n$. There are C_{n-1}^k such combinations. Hence $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$.

Calculating C_n^k

It's useful to write all C_n^k in a triangular form

$n=1$						C_1^0		C_1^1					
$n=2$					C_2^0		C_2^1		C_2^2				
$n=3$				C_3^0		C_3^1		C_3^2		C_3^3			
$n=4$			C_4^0		C_4^1		C_4^2		C_4^3		C_4^4		
$n=5$		C_5^0		C_5^1		C_5^2		C_5^3		C_5^5		C_5^5	
$n=6$	C_6^0		C_6^1		C_6^2		C_6^3		C_6^4		C_6^5		C_6^0
	$\dots$		$\dots$		$\dots$		$\dots$		$\dots$		$\dots$		$\dots$

Calculating C_n^k

We know that there's only 1 way to choose one item out of one

$$
\begin{array}{r|ccccccccccccc}
n=1 & & & & & & 1 & & 1 & & & & & \\
n=2 & & & & & C_2^0 & & C_2^1 & & C_2^2 & & & & \\
n=3 & & & & C_3^0 & & C_3^1 & & C_3^2 & & C_3^3 & & & \\
n=4 & & & C_4^0 & & C_4^1 & & C_4^2 & & C_4^3 & & C_4^4 & & \\
n=5 & & C_5^0 & & C_5^1 & & C_5^2 & & C_5^3 & & C_5^5 & & C_5^5 & \\
n=6 & C_6^0 & & C_6^1 & & C_6^2 & & C_6^3 & & C_6^4 & & C_6^5 & & C_6^0 \\
 & \dots & & \dots & & \dots & & \dots & & \dots & & \dots & & \dots
\end{array}
$$

Calculating C_n^k

Use the first line to calculate the second line: $C_2^0 = 1$, $C_2^k = C_1^{k-1} + C_1^k$

$$\begin{array}{r|ccccccccccccc}
n=1 & & & & & & 1 & & 1 & & & & & \\
n=2 & & & & & 1 & & 2 & & 1 & & & & \\
n=3 & & & & C_3^0 & & C_3^1 & & C_3^2 & & C_3^3 & & & \\
n=4 & & & C_4^0 & & C_4^1 & & C_4^2 & & C_4^3 & & C_4^4 & & \\
n=5 & & C_5^0 & & C_5^1 & & C_5^2 & & C_5^3 & & C_5^5 & & C_5^5 & \\
n=6 & C_6^0 & & C_6^1 & & C_6^2 & & C_6^3 & & C_6^4 & & C_6^5 & & C_6^0 \\
 & \ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots
\end{array}$$

Calculating C_n^k

Use the second line to calculate the third line: $C_3^0 = 1$, $C_3^k = C_2^{k-1} + C_2^k$

$$
\begin{array}{r|ccccccccccccc}
n=1 & & & & & & 1 & & 1 & & & & & \\
n=2 & & & & & 1 & & 2 & & 1 & & & & \\
n=3 & & & & 1 & & 3 & & 3 & & 1 & & & \\
n=4 & & & C_4^0 & & C_4^1 & & C_4^2 & & C_4^3 & & C_4^4 & & \\
n=5 & & C_5^0 & & C_5^1 & & C_5^2 & & C_5^3 & & C_5^5 & & C_5^5 & \\
n=6 & C_6^0 & & C_6^1 & & C_6^2 & & C_6^3 & & C_6^4 & & C_6^5 & & C_6^0 \\
 & \ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots
\end{array}
$$

Calculating C_n^k

Calculate the rest in similar manner. We obtain *the Pascal triangle*:

$$
\begin{array}{r|ccccccccccccc}
n=1 & & & & & & 1 & & 1 & & & & & \\
n=2 & & & & & 1 & & 2 & & 1 & & & & \\
n=3 & & & & 1 & & 3 & & 3 & & 1 & & & \\
n=4 & & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
n=5 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
n=6 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 \\
 & \dots & & \dots & & \dots & & \dots & & \dots & & \dots & & \dots
\end{array}
$$

Properties of binomial coefficients:

$$C_n^k = C_n^{n-k} \quad \text{(symmetry)}$$
$$C_n^0 + C_n^1 + \ldots + C_n^n = 2^n$$

Newton's binomial formula

$$
\begin{aligned}
a+b &= C_1^0 a + C_1^1 b, \\
(a+b)^2 &= (a+b)(C_1^0 a + C_1^0 b) = C_1^0 a^2 + (C_1^0 + C_1^1)ab + C_1^1 b^2 \\
&= C_2^0 a^2 + C_2^1 ab + C_2^2 b^2, \\
(a+b)^3 &= (a+b)(C_2^0 a^2 + C_2^1 ab + C_2^2 b^2) = C_2^0 a^3 + (C_2^0 + C_2^1)a^2 b + (C_2^1 + C_2^2)ab^2 + C_2^2 b^3 = \\
&= C_3^0 a^3 + C_3^1 a^2 b + C_3^2 ab^2 + C_3^3 b^3, \\
&\quad \ldots \\
(a+b)^n &= C_n^0 a^n + C_n^1 a^{n-1} b + C_n^2 a^{n-2} b^2 + \ldots + C_n^{n-1} ab^{n-1} + C_n^n b^n.
\end{aligned}
$$

Put $a = 1$, $b = 1$, immediately obtain $C_n^0 + C_n^1 + \ldots + C_n^n = 2^n$.
Put $a = 1$, $b = -1$, then $C_n^0 - C_n^1 + C_n^2 - C_n^3 + \ldots + (-1)^n C_n^n = 0$.

Theorem

$$C_n^k = \frac{n \cdot (n-1) \cdots (n-k+1)}{k \cdot (k-1) \cdots 1} \quad \textit{for } n \geqslant 1 \textit{ and } k = 1, 2, \ldots, n.$$

Proof

$$\begin{aligned} C_{n-1}^{k-1} + C_{n-1}^{k} &= \frac{\overbrace{(n-1)(n-2)\cdots(n-1-(k-1)+1)}^{k-1 \text{ factors}}}{(k-1)(k-2)\cdots 1} \\ &\quad + \frac{\overbrace{(n-1)(n-2)\cdots(n-1-k+1)}^{k \text{ factors}}}{k(k-1)(k-2)\cdots 1} \\ &= \frac{(n-1)(n-2)\cdots(n-1-(k-1)+1)}{(k-1)(k-2)\cdots 1}\left(1 + \frac{n-k}{k}\right) \\ &= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k(k-1)(k-2)\cdots 1}. \end{aligned}$$

$$C_n^k = \frac{n \cdot (n-1) \cdots (n-k+1)}{k \cdot (k-1) \cdots 1} = \frac{n!}{k!(n-k)!}$$

This formula is useful for large n, k. For this formula to be exact, assume $0! = 1$. Let's compute C_{100}^{50} using logarithms (base 10):

$$\begin{aligned}
\lg 100! &= 157{,}9700, \\
\lg 50! &= 64{,}4831, \\
\lg (50!)^2 &= 128{,}9662, \\
\lg C_{100}^{50} &= \lg \frac{100!}{50!50!} = \lg 100! - \lg (50!)^2 = 29{,}0038, \\
C_{100}^{50} &= 10^{29{,}0038} = 1{,}008788214949312 \cdot 10^{29}.
\end{aligned}$$

In fact, $C_{100}^{50} = 100\,891\,344\,545\,564\,193\,334\,812\,497\,256$, relative error is less than $0{,}012\%$.

Choosing from items of two types

Problem. Suppose we have a red balls and b blue balls in an urn and r balls are taken out. How many combinations contain exactly k red balls?

Solution. We see that there should be $(r-k)$ blue balls in a sample. There are C_a^k ways to choose k red balls and C_b^{r-k} ways to choose $r-k$ blue balls. Hence the total number of combinations is $C_a^k C_b^{r-k}$.

Choosing from items of two types

Problem. Suppose we have a red balls and b blue balls in an urn and r balls are taken out. How many combinations contain exactly k red balls?

Solution. We see that there should be $(r-k)$ blue balls in a sample. There are C_a^k ways to choose k red balls and C_b^{r-k} ways to choose $r-k$ blue balls. Hence the total number of combinations is $C_a^k C_b^{r-k}$.

Another property of binomial coefficients

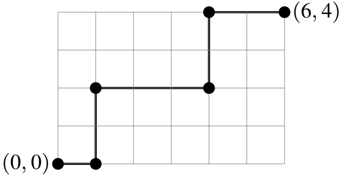
For integer a, b, and $r \leqslant \min\{a, b\}$

$$C_a^0 C_b^r + C_a^1 C_b^{r-1} + C_a^2 C_b^{r-2} + \ldots + C_a^r C_b^0 = C_{a+b}^r$$

Proof. A combination of r balls out of a red balls and b blue balls has either 0 red balls, or exactly 1 red ball, or exactly 2 red balls, and so on up to exactly r red balls (thus we have $r+1$ classes). Now we use summation theorem.

Paths on lattices

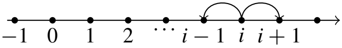
How many paths lead from $(0,0)$ to (m,n) going either up or right along the grid?



Obviously, every path consists of $m+n$ unit steps upwards or rightwards. If we choose m particular steps to go to the right, we completely define the path. Hence there are C^m_{m+n} different paths.

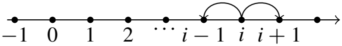
Random walk on a line

A particle starts from zero and make one unit step per time 1 second. It moves either left or right. How many paths return to zero after $2n$ seconds?



Random walk on a line

A particle starts from zero and make one unit step per time 1 second. It moves either left or right. How many paths return to zero after $2n$ seconds?



A path returns to zero after $2n$ seconds if an only if n steps are to the left and n steps are to the right. There are C_{2n}^{n} possible paths.

Recurrent equations

We had an example before: recurrent equation $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$ for binomial coefficients. We'll use the same idea to solve a couple more problems.

Barrel problem. You have an 8-gallon barrel and two buckets, one for 1 gallon and one for 2 gallons. In how many ways can you empty the barrel?

We can use 1-gallon bucket 8 times, or 2-gallon bucket 4 times; 1-gallon bucket 2 times and 2-gallon bucket 3 times; 2 gallon bucket, then 1-gallon bucket twice, then 2-gallon bucket twice etc. Denote F_N the number of ways to empty N gallon barrel. Obviously, $F_1 = 1$, $F_2 = 2$ $(2 = 1 + 1)$. For $N \geqslant 2$, first we may use 1-gallon bucket and then empty $(N - 1)$-gallon barrel; or use 2-gallon bucket and then empty $(N - 2)$-gallon barrel. The total number of ways is a sum of F_{N-1} and N_{N-2}:

$$F_N = F_{N-1} + F_{N-2}.$$

Then fill the table:

F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
1	2	3	5	8	13	21	**34**

The answer is in **bold face**.