MAXimal

added: 25 Aug 2011 13:55
Edited: 22 Oct 2011 19:29

The principle of inclusion-exclusion

Contents [hide]

The principle of inclusion-exclusion - this is an important combinatorial trick to calculate the size of any of the sets, or to calculate the probability of complex events.

Formulation of the principle of inclusion-exclusion

The wording

The principle of inclusion-exclusion is as follows:

To calculate the size of combining multiple sets, it is necessary to sum the sizes of these sets **separately** , then subtract the sizes of all **pairwise** intersections of these sets, add back all sorts of sizes of intersections **triples** sets, subtract dimensions intersections **quadruples** , and so on, up to the intersection **of all** sets.

Formulation in terms of sets

In mathematical form, the above wording is as follows:

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i| - \sum_{\substack{i,j:\\ 1\le i<j\le n}} |A_i \cap A_j| + \sum_{\substack{i,j,k:\\ 1\le i<j<k\le n}} |A_i \cap A_j \cap A_k| - \ldots + (-1)^{n-1} |A_1 \cap \ldots \cap A_n|.$$

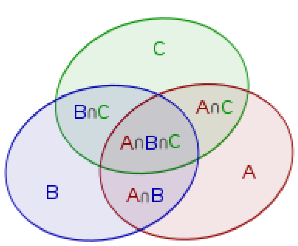
It can be written in a more compact, a sum of over subsets. We denote $B$the set whose elements are A_i. Then the principle of inclusion-exclusion takes the form:

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{C \subseteq B} (-1)^{size(C)-1} \left|\bigcap_{e \in C} e\right|.$$

This formula is credited Moivre (Abraham de Moivre).

Formulation using Venn diagrams

Let the diagram marked by three figures A, $B$and C:



Then the area of union $A \cup B \cup C$is the sum of squares A, $B$and Cminus twice the area covered $A \cap B$, $A \cap C$, $B \cap C$, but with the addition of three times the area covered $A \cap B \cap C$:

$$S(A \cup B \cup C) = S(A) + S(B) + S(C) - S(A \cap B) - S(A \cap C) - S(B \cap C) + S(A \cap B \cap C).$$

Similarly, it can be generalized to association $n$figures.

Formulation in terms of probability theory

If - this event - their probabilities, the probability of their union (ie what happens at least one of these events) is equal to:

If - this event - their probabilities, the probability of their union (ie what happens at least one of these events) is equal to:
$A_i (i = 1 \ldots n) \mathcal{P}(A_i)$

$$\mathcal{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathcal{P}(A_i) - \sum_{\substack{i,j:\\ 1 \le i < j \le n}} \mathcal{P}(A_i \cap A_j) + \\ + \sum_{\substack{i,j,k:\\ 1 \le i < j < k \le n}} \mathcal{P}(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1} \mathcal{P}(A_1 \cap \ldots \cap A_n).$$

This amount can also be written as a sum over subsets of the set Bwhose elements are the events A_i:

$$\mathcal{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{C \subseteq B} (-1)^{size(C)-1} \cdot \mathcal{P}\left(\bigcap_{e \in C} e\right).$$

Proof of inclusion-exclusion principle

To prove convenient to use the mathematical formulation in terms of set theory:

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{C \subseteq B} (-1)^{size(C)-1} \left|\bigcap_{e \in C} e\right|,$$

where B, we recall - is the set of A_i's.

We need to prove that any element contained in at least one of the sets A_i, take into account the formula exactly once. (Note that other elements not contained in either of A_i, in no way can be considered as missing in the right side of the formula).
Consider an arbitrary element $x$contained in exactly $k \ge 1$sets A_i. We show that it considers the formula exactly once.

Note that:

- in those terms, which $size(C) = 1$element $x$will consider exactly $k$once, with a plus sign;
- in those terms, which $size(C) = 2$element $x$will consider (with a minus sign) exactly C_k^2once - because it $x$will be counted only in those terms, which correspond to the two sets of $k$sets containing x;
- in those terms, which $size(C) = 3$element $x$will consider exactly C_k^3once, with a plus sign;
- $\ldots$
- in those terms, which $size(C) = k$element $x$will consider exactly C_k^konce, with the sign $(-1)^{k-1}$;
- in those terms, which $size(C) > k$element $x$will allow for zero times.

Thus, we need to calculate a sum of binomial coefficients :

$$T = C_k^1 - C_k^2 + C_k^3 - \ldots + (-1)^{i-1} \cdot C_k^i + \ldots + (-1)^{k-1} \cdot C_k^k.$$

The easiest way to calculate this amount by comparing it with the expansion of the binomial theorem in the expression $(1-x)^k$:

$$(1-x)^k = C_k^0 - C_k^1 \cdot x + C_k^2 \cdot x^2 - C_k^3 \cdot x^3 + \ldots + (-1)^k \cdot C_k^k \cdot x^k.$$

We see that when $x = 1$the expression $(1-x)^k$is nothing else than $1 - T$. Therefore, $T = 1 - (1-1)^k = 1$what we wanted to prove.

Use in solving problems

The principle of inclusion-exclusion is difficult to understand without learning good examples of its applications.

First, we consider three simple tasks "on paper", illustrating the application of the principle, and then consider the more practical problems that are difficult to solve without the use of the principle of inclusion-exclusion.

Of particular note is the problem "search number of ways," because it demonstrates that the principle of inclusion-exclusion can sometimes lead to the polynomial solutions, not necessarily exponential.

Simple task of permutations

Simple task of permutations

How many permutations of 0 up 9 such that the first element is more 1, and the last - less 8?

Count the number of "bad" permutations, ie those in which the first element ≤ 1 and / or last ≥ 8.

We denote X the set of permutations in which the first element ≤ 1 and through Y - whose last element ≥ 8. Then the amount of "bad" permutations on inclusion-exclusion formula equals:

$$|X| + |Y| - |X \cap Y|.$$

After spending a simple combinatorial calculations, we find that it is equal to:

$$2 \cdot 9! + 2 \cdot 9! - 2 \cdot 2 \cdot 8!$$

Subtracting this number from the total number of permutations $10!$, we get a response.

Simple task of (0,1,2)-sequences

How many sequences of length n, consisting only of numbers $0, 1, 2$, each number occurs at least once?

Again, we pass to the inverse problem, ie assume the number of sequences which are not present in at least one of the properties.

We denote by A_i($i = 0 \ldots 2$) the set of sequences in which the number is not found i. Then, by inclusion-exclusion by "bad" sequences is:

$$|A_0| + |A_1| + |A_2| - |A_0 \cap A_1| - |A_0 \cap A_2| - |A_1 \cap A_2| + |A_0 \cap A_1 \cap A_2|.$$

The dimensions of each of A_i equal apparent 2^n(such as sequences may occur only two types of numbers). Capacity of each pairwise intersection $A_i \cap A_j$ are 1(still available as only one digit). Finally, the cardinality of the intersection of all three sets is equal to 0(as available figures do not remain).

Remembering that we solve the inverse problem, we get the final **answer** :

$$3^n - 3 \cdot 2^n + 3 \cdot 1 - 0.$$

Number of integral solutions of the equation

Given the equation:

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 20,$$

where all $0 \leq x_i \leq 8$(where $i = 1 \ldots 6$).

Required to count the number of solutions.

Forget about the first limitation $x_i \leq 8$, and simply count the number of non-negative solutions of this equation. This is easily done through the binomial coefficients - we want to beat the 20 elements to 6 groups, ie distribute the 5 "walls" separating groups, 25 places:

$$N_0 = C_{25}^5$$

Now count on inclusion-exclusion formula by "bad" decisions, ie solutions of these in which one or more x_i larger 9.

Denoted by A_k(where $k = 1 \ldots 6$) the set of solutions of the equation in which $x_k \geq 9$, while all the other $x_i \geq 0$(for all $i \neq k$). To calculate the size of the set A_k, we note that we have essentially the same combinatorial problem that was solved by the two paragraphs above, but now 9 the elements excluded from consideration and exactly belong to the first group. Thus:

$$|A_k| = C_{16}^5$$

Similarly, the cardinality of the intersection of two sets A_k and the A_p number is:

$$|A_k \cap A_p| = C_7^5$$

The power of each of three or more sets is zero, since 20 the elements are not enough for three or more variables, more than or equal to 9.

Combining all this inclusion-exclusion formula and given that we solve the inverse problem, we finally get **the answer** :

$$C_{25}^5 - C_6^1 \cdot C_{16}^5 + C_6^2 \cdot C_7^5.$$

Number of relatively prime numbers in a given interval

Let numbers $n$and r. Required to count the number of numbers in the interval $[1; r]$prime to n.

Go straight to the inverse problem - do not count the number of relatively prime integers.

Consider all the prime divisors of $n$which we denote by p_i($i = 1 \ldots k$).

How many numbers in the interval $[1; r]$divisible by p_i? Their number is:

$$\left\lfloor \frac{r}{p_i} \right\rfloor$$

However, if we simply sum up these numbers, we get the wrong answer - some numbers to be added together several times (the ones that are divided to several p_i). Therefore it is necessary to use inclusion-exclusion formula.

For example, it is possible for 2^ka subset of all the bust p_i's, find them work, and add or subtract to the inclusion-exclusion formula next term.

The final **implementation** for counting the number of relatively prime numbers:

```
int solve (int n, int r) {
        vector<int> p;
        for (int i=2; i*i<=n; ++i)
                if (n % i == 0) {
                        p.push_back (i);
                        while (n % i == 0)
                                n /= i;
                }
        if (n > 1)
                p.push_back (n);

        int sum = 0;
        for (int msk=1; msk<(1<<p.size()); ++msk) {
                int mult = 1,
                        bits = 0;
                for (int i=0; i<(int)p.size(); ++i)
                        if (msk & (1<<i)) {
                                ++bits;
                                mult *= p[i];
                        }

                int cur = r / mult;
                if (bits % 2 == 1)
                        sum += cur;
                else
                        sum -= cur;
        }

        return r - sum;
}
```

Asymptotics of the solution is $O(\sqrt{n})$.

The number of integers in a given interval, multiple at least one of the given numbers

Given $n$number $a_i$and number r. Required to count the number of numbers in the interval $[1; r]$that is a multiple of at least one a_i.

An algorithm for solving almost identical to the previous task - make inclusion-exclusion formula on numbers a_i, ie each term in this equation - the number of integers divisible by the specified subset of numbers a_i(in other words, dividers for their least common multiple).

Thus, the decision boils down to is that for 2^na subset of properties to sort, for $O(n \log r)$their operations to find the least common multiple, and add or subtract from the answer of the next value.

The number of rows that satisfy a number of patterns

Given $n$patterns - lines of equal length, consisting only of letters and question marks. Also given a number k. Required to count the number of rows that satisfy exactly $k$patterns or, in another formulation, at least $k$patterns.

Note first that we can **easily calculate the number of rows** that satisfy all of the right patterns. To do this, simply "cross"

Note first that we can **easily calculate the number of rows** that satisfy all of the right patterns. To do this, simply "cross" these patterns: A look at the first character (whether in all the patterns in the first position, the question is whether or not at all - if the first character is uniquely defined) for the second character, etc.

Now learn how to solve **the problem the first option** when the desired line must satisfy exactly $k$patterns.

To do this, let's look over and zafksiruem specific subset Xsize patterns k- now we have to count the number of rows that satisfy this set of patterns and only him. For this we use the inclusion-exclusion formula: we sum over a superset of the set X, and either added to the current account, or subtract from it the number of rows that correspond to the current set:

$$ans(X) = \sum_{Y \supseteq X} (-1)^{|Y|-k} \cdot f(Y),$$

where $f(Y)$denotes the number of rows that correspond to a set of patterns Y.

If we sum $ans(X)$over all X, we get the answer:

$$ans = \sum_{X ~:~ |X|=k} ans(X).$$

However, by doing so we got the solution for the time of the order $O(3^k \cdot k)$.

Solution can be accelerated, noting that different $ans(X)$summation often conducted on the same sets Y.

Turn the inclusion-exclusion formula and will lead to the summation Y. Then it is easy to understand that many Ywill consider in the $C_{|Y|}^k$inclusion-exclusion formula, all with the same sign $(-1)^{|Y|-k}$:

$$ans = \sum_{Y ~:~ |Y| \geq k} (-1)^{|Y|-k} \cdot C_{|Y|}^k \cdot f(Y).$$

The decision came with the asymptote $O(2^k \cdot k)$.

We now turn to **the second embodiment of the problem** : when the search string must satisfy at least $k$patterns.

Clearly, we can simply use the solution of the first version of the problem and summarize responses from $k$before n. However, you will notice that all the arguments will continue to be true, but in this version of the problem by programming $X$is not only for those sets whose size is equal to k, and over all sets with size $\geq k$.

Thus, in the final formula before $f(Y)$will stand another factor: not one binomial coefficient with some familiar, and their sum:

$$(-1)^{|Y|-k} \cdot C_{|Y|}^k ~+~ (-1)^{|Y|-k-1} \cdot C_{|Y|}^{k+1} ~+~ (-1)^{|Y|-k-2} \cdot C_{|Y|}^{k+2} ~+~ \ldots ~+~ (-1)^{|Y|-|Y|} \cdot C_{|Y|}^{|Y|}.$$

Looking into Graham (Graham, Knuth, Patashnik. **"Concrete Mathematics"** [1998]), we see a well-known formula for the binomial coefficients :

$$\sum_{k=0}^m (-1)^k \cdot C_n^k = (-1)^m \cdot C_{n-1}^m.$$

Applying it here, we see that the whole sum of the binomial coefficients is minimized in:

$$(-1)^{|Y|-k} \cdot C_{|Y|-1}^{|Y|-k}.$$

Thus, for this version of the problem we got the solution with the asymptotic form $O(2^k \cdot k)$:

$$ans = \sum_{Y ~:~ |Y| \geq k} (-1)^{|Y|-k} \cdot C_{|Y|-1}^{|Y|-k} \cdot f(Y).$$

Number of ways

There is a field $n \times m$, some $k$cells which - impenetrable wall. On the field in the cell $(1, 1)$(lower left cell) is initially robot. The robot can only move right or up and in the end he has to get into the cage (n, m), avoiding all obstacles. Required to count the number of ways in which he can do it.

Assume that the size $n$and $m$very large (say, up 10^9), and the number k- a small (of the order 100).

To solve immediately for convenience **sort the** obstacles in the order in which we can work around them: that is, for example, coordinate x, and at equality - coordinate y.

Also learn how to solve the problem immediately without obstacles: ie learn to count the number of ways to reach from

Also learn how to solve the problem immediately without obstacles: ie learn to count the number of ways to reach from one cell to another. If one coordinate, we need to pass $x$the cells, and on the other - $y$cells, from simple combinatorics, we obtain a formula through the binomial coefficients :

$$C_{x+y}^{x}$$

Now count the number of ways to reach from one cell to another, avoiding all obstacles, you can use the **inclusion-exclusion formula** : count the number of ways to reach, stepping at least one obstacle.

For this example, you can iterate through the subset of those obstacles, which we do come, count the number of ways to do it (just multiplying the number of ways to reach from the start to the first cell of the selected obstacles, the first obstacle to the second, and so on), and then add or subtract a number from the response in accordance with the standard inclusion-exclusion formula.

However, this again is a non-polynomial solution - for the asymptotic behavior $O(2^k k)$. We show how to obtain **a polynomial solution** .

Will solve **dynamic programming** : learn how to calculate the number $d[i][j]$- the number of ways to reach from $i$the second point to the $j$second, while not stepping on any one obstacle (except for themselves $i$and j, of course). In total we will $k+2$point to the obstacles are added as the start and end cells.

If we for a moment forget about all the obstacles and simply count the number of paths from the cell $i$into the cell j, we thus take into account some of the "bad" way through obstacles. Learn to count the number of these "bad" ways. Iterate over the first obstacle $i < t < j$, to which we come, then the number of paths is equal to $d[i][t]$times the number of arbitrary paths $t$in j. Summing it all t, we count the number of "bad" ways.

Thus, the value $d[i][j]$we have learned to count in time $O(k)$. Consequently, the solution of the whole problem has asymptotics $O(k^3)$.

Number coprimes quadruples

Given the $n$numbers: $a_1, a_2, \ldots, a_n$. Required to count the number of ways to choose four numbers of them so that their combined greatest common divisor is equal to unity.

We solve the inverse problem - count the number of "bad" quadruples, ie such quadruples in which all numbers are divisible by $d > 1$.

We use the inclusion-exclusion formula, summing the number of fours divisible by the divisor d(but perhaps more dividers and divider)

$$ans = \sum_{d \geq 2} (-1)^{deg(d)-1} \cdot f(d),$$

where $deg(d)$- is the number of prime factorization of the number d, $f(d)$- the number of fours, dividers for d.

To calculate the function $f(d)$, you simply count the number of numbers divisible d, and binomial coefficients count the number of ways to choose four of them.

Thus, using the inclusion-exclusion formula, we summarize the number of fours divisible by primes, then subtract the number of quads that are divisible by the product of two primes, we add four divisible by three simple, etc.

Harmonic number of triples

Given the number $n \leq 10^6$. Required to count the number of triples $2 \leq a < b < c \leq n$that they are harmonic triplets, ie:

- either $\gcd(a,b) = \gcd(a,c) = \gcd(b,c) = 1$,
- or $\gcd(a,b) > 1, \gcd(a,c) > 1, \gcd(b,c) > 1$.

First, go straight to the inverse problem - ie count the number of non-harmonic triples.

Secondly, we note that any non-harmonic exactly two triple its numbers are in such a situation, that this number is relatively prime to the number one triple and not just a one with a different number threes.

Thus, the amount equal to the sum of non-harmonic triples over all numbers of 2up to $n$up the number of relatively prime to the current number of properties on the number of non-mutually prime numbers.

Now all that remains for us to solve the problem - is to learn to count numbers for each segment in $[2; n]$the amount of numbers relatively prime (or coprime) with him. Although this task has already been highlighted, the solution described above is not appropriate here - it requires the factorization of each of the numbers 2up n, and then iterate through all possible products of primes in the factorization.

Therefore we need a faster solution that calculates the answers to all numbers in the interval $[2; n]$immediately.

To do this, you can implement a **modification of the sieve of Eratosthenes** :

- First, we need to find all the numbers in the interval $[2; n]$, which are in a factorization, there is not included twice

- First, we need to find all the numbers in the interval $[2; n]$ in which no prime factorization is not included twice. In addition, for inclusion-exclusion formula, we need to know how simple factorization contains every such number.

  To do this we need to have arrays $deg[]$ that store for each number the number of primes in its factorization, and $good[]$ - containing for each number $true$ or $false$ - all simple enter it in power ≤ 1 or not.

  Thereafter, during the processing of the sieve of Eratosthenes next prime number, we go through all the numbers that are multiples of the current number, and increase $deg[]$ them, and all the numbers divisible by the square of the current simple - deliver $good = false$.
- Secondly, we need to calculate the answer for all the numbers from 2 before n, ie array $cnt[]$ - the number of numbers that are not prime to data.

  For this, we recall how the inclusion-exclusion formula - here we actually implement it the same, but with an inverted logic: if we iterate term and see in which inclusion-exclusion formula for what number this term included.

  Thus, suppose we have a number i for which $good[] = true$, ie this number is involved in inclusion-exclusion formula. Iterate through all the numbers that are multiples i, and to answer $cnt[]$ each of these numbers, we have to add or subtract value $\lfloor N/i \rfloor$. Zodiac - addition or subtraction - depends on $deg[i]$ if $deg[i]$ odd, then we must add, subtract otherwise.

Implementation :

```
int n;
bool good[MAXN];
int deg[MAXN], cnt[MAXN];

long long solve() {
	memset (good, 1, sizeof good);
	memset (deg, 0, sizeof deg);
	memset (cnt, 0, sizeof cnt);

	long long ans_bad = 0;
	for (int i=2; i<=n; ++i) {
		if (good[i]) {
			if (deg[i] == 0)  deg[i] = 1;
			for (int j=1; i*j<=n; ++j) {
				if (j > 1 && deg[i] == 1)
					if (j % i == 0)
						good[i*j] = false;
					else
						++deg[i*j];
				cnt[i*j] += (n / i) * (deg[i]%2==1 ? +1 : -1);
			}
		}
		ans_bad += (cnt[i] - 1) * 1ll * (n-1 - cnt[i]);
	}

	return (n-1) * 1ll * (n-2) * (n-3) / 6 - ans_bad / 2;
}
```

Asymptotics of this solution is $O(n \log n)$, as almost every number i it makes about n/i nested loop iterations.

The number of permutations with no fixed points

We prove that the number of permutations of length n without fixed points is the next number:

$$n! - C_n^1 \cdot (n-1)! + C_n^2 \cdot (n-2)! - C_n^3 \cdot (n-3)! + \ldots \pm C_n^n \cdot (n-n)!$$

and approximately equal to the number:

$$\frac{n!}{e}$$

(Moreover, if the expression is rounded to the nearest whole - you get exactly the number of permutations with no fixed points)

We denote A_k the set of permutations of length n with a fixed point at the position k $(1 \leq k \leq n)$.

We now use the inclusion-exclusion formula to calculate the number of permutations with at least one fixed point. To do this, we need to learn to count-size sets intersections A_i, they are as follows:

$$|A_p| = (n-1)!$$

$|A_p| = (n-1)!$,
$|A_p \cap A_q| = (n-2)!$,
$|A_p \cap A_q \cap A_r| = (n-3)!$,
$\ldots$,

because if we know that the number of fixed points is x, by the same token we know the position of $x$permutation elements, and all other $(n-x)$elements can stand anywhere.

Substituting this in the inclusion-exclusion formula and taking into account that the number of ways to select a subset of the size $x$of the $n$power-element set C_n^x, a formula for the number of permutations with at least one fixed point:

$$C_n^1 \cdot (n-1)! - C_n^2 \cdot (n-2)! + C_n^3 \cdot (n-3)! - \ldots \pm C_n^n \cdot (n-n)!$$

Then the number of permutations without fixed points is:

$$n! - C_n^1 \cdot (n-1)! + C_n^2 \cdot (n-2)! - C_n^3 \cdot (n-3)! + \ldots \pm C_n^n \cdot (n-n)!$$

Simplifying this expression, we obtain **the exact and approximate expressions for the number of permutations with no fixed points** :

$$n! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots \pm \frac{1}{n!} \right) \approx \frac{n!}{e}.$$

(Since the sum in brackets - are the first $n+1$members of the Taylor series expansion e^{-1})

In conclusion, it is worth noting that in a similar way to solve the problem when you want to fixed points was not among m the first elements of the permutation (instead of all, as we have just solved). Will result in the formula as the above exact formula, only with the amount will go up k, but not before n.

Problem in online judges

List of tasks that can be solved using the principle of inclusion-exclusion:

- UVA # 10325 **"The Lottery"** [Difficulty: Easy]
- UVA # 11806 **"Cheerleaders"** [Difficulty: Easy]
- TopCoder SRM 477 **"CarelessSecretary"** [Difficulty: Easy]
- TopCoder TCHS 16 **"Divisibility"** [Difficulty: Easy]
- SPOJ # 6285 NGM2 **"Another Game With Numbers"** [Difficulty: Easy]
- TopCoder SRM 382 **"CharmingTicketsEasy"** [Difficulty: Medium]
- TopCoder SRM 390 **"SetOfPatterns"** [Difficulty: Medium]
- TopCoder SRM 176 **"Deranged"** [Difficulty: Medium]
- TopCoder SRM 457 **"TheHexagonsDivOne"** [Difficulty: Medium]
- SPOJ # 4191 MSKYCODE **"Sky Code"** [Difficulty: Medium]
- SPOJ # 4168 SQFREE **"Square Free-integers"** [Difficulty: Medium]
- CodeChef **"Count Relations"** [Difficulty: Medium]

Literature

- Debra K. Borkovitz. **"Derangements and the Inclusion-Exclusion Principle"**

Лучшее вначале

Поделиться

Избранный

Присоединиться к обсуждению...

Константин Ольмезов · 21 день назад

Задача про число взаимно простых четвёрок может решаться аналогично для взаимно простых пятёрок, шестёрок и т. д.?

Ответить · Поделиться ›