Texts in Theoretical Computer Science. An EATCS Series

Extremal Combinatorics

With Applications in Computer Science

Bearbeitet von
Stasys Jukna

1. Auflage 2001. Buch. XVII, 375 S. Hardcover
ISBN 978 3 540 66313 3
Format (B x L): 15,5 x 23,5 cm
Gewicht: 706 g

1. Counting

We start with the oldest combinatorial tool — *counting*.

1.1 The binomial theorem

Given a set of n elements, how many of its subsets have exactly k elements? This number (of k-element subsets of an n-element set) is usually denoted by $\binom{n}{k}$ and is called the *binomial coefficient*. The following identity was proved by Sir Isaac Newton in about 1666, and is known as the *Binomial theorem*.

Binomial Theorem. *Let n be a positive integer. Then for all x and y,*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Proof. By definition, $\binom{n}{k}$ is precisely the number of ways to get the term $x^k y^{n-k}$ when multiplying $(x+y)(x+y)\cdots(x+y)$. □

The *factorial* of n is the product $n! \rightleftharpoons n(n-1)\cdots 2\cdot 1$. This is extended to all non-negative integers by letting $0! = 1$. The *k-th factorial* of n is the product of the first k terms:

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

Note that $\binom{n}{0} = 1$ (the empty set) and $\binom{n}{n} = 1$ (the whole set). In general, binomial coefficients can be written as quotients of factorials:

Proposition 1.1. $$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}.$$

Proof. Observe that $(n)_k$ is the number of (ordered!) strings $(x_1, x_2, \ldots, x_k)$ consisting of k different elements of a fixed n-element set: there are n possibilities to choose the first element x_1; after that there are still $n-1$ possibilities to choose the next element x_2, etc. Another way to produce such strings is to choose a k-element set and then arrange its elements in an arbitrary order. Since each of $\binom{n}{k}$ k-element subsets produces exactly $(k)_k = k!$ such strings, we conclude that $(n)_k = \binom{n}{k}k!$. □

There are a lot of useful equalities concerning binomial coefficients. In most situations, using their *combinatorial* nature (instead of algebraic, as given by the previous proposition) we obtain the desired result fairly easily. For example, if we observe that each subset is uniquely determined by its complement, then we immediately obtain the equality

$$\binom{n}{n-k} = \binom{n}{k}. \tag{1.1}$$

In a similar way other useful identities can be established (see Exercises for more examples).

Proposition 1.2. $\displaystyle\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$

Proof. The first term $\binom{n-1}{k-1}$ is the number of k-sets containing a fixed element, and the second term $\binom{n-1}{k}$ is the number of k-sets avoiding this element; their sum is the whole number $\binom{n}{k}$ of k-sets. □

For growing n and k, exact values of binomial coefficients $\binom{n}{k}$ are hard to compute. In applications, however, we are often interested only in their rate of growth, so that (even rough) estimates suffice. Such estimates can be obtained, using the Taylor series of the exponential function:

$$\mathrm{e}^t = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots \tag{1.2}$$

This, in particular, implies one useful estimate, which we will use later quite often:

$$1 + t < \mathrm{e}^t \quad \text{for all } t \in \mathbb{R},\ t \neq 0. \tag{1.3}$$

Proposition 1.3. $\displaystyle\left(\frac{n}{k}\right)^k \leqslant \binom{n}{k} < \left(\frac{\mathrm{e}n}{k}\right)^k.$

Proof. Lower bound:

$$\left(\frac{n}{k}\right)^k = \frac{n}{k} \cdot \frac{n}{k} \cdots \frac{n}{k} \leqslant \frac{n}{k} \cdot \frac{n-1}{k-1} \cdots \frac{n-k+1}{1} = \binom{n}{k}.$$

Upper bound. By (1.3) and binomial theorem,

$$\mathrm{e}^{nt} > (1+t)^n = \sum_{i=1}^{n} \binom{n}{i} t^i > \binom{n}{k} t^k.$$

Substituting $t = k/n$ we obtain

$$\mathrm{e}^k > \binom{n}{k} \left(\frac{k}{n}\right)^k,$$

as desired. □

Tighter (asymptotic) estimates can be obtained using the famous *Stirling formula* for the factorial:

$$n! = \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}\, \mathrm{e}^{\alpha_n}, \tag{1.4}$$

where $1/(12n+1) < \alpha_n < 1/12n$. This leads, for example, to the following elementary but very useful asymptotic formula for the kth factorial:

$$(n)_k = n^k \mathrm{e}^{-\frac{k^2}{2n} - \frac{k^3}{6n^2} + o(1)} \quad \text{valid for } k = o(n^{3/4}), \tag{1.5}$$

and hence, for binomial coefficients:

$$\binom{n}{k} = \frac{n^k \mathrm{e}^{-\frac{k^2}{2n} - \frac{k^3}{6n^2}}}{k!}(1 + o(1)). \tag{1.6}$$

1.2 Selection with repetitions

In the previous section we considered the number of ways to choose r *distinct* elements from an n-element set. It is natural to ask what happens if we can choose the same element repeatedly? In other words, we may ask how many integer solutions does the equation $x_1+\cdots+x_n = r$ have under the condition that $x_i \geqslant 0$ for all $i = 1, \ldots, n$? (Look at x_i as the number of times the ith element was chosen.) The following more entertaining formulation of this problem was suggested by Lovász, Pelikán, and Vesztergombi (1977).

Suppose we have r sweets (of the same sort), which we want to distribute to n children. In how many ways we can do this? Letting x_i denote the number of sweets we give to the ith child, this question is equivalent to that stated above.

The answer depends on how many sweets we have and how fair we are. If we are fair but have only $r \leqslant n$ sweets, then it is natural to allow no repetitions and give each child no more than one sweet (each x_i is 0 or 1). In this case the answer is easy: we just choose those r (out of n) children who will get a sweet, and we already know that this can be done in $\binom{n}{r}$ ways.

Suppose now that we have enough sweets, i.e., that $r \geqslant n$. Let us first be fair, that is, we want every child gets at least one sweet. We lay out the sweets in a single row of length r (it does not matter in which order, they all are alike), and let the first child pick them up from the left to right. After a while we stop him/her and let the second child pick up sweets, etc. The distribution of sweets is determined by specifying the place (between consecutive sweets) of where to start with a new child. There are $r-1$ such places, and we have to select $n-1$ of them (the first child always starts at the beginning, so we have no choice here). For example, if we have $r = 9$ sweets and $n = 6$ children, a typical situation looks like this:

$$\begin{array}{cccccccccccccc} \square & \curlywedge & \square & \square & \square & \curlywedge & \square & \curlywedge & \square & \square & \curlywedge & \square & \curlywedge & \square \\ & 2 & & & & 3 & & 4 & & & 5 & & 6 & \end{array}$$

Thus, we have to select an $(n-1)$-element subset from an $(r-1)$-element set. The number of possibilities to do so is $\binom{r-1}{n-1}$. If we are unfair, we have more possibilities:

Proposition 1.4. *The number of integer solutions to the equation*

$$x_1 + \cdots + x_n = r$$

under the condition that $x_i \geqslant 0$ *for all* $i = 1, \ldots, n$, *is* $\binom{n+r-1}{r}$.

Proof. In this situation we are unfair and allow that some of the children may be left without sweet. With the following trick we can reduce the problem of counting the number of such distributions to the problem we just solved: we borrow one sweet from each child, and then distribute the whole amount of $n+r$ sweets to the children so that each child gets at least one sweet. This way every child gets back the sweet we borrowed from him/her, and the lucky ones get some more. This "more" is exactly r sweets distributed to n children. We already know that the number of ways to distribute $n+r$ sweets to n children in a fair way is $\binom{n+r-1}{n-1}$, which by (1.1) equals $\binom{n+r-1}{r}$. □

1.3 Partitions

A *partition* of n objects is a collection of its mutually disjoint subsets, called *blocks*, whose union gives the whole set. Let $S(n; k_1, k_2, \ldots, k_n)$ denote the number of all partitions of n objects with k_i i-element blocks ($i = 1, \ldots, n$; $k_1 + 2k_2 + \ldots + nk_n = n$).

Proposition 1.5. $$S(n; k_1, k_2, \ldots, k_n) = \frac{n!}{k_1! \cdots k_n! (1!)^{k_1} \cdots (n!)^{k_n}}.$$

Proof. If we consider any arrangement (i.e., a permutation) of the n objects we can get such a partition by taking the first k_1 elements as 1-element blocks, the next $2k_2$ elements as 2-element blocks, etc. Since we have $n!$ possible arrangements, it remains to show that we get any given partition exactly

$$k_1! \cdots k_n! (1!)^{k_1} \cdots (n!)^{k_n}$$

times. Indeed, we can construct an arrangement of the objects by putting the 1-element blocks first, then the 2-element blocks, etc. However, there are $k_i!$ possible ways to order the i-element blocks and $(i!)^{k_i}$ possible ways to order the elements in the i-element blocks. □

1.4 Double counting

The *double counting* principle states the following "obvious" fact: if the elements of a set are counted in two different ways, the answers are the same.

In terms of matrices the principle is as follows. Let M be an $n \times m$ matrix with entries 0 and 1. Let r_i be the number of 1's in the ith row, and c_j be the number of 1's in the jth column. Then

$$\sum_{i=1}^{n} r_i = \sum_{j=1}^{m} c_j = \text{the total number of 1's in } M.$$

The next example is a standard demonstration of double counting. Suppose a finite number of people meet at a party and some shake hands. Assume that no person shakes his or her own hand and furthermore no two people shake hands more than once.

Handshaking Lemma. *At a party, the number of guests who shake hands an odd number of times is even.*

Proof. Let $P_1, \ldots, P_n$ be the persons. We apply double counting to the set of ordered pairs (P_i, P_j) for which P_i and P_j shake hands with each other at the party. Let x_i be the number of times that P_i shakes hands, and y the total number of handshakes that occur. On one hand, the number of pairs is $\sum_{i=1}^{n} x_i$, since for each P_i the number of choices of P_j is equal to x_i. On the other hand, each handshake gives rise to two pairs (P_i, P_j) and (P_j, P_i); so the total is $2y$. Thus $\sum_{i=1}^{n} x_i = 2y$. But, if the sum of n numbers is even, then evenly many of the numbers are odd. (Because if we add an odd number of odd numbers and any number of even numbers, the sum will be always odd). □

This lemma is also a direct consequence of the following general identity, whose special version for graphs was already proved by Euler. For a point x, its *degree* or *replication number* $d(x)$ in a family $\mathcal{F}$ is the number of members of $\mathcal{F}$ containing x.

Proposition 1.6. *Let $\mathcal{F}$ be a family of subsets of some set X. Then*

$$\sum_{x \in X} d(x) = \sum_{A \in \mathcal{F}} |A|. \tag{1.7}$$

Proof. Consider the *incidence matrix* $M = (m_{x,A})$ of $\mathcal{F}$. That is, M is a 0-1 matrix with $|X|$ rows labeled by points $x \in X$ and with $|\mathcal{F}|$ columns labeled by sets $A \in \mathcal{F}$ such that $m_{x,A} = 1$ if and only if $x \in A$. Observe that $d(x)$ is exactly the number of 1's in the x-th row, and $|A|$ is the number of 1's in the A-th column. □

Graphs are families of 2-element sets, and the degree of a vertex x is the number of edges incident to x, i.e., the number of vertices in its neighborhood. Proposition 1.6 immediately implies

Theorem 1.7 (Euler 1736)**.** *In every graph the sum of degrees of its vertices is two times the number of its edges, and hence, is even.*

The following identities can be proved in a similar manner (we leave their proofs as exercises):

$$\sum_{x \in Y} d(x) = \sum_{A \in \mathcal{F}} |Y \cap A| \quad \text{for any } Y \subseteq X. \tag{1.8}$$

$$\sum_{x \in X} d(x)^2 = \sum_{A \in \mathcal{F}} \sum_{x \in A} d(x) = \sum_{A \in \mathcal{F}} \sum_{B \in \mathcal{F}} |A \cap B|. \tag{1.9}$$

Turán's number $T(n, k, l)$ $(l \leqslant k \leqslant n)$ is the smallest number of l-element subsets of an n-element set X such that every k-element subset of X contains at least one of these sets.

Proposition 1.8. *For all positive integers $l \leqslant k \leqslant n$,*

$$T(n, k, l) \geqslant \binom{n}{l} \Big/ \binom{k}{l}.$$

Proof. Let $\mathcal{F}$ be a smallest l-uniform family over X such that every k-subset of X contains at least one member of $\mathcal{F}$. Take a 0-1 matrix $M = (m_{A,B})$ whose rows are labeled by sets A in $\mathcal{F}$, columns by k-element subsets B of X, and $m_{A,B} = 1$ if and only if $A \subseteq B$. Let r_A be the number of 1's in the A-th row and c_B be the number of 1's in the B-th column. Then, $c_B \geqslant 1$ for every B, since B must contain at least one member of $\mathcal{F}$. On the other hand, r_A is precisely the number of k-element subsets B containing a fixed l-element set A; so $r_A = \binom{n-l}{k-l}$ for every $A \in \mathcal{F}$. By the double counting principle,

$$|\mathcal{F}| \cdot \binom{n-l}{k-l} = \sum_{A \in \mathcal{F}} r_A = \sum_{B} c_B \geqslant \binom{n}{k},$$

which yields

$$T(n, k, l) = |\mathcal{F}| \geqslant \binom{n}{k} \Big/ \binom{n-l}{k-l} = \binom{n}{l} \Big/ \binom{k}{l},$$

where the last equality is another property of binomial coefficients (see Exercise 1.10). □

1.5 The averaging principle

Suppose we have a set of m objects, the ith of which has "size" l_i, and we would like to know if at least one of the objects is large, i.e., has size $l_i \geqslant t$ for some given t. In this situation we can try to consider the average size $\bar{l} = \sum l_i/m$ and try to prove that $\bar{l} \geqslant t$. This would immediately yield the result, because we have the following

Averaging Principle. *Every set of numbers must contain a number at least as large ($\geqslant$) as the average and a number at least as small ($\leqslant$) as the average.*

This principle is a prototype of a very powerful technique – the probabilistic method – which we will study in Part 4. The concept is very simple, but the applications can be surprisingly subtle. We will use this principle quite often.

To demonstrate the principle, let us prove the following sufficient condition that a graph is disconnected.

A (connected) *component* in a graph is a set of its vertices such that there is a path between any two of them. A graph is *connected* if it consists of one component; otherwise it is *disconnected*.

Proposition 1.9. *Every graph on n vertices with fewer than $n-1$ edges is disconnected.*

Proof. Induction by n. When $n = 1$, the claim is vacuously satisfied, since no graph has a negative number of edges.

When $n = 2$, a graph with less than 1 edge is evidently disconnected.

Suppose now that the result has been established for graphs on n vertices, and take a graph $G = (V, E)$ on $|V| = n + 1$ vertices such that $|E| \leqslant n - 1$. By Euler's theorem (Theorem 1.7), the average degree of its vertices is

$$\frac{1}{|V|}\sum_{x\in V} d(x) = \frac{2\,|E|}{|V|} \leqslant \frac{2(n-1)}{n+1} < 2.$$

By the averaging principle, some vertex x has degree 0 or 1. If $d(x) = 0$, x is a component disjoint from the rest of G, so G is disconnected. If $d(x) = 1$, suppose the unique neighbor of x is y. Then, the graph H obtained from G by deleting x and its incident edge has $|V| - 1 = n$ vertices and $|E| - 1 \leqslant (n-1) - 1 = n - 2$ edges; by the induction hypothesis, H is disconnected. The restoration of an edge joining a vertex y in one component to a vertex x which is outside of a second component cannot reconnect the graph. Hence, G is also disconnected. □

We mention one important inequality, which is especially useful when dealing with averages.

A real-valued function $f(x)$ is *convex* if

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y),$$

for any $0 \leqslant \lambda \leqslant 1$. From a geometrical point of view, the convexity of f means that if we draw a line l through points $(x, f(x))$ and $(y, f(y))$, then the graph of the curve $f(z)$ must lie below that of $l(z)$ for $z \in [x, y]$. Thus, for a function f to be convex it is sufficient that its second derivative is nonnegative.

Proposition 1.10 (Jensen's Inequality)**.** *If $0 \leqslant \lambda_i \leqslant 1$, $\sum_{i=1}^{r}\lambda_i = 1$ and f is convex, then*

$$f\Big(\sum_{i=1}^{r}\lambda_i x_i\Big) \leqslant \sum_{i=1}^{r}\lambda_i f(x_i).$$

Proof. Easy induction on the number of summands r. For $r = 2$ this is true, so assume the inequality holds for the number of summands up to r, and prove it for $r + 1$. For this it is enough to replace the sum of the first two terms in $\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_{r+1} x_{r+1}$ by the term

$$(\lambda_1 + \lambda_2)\left(\frac{\lambda_1}{\lambda_1 + \lambda_2} x_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2} x_2\right),$$

and apply the induction hypothesis. □

Jensen's inequality immediately yields the following useful inequality between the arithmetic and geometric means.

Proposition 1.11. *Let $a_1, \ldots, a_n$ be non-negative numbers. Then*

$$\frac{1}{n}\sum_{i=1}^{n} a_i \geqslant \left(\prod_{i=1}^{n} a_i\right)^{1/n}. \tag{1.10}$$

Proof. Let $f(x) = 2^x$, $\lambda_1 = \ldots = \lambda_n = 1/n$ and $x_i = \log_2 a_i$, for all $i = 1, \ldots, n$. By Jensen's inequality

$$\frac{1}{n}\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} \lambda_i f(x_i) \geqslant f\left(\sum_{i=1}^{n} \lambda_i x_i\right) = 2^{\left(\sum_{i=1}^{n} x_i\right)/n} = \left(\prod_{i=1}^{n} a_i\right)^{1/n}.$$

□

To give a less direct illustration, consider the following question: given a binary tree with m leaves, what can be said about the average length of its paths from a root to a leaf? This question may be restated in terms of so-called "prefix-free codes."

A string a is a *prefix* of a string b if $b = ac$ for some string c. A *prefix-free code* is a set $C = \{c_1, \ldots, c_m\}$ of 0-1 strings none of which is a prefix of another. Let l_i be the length of c_i. The famous Kraft inequality from Information Theory states that

$$\sum_{i=1}^{m} 2^{-l_i} \leqslant 1 \tag{1.11}$$

(see Exercise 1.24 for the proof). Together with Jensen's inequality, this implies that the average length $\bar{l} = \sum l_i/m$ of strings in C is at least $\log m$ (see Exercise 1.25). This fact may be also easily derived using the direct convexity argument.

Proposition 1.12. *If C is a prefix-free code then the average length of its strings is at least* $\log_2 |C|$.

Note that prefix-free codes are in one-to-one correspondence with binary trees: given a binary tree, the set of paths from its root to leaves forms a prefix-free code (just because no such path can be prolonged). Thus, the proposition says that the average depth of a binary tree with m leaves is at least $\log_2 m$.

Proof. Let $C = \{c_1, \ldots, c_m\}$ be a prefix-free code and let l_i be the length of c_i. Consider the total length $\ell(C) \rightleftharpoons l_1 + l_2 + \cdots + l_m$. Our goal is to show that $\ell(C) \geqslant |C| \cdot \log_2 |C|$.

We argue by induction on m. The basic cases $m = 1$ and $m = 2$ are trivial. Now take a prefix-free code C with more than two strings. Depending on what the first bit is, split C into two subsets C_0 and C_1. If we attach a fixed string as a prefix to all the strings in C, then the total length can only increase, whereas the number of strings remains the same. Therefore, we can assume w.l.o.g. that both C_0 and C_1 are non-empty.

Since the first bit contributes exactly $|C_0| + |C_1|$ to the total length $\ell(C)$, we have, by the induction hypothesis, that

$$\begin{aligned}\ell(C) &= \big(\ell(C_0) + |C_0|\big) + \big(\ell(C_1) + |C_1|\big) \\ &\geqslant |C_0| \cdot \log_2 |C_0| + |C_1| \cdot \log_2 |C_1| + |C|.\end{aligned}$$

For $x > 0$ the function $f(x) = x \log_2 x$ is convex (its second derivative is $f''(x) = (\log_2 e)/x > 0$). Applying Jensen's inequality to the previous estimate (with $x_1 = |C_0|$, $x_2 = |C_1|$ and $\lambda_1 = \lambda_2 = 1/2$), we conclude that

$$\begin{aligned}\ell(C) &\geqslant \big(|C_0| + |C_1|\big) \cdot \log_2 \frac{|C_0| + |C_1|}{2} + |C| \\ &= |C| \cdot \log_2 \frac{|C|}{2} + |C| = |C| \cdot \log_2 |C|,\end{aligned}$$

as desired. □

Exercises

1.1.$^-$ In how many ways can we distribute k balls to n boxes so that each box has at most one ball?

1.2.$^-$ Show that for every k the product of any k subsequent natural numbers is divisible by $k!$. *Hint*: Consider $\binom{n+k}{k}$.

1.3.$^-$ Show that the number of pairs (A, B) of distinct subsets of $\{1, \ldots, n\}$ with $A \subset B$ is $3^n - 2^n$.

Hint: Use the binomial theorem to evaluate $\sum_{k=0}^{n} \binom{n}{k}(2^k - 1)$.

1.4.$^-$ Prove that

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}.$$

Hint: Count in two ways the number of pairs (x, M), where M is a k-element subset of $\{1, \ldots, n\}$ and $x \in M$.

1.5.$^-$ Prove that $\sum_{k=1}^{n} k\binom{n}{k} = n2^{n-1}$.

Hint: Count in two ways the number of pairs (x, M) with $x \in M \subseteq \{1, \dots, n\}$.

1.6.$^-$ Show that $\binom{n}{0}^2 + \binom{n}{1}^2 + \binom{n}{2}^2 + \cdots + \binom{n}{n}^2 = \binom{2n}{n}$.

Hint: Apply the binomial theorem for $(1 + x)^{2n}$.

1.7.$^-$ There is a set of $2n$ people: n male and n female. A good party is a set with the same number of male and female. How many possibilities are there to build such a good party?

1.8.$^-$ Prove the Cauchy–Vandermonde identity: $\binom{m+n}{k} = \sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i}$.

Hint: Take a set of $m + n$ people (m male and n female) and make a set of k people (with i male and $k - i$ female).

1.9. Prove the following analogy of the binomial theorem for factorials:

$$(x + y)_n = \sum_{k=0}^{n} \binom{n}{k} (x)_k (y)_{n-k}.$$

Hint: Divide both sides by $n!$, and use the Cauchy–Vandermonde identity.

1.10.$^{(!)}$ Let $0 \leqslant l \leqslant k \leqslant n$. Show that

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

Hint: Count in two ways the number of all pairs (L, K) of subsets of $\{1, \dots, n\}$ such that $L \subseteq K$, $|L| = l$ and $|K| = k$.

1.11. Use combinatorics (not algebra) to prove that, for $0 \leqslant k \leqslant n$, $\binom{n}{2} = \binom{k}{2} + k(n - k) + \binom{n-k}{2}$.

Hint: $\binom{n}{2}$ is the number of edges in a complete graph on n vertices.

1.12. Prove Fermat's Little theorem: if p is a prime and if a is a natural number, then $a^p \equiv a \pmod{p}$. In particular, if p does not divide a, then $a^{p-1} \equiv 1 \pmod{p}$.

Hint: Apply the induction on a. For the induction step, use the binomial theorem to show that $(a + 1)^p \equiv a^p + 1 \pmod{p}$.

1.13. Let $0 < \alpha < 1$ be a real number, and αn be an integer. Using Stirling's formula show that

$$\binom{n}{\alpha n} = \frac{1 + o(1)}{\sqrt{2\pi\alpha(1 - \alpha)n}} \cdot 2^{n \cdot H(\alpha)},$$

where $H(\alpha) = -\alpha \log_2 \alpha - (1 - \alpha)\log_2(1 - \alpha)$ is the binary entropy function.
Hint: $H(\alpha) = \log_2 h(\alpha)$, where $h(\alpha) = \alpha^{-\alpha}(1 - \alpha)^{-(1-\alpha)}$.

1.14. Prove that, for $s \leqslant n/2$,

(1) $\displaystyle\sum_{k=0}^{s}\binom{n}{k} \leqslant \binom{n}{s}\left(1+\frac{s}{n-2s+1}\right)$;

(2) $\displaystyle\sum_{k=0}^{s}\binom{n}{k} \leqslant 2^{n\cdot H(s/n)}$.

Hint: To (1): observe that $\binom{n}{k-1}/\binom{n}{k} = k/(n-k+1)$ does not exceed $\alpha \rightleftharpoons s/(n-s+1)$, and use the identity $\sum_{i=0}^{\infty}\alpha^i = 1/(1-\alpha)$.
To (2): set $p = s/n$ and apply the binomial theorem to show that

$$p^s(1-p)^{n-s}\sum_{k=0}^{s}\binom{n}{k} \leqslant 1.$$

See also Corollary 23.6 for another proof.

1.15.$^{+}$ Prove the following estimates: If $k \leqslant k+x < n$ and $y < k \leqslant n$, then

$$\left(\frac{n-k-x}{n-x}\right)^x \leqslant \binom{n-x}{k}\binom{n}{k}^{-1} \leqslant \left(\frac{n-k}{n}\right)^x \leqslant \mathrm{e}^{-(k/n)x} \tag{1.12}$$

and

$$\left(\frac{k-y}{n-y}\right)^y \leqslant \binom{n-y}{k-y}\binom{n}{k}^{-1} \leqslant \left(\frac{n}{k}\right)^y \leqslant \mathrm{e}^{-(1-k/n)y}.$$

1.16.$^{+}$ Prove that if $1 \leqslant k \leqslant n/2$, then

$$\binom{n}{k} \geqslant \gamma\cdot\left(\frac{n\mathrm{e}}{k}\right)^k, \text{ where } \gamma = \frac{1}{\sqrt{2\pi k}}\,\mathrm{e}^{-k^2/n-1/(6k)}. \tag{1.13}$$

Hint: Use Striling's formula to show that

$$\binom{n}{k} \geqslant \frac{1}{\sqrt{2\pi}\,\mathrm{e}^{1/(6k)}}\left(\frac{n}{k}\right)^k\left(\frac{n}{n-k}\right)^{n-k}\left(\frac{n}{k(n-k)}\right)^{1/2},$$

and apply the estimate $\ln(1+t) \geqslant t - t^2/2$ valid for all $t \geqslant 0$.

1.17. In how many ways can we choose a subset $S \subseteq \{1,2,\ldots,n\}$ such that $|S| = k$ and no two elements of S precede each other, i.e., $x \neq y+1$ for all $x \neq y \in S$?

Hint: If $S = \{a_1,\ldots,a_k\}$ is such a subset with $a_1 < a_2 < \ldots < a_k$, then $a_1 < a_2 - 1 < \ldots < a_k - (k-1)$.

1.18. Let $k \geqslant 2n$. In how many ways can we distribute k sweets to n children, if each child is supposed to get at least 2 of them?

1.19. *Bell's number* B_n is the number of all possible partitions of an n-element set X (we assume that $B_0 = 1$). Prove that $B_{n+1} = \sum_{i=1}^{n}\binom{n}{i}B_i$.

Hint: For every subset $A \subseteq X$ there are precisely $B_{|X-A|}$ partitions of X containing A as one of its blocks.

1.20. Let $|N| = n$ and $|X| = x$. Show that there are x^n mappings from N to X, and that $S(n,k)x(x-1)\cdots(x-k+1)$ of these mappings have a range of cardinality k; here $S(n,k)$ is the Stirling number (the number of partitions of an n-element set into exactly k blocks).

Hint: We have $x(x-1)\cdots(x-k+1)$ possibilities to choose a sequence of k elements in X, and we can specify $S(n,k)$ ways in which elements of N are mapped onto these chosen elements.

1.21.$^-$ Let $\mathcal{F}$ be a family of subsets of an n-element set X with the property that any two members of $\mathcal{F}$ meet, i.e., $A \cap B \neq \emptyset$ for all $A, B \in \mathcal{F}$. Suppose also that no other subset of X meets all of the members of $\mathcal{F}$. Prove that $|\mathcal{F}| = 2^{n-1}$. *Hint*: Consider sets and their complements.

1.22. Let $\mathcal{F}$ be a family of k-elements subsets of an n-element set X such that every l-element subset of X is contained in at least one member of $\mathcal{F}$. Show that $|\mathcal{F}| \geqslant \binom{n}{l}/\binom{k}{l}$. *Hint*: Argue as in the proof of Proposition 1.8.

1.23 (Sperner 1928). Let $\mathcal{F}$ be a family of k-element subsets of $\{1,\ldots,n\}$. Its *shadow* is the family of all those $(k-1)$-element subsets which lie entirely in at least one member of $\mathcal{F}$. Show that the shadow contains at least $k|\mathcal{F}|/(n-k+1)$ sets. *Hint*: Argue as in the proof of Proposition 1.8.

1.24. Prove the Kraft inequality (1.11).

Hint: Let $l = \max l_i$, and let A_i be the set of vectors in the cube $\{0,1\}^l$ for which c_i is a prefix. These sets are disjoint and each A_i is a 2^{-l_i}-fraction of the cube.

1.25.$^-$ Use Jensen's inequality to derive Proposition 1.12 from the Kraft inequality (1.11). *Hint*: Apply Jensen's inequality with $\lambda_i = 1/m$ and $f(x) = 2^{-x}$.

1.26 (Quine 1988). The famous Fermat's Last Theorem states that if $n > 2$, then $x^n + y^n = z^n$ has no solutions in nonzero integers x, y and z. This theorem can be stated in terms of sorting objects into a row of bins, some of which are red, some blue, and the rest unpainted. The theorem amounts to saying that when there are more than two objects, the following statement is newer true: *The number of ways of sorting them that shun both colors is equal to the number of ways that shun neither.* Show that this statement is equivalent to Fermat's equation $x^n + y^n = z^n$.

Hint: Let n be the number of objects, z the number of bins, x the number of bins that are not red and y the number of bins that are not blue.