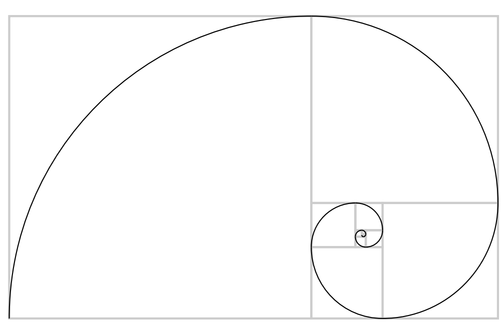
Combinatorics and Applications

Chapter 1: Enumeration

Jenny Zito, Mike Wertheimer, Cheyne Homberger



Contents

– Chapter 1 –

Introduction to Enumeration

Predicting gambling odds provided the early motivation for wanting to answer enumeration questions. When rolling three dice, it is profitable for a gambler to know how many different rolls can occur and in how many ways (and hence how likely) each different roll can occur. The letters between Pascal and Fermat include one of the early discussions of how to calculate the fair division of a gambling pot when a game is terminated early with a certain number of wins on each side.

In the current world, the most powerful motivation for understanding enumeration problems is applications to the design and evaluation of computer algorithms. Evaluation of what is feasible to compute often relies on combinatorial topics such as enumeration, combinatorial structures, and combinatorial algorithms.

What is in this Chapter: The goal of this chapter is to introduce the reader to combinatorial enumeration. The central thread of the chapter is a discussion of four enumeration problems in which we count the number of ways to select k objects from n objects in the cases where we distinguish (or don't distinguish) the order of the selection of the k objects and choose the objects with (or without) repetition from the set of n objects. See the examples on the next few pages. These sample problems will allow the reader to pick up many of the fundamental concepts and skills of combinatorial enumeration and serve as a starting point to introduce many other significant combinatorial enumeration problems and techniques later in the text. (And of course, give the reader the ability to determine odds in games of chance.)

Because of the intimate relationship between counting and probability, we will use many terms normally introduced in a probability course, such as: sampling, repeated sampling, expectation, probability, and characteristic functions. We'll define these as they arise for those who are rusty or as yet uninformed. Some other topics we will see in this chapter are: equivalence relations, partitions, combinatorial style proofs, the equivalence of some basic enumeration problems, and a few extensions of the four main problems of this chapter, such as circular permutations and multinomial coefficients. We will also deal briefly with the notation and methods for estimating the asymptotic behavior of functions, such as Stirling's formula which estimates n factorial for large values of n.

The topics in this chapter can be attempted without developing much heavy combinatorial machinery, but in the course of our discussion we will point out other problems to be tackled later in the text that will require bigger hammers. Many of these harder enumeration problems are just as easy to state as the easy problems, so we will try to point out what makes them harder. At the end of this chapter we'll list the problems and solutions we handled so far, as well as the problems we brought up but won't solve until later in the text.

We encourage readers to take an active part as they read, pausing in their reading to generate their own examples and proof attempts.

---

§ 1.1 Four Problems: Selecting k Things from n Things

Okay, let's start counting. The backbone of this chapter will be to count all the ways of picking k elements from a set of n elements in four different situations, depending on whether or not the k elements are required to be distinct elements of the n set and whether or not the order in which we select the k elements is considered. We will fill in the four boxes in the following chart:

	All k Elements Distinct (Sampling Without Replacement)	**Repeated Elements Allowed (Sampling With Replacement)**
Order of Choice Important	BOX 1 k-permutation	BOX 2 k-sample (ordered list)
Order of Choice Unimportant	BOX 3 k-combination (subset)	BOX 4 k-multiset

Figure 1.1: The number of ways to select k elements from a set of n elements.

First, to become familiar with what we are counting in each of these four boxes, let's look at an example with $k = 3$ and $n = 4$. We will select three elements from the four element set $\{\spadesuit, \heartsuit, \diamondsuit, \clubsuit\}$.

Notation

The standard convention to distinguish ordered lists from subsets of items (where order is not important), is to use parentheses around ordered lists and curly braces around sets. Thus

$$(1, 4, 5) \neq (4, 1, 5), \text{ whereas } \{1, 4, 5\} = \{4, 1, 5\}.$$

Ordered lists in which repeated elements are allowed are often called *samples*. (We'll do our best to use widely accepted notation and terminology, but you will see variations in almost every textbook you pick up.) Ordered lists without any repeated elements will be called *permutations*, and frequently, if there are only $k < n$ elements in the permutation (selected from a set of n elements), it will be called a *k-permutation of an n-set*.

We will also use the curly braces for multisets. Like a set, a *multiset* is an unordered collection of elements, but in contrast to sets, multisets may have multiple copies of the same object. Usually multisets, like sets are written using curly braces as delimiters. If you find it useful to have a different delimiter for multisets, you may want to adopt the suggested by Scheinerman in [4] of using angle braces (e.g. $< 2, 2, 3 >$) for multisets. He also suggests the symbol $\uplus$ for multiset union, so you know to keep track of multiple copies of elements.

For purposes of this text we will use curly braces for both sets and multisets, since it should be clear from the context when we are using sets and when we are allowing multisets.

Since order is unimportant in both sets and multisets, one usually makes the choice to list the elements inside the curly braces in nondecreasing order for ease of reading and comparison (e.g., use $\{1, 2, 2, 5, 6, 6, 6\}$ instead of $\{6, 2, 5, 1, 6, 6, 2\}$).

Because the set $\{1, 2, 3, \ldots, n\}$ appears so frequently in the text, we will find it useful to have a special notation for it

$$[n] = \{1, 2, 3, \ldots, n\}.$$

Now, armed with parenthetical and curly delimiters, let us generate examples for the counting problems described in the four boxes on page 6.

Example 1.1 (3-permutations of a 4-set)**.** In BOX 1, repeated elements are not allowed, and the order of elements in the list is important. These are called k-permutations of an n-set. There are twenty-four 3-permutations with entries from a 4-set.

$$\begin{array}{cccc}
(\spadesuit,\heartsuit,\diamondsuit) & (\heartsuit,\spadesuit,\diamondsuit) & (\diamondsuit,\spadesuit,\heartsuit) & (\clubsuit,\spadesuit,\heartsuit) \\
(\spadesuit,\heartsuit,\clubsuit) & (\heartsuit,\spadesuit,\clubsuit) & (\diamondsuit,\spadesuit,\clubsuit) & (\clubsuit,\spadesuit,\diamondsuit) \\
(\spadesuit,\diamondsuit,\heartsuit) & (\heartsuit,\diamondsuit,\spadesuit) & (\diamondsuit,\heartsuit,\spadesuit) & (\clubsuit,\heartsuit,\spadesuit) \\
(\spadesuit,\diamondsuit,\clubsuit) & (\heartsuit,\diamondsuit,\clubsuit) & (\diamondsuit,\heartsuit,\clubsuit) & (\clubsuit,\heartsuit,\diamondsuit) \\
(\spadesuit,\clubsuit,\heartsuit) & (\heartsuit,\clubsuit,\spadesuit) & (\diamondsuit,\clubsuit,\spadesuit) & (\clubsuit,\diamondsuit,\spadesuit) \\
(\spadesuit,\clubsuit,\diamondsuit) & (\heartsuit,\clubsuit,\diamondsuit) & (\diamondsuit,\clubsuit,\heartsuit) & (\clubsuit,\diamondsuit,\heartsuit)
\end{array}$$

Example 1.2 (3-samples of a 4-set)**.** In BOX 2, repeated elements are allowed, and the order of elements in the list is important. These are called k-long ordered lists of an n-set, or k-samples of an n-set. There are sixty-four 3-long ordered lists with entries from a 4-set.

$$\begin{array}{cccc}
(\spadesuit,\spadesuit,\spadesuit) & (\heartsuit,\spadesuit,\spadesuit) & (\diamondsuit,\spadesuit,\spadesuit) & (\clubsuit,\spadesuit,\spadesuit) \\
(\spadesuit,\spadesuit,\heartsuit) & (\heartsuit,\spadesuit,\heartsuit) & (\diamondsuit,\spadesuit,\heartsuit) & (\clubsuit,\spadesuit,\heartsuit) \\
(\spadesuit,\spadesuit,\diamondsuit) & (\heartsuit,\spadesuit,\diamondsuit) & (\diamondsuit,\spadesuit,\diamondsuit) & (\clubsuit,\spadesuit,\diamondsuit) \\
(\spadesuit,\spadesuit,\clubsuit) & (\heartsuit,\spadesuit,\clubsuit) & (\diamondsuit,\spadesuit,\clubsuit) & (\clubsuit,\spadesuit,\clubsuit) \\
(\spadesuit,\heartsuit,\spadesuit) & (\heartsuit,\heartsuit,\spadesuit) & (\diamondsuit,\heartsuit,\spadesuit) & (\clubsuit,\heartsuit,\spadesuit) \\
(\spadesuit,\heartsuit,\heartsuit) & (\heartsuit,\heartsuit,\heartsuit) & (\diamondsuit,\heartsuit,\heartsuit) & (\clubsuit,\heartsuit,\heartsuit) \\
(\spadesuit,\heartsuit,\diamondsuit) & (\heartsuit,\heartsuit,\diamondsuit) & (\diamondsuit,\heartsuit,\diamondsuit) & (\clubsuit,\heartsuit,\diamondsuit) \\
(\spadesuit,\heartsuit,\clubsuit) & (\heartsuit,\heartsuit,\clubsuit) & (\diamondsuit,\heartsuit,\clubsuit) & (\clubsuit,\heartsuit,\clubsuit) \\
(\spadesuit,\diamondsuit,\spadesuit) & (\heartsuit,\diamondsuit,\spadesuit) & (\diamondsuit,\diamondsuit,\spadesuit) & (\clubsuit,\diamondsuit,\spadesuit) \\
(\spadesuit,\diamondsuit,\heartsuit) & (\heartsuit,\diamondsuit,\heartsuit) & (\diamondsuit,\diamondsuit,\heartsuit) & (\clubsuit,\diamondsuit,\heartsuit) \\
(\spadesuit,\diamondsuit,\diamondsuit) & (\heartsuit,\diamondsuit,\diamondsuit) & (\diamondsuit,\diamondsuit,\diamondsuit) & (\clubsuit,\diamondsuit,\diamondsuit) \\
(\spadesuit,\diamondsuit,\clubsuit) & (\heartsuit,\diamondsuit,\clubsuit) & (\diamondsuit,\diamondsuit,\clubsuit) & (\clubsuit,\diamondsuit,\clubsuit) \\
(\spadesuit,\clubsuit,\spadesuit) & (\heartsuit,\clubsuit,\spadesuit) & (\diamondsuit,\clubsuit,\spadesuit) & (\clubsuit,\clubsuit,\spadesuit) \\
(\spadesuit,\clubsuit,\heartsuit) & (\heartsuit,\clubsuit,\heartsuit) & (\diamondsuit,\clubsuit,\heartsuit) & (\clubsuit,\clubsuit,\heartsuit) \\
(\spadesuit,\clubsuit,\diamondsuit) & (\heartsuit,\clubsuit,\diamondsuit) & (\diamondsuit,\clubsuit,\diamondsuit) & (\clubsuit,\clubsuit,\diamondsuit) \\
(\spadesuit,\clubsuit,\clubsuit) & (\heartsuit,\clubsuit,\clubsuit) & (\diamondsuit,\clubsuit,\clubsuit) & (\clubsuit,\clubsuit,\clubsuit)
\end{array}$$

Example 1.3 (3-subsets of a 4-set)**.** In BOX 3, elements are distinct, and the order of elements is unimportant. These are called the k-subsets of an n-set or the k-combinations of an n-set. There are four 3-subsets of a 4-set.

$$\{\spadesuit,\heartsuit,\diamondsuit\} \quad \{\spadesuit,\heartsuit,\clubsuit\} \quad \{\spadesuit,\diamondsuit,\clubsuit\} \quad \{\heartsuit,\diamondsuit,\clubsuit\}$$

Example 1.4 (3-multisets of a 4-set)**.** In BOX 4, elements are not required to be distinct, and the order of elements is unimportant. These are called the k-multisets of an n-set. There are twenty 3-multisets of a 4-set.

$$\begin{array}{ccccc} \{\spadesuit,\spadesuit,\spadesuit\} & \{\spadesuit,\spadesuit,\heartsuit\} & \{\spadesuit,\spadesuit,\diamondsuit\} & \{\spadesuit,\spadesuit,\clubsuit\} & \{\spadesuit,\heartsuit,\heartsuit\} \\ \{\spadesuit,\heartsuit,\diamondsuit\} & \{\spadesuit,\heartsuit,\clubsuit\} & \{\spadesuit,\diamondsuit,\diamondsuit\} & \{\spadesuit,\diamondsuit,\clubsuit\} & \{\spadesuit,\clubsuit,\clubsuit\} \\ \{\heartsuit,\heartsuit,\heartsuit\} & \{\heartsuit,\heartsuit,\diamondsuit\} & \{\heartsuit,\heartsuit,\clubsuit\} & \{\heartsuit,\diamondsuit,\diamondsuit\} & \{\heartsuit,\diamondsuit,\clubsuit\} \\ \{\heartsuit,\clubsuit,\clubsuit\} & \{\diamondsuit,\diamondsuit,\diamondsuit\} & \{\diamondsuit,\diamondsuit,\clubsuit\} & \{\diamondsuit,\clubsuit,\clubsuit\} & \{\clubsuit,\clubsuit,\clubsuit\} \end{array}$$

Over the course of the next few sections we will generate the formulas in terms of k and n for the number of ways can we select k-elements from an n-set when the k elements form: a sample (also called an ordered list), a permutation, a combination (also called a subset), and a multiset.

At this point it is a useful exercise for the reader to take a few minutes to try to write a formula for the number of objects in each of the four boxes in terms of k and n.

We'll start by solving the two ordered selection problems in our chart (BOX 1 and BOX 2) and a few minor variations. Then we'll introduce the basic probability theory needed to solve a few more problems involving ordered selection. After that we'll go on to the two unordered selection problems from our chart (BOX 3 and BOX 4) and a few variations. Then we'll end the chapter with a summary of what we have counted and a list of more difficult problems that we have introduced without solving.

§ 1.2 Distinct, Ordered Selections: Permutations

We had two problems where order of selection was important: the k-permutations (no repetitions allowed) and the k-samples or ordered lists (repetition allowed).

Let's start with *permutations*, ordered lists where no repeats are allowed (BOX 1 from page 6). Let S be an n-set. We seek the number of ways of selecting k *distinct* elements of S where two selections $(a_1, a_2, \ldots a_k)$ and $(b_1, b_2, \ldots, b_k)$ are considered the same if and only if $a_i = b_i$ for $1 \le i \le k$. We call such an object a k-permutation of an n-set and denote the number of them by $P(n, k)$.

Theorem 1.5. *The number of k-permutations of an n-set (selecting without replacement, with order important) is given by*

$$P(n,k) = \underbrace{n(n-1)(n-2)\cdots(n-k+1)}_{k-terms}.$$

Proof: *The number of ways of selecting one element from a j-set is clearly j so that $P(j, 1) = j$ for all $j > 0$. Now we proceed by induction. Assume the formula holds for $P(j, i)$, $1 \le j < n$, $1 \le i < k$. Evidently*

$$\begin{aligned} P(n,k) &= P(n,1)P(n-1,k-1) \\ &= n(n-1)(n-2)\cdots(n-k+1), \end{aligned}$$

as we choose the first element $P(n,1)$ ways and choose the remaining $k-1$ elements from the remaining $n-1$ elements in $P(n-1,k-1)$ ways. qed

Factorials: The special term $P(n,n) = n(n-1)\cdots 1$ is written $n!$ (and spoken as "n factorial". It counts the number of permutations on n objects. Using this notation, we can write the formula for the number of k-permutations of an n-set that goes in BOX 1 of our chart on page 6 as

$$P(n,k) = \frac{n!}{(n-k)!}.$$

The factorial notation only makes sense, as we have defined it, for positive integers. It will be convenient (and correct in a sense you can eventually convince yourself) to extend the definition so that $0! = 1$. In Section **??** and some exercises we investigate several ways the factorial function can be generalized so that the function is defined over the rationals or even over the real numbers (with the exception of the negative integers).

Circular Permutations: For a set of n distinct objects, how many distinct circular permutations of length k are there?

A circular permutation is an arrangement of distinct objects in a circle, with two arrangements considered equivalent if one can be rotated to become the other. In terms of lists, define a *left-circular shift* of an n-long ordered list $(a_1, a_2, a_3, \ldots, a_n)$ to be the list $(a_2, a_3, \ldots, a_n, a_1)$ resulting when the left-most item is moved to the end of the list, and all other items are moved one position to the left. Two permutations are considered to be the same *circular permutation* if a series of circular shifts of one of the permutations will make it equal to the other permutation. For example: the linear permutations $(2,5,8,4)$, $(5,8,4,2)$, $(8,4,2,5)$, and $(4,2,5,8)$ are circular shifts of each other and hence correspond to the same circular permutation.

In fact, since the k elements of a k-permutation of an n-set are distinct, each of the k possible shifts of the permutation are different standard k-permutations of an n-set corresponding to the same circular k-permutation of the n-set. Hence, the number of circular permutations of length k with elements selected from an n-set is

$$\frac{P(n,k)}{k}.$$

§ 1.3 Approximation Notation and Methods

One of the questions we will often ask is "how fast does a function grow?". Since many of the functions we will see are defined in terms of the factorial function, it is helpful to know how fast the factorial function grows. A useful approximation for $n!$ is given by:

$$\textbf{Stirling's Approximation Formula:} \quad n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \tag{1.1}$$

The tilde symbol can be verbalized as "is asymptotically equal to" and is defined in Section 1.3, below.

Isn't it curious that there is a formula asymptotically relating $n!$, e, and π?

Often it will not be important (or possible) to obtain the exact value of some combinatorial quantity. Instead, we might ask for an *approximation* that is sufficiently close to truth to be meaningful. This is especially true for counting large numbers of objects or for attempting to characterize an entire family of objects. An understanding of the term "approximately" is crucial in order to fully exercise the implications of some particular estimation.

We begin by introducing the following terminology. Suppose f and g are arbitrary functions defined on the positive integers. We say

$$f(n) \sim g(n) \quad \Leftrightarrow \quad \lim_{n\to\infty} \frac{f(n)}{g(n)} = 1 \quad \text{“}f \text{ is asymptotic to } g\text{”} \tag{1.2}$$

Unfortunately, equation (1.2) does not say much about the rate of convergence of the limits. Thus, for example, we can say $n^3+10^{10}n^{2.999}+1 \sim n^3$ even though this approximation may be worthless for small values of n.

The "big O" notation is an extremely convenient and compact way to give a precise meaning to a particular approximation. In particular, if f and g are arbitrary functions defined on the positive integers, we say:

$$f(n) \in O(g(n)) \quad \text{if there exist constants } M \text{ and } N,$$
$$\text{such that for } n > N, \quad |f(n)| \leq M \cdot g(n)$$

The actual values of the constants M and N are generally not specified and in some cases, may not even be known. However, it is important to remember that they can play a very important role in influencing the accuracy of a given approximation. Despite these concerns, the $O-$notation is generally very useful in obtaining accurate estimates for the behavior of complex functions, mostly by extracting the dominant portion of the function and suppressing the irrelevant details.

Example 1.6 (Growth of Polynomials)**.** One of the most useful applications of the $O-$notation is the following:

$$P(n) = a_m n^m + \cdots + a_1 n + a_0 \in O(n^m) \tag{1.3}$$

for constants $a_o, a_1, \ldots, a_m$.

This follows since

$$\begin{aligned} |P(n)| &\leq \sum_{k=0}^{m} \left|a_k n^k\right| = n^m \sum_{k=0}^{m} \left|a_k \frac{n^k}{n^m}\right| \\ &\leq n^m \sum_{k=0}^{m} |a_k| \end{aligned}$$

for $n \geq 1$. Thus we may take $N = 1$ and $M = |a_m| + \cdots + |a_0|$ to prove (1.3). To make a stronger statement we can use (1.3) again to write:

$$P(n) = a_m n^m + O(n^{m-1}).$$

The use of a plus sign with a "big O" class, as in the equation above, is a common shorthand notation for:

$$P(n) = a_m n^m + r(n) \quad \text{where} \quad r(n) \in O(n^{m-1}).$$

While this equation does not give us any handle on the hidden coefficients $a_k,\ k < m$, it does accurately describe the asymptotic behavior of $P(n)$ while also providing useful information on the growth of the error.

Growth of Harmonic Numbers: One of the most striking, and useful, approximations involving the $O-$notation concerns the *harmonic numbers*:

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

It should not be surprising that H_n acts like $\ln n$ as we can quickly convince ourselves that:

$$\ln(n+1) \;=\; \int_1^{n+1} \frac{1}{x}\,dx \;<\; H_n \;<\; 1 + \int_1^{n} \frac{1}{x}\,dx \;=\; 1 + \ln n$$

by using a little bit of calculus ($\int_1^n \frac{1}{x}\,dx = \ln n$) and examining the graph for $y = 1/x$ bounded above or below with a series of rectangles of width 1 and heights $1, 1/2, 1/3, \ldots 1/n$.

Thus, we can say

$$H_n = \ln n + r(n) \quad \text{where} \quad r(n) \in O(1)$$

or using the shorthand notation we have

$$H_n = \ln n + O(1).$$

In fact, a more powerful result is known:

$$H_n = \ln n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} - r(n), \quad \text{where} \quad 0 < r(n) < \frac{1}{252n^6} \tag{1.4}$$

where $\gamma = .5772156649...$ is called the *Euler-Mascheroni constant* or the *Euler constant*.[1] The more customary result regarding the harmonic numbers is simply

$$H_n = \ln n + \gamma + O\left(\frac{1}{n}\right) \tag{1.5}$$

a remarkably concise *and* accurate approximation that has many applications. We will see the harmonic numbers and the Euler-Mascheroni constant again when we consider the Coupon Collectors problem in Section ??.

[1] Although the Euler-Mascheroni constant seems to appear in a variety of equations, as far as I (JSZ) can tell they all boil down to being the limiting difference between the harmonic series and the natural logarithm. Mascheroni is pronounced similarly to "Mahs-keroni".

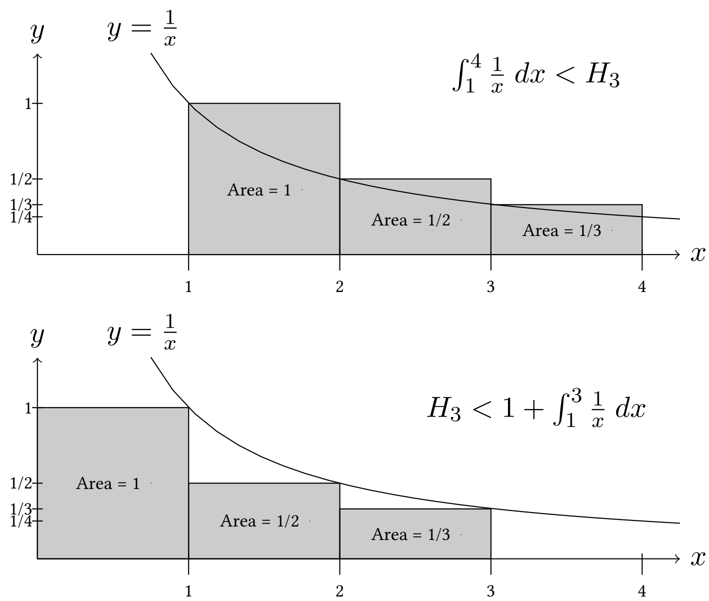


Figure 1.2: The shaded area is the harmonic number $H_3 = 1 + \frac{1}{2} + \frac{1}{3}$ and is bounded above and below in terms of the integral of $\frac{1}{x}$.

Stirling Approximation Formula: It is often very difficult to obtain exact numerical values for expressions involving factorials (such as the binomial coefficients, which we will see often in this text). As such, it is very useful to have a simple, and accurate, approximation for $n!$, which is easily computed. *Stirling's Formula* provides just such an approximation. While its complete proof is beyond our scope we shall give a quick derivation of a slightly easier bound

$$n! < e\sqrt{n}\left(\frac{n}{e}\right)^n \text{ where } e \cong 2.7 \text{ versus } \sqrt{2\pi} \cong 2.5. \tag{1.6}$$

To begin, we see

$$\ln(n!) = \ln(n) + \ln(n-1) + \cdots + \ln(1).$$

Examining the graph in Figure 1.3 for $\ln(x)$ and using the fact that the $\ln$ function is concave it is evident that

$$\begin{aligned}\int_1^n \ln x \, dx &> \frac{1}{2}\ln(2) + \frac{1}{2}(\ln(3) + \ln(2)) + \ldots + \frac{1}{2}(\ln(n) + \ln(n-1)) \\ &= \ln(2) + \ln(3) + \ldots + \ln(n-1) + \frac{1}{2}\ln(n) \\ &= \ln(n!) - \frac{1}{2}\ln(n).\end{aligned}$$

Using $\int \ln x dx = x\ln x - x$ to explicitly evaluate the integral on the left gives the bound

$$\ln n! < \frac{1}{2}\ln n + n\ln n - n + 1. \tag{1.7}$$

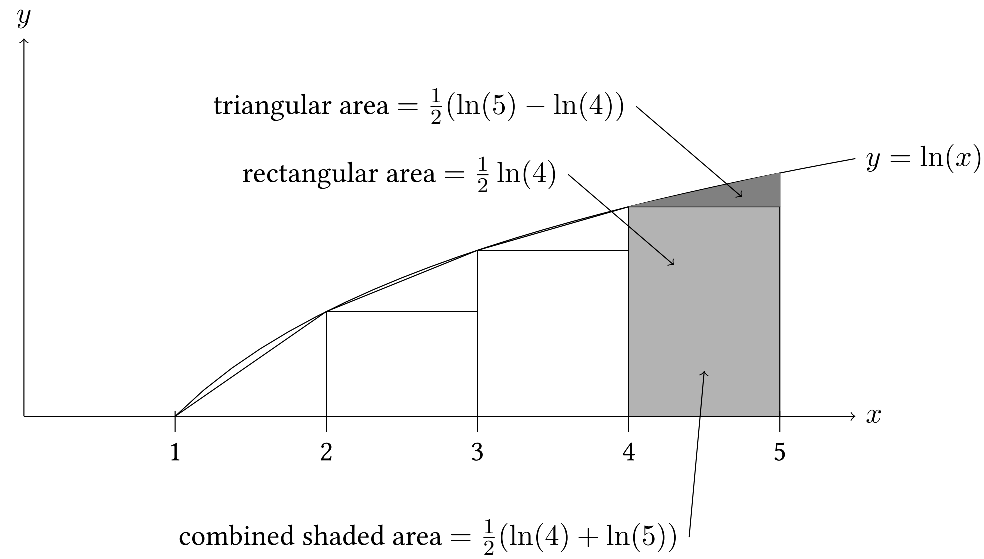


Figure 1.3: The ln function is concave so $\int_k^{k+1} \ln(x) > \frac{1}{2}\big(\ln(k) + \ln(k+1)\big)$ for $k > 1$.

Then we exponentiate both sides to get the desired bound

$$n! < e\sqrt{n}\Big(\frac{n}{e}\Big)^n. \tag{1.8}$$

Note that it is not difficult to show that $\lim_{n\to\infty} \frac{n!}{\sqrt{n}(\frac{n}{e})^n}$ exists. However, it is less elementary to show that the limit is, in fact,

$$\lim_{n\to\infty} \frac{n!}{\sqrt{n}(\frac{n}{e})^n} = \sqrt{2\pi}.$$

The full strength of the result is given by:

STIRLING'S FORMULA:

$$n! = \sqrt{2\pi n}\Big(\frac{n}{e}\Big)^n\bigg(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} - \frac{571}{2488320n^4} + O\Big(\frac{1}{n^5}\Big)\bigg). \tag{1.9}$$

In practice it is common to use the "shortened form" of (1.9):

$$n! \sim \sqrt{2\pi n}\,\left(\frac{n}{e}\right)^n. \tag{1.10}$$

It is important to note that the difference between the two sides of (1.10) increases without bound, but that the ratio of the two sides approaches one. Thus the *percentage* error decreases, and in fact it does so quite rapidly as the following table indicates:

n	n!	Stirling's Approximation (Equation 1.10)	Percentage Error
1	1	0.92214	8.44376
2	2	1.91900	4.22071
3	6	5.83621	2.80645
4	24	23.5062	2.10083
5	120	118.019	1.67840
6	720	710.078	1.39728
7	5040	4980.40	1.19678
8	40320	39902.4	1.04657
9	362880	359537.0	0.92984
10	$3.62880 \cdot 10^6$	$3.59870 \cdot 10^6$	0.83653
11	$3.99168 \cdot 10^7$	$3.96156 \cdot 10^7$	0.76024
12	$4.79001 \cdot 10^8$	$4.75687 \cdot 10^8$	0.69670
13	$6.22702 \cdot 10^9$	$6.18724 \cdot 10^9$	0.64295
14	$8.71782 \cdot 10^{10}$	$8.66610 \cdot 10^{10}$	0.59691
15	$1.30767 \cdot 10^{12}$	$1.30043 \cdot 10^{12}$	0.55702

Figure 1.4: Stirling's Approximation to n!.

§ 1.4 Ordered Selections with Repeats Allowed: Samples

Now let's find the formula for *samples* (BOX 2 from the chart on page 6), where order of selection is important, but now repetition of elements is allowed. This is also called sampling with replacement. Two k-samples $(a_1, a_2, \ldots, a_k)$ with $a_i \in [n]$ and $(b_1, b_2, \ldots, b_k)$ with $b_i \in [n]$ are the same if and only if $a_i = b_i$ for $i = 1$ to k. These objects are also called *ordered lists.*

Theorem 1.7. *The number of k-samples of an n-set (selecting with replacement, with order important) is n^k.*

Proof: *Since we are sampling with replacement from the n-set, the values of each of the k-samples is independent of any of the other samples taken. Thus the total number of k-samples will be the product of k terms, each of which is the number of ways to take a single sample from an n-set (which is of course n). Thus the total number of k-samples of an n-set is*

$$\prod_{k=1}^{n} n = n^k.$$

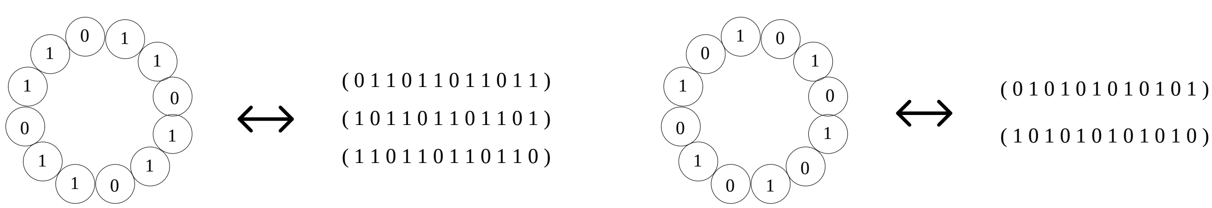
qed

The Necklace Problem: If we ask about circular arrangements when sampling with replacement, the problem of counting the distinct arrangements becomes much more difficult than the problem of circular permutations that we counted on page 10 and is called the *necklace problem.* For example, say that we have two colors of beads, red and green, and we are filling a display case with all the different necklaces with 4 beads. How many are there? When we looked at a k-long permutation laid out in a circle, there are k standard (non-circular) permutations that correspond to the same circular permutation. Can we do something similar here with all the 6-long linear patterns using R and G? Let's write out each set of linear patterns that corresponds to each different necklace. We have:

$$\{RRRR\},\ \{RRRG, RRGR, RGRR, GRRR\},\ \{RRGG, RGGR, GGRR, GRRG\}$$

$$\{RGRG, GRGR\},\ \{RGGG, GRGG, GGRG, GGGR\},\ \text{and } \{GGGG.\}$$

Unlike the case of permutation, a sample (ordered list) can have a repeated pattern of length d, and when shifted by d steps we get back a list that is identical to the original. This means that when repeats are allowed, we no longer have the case that each k-long circular pattern corresponds to exactly k different linear patterns. For example, one of these 12-long necklaces corresponds to two linear patterns, and the other corresponds to three linear patterns. In the extreme case where all



the beads in the necklace are identical, all the circular shifts are indistinguishable and there is only one non-circular sample that corresponds to this necklace. In Section **??** we solve the necklace problem using a more advanced technique called Möbius inversion.

To formalize the special properties that allow us to count the number of circular permutations (no repeats allowed), but are not available to help when we try to solve the necklace problem (repeats allowed), we will want to define equivalence relations and partitions, both of which will be useful in combinatorial proofs later in the text as well.

Equivalence Relation. An *equivalence relation* (written as $\equiv$) on a set S is a binary relationship that is

- reflexive ($a \equiv a \quad \forall a \in S$),
- symmetric ($a \equiv b \iff b \equiv a \quad \forall a, b \in S$), and
- transitive ($a \equiv b$ and $b \equiv c$, implies $a \equiv c \quad \forall a, b, c \in S$).

The subset of elements in the set S that are equivalent to an element a is called the *equivalence class of* a, and we will denote this by $|a|$. (In some books this is denoted by $\overline{a}$ or by $[a]$.) One can quickly verify from the definition of the equivalence relation that

$$a \equiv b \iff |a| = |b|,$$

which is the basis for the following:

Equivalence Classes Partition the Underlying Set. Let S be a set with an equivalence relation $\equiv$. Form subsets $S_i \subset S$ as follows. Pick any element $a_1 \in S$ and set

$$S_1 = |a_1| = \text{ the equivalence class of } a_1.$$

Pick any element not yet in the union of already selected equivalence classes

$$a_i \in S \setminus \bigcup_{j=1}^{i-1} S_j,$$

and define the next subset S_i to be the corresponding equivalence class

$$S_i = |a_i| = \text{ the equivalence class of } a_i,$$

until we have no remaining elements outside of $\{S_1, S_2, \ldots, S_k\}$. Then the S_i are disjoint and their union is the set S. This is expressed more succinctly if we define "set partitions".

Set Partitions. A set S has a *partition* $P = \{S_1, S_2, \cdots, S_t\}$ if the S_i's are subsets of S ($S_i \subseteq S$) that are disjoint ($S_i \cap S_j = \emptyset$) for $i \neq j$ and whose union is the whole set ($S_1 \cup S_2 \cup \cdots \cup S_t = S$). The subsets S_i are called the *parts*, and unless otherwise specified, parts are nonempty.

For example, the set $\{4, 5, 6, 7, 8, 9\}$ has a partition $P = \{\{4, 9\}, \{5\}, \{6, 7, 8\}\}$, which has three parts. Notice that since the order of the subsets within the partition is unimportant, we chose to list first the subset with the smallest entry. This ordering is called the *standard ordering* or *canonical ordering*.

We leave it for the student to verify from the definition of equivalence relation for a set that the equivalence classes of the set under the equivalence relation form a partition of that set (Exercise 1.1 on page 42). This works for finite or infinite sets S.

Circular and Linear Permutations and Equivalence Relations. Now we can return to the example of counting circular permutations. Define an equivalence relation on the set S of linear k-permutations of an n-set, such that $\sigma \equiv \lambda$ if there is a circular shift of σ that gives you λ. To make this more precise, let us say π_i circular shifts the elements of a linear permutation i spaces to the left, thus if $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_k)$, we have

$$\pi_i(\sigma) = (\sigma_{i+1}, \sigma_{i+2}, \ldots, \sigma_k, \sigma_1, \sigma_2, \ldots, \sigma_i).$$

Because each of the elements in a k-permutation are distinct, each of the k possible circular shifts of that permutation are different linear permutation. Thus the equivalence classes of k-permutations under the equivalence relation of cirular shifting each have k elements. Since the equivalence classes form a partition of the set of linear permutations we can count the number of equivalence classes by dividing the size of the set by the size of the equivalence classes. This gives $P(n,k)/k$ equivalence classes each of which corresponds to a different circular permutation.

We could not use this same proof technique to count the number of k long circular patterns when repeats are allowed because equivalence classes under shifting no longer all have the same size. For example under circular shifting this equivalence class has three members:

$$(R,R,G,R,R,G) \equiv (R,G,R,R,G,R) \equiv (G,R,R,G,R,R),$$

whereas this equivalence class has two members:

$$(G,R,G,R,G,R) \equiv (R,G,R,G,R,G).$$

Counting Set Partitions and Unordered Set Partitions. The number of partitions of an r-set into t nonempty parts are counted by the Stirling numbers of the second kind, which are also called the *subset Stirling numbers* [2]. There are many notations for Stirling numbers. My favorite is

$$\left\{ {r \atop t} \right\},$$

which is spoken as "r subset t" because it partitions $[r]$ into t subsets [2].

Once we have developed the technique of Inclusion-Exlusion in Chapter **??** on Inversion Theory, we prove the following formula for the Stirling subset numbers (see Equation **??** on page **??**):

$$\left\{ {r \atop t} \right\} = \sum_{i=0}^{t} \frac{(-1)^i (t-i)^r}{i!(t-i)!}. \tag{1.11}$$

Although, we don't have the tools yet to prove this formula, it is important to recognize when we are counting set partitions. We will see more on the Stirling numbers in Section **??** where we discuss the Bell numbers. The Bell numbers are the sum over t of the Stirling subset numbers, thus count all the possible partitions of $[r]$ (with no restriction on the number of subsets in the partition).

When we were defining set partitions we stated that the order that we list the parts does not change the partition, thus we use curly braces as the outer delimiters and have that

$$\{\{6\},\{1,2,4\},\{3,5\}\} = \{\{3,5\},\{1,2,4\},\{6\}\}\,.$$

We will also run into another object of interest called *ordered set partitions*, in which a set S is partitioned into disjoint parts S_i whose union is the set, but now in contrast to the standard set partition, the order of the parts matters. For ordered set partitions we use parentheses as the outer delimiters and we have that:

$$(\{6\},\{1,2,4\},\{3,5\}) \neq (\{3,5\},\{1,2,4\},\{6\})\,.$$

Now we can use an equivalence relation argument to relate the number of standard set partitions with t nonempty parts to the number of ordered set partitions with t nonempty parts. We claim that permuting the parts of ordered set partitions is an equivalence relation and that each equivalence class has exactly $t!$ members. The cannonical form of a standard set partition corresponds to exactly one of these permutations. Thus

$$\left\{ \begin{matrix} r \\ t \end{matrix} \right\} = \text{ the number of standard set partitions of an } r\text{-set into } t \text{ nonempty parts, and}$$

$$t! \left\{ \begin{matrix} r \\ t \end{matrix} \right\} = \text{ the number of ordered set partitions of an } r\text{-set into } t \text{ nonempty parts.}$$

Ordered objects such as permutations and ordered set partitions arise in the enumeration of various types of functions. Function enumeration provides another classical group of enumeration problems, since many problems can be rephrased in terms of functions. Let's take a look.

Enumerating Functions. A *function* f from a set X to a set Y is an association such that for each $x \in X$ there is a particular $y \in Y$ associated to x by the function. We write $f(x) = y$ and call x the *pre-image* of y and y the *image* of x. A function $f : X \to Y$ is called *one-to-one* or *injective* if all the images of f are distinct; that is, if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$. A function $f : X \to Y$ is called *onto* or *surjective* if every element of Y has a pre-image; that is, for each $y \in Y$ there exists an $x \in X$ such that $f(x) = y$. Each function is equivalent to an r-long table that gives the y corresponding to $f(x)$ for each $x \in X$, where $r = |X|$ is the number of elements in X.

How many functions are there from an r-set to a t-set? In this case, for each of the r different objects $x \in X$, we can choose independently which $y \in Y$ will be $f(x)$. Thus, there are t^r possible functions from an r-set to a t-set.

How many one-to-one functions are there from an r-set to a t-set? In this case the list of images y_i must be distinct where x_i are the elements of X and y_i the associated images $y_i = f(x_i)$. This is equivalent to choosing an ordered list $(y_1, y_2, \ldots y_r)$ of $r = |X|$ elements from Y in which all the elements must be distinct. Thus, there are $P(t, r)$ one-to-one functions from an r-set to a t-set. And as we would expect $P(t, r) = 0$ if $t < r$.

How many onto functions are there from an r-set to a t-set? Now we must have one or more pre-images of every element y. For convenience, let $X = [r] = \{1, 2, \ldots, r\}$ and $Y = [t] = \{1, 2, \ldots, t\}$. To describe the function, we need to specify sets S_i where $x \in S_i$ if $f(x) = i$. The onto restriction means that none of the S_i are empty. Thus a function is completely determined by making an ordered list $(S_1, S_2, \ldots, S_t)$, where each S_i is a nonempty subset of $X = [r]$, and each element of X appears in exactly one of the subsets S_i.

We can see that this problem is related to counting the number of ordered partitions of a set, since we can uniquely describe each onto function as a list of disjoint subsets $(S_1, S_2, \ldots, S_t)$ where S_i is the subset elements in $X = [r]$ that map to i in $Y = [t]$. Thus the number of onto functions from $[r] = \{1, 2, \ldots, r\}$ to $[t] = \{1, 2, \ldots, t\}$ is

$$t! \left\{ {r \atop t} \right\}.$$

---

§ 1.5 Probability Theory

Now that we know how many k-permutations and k-samples there are, let's look at several "typical" enumerative questions involving ordered selection, probabilities, and expectations.

There is a large variation for readers of this text in how much previous exposure they have had to probability theory, so we have included a short introduction to all the necessary concepts relating to random variables. Readers who are less familiar with random variables might also want to refer to sections on random variables, expectation, or moments in any good text on probability such as Feller's beautifully written **Introduction to Probability Theory** (see Chapter I. Samples Spaces, Chapter IX. Random Variables; Expectations in [1]).

Sample Spaces and Random Variables

A fundamental starting point underlying modern probability theory the idea of an *experiment* or observation in which some set of results/outcomes are possible, where each outcome has an associated probability. Here are a few example experiments we'll return to over the course of our discussion.

Length of Longest Runs: zeros in 20-bit binary numbers. Choose a random 20-bit binary number and record the length of the longest sequence of adjacent zeros.

Distribution of Outcomes: n flips of biased coin. Flip a biased coin n times and record the number of heads.

Collect One of Each Kind: fair dice with distinct faces. Roll a die repeatedly and record the number of rolls necessary before every value has appeared at least once. Assume that in this example, each side of the die has a different value on it and the die is fair (equally likely to land with any particular face up).

Appearance of Particular Event: selecting an ace from a deck. Select a random card from a deck and record whether the card is an ace.

Typically we have a set $\mathcal{S}$ of objects called the *sample space* and to each object $s \in \mathcal{S}$ is associated a probability prob(s). For most of our purposes the set will be finite and on the few occasions when it is infinite it will still be a discrete set (a countable set). Most typically we think of the sample space as consisting of the possible outcomes of our experiment and the probability of each element in the sample space as being the probability of that outcome in our experiment. Thus we have

$$\sum_{s \in \mathcal{S}} \text{prob}(s) = 1,$$

the sum of the probabilities over the entire sample space is 1. If every object in the set has equal probability, we call the probability distribution *flat* or *uniform*. What are the sample spaces from the list of probability experiments we just gave above?

Length of Longest Runs The sample space from this particular example is all 20-bit binary numbers and each binary number appears with equal probability ($1/2^{20}$).

Distribution of Outcomes The sample space from this example is all n-long sequences composed of **H** and **T** for heads and tails (e.g. $HHTHT \ldots TTHTH$). Since the coin is biased it has a probability p_H of turning up heads which may not be equal to $1/2$. The probability of the coin turning up tails will be $p_T = 1 - p_H$. The probability associated with a particular n-long sequence of n_H heads and $n_T = n - n_H$ tails will be $p_H^{n_H} p_T^{n_T}$.

Collect One of Each Kind The sample space is strings of values from the faces of the die, where the string ends as soon as each value appears at least once. For example, for a standard 6-sided die we would have in our sample space strings such as 156234, 6352452635431, and 3424121235232232315456. If the die has f equally likely faces, then the probability associated with any string in the sample space of length l will be $1/f^l$ (see Exercise 1.4 for hints to proving this probability distribution). Thus for the 6-sided die the probability of the string 6352452635431 would be $1/6^{13}$.

Appearance of Particular Event In this example we have several choices of how we could describe the sample space and associated probabilities (e.g. is the sample the value of the card or perhaps the value and suit of the card?). Let us choose that the samples are the individual cards including both the value of the cards $A, 2, 3, \ldots, 10, J, Q, K$ and the suit $\clubsuit, \diamondsuit, \heartsuit, \spadesuit$ and each card has probability $1/52$.

A random variable is a function from a sample space. In combinatorics we most often consider random variables which count something, and hence take on values in the nonnegative integers. The name "random variable" is somewhat confusing. The description "random" refers to the random choice of an element from the sample space. Let's consider the random variables in our continuing examples.

Length of Longest Runs In this case the value of the random variable for each 20-bit binary number is the length of the longest sequence of adjacent zeros in that binary number.

Distribution of Outcomes In this case the value of the random variable of an n-long sequence of heads and tails is n_H, the number of heads.

Collect One of Each Kind In this case the value of the random variable of a string from our sample space of roll results is the length l of that string.

Appearance of Particular Event In this case the value of the random variable for a card is 1 if the card is an ace and 0 if the card is not an ace.

Random variables that take on only the values 0 or 1, are called *characteristic functions*. This name is used because the set of elements in the sample space for which the random variable is 1 can be considered the set satisfying some characteristic or property. We will see a little further on that characteristic functions are useful for enumeration.

Random Variable Notation. Usually random variables are denoted by boldface capital letters, most frequently $\mathbf{X}$. The specific values that the random variable can take are usually denoted by lower case letters.

Thus with this notation, we can see that given a sample space S and a property P that elements $s \in S$ either have or don't have, the *characteristic function* $\mathbf{X}$ on the set S takes on the values $\mathbf{X}(s) = 1$ if s has property P and $\mathbf{X}(s) = 0$ if s does not have property P.

One place we will see characteristic functions used is in Bernoulli trials. Before we define Bernoulli trials we will need to discuss independence.

Independence. Two events A and B are said to be statistically independent if the probability that both events occur simultaneously $\operatorname{prob}\{AB\}$ is equal to the product of the probability that A occurs $\operatorname{prob}\{A\}$ and the probability that B occurs $\operatorname{prob}\{B\}$, thus

$$\operatorname{prob}\{AB\} = \operatorname{prob}\{A\} \operatorname{prob}\{B\}.$$

For example, the probability of picking the queen of hearts from a deck of cards is $1/52$, which is precisely the product of the probability of picking a queen $(1/13)$ and the probability of picking a heart $(1/4)$.

Bernoulli trials are repeated independent trials where only two outcomes are possible and the probability of each outcome remains fixed throughout the trial. The standard notation is to use

p for the "probability of success" and $q = 1 - p$ for the "probability of failure". We can define a sample space consisting of all possible n-long Bernoulli trials, denoted as a string of n characters, with S standing for success and F standing for failure. The probability associated with a particular string will just be the product of the corresponding probabilities

$$\text{prob}(SSFSFFFSS) = p \cdot p \cdot q \cdot p \cdot q \cdot q \cdot q \cdot p \cdot p.$$

Thus we see that the example we have been looking at with n biased coin flips is an example of a Bernoulli trial.

We will discuss the probability distribution that results from Bernoulli trials that is called the binomial distribution later in this chapter (in Section 1.19 starting on page 29).

Expected Value, Variance, and Standard Deviation

The *expected value* $\text{E}(\text{X})$ of a random variable X is the weighted sum[2] of the value of the random variable over the sample space $\mathcal{S}$

$$\text{E}(\text{X}) = \sum_{s \in \mathcal{S}} \text{prob}(s)\text{X}(s). \tag{1.12}$$

When the probability distribution is flat, the expected value of a random variable is just the average value. The expected value is also called the mean value.

Now let's look at an example in which we calculate an expected value.

Example 1.8 (Fixed Points of Permutations)**.** Let $\mathcal{S}_n$ be the set of all permutations of $[n]$ with the flat probability distribution $\text{prob}(\pi) = \frac{1}{n!}$ for all permutations $\pi \in \mathcal{S}_n$. We say that i is a *fixed point* of a permutation π if $\pi(i) = i$ (the value in position i of the permutation is i). For example, the permutation (361542) has no fixed points; the permutation (624351) has two fixed points, 2 and 5.

Now we'd like to know: What is the expected number of fixed points of a random permutation of $[n]$?

Let $\text{X}(\pi)$ be the number of fixed points of the permutation π. Let us calculate

$$\text{E}(\text{X}) = \sum_{\pi \in \mathcal{S}_n} \text{prob}(\pi)\text{X}(\pi).$$

For small n we can list the permutations, count their fixed points, and average the result over all $n!$ permutations. For example setting $n = 3$ we list the permutations (the fixed points are in bold face)

$$\mathcal{S}_3 = \{(\mathbf{123}), (\mathbf{1}32), (21\mathbf{3}), (231), (312), (3\mathbf{2}1)\}.$$

[2] In this book we will be considering *discrete* random variables. The formula for the expected value would contain an integral if we were considering continuous random variables.

And we have that the expected number of fixed points in a random permutation of [3], which is also the average since the distribution is flat, is

$$\begin{aligned} \mathrm{E}_3 &= \mathrm{E}(\text{number of fixed points in a permutation of } [3]) \\ &= \sum_{\pi \in \mathcal{S}_3} \frac{1}{3!}(\text{number of fixed points in } \pi) = \frac{3+1+1+0+0+1}{3!} = 1. \end{aligned}$$

For general n, the expectation of the number of fixed points in a random permutation of $[n]$ is

$$\begin{aligned} \mathrm{E}_n &= \mathrm{E}\,(\text{number of fixed points in a permutation of } [n]) \\ &= \sum_{\pi \in \mathcal{S}_n} \frac{1}{n!}(\text{ number of fixed points in } \pi) \\ &= \sum_{\pi \in \mathcal{S}_n} \frac{1}{n!} \sum_{i=1}^{n} \chi_i(\pi), \qquad \text{where } \chi_i(\pi) = \begin{cases} 1 & \text{if } i \text{ is a fixed point of } \pi \\ 0 & \text{otherwise} \end{cases} \\ &= \frac{1}{n!} \sum_{i=1}^{n} \sum_{\pi \in \mathcal{S}_n} \chi_i(\pi) \\ &= \frac{1}{n!} \sum_{i=1}^{n} \left(\begin{array}{c} \text{the number of permutations of } [n] \\ \text{that have } i \text{ as a fixed point} \end{array} \right). \end{aligned}$$

How many permutations of $[n]$ have i as a fixed point? To create all the permutations with i as a fixed point, we simply fix i in place and make all permutations of the remaining $(n-1)$ elements. Thus there are $(n-1)!$ permutations with i as a fixed point, giving us that

$$\mathrm{E}_n = \frac{1}{n!} \sum_{i=1}^{n} (n-1)! = \frac{1}{n!} n(n-1)! = 1,$$

thus proving that the expected number of fixed points of a permutation is 1.

The fact that the expected number of fixed points is independent of the size of the permutation often surprises people.

Characteristic Functions: The function χ_i used in the calculation for the expected number of fixed points of a permutation above is an example of a characteristic function. For a given set S and a property P of elements in the set, a *characteristic function* $\chi_P : S \to \{0, 1\}$ has the value 1 for each $s \in S$ that has the property P and 0 for each $s \in S$ that does not have the property P.

$$\chi_P(s) = \begin{cases} 1 & \text{if } s \text{ has property } P \\ 0 & \text{if } s \text{ does not have property } P \end{cases}$$

Characteristic functions often show up in proofs right before a switch in the order of summation occurs. For example, in the calculation for the expected number of fixed points of a permutation, $\chi_i(\pi)$ shows up right before a switch in the order of summation over the permutations π and the position of the fixed point i.

Linearity of Expectation: If a random variable $\mathbf{Z}$ is the sum

$$\mathbf{Z} = \mathbf{X} + \mathbf{Y}$$

of two random variables $\mathbf{X}$ and $\mathbf{Y}$, then the expectation $E(\mathbf{Z})$ of $\mathbf{Z}$ is the sum of the expectation of $\mathbf{X}$ and the expectation of $\mathbf{Y}$. That is,

$$E(\mathbf{Z}) = E(\mathbf{X}) + E(\mathbf{Y}) \quad for \mathbf{Z} = \mathbf{X} + \mathbf{Y}. \tag{1.13}$$

The proof of linearity of the expectation follows directly by substituting in the definition of expectation and regrouping terms. (Which is worthwhile writing out once for students new to the concept and is included as the first part of Exercise 1.23 on page 46).

Fixed Points of a Permutation, Again: Notice that in the derivation of the expected number of fixed points of a permutation on page 24, could also have been expressed in terms of the linearity of expectations. For the sample space of permutations S_n, define random variable χ_i for $i = 1, 2, \ldots, n$ to be the characteristic function

$$\chi_i(\pi) = \begin{cases} 1 & \text{if } i \text{ is a fixed point of } \pi \\ 0 & \text{otherwise} \end{cases}$$

Then the expectation $E(X_i)$ is the number of permutations with i fixed divided by the total number of permutations, and hence is

$$E(X_i) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

We set a random variable

$$X = \sum_{k=1}^{n} \chi_i$$

which we can see gives the number of fixed points for a permutation. Now we use the linearity of expectation, to get the expected number of fixed points

$$E(X) = \sum_{k=1}^{n} E(\chi_i) = \sum_{k=1}^{n} \frac{1}{n} = 1.$$

Variance and Standard Deviation: Another common topic in probability theory is how tightly values cluster around the mean $E(X)$. The *variance* of a random variable $\mathrm{Var}(X)$ is the expectation of the squared deviation of that variable from its expected value or mean. For sample space S we have

$$\mathrm{Var}(X) = \sum_{s \in \mathcal{S}} (X(s) - \mathrm{E}(X))^2 \mathrm{prob}(s) = \mathrm{E}\left((X - \mathrm{E}(X))^2\right) = \mathrm{E}(X^2) - \mathrm{E}(X)^2. \tag{1.14}$$

Any readers unfamilar with variance should try to show that the second part of this formula is equal to the first.

It might at first have seemed a more natural measure of the distance from the mean to ask for the expected value of the absolute distance from the mean, however absolute value is more complicated to calculate, so tradition has settled on the above definition for variance.

For example, say we are rolling a standard 4-sided die, which is equally likely to land on any one of its faces. Then the expected value is

$$\mathrm{E}(X) = \frac{(1+2+3+4)}{4} = 2.5$$

and the variance is

$$\mathrm{Var}(X) = \frac{((1-2.5)^2 + (2-2.5)^2 + (3-2.5)^2 + (4-2.5)^2)}{4} = 1.25.$$

The *standard deviation* $\sigma(X)$ is the square root of the variance.

Linearity of Variance for Independent Random Variables: If a random variable $\mathbf{Z}$ is the sum

$$\mathbf{Z} = \mathbf{X} + \mathbf{Y}$$

of two independent random variables $\mathbf{X}$ and $\mathbf{Y}$, then then the variance $\mathbf{X}$ and variance of $\mathbf{Y}$:

$$\mathrm{Var}(\mathbf{Z}) = \mathrm{Var}(\mathbf{X}) + \mathrm{Var}(\mathbf{Y}) \ \ for \mathbf{Z} = \mathbf{X} + \mathbf{Y}, \text{ where } \mathbf{X} and \mathbf{Y} are independent. \tag{1.15}$$

The proof of the linearity of variance for independent random variables is the second point of Exercise 1.23 on page 46.

The Mean and Variance of the Binomial Distribution

Let X_n be a random variable whose value is the number of successes after n independent Bernoulli trials, each with probability of success p. Then we can express X_n in terms the random variable for a single trial X_1

$$X_n = \sum k = 1^n X_1.$$

This is a sum of independent random variables so both expectation and variance will be linear.

The expected value of a single trial is just the probability of success p

$$E(X_1) = p,$$

and the variance of the number of successes for a single trial is

$$\mathrm{Var}(X_1) = (1-p)^2 p + (0-p)^2(1-p) = p - 2p^2 + p^3 + p^2 - p^3 = p(1-p).$$

Thus the expected number of successes for n independent Bernoulli trials carried out with probability of success p is:

$$E(X_n) = \sum k = 1^n E(X_1) = np, \tag{1.16}$$

and the variance in the number of successes is

$$\mathrm{Var}(X_n) = \sum k = 1^n \mathrm{Var}(X_1) = np(1-p). \tag{1.17}$$

§ 1.6 Unordered and Distinct Selections: Subsets or Combinations

Now we return to our task of determining the number of k selections from an $n-$set. We completed the cases when the order of selection mattered, now we want to look at the cases (repeats allowed or not allowed) when the order of selection is unimportant (the last two remaining boxes of our chart from page 6).

First let's examine the case of unordered selection where we select without replacement (BOX 3 from page 6). These selections are called the *k-combinations* (or *k-subsets*) of an n-set.

Both k-combinations and k-permutations have k distinct elements. A k-element subset composes one k-combination but is associated with $k!$ different k-permutations, one permutation for each of the possible different orderings of the elements in the subset.

Example 1.9 (Subsets Associated with Permutations)**.** Using the examples from the beginning of the chapter we associate each of the 3-subsets of 4 (listed on the top row) with the $3!$ different 3-permutations of 4 (listed below the subset) that have the same elements as the subset.

$$
\begin{array}{cccc}
\{\spadesuit, \heartsuit, \diamondsuit\} & \{\spadesuit, \heartsuit, \clubsuit\} & \{\spadesuit, \diamondsuit, \clubsuit\} & \{\heartsuit, \diamondsuit, \clubsuit\} \\
\\
(\spadesuit, \heartsuit, \diamondsuit) & (\spadesuit, \heartsuit, \clubsuit) & (\spadesuit, \diamondsuit, \clubsuit) & (\heartsuit, \diamondsuit, \clubsuit) \\
(\spadesuit, \diamondsuit, \heartsuit) & (\spadesuit, \clubsuit, \heartsuit) & (\spadesuit, \clubsuit, \diamondsuit) & (\heartsuit, \clubsuit, \diamondsuit) \\
(\heartsuit, \spadesuit, \diamondsuit) & (\heartsuit, \spadesuit, \clubsuit) & (\diamondsuit, \spadesuit, \clubsuit) & (\diamondsuit, \heartsuit, \clubsuit) \\
(\heartsuit, \diamondsuit, \spadesuit) & (\heartsuit, \clubsuit, \spadesuit) & (\diamondsuit, \clubsuit, \spadesuit) & (\diamondsuit, \clubsuit, \heartsuit) \\
(\diamondsuit, \spadesuit, \heartsuit) & (\clubsuit, \spadesuit, \heartsuit) & (\clubsuit, \spadesuit, \diamondsuit) & (\clubsuit, \heartsuit, \diamondsuit) \\
(\diamondsuit, \heartsuit, \spadesuit) & (\clubsuit, \heartsuit, \spadesuit) & (\clubsuit, \diamondsuit, \spadesuit) & (\clubsuit, \diamondsuit, \heartsuit)
\end{array}
$$

We are able to partition all the permutations into sets corresponding to the subset of elements contained in those permutations. Since we know from Theorem 1.5 that there are $P(n,k) = \frac{n!}{(n-k)!}$ k-permutations, and we know that each k-subset is associated with $k!$ different permutations we are able to conclude the following:

Theorem 1.10. *The number of k-combinations (k-subsets) of an n-set (select without replacement, order unimportant) is*

$$\frac{P(n,k)}{P(k,k)} = \frac{n!}{k!(n-k)!}.$$

This number, the number of ways to choose a k-subset from an n-set, is spoken as "n choose k" and given the notation

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

The terms $\binom{n}{k}$ are also called the *binomial coefficients* for their role in the binomial equation. We will see in the following theorem that the coefficient of $x^k y^{(n-k)}$ in the expansion of the formula $(x+y)^n$ is $\binom{n}{k}$.

Theorem 1.11 (Binomial Theorem)**.**

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{(n-k)}. \tag{1.18}$$

Proof: *(Combinatorial Style Proof) We write the binomial as a product of n copies of the term $(x+y)$:*

$$\underbrace{(x+y)(x+y)\dots(x+y)}_{n \text{ copies}} = \sum b_k x^k y^{(n-k)}.$$

We get a contribution to the coefficient b_k of the term $x^k y^{(n-k)}$ on the RHS of the equation for every way we can choose which k terms on the LHS contribute a factor of x (with the remaining $(n-k)$ terms contributing a factor of y). Thus b_k is the number of ways to pick a k-set from an n-set, $\binom{n}{k}$.
qed

The binomial theorem was first proven by Newton. We will discuss an extension of the binomial theorem to the reals (replacing integer n with a real number) in Section **??**.

Note: It follows immediately from the formula for the binomial coefficients that

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

Another way to quickly prove this identity is to notice that we have a bijection between the k-subsets of an n-set and the $(n-k)$-subsets of an n-set, given by taking a set's complement.

The method used above in the proof of Theorem 1.10 is an example of one of the basic methods for combinatorial proofs (the proof given above is on the informal side). The general idea behind a *combinatorial proof* is to use counting tricks to prove the desired result. The tricks take many forms. To prove a combinatorial identity one might cleverly show that the two sides of the desired identity count the same set of objects. To prove that a set of objects has a certain size we might count a different set of objects then show how this set relates to our goal set. In the proof of Theorem 1.10, we take objects we have already counted, and partition these objects into *equal sized parts*. Then we are able to calculate the number of k-combinations by dividing our original count by the size of the parts, $k!$, to get the number of k-combinations, $\binom{n}{k}$.

This is a good place for any readers who want a quick exercise in how to use equivalence classes to create a combinatorial proof to do Exercise 1.30 on page 48, which gives a more formal combinatorial proof of Theorem 1.10 by counting the number of k-subsets of an n-set.

Binary Vectors. It is worthwhile to look at another class of familiar objects that are enumerated by the binomial coefficients. We initially derived the equation for BOX 3 from page 6 as a way of counting the k-combinations of an n-set (Theorem 1.10). Another common class of objects that is enumerated by the binomial coefficients $\binom{n}{k}$ is the set of n-long binary strings with exactly k ones. To see that these two classes of objects (k-combinations of an n-set and n-long binary strings of density k) result in the same count, we can give a bijective mapping between the k-subsets of an n-set and the n-long binary strings with exactly k ones. The mapping is to interpret the k-subset as the set of indices in the binary string where there are ones. A one in the ith position of the binary string corresponds to element i being in the subset. For example, $\{1, 3, 4\}$ maps to $(1, 0, 1, 1, 0)$ and $\{2, 3, 5\}$ maps to $(0, 1, 1, 0, 1)$. Since the ith entry of the binary string is a 1 if and only if i is an element of the k-subset, the binary string must have exactly k ones. We also easily see that this map is a bijection, thus we conclude that there are $\binom{n}{k}$ binary strings of length n that have exactly k ones.

The binary string corresponding to a subset is also called the *characteristic vector* of the subset. One way of thinking about a characteristic vector of a subset is that it is a vector containing the table of values of the characteristic function of that subset. Remember, (or if you don't remember go look on page 24), that the characteristic function F_S of a subset S of the set R takes on the value 1 on any element $r \in S$ and the value 0 on elements of r that are not in S. So in this case our set is the numbers 1 to n and each subset S corresponds to a characteristic vector $(F_S(1), F_S(2), \ldots, F_S(n))$ where $F_S(i)$ is 1 if $i \in S$ and $F_S(i)$ is 0 if $i \notin S$.

Bernoulli Trials and the Binomial Distribution

Remember from page 22 that Bernoulli trials are repeated independent trials with only two possible outcomes (success or failure) with fixed probability of each outcome throughout the trials. If p is the "probability of success" and $q = 1 - p$ is the "probability of failure" then the probability that the outcome of a sequence of failures F and successes S is just the product of the corresponding probabilities. For example

$$\text{prob}(F, F, S, F, S, S, S) = q \cdot q \cdot p \cdot q \cdot p \cdot p \cdot p.$$

What is the probability of k successes in an n-long Bernoulli trial? Well, there are $\binom{n}{k}$ ways to choose the order of a sequence containing k successes S and $(n-k)$ failures F. So the probability that the n-long Bernoulli trial has k successes is

$$\binom{n}{k} p^k q^{(n-k)} \quad 0 \leq k \leq n. \tag{1.19}$$

The probability distribution given by equation 1.19 is called the *binomial distribution*. We showed in Equations 1.16 and 1.17 that the expected number of successes for an n-long Bernoulli trial is np and that the variance in the number of successes is $np(1 - p)$.

The Multinomial Distribution

Now we introduce a generalization of the binomial coefficients.

Multinomial Coefficients. The multinomial coefficients enumerate n-long strings with a fixed number of digits of each given type, or equivalently, ordered partitions of an n-set into parts of specified sizes (parts may be empty).

Theorem 1.12. *The number of ordered partitions $P = (S_1, S_2, \cdots, S_k)$ of an n-set S with (possibly empty) parts of size $|S_i| = r_i$ is*

$$\frac{n!}{r_1! r_2! \cdots r_k!}.$$

Proof: *We repeatedly apply Theorem 1.10. First we select which r_1 elements of the n set are to be in S_1; there are $\binom{n}{r_1}$ of these. Then we select which r_2 elements of the remaining $n - r_1$ are to be in S_2; there are $\binom{n-r_1}{r_2}$ of these. Continuing in this fashion, we find that the number of ordered partitions $(S_1, S_2, \ldots, S_k)$ is given by*

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\binom{n-r_1-r_2}{r_3}\cdots\binom{n-r_1-\cdots-r_{k-1}}{r_k} = \frac{n!}{r_1! r_2! \cdots r_k!}.$$

qed

This term is called a *multinomial coefficient* and is denoted by

$$\binom{n}{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1! r_2! \cdots r_k!}.$$

This multinomial coefficient notation is used when $k \geq 3$. For the case when $k = 2$ we have the binomial coefficients, and we revert to the standard binomial coefficient notation, where only one of the two part sizes is given and the other part size is implied.

Since binomial coefficients count the number of n-long binary strings with a fixed number of 1's and 0's, we should not be surprised to find that the multinomial coefficients have a second interpretation in terms of n-long strings with fixed numbers of each kind of digit. For example, we might ask how many 6-digit ternary numbers have exactly three 0's, two 1's, and one 2.

Corollary 1.13. *The number of n-long strings of digits with exactly r_i digits that are "i" for $i = 1, 2, \ldots, k$ where $r_1 + r_2 + \cdots + r_k = n$ is*

$$\binom{n}{r_1, r_2, \ldots, r_k}.$$

Proof: *We'll give a bijection between strings having r_i digits that are "i" and ordered partitions. Given an ordered partition $(S_1, S_2, \ldots, S_k)$ of the set*

$$S = \{1, 2, \ldots, n\} = S_1 \cup S_2 \cup \cdots \cup S_k$$

where $|S_i| = r_i$, we construct a string $(d_1, d_2, \ldots, d_n)$ where $d_j = i$ if $j \in S_i$.

Conversely, given a string of digits

$$d_1, d_2, \ldots, d_n$$

with r_i digits that are "i" we construct an ordered partition $(S_1, S_2, \ldots, S_k)$ of

$$S = \{1, 2, \ldots, n\} = S_1 \cup S_2 \cup \cdots \cup S_k$$

by setting S_i to be the set of indices j where $d_j = i$. This establishes a 1–1 correspondence between the number of ordered partitions of S into k parts (some possibly empty) as specified by Theorem 1.12, and n-long strings of digits with exactly r_i digits that are i for $i = 1, 2, \ldots, k$, proving the corollary.

qed

The *multinomial distribution* generalizes the binomial distribution. It describes a situation in which multiple independent trials are carried out which have k possible outcomes, where the i-th outcome has fixed probability p_i. Let n be the total number of trials and x_i be the number of outcomes of type i during the trials. We have that the sum of the probabilities is one, $\sum_{i=1}^{k} p_i = 1$, and that the sum of the trials is n, $\sum_{i=1}^{k} x_i = n$. The number of different orderings of a trial sequences with this specific set of counts of each type $\{x_i\}$, is the multinomial coefficient

$$\binom{n}{x_1 x_2 \cdots x_k} = \frac{n!}{x_1! x_2! \cdots x_k!}.$$

A particular sequence with the specific set of counts of each type, $\{x_i\}$, appears with the product of the probabilities of each outcome, thus is:

$$\prod_{i=1}^{k} {p_i}^{x_i}.$$

Thus, the probability that the number of results of type i is x_i for $i = 1, 2, \ldots, k$ after n independent trials of fixed probabilities p_i, is:

$$\binom{n}{x_1 x_2 \cdots x_k} \prod_{i=1}^{k} {p_i}^{x_i}. \tag{1.20}$$

This is called the multinomial distribution. For those who have had a statistics class or are familiar with the terminology, this equation gives the *probabilty mass function* for the multinomial distribution..

We see that when there are only two kinds of outcome ($k = 2$), the multinomial situation describes a set of n Bernoulli trials with probability $p = p_1$ of an successful outcome and probability $p - 1 = p_2$ of failure. Equation 1.20 reduces to the binomial distribution we first saw in Equation 1.19 and again in more detail in Section **??**.

We can use what we know about binomial distributions to solve for values of expectation and variance in the multinomial distribution. We showed that for a random variable X that gives the number of successes in a n Bernoulli trials with probability of success p, that the expectation of and variance of X are

$$E(X) = np \qquad \mathrm{Var}(X) = np(1-p).$$

Thus in the case of the multinomial distribution we can look at a random variable X_i that gives the number of occurances of outcome i where outcome i appears with probability p_i and call this outcome i a "success" and all the other outcomes as "failure" which occur with probability $\sum_{j \neq i} p_j = 1 - p_i$. Then we have reduced to the binomial distribution and can conclude that the expectation of and variance of X_i are

$$E(X_i) = np_i \qquad \mathrm{Var}(X) = np_i(1-p_i).$$

Examples with Binomial and Multinomial Coefficients

Let's practice what we have learned about binomial and multinomial coefficients on a few familiar examples.

Example 1.14 (Poker and Bridge Hands)**.** Each card in a standard deck of 52 cards is uniquely labeled by one of the symbols {2–10, J, Q, K, A} together with one of the four suits {clubs, diamonds, hearts, spades}.

1. A bridge hand consists of 13 cards, order unimportant. Hence there are

$$\binom{52}{13} = 635,013,559,600$$

possible bridge hands.

2. At a bridge table, each of four players sits along one of the sides of a square table and is dealt 13 cards. The number of different bridge situations is given by

$$\binom{52}{13, 13, 13, 13} = (5.3645\ldots)10^{28}.$$

(If it takes one minute to deal a round of bridge, it would take over 10^{23} *years* to deal all the possible situations. It is probably safe to say that humankind will never get a chance to play all the possible bridge hands!)

3. In poker, five cards are dealt to each player (after considerable haggling about the rules of betting, the wild cards, etc.). *What is the probability that a poker hand contains precisely one pair of identically numbered cards*? First we count the total number of hands that contain exactly one pair. There are 13 choices for the numerical value of the pair. Then there are

$$\binom{12}{3}$$

ways of completing the hand numerically. Now we must assign the suits to the cards. There are $\binom{4}{2}$ ways of selecting the suits for the matching pair (we must sample from the suits without replacement, order unimportant) and four choices for each of the suits of the remaining three cards. This gives a total of

$$13 \cdot \binom{12}{3} \cdot \binom{4}{2} \cdot 4^3 = 1098240.$$

Dividing this value by $\binom{52}{5}$ (the number of possible poker hands) we get the probability of precisely one pair to be **0.423**.

4. Multinomial coefficients are very useful in counting the number of words that can be created with fixed letter counts. For example, "in how many ways can the word *Mississippi* be rearranged?" *Mississippi* contains four distinct letters $\{$i, m, p, s$\}$ that appear with frequencies $\{4, 1, 2, 4\}$, respectively. So *Mississippi* is but one ordering of a set of 11 objects of which there are 4 of one type, 1 of a second type, 2 of a third type, and 4 of a fourth type. The total number of orderings is then

$$\binom{11}{4, 1, 2, 4} = 11550.$$

Hypergeometric Distribution

Given a fixed size overall population N which consists of individuals in two groups of size K_1 and K_2, we take n samples from the overall population without replacement, and find that our sample consists of k_1 individuals from group 1 and k_2 individuals from group 2. Thus we have $K_1 + K_2 = N$ and $k_1 + k_2 = n$. We can count all the possible samples that have k_1 objects from group 1 and k_2 objects from group 2, by multiplying the number of ways to select k_1 things from K_1 things, times the number of ways to select k_2 things from K_2 things. It is also easy to count the total number of ways to pick n individuals from N individuals. The ratio of these two counts gives us the probability that our random sample of n things from the N set resulted in k_1 of the first group and k_2 of the second:

$$\frac{\binom{K_1}{k_1}\binom{K_2}{k_2}}{\binom{N}{n}}. \tag{1.21}$$

Equation 1.21 is called the *hypergeometric distribution* (by statisticians it would be called the probability mass function for the hypergeometric distribution).

The binomial distribution and its generalization the multinomial distribution are often confused with the hypergeometric distribution and its generalization the multivariate hypergeometric distribution. In one manner, confusing these distributions is surprising, because the sampling for binomial/multinomial is independent, whereas sampling for hypergeometric/multivariate hypergeometric is not. On the other hand, confusing the two distributions is not surprising since the binomial/multinomial distributions have the essentially the same expected value as the hypergeometric/multivariate hypergeometric, and their variances are close when the number of samples is small compared the overall population N.

We won't derive the mean and variance of the hypergeometric distribution here, but state them so they can be compared to the mean and variance of the binomial distribution. Set a random variable X to be the number of objects from group 1 that were selected in our sample of n objects in this hypergeometric distribution, then the expectation of X is exactly the same as the expectation in the binomial distribution (the number of the expected number of successes in Bernoulli trials) when the probability p is set to the fraction of the population in group 1, $p = \frac{K_1}{N}$:

$$\mathrm{E}(\mathrm{X}) = n\frac{K_1}{N}$$

The variance of the hypergeometric distribution approaches the variance of the binomial distribution when the size of the overall population N is very large compared to the number of samples n and is given by

$$\mathrm{Var}(\mathrm{X}) = n\left(\frac{K_1}{N}\right)\left(\frac{K_2}{N}\right)\left(\frac{N-n}{N-1}\right)$$

which you can see approaches the binomial variance $p(1-p)n$ by substituting in $p = \frac{K_1}{N}$ and $(1-p) = \frac{K_2}{N}$ and noticing that the last fraction $(N-n)/(N-1)$ is close to one when $n \ll N$.

§ 1.7 Unordered Selection with Replacement, Multisets

The case where we make an unordered selection but do so with replacement, is perhaps the most interesting of all. This was BOX 4 from our chart on page 6. These are called the *k-multisets of an n-set*. In some texts k-multisets of an n-set are called unordered k-selections of an n-set with replacement allowed. There are also texts in which multisets are called *distributions*.

We will use curly braces as delimiters on multisets as well as sets, but try to make it clear in the text that we are discussing multisets.

Example 1.15 (Rolls of Two Dice)**.** What are the possible rolls of two dice?
We can view this as selecting twice, with replacement, from the set $[6] = \{1, 2, 3, 4, 5, 6\}$ where order is unimportant. These are the 2-multisets of a 6-set. This example is small, so we can explicitly construct all the solutions:

{1,1}	{1,2}	{1,3}
{1,4}	{1,5}	{1,6}
{2,2}	{2,3}	{2,4}
{2,5}	{2,6}	{3,3}
{3,4}	{3,5}	{3,6}
{4,4}	{4,5}	{4,6}
{5,5}	{5,6}	{6,6}

These are not equally likely! The multiset $\{a, a\}$ occurs with probability $\frac{1}{36}$, while $\{a, b\}$ where $a \neq b$ occurs with probability $\frac{2}{36}$. As a quick check, notice that $6\frac{1}{36} + 15\frac{2}{36} = 1$. It may come as a surprise that the number of solutions, twenty-one, is the binomial coefficient $\binom{7}{2}$.

That, in general, the number of k-multisets of an n-set are binomial coefficients is rather remarkable, and we will show that:

Theorem 1.16. *The number of k-multisets of an n-set is*

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

There are many proofs of this result, notably one using generating functions (the subject of Chapter **??**), but perhaps the best known are the following two proofs using bijections.

Proof: **(by bijective mapping to k-subsets of an (n+k-1)-set)**
Assume without loss of generality that $[n] = \{1, 2, \ldots, n\}$ is the n-set. Since order in our selection is unimportant, rearrange the k elements into increasing order so that

$$a_1 \leq a_2 \leq \cdots \leq a_k.$$

Now construct the following subset of $\{1, 2, \ldots, n+k-1\}$:

$$\{a_1 + 0, a_2 + 1, a_3 + 2, \ldots, a_k + k - 1\}.$$

This construction induces a 1–1 mapping between k-subsets of $\{1, 2, \ldots, n+k-1\}$ and unordered k-selections, with replacement, of $\{1, 2, \ldots, n\}$. By Theorem 1.10, the number of such objects is

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

qed

Proof: **(by bijective mapping to (n+k-1)-long binary strings with k 0's)**
This proof has a visual trick. To show the bijection between $(n+k-1)$-long binary strings with $(n-1)$ ones and k zeros and k-multisets of an n-set, start by writing the binary string inside a pair of parentheses. For example:

$$(0\,0\,1\,0\,0\,0\,1\,1\,0\,1)\,.$$

Now we can make this string look like a series of balls in bins by thinking of the 1's and parentheses as barriers between bins and the 0's as balls:

(◯◯|◯◯◯| |◯|).

We will have n bins and k balls. Numbering the bins from the left, we count the number of balls b_i in bin i. Then we construct a multiset that has exactly b_i i's. Visually, one can think of this as labeling the balls in bin i with an i.

In our example above, $b_1 = 2$, $b_2 = 3$, $b_3 = 0$, $b_4 = 1$, and $b_5 = 0$, giving the multiset

$$\{1, 1, 2, 2, 2, 4\}.$$

This procedure gives us the desired bijection between $(n + k - 1)$-long binary strings with $(n - 1)$ 1's and k 0's and k-multisets of an n-set. There are more examples in Figure 1.5 of this bijection. qed

$\{1,2,3,3,3,3,4,4,4\}$ — (0,1,0,1,0,0,0,0,1,0,0,0)

$\{1,1,1,1,1,2,4,4,4\}$ — (0,0,0,0,0,1,0,1,1,0,0,0)

$\{3,3,3,4,4,4,4,4,4\}$ — (1,1,0,0,0,1,0,0,0,0,0,0)

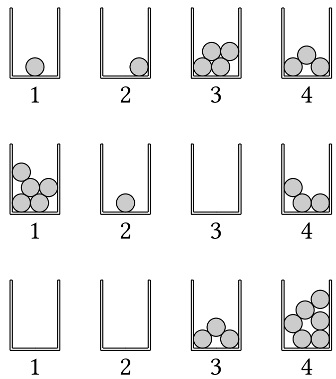


Figure 1.5: Examples with $k = 9$ and $n = 4$ showing the bijection between: k-multisets of an n-set (on the left), distributing k unlabeled balls into k distinctly labeled bins (in the center), and binary strings of length $n + k - 1$ with exactly n zeros and $k - 1$ ones (on the right).

Thus the number of ways to select a k-multiset from an n-set (BOX 4 from page 6) is

$$\binom{k+n-1}{k}.$$

The second proof of this result gives us the following corollary:

Corollary 1.17. *The number of ways to distribute k unlabeled balls into n distinctly labeled bins is*

$$\binom{k+n-1}{k}.$$

Example 1.18 (Multiset)**.** In a throw of r standard six-sided dice, how many distinct rolls are there?
In the example on p. 34 we showed that there were 21 rolls when $r = 2$. Now we are able to solve the general case using Theorem 1.16, giving us that there are

$$\binom{r+6-1}{r} = \binom{r+5}{r} = \binom{r+5}{5}$$

distinct rolls. We are making an unordered r-selection from $\{1, 2, \ldots, 6\}$ with replacement (an r-multiset of [6]).

§ 1.8 Summary of Equivalent Enumeration Problems

We can now fill in our chart from page 6 with all the ways to select k things from n things.

Figure 1.6: Selecting k objects from a n-et.

	All k Elements Distinct (Sampling Without Replacement)	**Repeated Elements Allowed (Sampling With Replacement)**
Order of Choice Important	BOX 1 k-permutation $P(n, k) = \frac{n!}{(n-k)!}$	BOX 2 k-sample (ordered list) n^k
Order of Choice Unimportant	BOX 3 k-combination (subset) $\binom{n}{k}$	BOX 4 k-multiset $\binom{k+n-1}{k}$

Figure 1.7: The number of ways to select k elements from a set of n elements.

Before we summarize facts we have learned in our introduction to enumeration, let's look at several of the classical ways to phrase these and related problems. We have already looked at two classical groups of problems: the k-selections from an n-set, and various types of functions. Now we will look at counting: solutions to equations, monotonic sequences, and distributions of balls in bins. We include these examples here because they are problems that arise frequently and that have a significant intersection with what we have already solved in this chapter.

We highly recommend that as you read through the following discussion of new enumeration problems and how they relate to old ones, that you write out small examples and draw pictures in order to better understand what is going on. There is a table at the end of the chapter summarizing all the problems in this chapter that count the equivalent objects.

Enumerating Equation Solutions. Now let's consider several equations of the form

$$a_1 + a_2 + \cdots + a_s = t$$

where the number summands s and total t are fixed and we are given some restriction on the a_i's. We will usually restrict the total t and each summand a_i's to be non-negative integers and we want to count how many possible solutions $(a_1, a_2, \ldots, a_s)$ exist. Such a set of a_i's is called an *integer partition* of t.

The subject of integer partitions will be dealt with in more depth once we have introduced generating functions. We have the machinery to solve some of the simpler problems already.

Nonnegative Summands: Given t and s, how many integer valued solutions $(a_1, a_2, \ldots, a_s)$ are there to the equation

$$a_1 + a_2 + \cdots + a_s = t, \qquad \text{when } a_i \geq 0?$$

One nice equivalent representation of this problem is to count the number of ways we can write a list consisting of t 1's and $s-1$ $+$ signs. We map such a string to the equation solution by counting the number of 1's between the $i-1$st and ith $+$ sign. Here are two examples:

$$(1\,1 + 1 + \ +1\,1\,1) \longleftrightarrow 2 + 1 + 0 + 3 = 6$$

$$(+1\,1\,1\,1 + 1 + 1) \longleftrightarrow 0 + 4 + 1 + 1 = 6$$

We already know how to count all binary strings of length $t + (s-1)$ with exactly t 1's, thus we know that there are

$$\binom{t+s-1}{t}$$

solutions to $a_1 + a_2 + \cdots + a_s = t$ with nonnegative summands.

This proof is almost identical to the second proof of the number of k-multisets of an n-set given in Theorem 1.16, with the plus signs taking the place of the barriers. Thus we have:

Corollary 1.19. *The number of k-multisets of an n-set is equal to the number of integer solutions $(a_1, a_2, \ldots, a_n)$ to the equation*

$$a_1 + a_2 + \cdots + a_n = k, \qquad \textit{when } a_i \geq 0$$

and is

$$\binom{n+k-1}{k}.$$

Positive Summands: Given t and s, how many integer solutions $(a_1, a_2, \ldots, a_s)$ are there to the equation

$$a_1 + a_2 + \cdots + a_s = t, \qquad \text{when } a_i > 0?$$

We can solve this problem two easy ways. Either we convert to the last problem by setting $b_i = a_i - 1$ and counting solutions with nonnegative summands to

$$b_1 + b_2 + \cdots + b_s = t - s,$$

of which there are

$$\binom{(t-s)+(s-1)}{(t-s)} = \binom{t-1}{s-1},$$

or we can consider arrangements of t 1's and $(s-1)$ plus signs where the plus signs are only allowed to fall in the $t-1$ spaces between the 1's,

$$\underbrace{1\_1\_1\_1\_1\_1\_1\_1\_1}_{(t-1)\text{ spaces between 1's}} \quad .$$

This allows us to immediately conclude that there are

$$\binom{t-1}{s-1}$$

solutions to $a_1 + a_2 + \cdots + a_s = t$ with positive summands.

Balls and Bins. Many of the problems we have discussed can be rephrased in terms of counting the number of ways that some fixed number of balls can be placed into a fixed number of bins. There are many variants depending on whether or not the bins are labeled, the balls are labeled, some bins are empty, and some bins may have only one ball in them.

When doing this problem on the white board we usually have some not-so-helpful suggestions from students about what letters to use for the balls and bins "How about b for balls and b for bins?", "Or you could use b for balls and c for cups?", and "Why not use c for coins and b for bins?". I tried using (with a nod to Dr. Seuss) f for foxes and s for socks, but foxes are hard to draw quickly in examples. This generated another helpful student suggestion "How about b for beetles and b for bottles?"

We shall settle on using r for rocks, which are easy to draw, and s for socks.

Here is a summary of the results:

Notice from the comments in the table, that Chapter **??**: Inversion will allow us to solve several of the problems that were too difficult for this chapter: how to count the number of necklaces (repetition allowed) and how to count partitions (the Stirling subset numbers), and how to count onto functions, which we noted in this chapter were related to the number of partitions.

We can also see that we picked to count something with relatively little structure (a set is just a pile of stuff, and an ordered list just puts it in a simple line). We got a little more structured by looking at partitions and circular arrangements, but there are much more complicated objects that are useful to enumerate, such as chemical isomers (trees) and arrays (Latin squares and designs).

We need to get some more tools into our tool belts before tackling the next set of problems. We'll see more on binomial coefficients and the binomial theorem when we introduce generating functions.

Number of ways to distribute r rocks into s socks					
		s labeled socks		s unlabeled socks	
		allow empty socks	no empty socks ($r \geq s$)	allow empty socks	no empty socks ($r \geq s$)
r labeled rocks	allow multiple rocks in socks	s^r	$s! \left\{ r \atop s \right\}$	$\sum_{m=1}^{s} \left\{ r \atop m \right\}$	$\left\{ r \atop s \right\}$
	at most 1 rock in each sock ($r \leq s$)	$P(s, r)$	$s!$	1 if $r \leq s$, 0 if $r > s$	1 if $r = s$, 0 if $r \neq s$
r unlabeled rocks	allow multiple rocks in socks	$\binom{s+r-1}{r}$	$\binom{r-1}{s-1}$	integer partitions of r into at most s parts	integer partitions of r into exactly s parts
	at most 1 rock in each sock ($r \leq s$)	$\binom{s}{r}$	1 if $r = s$, 0 if $r \neq s$	1 if $r \leq s$, 0 if $r > s$	1 if $r = s$, 0 if $r \neq s$

Figure 1.8: This table enumerates the ways to distribute rocks in socks under various conditions.

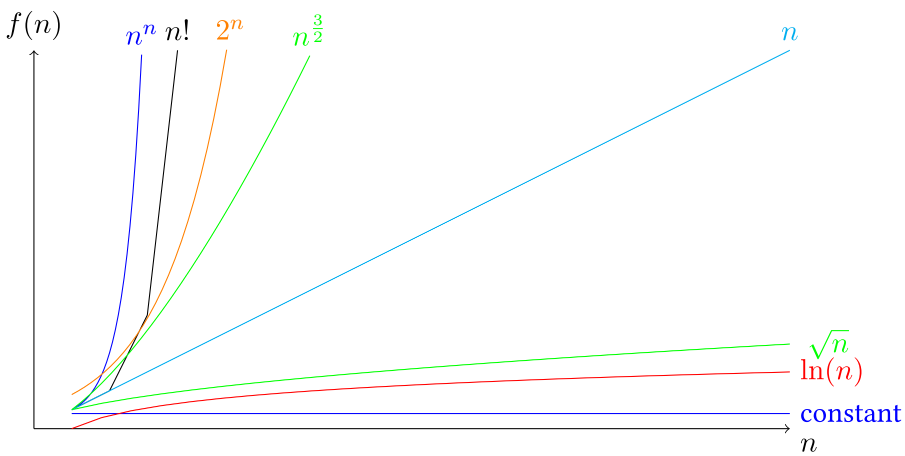
Enumeration Summary	
Ordered lists of k things select from n things, with repetition allowed.	n^k
Ordered lists of k things selected from n things, with no repetition allowed.	$P(n,k) = \frac{n!}{(n-k)!}$
Subset of k things selected from n things.	$\binom{n}{k}$
Multiset of k things selected from n things.	$\binom{n+k-1}{k}$
Circular permutation of k things selected from n things.	$\frac{P(n,k)}{k}$
n-long binary strings with exactly k ones.	$\binom{n}{k}$
Distribution of k unlabeled objects into n distinctly labeled bins.	$\binom{n+k-1}{k}$
Rolls of k n-sided dice (multiset with k face values)	$\binom{n+k-1}{k}$
Necklaces of length n chosen from r-colors of beads with repetition allowed.	Solution in Chapter **??** on Inversion.
Functions from $[r]$ to $[t]$.	t^r
Bijective functions from $[r]$ to $[t]$.	$P(t,r)$
Stirling subset numbers, defined as partitions of a subset of size r into s parts from Equation **??** in Chapter **??** on Inversion	$\left\{ {r \atop s} \right\} = \sum_{i=0}^{s} \frac{(-1)^i (s-i)^r}{i!(s-i)!}$
Onto functions from $[r]$ to $[t]$.	$t! \left\{ {r \atop t} \right\}$
Solutions to $a_1 + a_2 + \cdots + a_k = n$ with nonnegative summands.	$\binom{n+k-1}{n}$
Solutions to $a_1 + a_2 + \cdots + a_k = n$ with positive summands.	$\binom{n-1}{k-1}$

§ 1.9 Exercises

Exercise 1.1 (+): The definition of a partition was given as follows. A set S has a *partition* $P = \{S_1, S_2, \cdots, S_t\}$ if the S_i's are subsets of S ($S_i \subseteq S$) that are disjoint ($S_i \cap S_j = \emptyset$) for $i \neq j$ and whose union is the whole set ($S_1 \cup S_2 \cup \cdots \cup S_t = S$). The subsets S_i are called the *parts*, and unless otherwise specified, parts are nonempty. The order of the parts does not change the partition. Show that the parts of a partition allow us to define an equivalence relation on S where each part is an equivalence class. A set S has an *ordered partition* $P = (S_1, S_2, \cdots, S_t)$ if the S_i's are subsets of S ($S_i \subseteq S$) that are disjoint ($S_i \cap S_j = \emptyset$) for $i \neq j$ and whose union is the whole set. Now the order of the parts does change which ordered partition we are looking at. Show that there are exactly $t!$ ordered partitions for each ordered partition.

▷

Exercise 1.2 (+): Asymptotic Growth Rates



It is good to have a general sense of the relative growth of functions of n as n gets large. As in the figure above, it may be the case that for small or moderate values of n that $f(n)$ is greater than $g(n)$, but beyond some point $g(n)$ is much larger than $f(n)$. Let the notation $f(n) \prec g(n)$ mean $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$

For example, using $\prec$ we can make the chain of relationships

$$3 \prec 2 + n \prec n^4 \prec (.02)n^5.$$

If ε and c are arbitrary constants with

$$0 < \varepsilon < .1, \text{ and } 1 < c$$

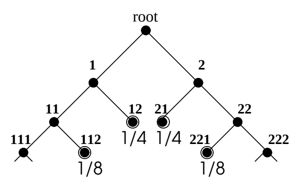
arrange the following functions into a chain of relationships.

n^c, 3^{-n}, c^2, $n^{\log n}$, ε^n, n^{8n}, $\log n$, c^{5n}, $\log\log n$, n^{ε}, and 2^{c^n}.

NOTE on "little-o": We write $f(n) \in o(g(n))$ and say "$f(n)$ is little-o of $g(n)$" if $g(n)$ grows much faster than $f(n)$, or more formally, if for every positive constant C there exists a constant N such that $|f(n)| \leq C|g(n)|$ for all $n \geq N$. If $g(n)$ becomes nonzero above some point then the relation $f(n)$ is little-o of $g(n)$ is equivalent to relation used in this problem: $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$.

Exercise 1.3 (): List the following functions in increasing order in terms of their asymptotic growth rates as $n \to \infty$, (for example $5 < 2^n < 1.4^{n^3}$):
$n^{\ln n}$, $\ln n$, 10, n^{2n}, $n^{\frac{1}{2}}$, 2^{3^n}, 5^n, $\ln \ln n$, n^6, and $e^{\sqrt{\ln n \ln \ln n}}$.

Exercise 1.4 (): On page 20 we describe a probability experiment in which a fair f-sided die with distinct values on all of its faces is rolled until every face value has appeared at least once and then the number of rolls is recorded. We claimed on page 21, that a sequence of length l in the sample space has associated probability $1/f^l$. As a warm up problem consider what happens if $f = 2$ and create a tree of options showing the growth of sequences and the associated probabilities. Circled nodes correspond to sequences in the sample space. After the second roll of a die, we are



equally likely to see the sequence 11, 12, 21, or 22. Two of the sequences are in the sample space (12 and 21) so we record the probability of $1/4$ for each of these and continue to roll again for the incomplete sequences. How would you argue that the sum of all the associated probabilities in our sample space is 1? One way to make the arguments in this problem exact is to take **all** sequences of length n and make an equivalence class of any sequences that start with the same sequence from our sample space (e.g. $\underline{2}1111 \equiv \underline{2}1112 \equiv \underline{2}1121 \cdots$).

Exercise 1.5 (+): State the definition of a Bernoulli trial. Give an example of a real life problem that can be described as a Bernoulli trial.

Exercise 1.6 (+):

$$\{(123), (132), (213), (231), (312), (321), (111), (222), (333)\}$$

in which each each triplet has associated probability $1/9$. Now consider the three random variables associated with the values in the first, second, and third positions respectively i.e. $X_1(abc) = a, X_2(abc) = b, X_3(abc) = c$. Which of the pairs of random variables are pairwise independent? Are X_1, X_2, and X_3 mutually independent?

Exercise 1.7 (+): **Poker hands:** Calculate the probability of getting the following poker hands: a full house, a flush, and a straight flush. A full house (e.g. K♡, K♢, K♠, 3♡, 3♣) consists of three cards of one value and two cards of another value. A flush (e.g. Q♠,9♠, 8♠, 5♠,4♠) is a set of five cards that are all in one suit and not all in one continuous sequence. A straight flush (e.g. J♡, 10♡, 9♡, 8♡, 7♡) is a set of five cards that are all in one suit and in a continuous sequence (the Ace can be high or low, but not in the middle of a sequence).

Exercise 1.8 (++): **Boolean Functions:** The set of n-long binary vectors is denoted by $\mathbb{B}^n$. For example $\mathbb{B}^1 = \{0, 1\}$ and $\mathbb{B}^2 = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. A *Boolean function* f on n variables maps

$$f : \mathbb{B}^n \to \mathbb{B}^1.$$

The *truth table* for such a Boolean function is the table of the outputs for all the 2^n possible inputs and thus completely determines the function.

(a)How many Boolean functions on n variables are there?

(b)A *self dual* Boolean function is one for which the truth table is unaffected if all the 1's and 0's in the domain are interchanged (that is, $f(x) = f(\overline{x})$). How many self dual Boolean functions on n variables are there?

(c)A *symmetric* Boolean function is one for which the truth table is unaffected if one arbitrarily permutes the variables (that is, $f(x) = f(\pi(x))$). How many *symmetric* Boolean functions are there?

Answer

Answer 1:

(a)A Boolean function on n variables is determined by a 2^n-long ordered list of zeros and ones. Therefore the total number of possibilities is 2^{2^n}.

(b)The fact $f(x) = f(\bar{x})$ implies that the 2^{n-1} function values where the first input bit is one are determined by 2^{n-1} function values when this bit is zero. We consider this a Boolean function on $n-1$ variables, so the answer using (a) is $2^{2^{n-1}}$.

(c)For a symmetric function, the function value only depends on the number of ones in the input, which varies from 0 to n. This is an ordered list of length $(n+1)$, so the number of possibilities is 2^{n+1}.

Part (b) Answer 2:

(a)A Boolean function f on n-variables has a domain of size 2^n (the number of binary vectors of length n). f takes one of two possible values for each element in the domain. Therefore, the number of Boolean functions on n variables is 2^{2^n}. For $a \in \mathbb{B}^n$, let $\overline{a} \in \mathbb{B}^n$ be the vector with 0s and 1s swapped. This allows us to form a partition of $\mathbb{B}^n$ into subsets of size 2:

$$\mathbb{B}^n = \bigcup_{a \in \mathbb{B}^n} \{a, \overline{a}\}$$

Let $\sim$ be the equivalence relation defined by this partition. Then any *self-dual* Boolean function f is equivalent to a function defined on the set of equivalence classes. Since the number of such equivalence classes is $2^n/2 = 2^{n-1}$, the number of *self-dual* boolean functions is $2^{2^{n-1}}$.

(b)If a Boolean function is unaffected by an arbitrary permutation of the variables (i.e., $f(\pi(x)) = f(x)$ for all $\pi \in S_n$), then it must be the case that $f(x) = 1$ for all x, or $f(x) = 0$ for all x. Thus, there are 2 *symmetric* Boolean functions.

▷ **Exercise 1.9** (+): **Polybit Statistics:** Suppose we observe an n-long binary stream. To test its randomness we first compute a *monobit statistic*; that is, we count the number of ones and zeros in the sequence. For example, the string $(0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1)$ has monobit statistic

$(n_0, n_1) = (8, 6)$. Clearly there are $n + 1$ possible results, since the number of ones can take any value—although not equally likely—in $\{0, 1, 2, \ldots, n\}$.
Now suppose that we divide our n-long binary stream into k-long non-overlapping polybits (assuming n is evenly divisible by k) and make counts of how many times we see every possible k-long binary string. We call the table of these counts the *polybit statistic*. For example if $k = 2$, then the string $(0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1)$ has breaks into non-overlapping polybits $(00, 10, 01, 10, 10, 10, 01)$ and the polybit statistic is $(n_{00}, n_{01}, n_{10}, n_{11}) = (1, 2, 4, 0)$.
How many possible tables of polybit statistics are possible given n and k (assuming n is divisible by k)?
Of course, we don't expect each polybit statistic to be equally likely. What is the expected value of the polybit statistic for random binary data? If you are having trouble doing this for general n and k solve the problem for a specific n and k that keeps the calculation reasonably small.

Exercise 1.10 (+): **Motivations:** Describe a real life combinatorics problem. The amount of detail can be anything from a paragraph to several pages. This can be a solved or an unsolved problem. If you have information about the solution, mention that as well. The problem should be something that either you are interested in yourself, or you think might be of general interest to other students. If you can't think of a problem yourself, feel free to ask some of the experienced people around you or search the web. Some general areas you might look for are: combinatorial algorithms (ways to list objects or search for objects), combinatorial structures (such as permutations, combinations, Latin squares, or designs), or graph theory and networks.

Exercise 1.11 (+): **Runs in Permutations:** If you want to determine if a string of data is random, there are various statistics you can calculate to see if the statistics for your string are unusual. A good general reference for this is [3]. One statistic that we can compute on permutations is the number of runs. Reading a permutation from left to right, a run is a maximal sequence of adjacent elements that are increasing. Thus (12345) has one run consisting of the entire permutation, and (42385617) has four runs: (4), (238), (56), and (17). (Two other references for runs are [2] and [5]; pp. 30-31 in Skiena's Implementing Discrete Mathematics is especially good if you are a Mathematica user.)

(a)Write a routine that will produce all the permutations of n for reasonably small n.

(b)Write a routine that will take a permutation and tell you how many runs are in that permutation.

(c)Make a table for small n (including at least $n = 1, 2, \ldots, 6$) and $k = 1, \ldots n$ that gives the number of permutations of $[n]$ that have exactly k runs.

(d)Calculate the expected number of runs for each value of n in your table.

(e)Use your results to guess a formula for the expected value of the number of runs in a permutation of $[n]$. Prove your formula.

Exercise 1.12 (+): Give a bijection between $k-$subsets of an $n-$set and $(n - k)-$subsets of that $n-$set that proves the identity

$$\binom{n}{k} = \binom{n}{n-k} \quad \text{for integer } n \text{ and } k \text{ with } 0 \leq k \leq n.$$

Exercise 1.13 (+): How many 12-long vectors with entries in the set $\{0, 1, 2\}$ have exactly five 0's and exactly three 1's?

Exercise 1.14 (+): How many non-negative integers less than 10^n have their digits in nondecreasing order?

Solution (Exercise 1.14) — Answer 1:
By extending to the left with zeros, we may restate this as the number of n-long strings of digits in which the digits never decrease from left to right. This in turn is the same as the number of n-multisets taken from a set of size ten, which is $\binom{n+9}{9}$.
Answer 2:

How many non-negative integers less than 10^n have their digits in nondecreasing order?
The general n-long integer has the following form:

$$000\cdots0111\cdots12\cdots23\cdots34\cdots45\cdots56\cdots67\cdots78\cdots89\cdots9$$

If we let each digit to represent one of ten possible bins, then we want a total of n objects in all the bins. We will then need $10 - 1 = 9$ dividers to distinguish among the digits. Therefore each integer is determined by choosing 9 dividers from $n + 9$ slots. The solution is $\binom{n+9}{9}$.

Exercise 1.16 (+): How many 9-digit sequences can be formed using the digits 2, 7, and 8?

Exercise 1.17 (+): Given k (standard six sided) dice, how many distinct throws are there?

Exercise 1.18 (+): How many subsets of six integers chosen without replacement from the set $[20]$ are there that contain no consecutive integers?

Exercise 1.19 (+): How many distinct 5-digit codes can be made using the digits $0 - 9$ such that the first digit is odd and the last is not 0?

▷ **Exercise 1.20** (): What is the coefficient of x^2y^3 in $(x + y)^5$?

Exercise 1.21 (++): What is the coefficient of $x_1^2x_3x_4^4$ in $(x_1 + x_2 + x_3 + x_4 + x_5)^7$?

▷ **Exercise 1.22** (++): A combinatorial identity can be given a combinatorial proof by finding a way to argue that both sides of the identity count the same thing (in two different ways). Can you think of a combinatorial proof of the following identity?

$$\binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}^2$$

▷ **Exercise 1.23** (): Show that if a random variable $\mathbf{Z}$ is the sum

$$\mathbf{Z} = \mathbf{X} + \mathbf{Y}$$

of two random variables $\mathbf{X}$ and $\mathbf{Y}$, that the expectation $E(\mathbf{Z})$ of $\mathbf{Z}$ is additive. That is,

$$E(\mathbf{Z}) = E(\mathbf{X}) + E(\mathbf{Y}).$$

If the two random variables $\mathbf{X}$ and $\mathbf{Y}$ are independent, show that the variance $var(\mathbf{Z})$ of $\mathbf{Z}$ is also additive. That is,

$$var(\mathbf{Z}) = var(\mathbf{X}) + var(\mathbf{Y}).$$

Exercise 1.24 (): The example on page 23 calculates that the expected number of fixed points of a permutation is 1, regardless of the size n of the permutation. Notice that the probability that 1 is a fixed point is just $\frac{1}{n}$, since all elements are equally likely to end up in position 1.
What is the probability that 2 is a fixed point in one of the $(n-1)!$ permutations where 1 is a fixed point? What is the probability that 2 is a fixed point in one of the $n! - (n-1)!$ permutations where 1 is not a fixed point? Do your answers fit in with the probability that 2 is a fixed point in a random permutation?
You may use the result of Problem 1.23.

Exercise 1.25 (): Suppose the successful outcome of an experiment occurs with probability p.

(a)Show that the probability of at least one success in N independent trials is

$$1 - (1-p)^N = \sum_{k \geq 1} \binom{N}{k} p^k (-1)^{k+1}.$$

(b)Show that the terms $\binom{N}{k} p^k$ are *monotonically decreasing* if and only if $Np < 1$.

(c)If $Np < 1$, show that

$$1 - (1-p)^N = Np + \mathrm{O}\left(N^2 p^2\right).$$

Exercise 1.26 (): Many states have lotteries and often the game with the biggest jackpot involves picking a sequence of numbers that matches a drawing of balls. In "Powerball" which is being played in 44 states (in 2014) the first five numbers are drawn without replacement from a set of red balls numbered 1 to 59 and the order of these balls does not matter, then a sixth ball is drawn from a separate set of blue balls numbered 1 to 35. [3] If no one matches all the numbers from one drawing then the big jackpot continues to accumulate until there is a drawing with a winner. If more than one person matches the numbers, the pot is split.

(a)If we call two tickets the same if they are the same after sorting the first five numbers into increasing order, how many different tickets are there?

(b)If you assume you don't have to share the pot and you only want to play when the expected value of the payoff is at least as big as the cost of the 1 dollar ticket, how big should the pot be before you start playing?

(c)If more than one person picks the winning number the pot is split, so the situation is more complicated. Collected statistics show that certain numbers are selected by players more often, such as numbers corresponding to people's birthdays. Assume that we have collected some statistics on what numbers lottery players select more frequently and we know that:

[3] Twenty percent of the money collected goes into the big jackpot, thirty percent into smaller jackpots, and fifty percent is kept by the state. A big jackpot winner usually pays about half of their winning to the state, thus about ten percent of the money collected gets to be kept by big jackpot winners.

a)For the first five spots, determined by the red balls numbered 1-59, lottery players choose numbers between 1 and 9 with probability $\frac{1}{18}$, and choose numbers between 10 and 59 with probability $\frac{1}{100}$.

b)For the last spot, determined by the yellow balls numbered 1-36, lottery players choose numbers between 1 and 12 with probability $\frac{1}{24}$, and choose numbers between 13 and 36 with probability $\frac{1}{48}$.

Say the number of players for the coming drawing is a million and the size of the jackpot is up to four hundred million. How does the expected payoff of someone who plays six numbers that are all from the popular list of numbers compare with the expected payoff of someone who plays six numbers that are all from the less frequently played numbers?

Exercise 1.27 (): You and agent Scully are put on a team to analyze the function of a machine found in an alien space craft. You are in charge of describing the output of one of the component "black boxes" inside the machine. Before you take this box apart physically, you observe that when you trigger the input switch on this box it outputs a list of ten numbers, each of the numbers is a four digit binary number (hence between 0 and 15). Scully is in charge of analyzing the next component of the machine, which will take the ten numbers from your box as input. She asks you whether you think that your box produces a permutation (no repeats allowed) or a sample (repeats allowed). You try triggering the input switch a number of times, and each time you observe that the ten output numbers contain no repeats (hence is a permutation), so you report that you believe your box produces a permutation.
Now she asks you how sure you are. You say that if this were producing samples, then you would only expect to see the output you saw with probability p. What is p? Say you triggered the input switch four times, and each time saw no repeats in the ten outputs. Use a calculator or a computer. Do you need to be careful about overflow errors?

Exercise 1.28 (): (Quicky) How many people must be selected to guarantee that at least three have the same first and last initials?

Exercise 1.29 (++): (a)Let $\mathbb{Z}^+$ be the positive integers. Consider the grid of points with integer coordinates in the plane

$$(x, y) \in \mathbb{Z}^+ \times \mathbb{Z}^+.$$

How many paths are there from the point $(0, 0)$ to the point (m, n), with steps consisting of moves $(0, 1)$ and $(1, 0)$?

(b)How many paths are there from $(0, 0, 0)$ to (m_1, m_2, m_3) in the three dimensional lattice $\mathbb{Z}^+ \times \mathbb{Z}^+ \times \mathbb{Z}^+$, with steps in the unit vector directions $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$?

(c)What about the k-dimensional version of this question?

Exercise 1.30 (): Verify that the k-permutations of an n set form an equivalence class under permutation of the order of the k elements in the k-permutation. Make the combinatorial style proof of Theorem 1.10 more rigorous than as it was presented in the text, by writing a short formal proof in the language of equivalence relations and equivalence classes.
HINT: Define a relation ($\equiv$) on the set of k-permutations S, that is defined as follows: if σ and $\rho \in S$, then say that $\sigma \equiv \rho$ if each k-permutation contains the same subset of k elements. Then

show that the relation ($\equiv$) is an equivalence relation and that every equivalence class has the same number of permutations as members.

Exercise 1.31 (): Remember that in a Bernoulli trial the probability of k successes in an n-long trial is:

$$\text{prob}(k; n) = \binom{n}{k} p^k (1-p)^{(n-k)} \quad 0 \leq k \leq n.$$

Use the binomial theorem to show that:

$$\sum_{k=0}^{n} \text{prob}(k; n) = 1.$$

Say we have a biased coin that comes up heads $\frac{1}{4}$-th of the time and we flip coin 100 times. What is the expected number of successes? What is the variance in the expected number of successes? What is the probability that we see at most 5 successes during our 100 trials? [See page 29 and page 22 for the definition of Bernoulli trials and of the binomial distribution. The definition of expected value is on page 23 and the definition of variance is on page 25. The expected value and variance for the number of successes in n Bernoulli trials are in Equations 1.16 and 1.17. The binomial theorem is on page 28.]

Exercise 1.32 (): Letti is given a set of three "lucky" coins by the magician at her birthday party. On the head side they have a picture of an animal (wolf, bear, hawk) and on the tail side there is a picture of a tree. When she plays with the coins later she thinks that they seem to land with the heads side up more often. She wants to test her hypothesis and does the following experiment. She tosses the three coins and records the total number of heads. She repeats this experiment a dozen times. What is the sample space in this experiment? What are the probabilities associated with elements of the space if the coins are fair? What are the probabilities associated with the space if the wolf is biased to land head up with a probability of p_w and the bear with a probability of p_b and the hawk with p_h?
The number of heads that showed up in her tests were: 2,3,1,3,2,0,2,3,3,2,1,3. Calculate the average number of heads that showed up. What would be the expected number of heads to show up in a single toss of the three coins if the coins were biased? If $p_w = .8$, $p_b = .6$, and $p_h = .4$, what is the expected value of the number of heads when the three coins are tossed?
If you were trying to give Letti a simple test to decide whether the coins are biased what would you suggest?

Exercise 1.33 (): Use Stirling's Formula to prove

$$\binom{2n}{n} \sim \frac{1}{\sqrt{\pi n}} 2^{2n}.$$

More generally, find an approximate expression for $\binom{(x+y)}{y}$ for large x and y.

Exercise 1.34 (): The fraction of k-samples of an n-set that have distinct values, that is, that are k-permutations is given by the fraction

$$\frac{P(n,k)}{n^k} = \prod_{j=0}^{k-1} \frac{n-j}{n}.$$

[A. k large or small] What happens when $k = 1$ and $k = 2$? When $n = k$, we have that the fraction of n-samples of an n-set that are n-permutations is

$$\frac{n!}{n^n}.$$

Using Stirling's Formula, show that when n is large and $n = k$ that the fraction is asymptotically

$$\frac{n!}{n^n} \sim \frac{\sqrt{2\pi n}}{e^n}.$$

What happens when $n - k = c$ where c is a small constant?

[B. $k \ll n$] It is sometimes a useful trick to use a truncated Taylor series expansion as an approximation for a function, or the function as the approximation for the first few terms of the series. For example, when x is sufficiently small we have that

$$e^x \sim 1 + x. \tag{1.22}$$

This is particularly useful when we have a product of terms, that are all close to one (or close to the same size), since we have that for small x_i that

$$\prod_i (1 + x_i) \sim e^{\sum_i x_i}. \tag{1.23}$$

If we take k samples (repetition is allowed) from an n-set, the probability that this k-sample is a k-permutation (contains no repeats) is

$$\begin{aligned} \text{prob}\,(n,k) &= \frac{\text{the number of } k\text{-permutations of an } n\text{-set}}{\text{the number of } k\text{-samples of an } n\text{-set}} = \frac{P(n,k)}{n^k} \\ &= \left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)\cdots\left(\frac{n-k+1}{n}\right) = \prod_{j=0}^{k-1}\left(\frac{n-j}{n}\right). \end{aligned}$$

Assume k is small in comparison to n. Use Equation 1.23 to get an approximation for $\text{prob}\,(n,k)$, then use Equation 1.22 to remove e from your answer. Can you replace the requirement $k \ll n$ with a more specific relationship of the form $k^a \ll n^b$ for two constants a and b which more precisely reflect what is required for your approximation for $\text{prob}\,(n,k)$ to be reasonable?

Exercise 1.35 (): The flowershop problem: A person comes into a flower a flowershop with d dollar bills and wants to buy a bouquet. Each flower costs 1 dollar. There are f distinct types of flowers (e.g. roses, daisies, tulips, etc.).

(a)How many different bouquets are possible using all d dollars?

(b)How many different bouquets are possible using at most d dollars?

(c)How many different bouquets have at least one of each distinct type of flower and have d flowers total?

Exercise 1.36 (++): Monotonic Sequences Now let's look at a few enumeration problems involving monotonic sequences.

(a)**Monotonic Sequences with Nondecreasing Terms:**
Given n and k, how many lists of integers $(a_1, a_2, \ldots, a_k)$ satisfy the inequalities

$$1 \leq a_1 \leq a_2 \leq \cdots \leq a_k \leq n?$$

For example, if $n = 3$ and $k = 3$ the solutions are

$$(a_1, a_2) = (1,1), (1,2), (1,3), (2,2), (2,3), (3,3).$$

(b)**Monotonic Sequences with Increasing Terms:**
Given n and k, how many lists of integers $\{a_1, a_2, \ldots, a_k\}$ satisfy the inequalities

$$1 \leq a_1 < a_2 < \cdots < a_k \leq n?$$

Exercise 1.37 (): Line codes are used as transmission codes. They do not to hide information, but rather help transmit the information from one place to another with fewer errors. A specific line code for fixed integers m and n (greater than m) will take all m-bit words and map them to a subset of the n-bit words, in a one-to-one fashion by an encoding map E. To send a word m-bit word W_m, it is mapped to an n-bit word $W_n = E(W_m)$ that is transmitted. The line code must also specify a decoding map D from all n-bit strings back to the smaller space of m-bit strings, because when transmission errors occur, the sender can receive a word that is not in the image of the map E. Line codes from m to n bits are designated by $mBnB$, for example the Manchester Code is a $1B2B$ Line Code that has encoding map: $0 \to 00$, $1 \to 11$ and decoding map D: $00 \to 0$, $11 \to 1$, $01 \to 1$, $10 \to 0$.

(a)How many ways are there to pick a subset S of the 2^n numbers in the bigger space, to map the smaller space of 2^m numbers if the encoding E must be 1-to-1.

(b)Once you have selected S how many encodings E are there to that particular subset S?

(c)The decoding map D, must be the inverse of E on the 2^m elements in the image of E. How many decoding maps D are there for a particular encoding E?

(d)Putting these answers together how many $mBnB$ line codes are there?

Exercise 1.38 (): Bar Code Scanners or Postal Codes
As you can see from problem **??** the number of line code transmission codes is huge. The choice of which a transmission code to use depends on man properties including requirements to keep transmission power low and ease of sychronization of the sender and receiver. The power need is reduced by having low DC balance (roughly equal number of ones and zeros). The Hamming Code of a string is the number of ones in the string, so a balanced number of ones and zeros corresponds to having Hamming Code close to $\frac{n}{2}$. The ability to synch timing between the sender and the receiver is increased by having short max run length (longest strings of consecutive zeros or ones). Bar code scanners use a transmission code that map the 10 digits to 5-bit binary string.

(a)How many 5-bit binary strings are there?

(b)How many 5-bit binary strings have Hamming Code 2 or 3? List them.

(c)Which four of these codes produce the longest length runs

Note: The teacher eventually plans to check if this is bar code scanners or postal codes and then wrap up this problem with a description of what they actually do by using odd and even code words and perhaps talk about using a majority vote decoder or a bit-wise complement assumption on the code and why you would do this.

Solutions

Solution (Exercise 1.2) — Hint: For some comparisons it is helpful to instead compare the log of both functions.

$$\epsilon^n < \frac{1}{3^n} < c^2 < \log(\log n) < \log n < n^\epsilon < n^c < n^{\log n} < c^{5n} < n^{8n} < 2^{c^n}$$

•Since $\lim_{n\to\infty} \epsilon^n = \lim_{n\to\infty} \frac{1}{3^n} = 0$, we have $\frac{1}{3^n} < c^2$.

•Since c^2 is constant and $\log(\log n)$ is increasing, $c^2 < \log(\log n)$.

•$\lim_{n\to\infty} \frac{\log(\log n)}{\log n} = \lim_{n\to\infty} \frac{1/(\log n)\cdot 1/n}{1/n} = 0$, so we have that $\log(\log n) < \log n$.

•$\lim_{n\to\infty} \frac{\log n}{n^\epsilon} = \lim_{n\to\infty} \frac{1/n}{\epsilon n^{\epsilon-1}} = 0$ implies that $\log n < n^\epsilon$.

•Since $\epsilon < 1 < c$, we have $n^\epsilon < n^c$.

•Since $\lim_{n\to\infty} \log n > c$, we have $n^c < n^{\log n}$.

•$\lim_{n\to\infty} \frac{n^{\log n}}{c^{5n}} = \lim_{n\to\infty} \frac{\log n \log n}{5n \log c} = \lim_{n\to\infty} \frac{2(1/n)\log n}{5 \log c} = \frac{2}{\log c} \lim_{n\to\infty} \frac{\log n}{n} = \frac{2}{\log c} \lim_{n\to\infty} \frac{1}{n} = 0$. Thus, $n^{\log n} < c^{5n}$.

•$\lim_{n\to\infty} \frac{c^{5n}}{n^{8n}} = 0$ since $c < n$ for sufficiently large n. Thus $c^{5n} < n^{8n}$.

•$\lim_{n\to\infty} \frac{n^{8n}}{2^{c^n}} = \lim_{n\to\infty} \frac{8n \log n}{c^n \log 2} = \lim_{n\to\infty} \frac{8+8\log n}{c^n \log c \log 2} = \lim_{n\to\infty} \frac{8/n}{c^n \log c \log c \log 2} = 0$ so $n^{8n} < 2^{c^n}$.

Solution (Exercise 1.9) — Because we don't keep track of where in the binary string we got each polybit count that contributed to a table entry, the order of the polybits in the string is unimportant. So we actually want to count the number of multisets that have n/k entries coming from the set of 2^k k-long binary strings. For example, when $k = 2$, we would be picking an $n/2$ multiset from the 4-set $\{00, 01, 10, 11\}$.
Thus the number of tables of polybit counts should be the number of ways to pick an n/k-multiset from a 2^k-set

$$\binom{2^k + \frac{n}{k} - 1}{\frac{n}{k}} = \binom{2^k - 1 + \frac{n}{k}}{2^k - 1}.$$

Solution (Exercise 1.11) — Part (e): The expected number of runs in a permutation of length n is $(n+1)/2$. Two possible proofs are given.

Proof.(Version 1) Let X be a random variable so that $X(\pi) =$ the number of run of $\pi \in S_n$. Then,

$$E(X) = \frac{1}{n!} \sum_{\pi \in S_n} X(\pi).$$

To finish the proof, we must count the total number of runs of each permutation $\pi \in S_n$.
Note that each permutation and its reverse may be paired together yielding $n+1$ total runs. Since there are $n!/2$ such distinct pairings, we can write

$$E(X) = \frac{1}{n!} \left(\frac{n!}{2}(n+1) \right),$$

and so

$$E(X) = \frac{n+1}{2}.$$

qed

Proof.(Version 2) Let R_n be the expected number of runs for a permutation in S_n. It is easily verified that $R_1 = 1$. Now we would like to show

$$R_{n+1} = R_n + \frac{1}{2}.$$

Observe that the permutation $(a_1, a_2, \ldots, a_{n+1}) \in S_{n+1}$ either has as many runs as $(a_2, \ldots, a_{n+1}) \in S_n$ or it has one more run. This depends on whether $a_1 < a_2$ or $a_1 > a_2$. Over all the permutations of n numbers, $\frac{1}{2}$ of the time $a_1 < a_2$ and $\frac{1}{2}$ of the time $a_1 > a_2$. Thus,

$$R_{n+1} = R_n + \frac{1}{2},$$

and this implies that

$$R_n = \frac{n+1}{2}.$$

qed

Solution (Exercise 1.15) — The number of sub-collections = $\sum_{k=0}^{n}$ number of sub-collections containing k of the distinct objects. So, $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

Solution (Exercise 1.16) — The number of 9-digit sequences is

$$\sum_{r_1+r_2+r_3=9} \binom{9}{r_1 r_2 r_3}.$$

This implies that

$$(1+1+1)^9 = \sum_{r_1+r_2+r_3=9} \binom{9}{r_1 r_2 r_3} 1^{r_1} 1^{r_2} 1^{r_3}$$

which gives

$$\sum_{r_1+r_2+r_3=9} \binom{9}{r_1 r_2 r_3} = 3^9 = 19683.$$

Solution (Exercise 1.17) — We are using the formula $\binom{k+n-1}{k}$ that finds the number of ways k objects can be chosen from n objects where repetition is allowed and order does not matter. If there are k 6-sided dice, $n = 6$, so the answer is $\binom{k+6-1}{k} = \binom{k+5}{k}$, assuming that each die is indistinguishable.

Solution (Exercise 1.18) — Consider the set S of all ordered 6-tuples with distinct elements from $[15]$ and the set T of all ordered 6-tuples with distinct non-consecutive elements from $[20]$ (T is the one we are interested in). Given $a = (a1, ..., a6)$ in S, the function f defined by $f(a) = (a1, a2 + 1, a3 + 2, a4 + 3, a5 + 4, a6 + 5)$ maps a onto an ordered 6-tuple from $[20]$ with no consecutive elements and thus f is well defined as a function $S \to T$. The function f is clearly one-to-one and, conversely, if $b = (b1, ..., b6)$ is in T, the 6-tuple $c = (b1, b2 - 1, b3 - 2, b4 - 3, b5 - 4, b6 - 5)$ is an element from S such that $f(c) = b$. Therefore, f is invertible and the cardinality of the set we seek (that is, $T = f(T)$) is the same as the cardinality of S, which is $\binom{15}{6}$ (trivially!).
In general, if $N(n, k)$ is the number of k-selections from $[n]$ with no consecutive integers, we have $N(n, k) = \binom{n-k+1}{k}$ (the above proof generalizes easily).

Solution (Exercise 1.19) — There are 5 choices for the first digit: $\{1, 3, 5, 7, 9\}$. There are 10 choices each for the second, third, and fourth digits. Finally, there are 9 choices for the last digit. Multiplying together gives us $5 \cdot 10 \cdot 10 \cdot 10 \cdot 9 = 45000$ distinct codes.

Solution (Exercise 1.20) — The coefficient of x^2y^3 in $(x+y)^5$ is given by the binomial coefficient $\binom{5}{2} = 10$.

▷ -

Solution (Exercise 1.21) — Answer 1:
We form each $x_1^2 x_3 x_4^4$ term by selecting two x_1's, one x_3, and four x_4 out of seven choices. The answer is thus the multinomial coefficient

$$\binom{7}{2,1,4} = \frac{7!}{2!\,1!\,4!} = 105$$

Answer 2:
Imagine all seven terms written side by side, and multiplying out by hand. How do we get a $x_1^2 x_3 x_4^4$ term? We simply choose 2 terms from the seven possible factors for the x_1^2, then choose the x_3

from one of the remaining 5 factors. After that, the x_4^4 will come from the remaining 4 factors. Thus the coefficient is

$$\binom{7}{2}\binom{5}{1} = \frac{7\cdot 6}{2}\cdot\frac{5}{1} = 105$$

Solution (Exercise 1.22) — Hint: Count the ways to pick pick an n person committee from a group consisting of n men and n women.

Second Hint: Let k be the number of number of men and remember $\binom{n}{k} = \binom{n}{n-k}$.

Answer 1:

We count the number of $2n$-long binary strings with exactly n ones. This is clearly $\binom{2n}{n}$, the left hand side. But this is also the number of ways of concatenating a n-long binary string with k ones, $0 \le k \le n$, with another n-long string with $n-k$ ones. For each k this number is $\binom{n}{k}\binom{n}{n-k} = \binom{n}{k}^2$. To get the total number we thus sum this term over all k, which is the expression on the right.

Answer 2:

The left-hand side is the coefficient of $x^n y^n$ in the expansion of $(x+y)^{2n}$. On the other hand, we can write

$$(x+y)^{2n} = [(x+y)^n]^2 = \left[\binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \cdots + \binom{n}{n}y^n\right]^2$$

In order to compute the coefficient of $x^n y^n$ in the right-hand term above, we group terms in pairs:

$$\binom{n}{0}\binom{n}{n} + \binom{n}{1}\binom{n}{n-1} + \cdots + \binom{n}{n}\binom{n}{0}$$

Using the fact that $\binom{n}{k} = \binom{n}{n-k}$, we get the right-hand side of the original equation.

Solution (Exercise 1.24) — In the $(n-1)!$ permutations where 1 is a fixed point, the probability that 2 is also fixed is $1/(n-1)$, since 2 is equally likely to assume any of the $(n-1)$ values that remain. Among the the $n! - (n-1)! = (n-1)(n-1)!$ permutations where 1 is *not* a fixed point, let π be one which fixes 2. Then $\pi(1) \ge 3$ since $\pi(1) \in \{1, 2\}$ gives a contradiction. For the $(n-2)$ values $j = 3$ to n, the number of permutations for which $\pi(1) = j$ and $\pi(2) = 2$ is $(n-2)!$, as the other values are free. We conclude that $(n-2)(n-2)!$ of these $(n-1)(n-1)!$ permutations fix 2, so the probability 2 is a fixed point when 1 is not is

$$\frac{(n-2)(n-2)!}{(n-1)(n-1)!} = \frac{(n-2)}{(n-1)^2}$$

This is consistent with the probability of $1/n$ that 2 is a fixed point, since by Bayes' theorem

$$\begin{aligned} p(\pi(2)=2) &= p(\pi(2)=2 \text{ and } \pi(1)=1) + p(\pi(2)=2 \text{ and } \pi(1)\neq 1) \\ &= p(\pi(2)=2|\pi(1)=1)\, p(\pi(1)=1) + p(\pi(2)=2|\pi(1)\neq 1)\, p(\pi(1)\neq 1) \\ &= \left(\frac{1}{n-1}\right)\left(\frac{1}{n}\right) + \left(\frac{n-2}{(n-1)^2}\right)\left(\frac{n-1}{n}\right) \\ &= \frac{1}{n(n-1)} + \frac{n-2}{n(n-1)} \\ &= \frac{1}{n} \end{aligned}$$

The expected value is always linear, regardless of dependence, by the definition:

$$\begin{aligned} E(X+Y) &= \sum_{s\in S} p(s)(X+Y)(s) = \sum_{s\in S} p(s)(X(s)+Y(s)) \\ &= \sum_{s\in S} p(s)X(s) + \sum_{s\in S} p(s)Y(s) = E(X)+E(Y) \end{aligned}$$

Thus by induction it's always true that $E(\sum_{k=1}^{n} X_i) = \sum_{k=1}^{n} E(X_i)$. Note that by precisely the same proof, the expected number of fixed points of a randomly selected function from $[N]$ to $[N]$ is also 1.

Answer 2: Let F_i be the event that a permutation $\pi \in S_n$ fixes the number i. Let N_i be the event that $\pi(i) \neq i$. Since $P(F_i) = 1/n$, we know that $P(N_i) = 1 - (1/n) = (n-1)/n$. The probability that 2 is fixed point in one of the $(n-1)!$ permutations where 1 is a fixed point is conditional probability: $P(F_2|F_1)$. Similarly, the probability that 2 is a fixed point in one of permutations where 1 is *not* a fixed point is $P(F_2|N_1)$.
We can use the definition of conditional probability to compute these:

$$P(A|B) = \frac{P(A\cap B)}{P(B)}$$

We need to compute $P(F_2 \cap F_1)$ and $P(F_2 \cap N_1)$. The number of permutations that fix both 1 and 2 is clearly $(n-2)!$, the number of permutations on the remaining $n-2$ elements. Therefore

$$P(F_2 \cap F_1) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$$

In order to compute the number of permutations that fix 2 and move 1, we imagine n slots, and insert 2 into the 2nd position. There are $n-2$ choices for slot number 1, since 2 is already taken, and 1 needs to go somewhere else. Finally, the remaining $n-2$ slots can be filled in any order with the remaining numbers, which gives $(n-2)!$. Therefore

$$P(F_2 \cap N_1) = \frac{(n-2)(n-2)!}{n!} = \frac{n-2}{n(n-1)}$$

Now we can compute the conditional probabilities:

$$P(F_2|F_1) = \frac{P(F_2 \cap F_1)}{P(F_1)} = \frac{1}{n(n-1)} \cdot \frac{n}{1} = \frac{1}{n-1}$$

$$P(F_2|N_1) = \frac{P(F_2 \cap N_1)}{P(N_1)} = \frac{n-2}{n(n-1)} \cdot \frac{n}{n-1} = \frac{n-2}{(n-1)^2}$$

For completeness, let's go ahead and compute the conditional probabilities $P(N_2|F_1)$ and $P(N_2|N_1)$. By symmetry, we can assume that $P(N_2 \cap F_1) = P(F_2 \cap N_1)$; that is, the probability that 1 is fixed and 2 is not fixed is precisely the same as if 1 were not fixed and 2 were fixed. Therefore

$$P(N_2|F_1) = \frac{P(N_2 \cap F_1)}{P(F_1)} = \frac{n-2}{n(n-1)} \cdot \frac{n}{1} = \frac{n-2}{n-1}$$

In order to compute $P(N_2|N_1)$, we need to compute $P(N_2 \cap N_1)$, the probability that neither 1 nor 2 is fixed. Let $\pi \in S_n$ be a random permutation with $\pi(1) \neq 1$ and $\pi(2) \neq 2$. Let's divide all such permutations into two groups, depending on whether $\pi(1) = 2$ or $\pi(1) \neq 2$. If $\pi(1) = 2$, then all the numbers, including 2, are free to roam among the remaining $n-1$ slots. This gives us $(n-1)!$ such permutations. In the second case, since $\pi(1) \neq 2$, there are only $n-2$ choices for 1. After 1 is determined, there are then $n-2$ choices for 2: neither slot 2, nor the slot taken by 1, are available, and these are distinct in this case. There are $(n-2)!$ choices for the remaining numbers. Adding together both cases, we get

$$P(N_2 \cap N_1) = \frac{(n-1)! + (n-2)^2(n-2)!}{n!} = \frac{n^2-3n+3}{n(n-1)}$$

We can now compute the final conditional probability:

$$P(N_2|N_1) = \frac{P(N_2 \cap N_1)}{P(N_1)} = \frac{n^2-3n+3}{n(n-1)} \cdot \frac{n}{n-1} = \frac{n^2-3n+3}{(n-1)^2}$$

As a check on these conditional probabilities, we can compute

$$P(F_2|F_1) + P(N_2|F_1) = \frac{1}{n-1} + \frac{n-2}{n-1} = 1$$

$$P(F_2|N_1) + P(N_2|N_1) = \frac{n-2}{(n-1)^2} + \frac{n^2-3n+3}{(n-1)^2} = 1$$

It is not an accident that the expected number of fixed points of an n-long permutation is equal to the sum over $i = 1$ to n of the expected number of permutations with i as a fixed point. In general, the expected value of a sum of random variables is the sum of the expectations. To prove this, it suffices to show this for two random variables X and Y. Let S_x and S_y be the sample spaces for

X and Y, respectively, and let $Z = X + Y$. To simplify notation, let x_i, y_j, and z_k denote $X(i)$, $Y(j)$, and $Z(k)$. Then

$$\begin{aligned}
E(Z) &= \sum_{k\in S_z} p(z_k) z_k \\
&= \sum_{i\in S_x}\sum_{j\in S_y} p(x_i, y_j)(x_i + y_j) \\
&= \sum_{i\in S_x}\sum_{j\in S_y} p(x_i)p(y_j)(x_i + y_j) \\
&= \sum_{i\in S_x}\sum_{j\in S_y} \Big[p(x_i)p(y_j)(x_i) + p(x_i)p(y_j)(y_j)\Big] \\
&= \sum_{i\in S_x}\Big[\sum_{j\in S_y} p(y_j)\Big]p(x_i)x_i + \sum_{j\in S_y}\Big[\sum_{i\in S_x} p(x_i)\Big]p(y_j)y_j \\
&= \sum_{i\in S_x} p(x_i)x_i + \sum_{j\in S_y} p(y_j)y_j \\
&= E(X) + E(Y)
\end{aligned}$$

Solution (Exercise 1.28) — Each person has a first and last initial, so there are 26 choices for the first initial and 26 choices for the second initial, yielding $26^2 = 676$ possible combinations. Set up 676 boxes for each different initials pair. Fill each box with two people, or $676 \cdot 2 = 1352$ people chosen so far. Choosing one more person guarantees that at least one box (pair of initials) contains at least three people (Pigeonhole Principle). Then the answer is 1353 people.

Solution (Exercise 1.29) — Answer 1:

(a)Among $m+n$ steps, we must choose n to be $(0,1)$ and the remaining $(1,0)$. The total number is thus $\binom{m+n}{n}$.

(b)Now there are $m_1 + m_2 + m_3$ steps from which m_1 are chosen as $(1,0,0)$, m_2 as $(0,1,0)$, and m_3 as $(0,0,1)$. The answer is thus the multinomial $\binom{m_1+m_2+m_3}{m_1,m_2,m_3}$.

(c)Exactly as in part (b), the answer is the multinomial coefficient

$$\binom{\sum_{j=1}^{k} m_j}{m_1, m_2, \ldots, m_k}$$

Answer 2:

(a)Let $\mathbb{Z}^+$ be the positive integers. Consider the grid of points with integer coordinates in the plane, $(x,y) \in \mathbb{Z}^+ \times \mathbb{Z}^+$. How many paths are there from the point $(0,0)$ to the point (m,n) with steps consisting of moves $(0,1)$ and $(1,0)$?
There are a total of $m+n$ steps that need to be taken to get from $(0,0)$ to (m,n). From these, we need to choose which to be our n "up" steps. Once this has been determined, the remaining steps will be the m "sideways" steps. Therefore, the answer is $\binom{m+n}{n}$.

(b)How many paths are there from $(0, 0, 0)$ to (m_1, m_2, m_3) in the three dimensional lattice $\mathbb{Z}^+ \times \mathbb{Z}^+ \times \mathbb{Z}^+$ with steps in the unit vector directions $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$?
From the total of $m_1 + m_2 + m_3$, first choose m_1 steps in the x direction. Then from the remaining $m_2 + m_3$ steps, choose m_2 in the y direction. The remaining m_3 steps will be in the z direction. The solution is

$$\binom{m_1 + m_2 + m_3}{m_1}\binom{m_2 + m_3}{m_2} = \binom{m_1 + m_2 + m_3}{m_1 \quad m_2}$$

(c)What about the k-dimensional version of this question?
In k dimensions, first choose m_1 steps in the x_1 direction from the $m_1 + \cdots + m_k$ steps, then choose m_2 steps in the x_2 direction from the $m_2 + \cdots + m_k$ steps, and so on. The total is given by the multinomial coefficient

$$\binom{m_1 + m_2 + \cdots + m_k}{m_1 \quad m_2 \quad \cdots \quad m_{k-1}}$$

Solution (Exercise 1.31) — Answer 1:
Show that the expected number of successes for an n-long Bernoulli trial is np, and that the variance is $np(1 - p)$.
Let x_n be a random variable for the number of successes in an n-long Bernoulli trial. Let $0 <= p <= 1$ be the probability of success of a single trial. Then $P(x_n = k) = \binom{n}{k} p^i (1 - p)^{n-k}$. We compute:

$$\begin{aligned}
E(x_n) &= \sum_{k=0}^{n} kP(s_n = k) \\
&= \sum_{k=0}^{n} k\binom{n}{k} p^i (1 - p)^{n-k} \\
&= \sum_{k=0}^{n} \frac{kn!}{k!(n-k)!} p^k (1 - p)^{n-k} \\
&= np\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1 - p)^{n-k} \\
&= np\sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1 - p)^{n-k}
\end{aligned}$$

Introduce the change of variable $t = k - 1$, and use the binomial theorem:

$$\begin{aligned}
E(x_n) &= np\sum_{t=0}^{n-1} \binom{n-1}{t} p^t (1 - p)^{n-1-t} \\
&= np\left[p + (1 - p)\right]^{n-1} \\
&= np
\end{aligned}$$

To compute the variance we first compute $E(x_n^2)$:

$$\begin{aligned} E(x_n^2) &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\ &= \end{aligned}$$

Finally, we compute the variance:

$$\begin{aligned} V(x_n) &= E(x_n^2) - [E(x_n)]^2 \\ &= n^2p^2 - np^2 + np - (np)^2 \\ &= -np^2 + np = np(1-p) \end{aligned}$$

Answer 2:
First suppose that we have a single Bernoulli trial with probability p of success. The expected value for the number of successes of this trial will be equal to :

$$1 \times p + 0 \times (1-p) = p.$$

The variance of this trial will be :

$$(1-p)^2 \times p + (0-p)^2 \times (1-p) = p(1-p).$$

The binomial distribution is the sum of n independent Bernoulli trials. The expected value of a sum of independent experiments is the sum of the expected values of the experiments so the expected value of binomial distribution is np. Similarly, the variance of a sum of independent experiments is the sum of the variance of the independent experiments so the variance of the binomial distribution is $np(1-p)$.

Solution (Exercise 1.33) — Recall that Stirling's approximation gives $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$. Then,

$$\begin{aligned} \binom{2n}{n} &= \frac{(2n)!}{n!n!} \\ &\sim \frac{\sqrt{2\pi 2n}\left(\frac{2n}{e}\right)^{2n}}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \sqrt{2\pi n}\left(\frac{n}{e}\right)^n} \\ &= \frac{2\sqrt{\pi n}(2n)^{2n}e^{2n}}{2\pi n n^{2n} e^{2n}} \\ &= \frac{2^{2n}\sqrt{\pi n}}{\pi n} \\ &= \frac{2^{2n}}{\sqrt{\pi n}} \end{aligned}$$

More generally,

$$\begin{aligned}
\binom{x+y}{y} &= \frac{(x+y)!}{x!(x+y-y)!} \\
&\sim \frac{\sqrt{2\pi(x+y)}\left(\frac{x+y}{e}\right)^{x+y}}{\sqrt{2\pi y}\left(\frac{y}{e}\right)^{y}\sqrt{2\pi x}\left(\frac{x}{e}\right)^{x}} \\
&= \frac{\sqrt{2\pi(x+y)}(x+y)^{x+y}e^{x+y}}{2\pi\sqrt{y}y^{y}\sqrt{x}x^{x}e^{x+y}} \\
&= \frac{\sqrt{2\pi}(x+y)^{x+y+1/2}}{2\pi y^{y+1/2}x^{x+1/2}} \\
&= \frac{(x+y)^{x+y+1/2}}{\sqrt{2\pi}y^{y+1/2}x^{x+1/2}}
\end{aligned}$$

Solution (Exercise 1.34) — (a)When $k = n$, the product is $\frac{n!}{n^n}$. Then by Stirling's formula,

$$\prod_{j=0}^{k-1} \frac{n-j}{n} \sim \frac{n!}{n^n} \sim \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{n^n} = \frac{\sqrt{2\pi n}}{e^n}.$$

For k sufficiently small compared to n we have that

$$\frac{(n-k)}{n} = 1 - \frac{k}{n} \sim e^{-\frac{k}{n}}.$$

This follows by either truncating the Taylor series for $e^{-\frac{k}{n}}$ or by truncating the Taylor series for the natural logarithm in the following equality

$$1 - \frac{k}{n} = e^{ln(1-\frac{k}{n})} = e^{-\frac{k}{n}-\frac{k^2}{2n^2}\cdots}.$$

We will see in Exercise 1.34 of the Chapter on asymptotic approximation that if k^2 is small compared to n that the probability $\mathrm{prob}(n,k)$ that an k-sample of an n-set being a k-permutation is approximately

$$\frac{(n-k)}{n} = 1 - \frac{k}{n} \sim e^{-\frac{k}{n}}.$$

(Someone could put in a few more details here.) The answer to the problem is

$$\sim \prod_{j=0}^{k-1} e^{\frac{-j}{n}} = e^{\sum_{j=0}^{k-1}\frac{-j}{n}} = e^{-\frac{1}{n}\left(\frac{(k-1)(k-2)}{2}\right)} \sim e^{-\frac{1}{n}\frac{k^2}{2}} \sim 1 - \frac{k^2}{2n}.$$

When is this approximation correct? When the truncated terms are small, that is when we have small

$$\sum_{j=0}^{k-1} \frac{j^2}{n^2} \sim \frac{k^3}{n^2}.$$

Which is certainly true if $k^2 \ll n$.

Solution (Exercise 1.36) — (a)This is just the k-multisets of a n-set where each list is given in standard sorted order to satisfy the inequalities. Thus there are

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

(b)For this problem, we just choose any k distinct integers between 1 and n and then list them in sorted order. Thus there are

$$\binom{n}{k} \text{ solutions.}$$

Solution (Exercise 1.37) — (a)There are $\binom{2^n}{2^m}$ ways are there to pick a subset S of the 2^n numbers in the bigger space of the right size to map the smaller space of 2^m numbers into so that the encoding E can be 1-to-1.

(b)There are $(2^m)!$ of an encoding E once you have selected the subset S.

(c)There are 2^m possible preimages for each of the $(2^n - 2^m)$ elements not in the image, so there are $(2^m)^{(2^n-2^m)}$ possible choices for the remaining part of the decoding D map.

(d)Putting these all together there are:

$$\binom{2^n}{2^m}(2^m)!(2^m)^{(2^n-2^m)}$$

different $mBnB$ line codes.

Solution (Exercise 1.38) — (a)There are 2^5 5-bit binary strings.

(b)There are $\binom{5}{2} + \binom{5}{3}$ 5-bit binary strings which have Hamming Code 2 or 3. The Hamming Code 2 strings are: 11000, 01100, 00110, 00011, 10100, 01010, 00101, 10010, 01001, and 10001. The Hamming Code 3 strings are the complementary strings.

(c)The four codes which produce the longest length runs are: 11000, 00011, 11100, and 00111.

Bibliography

[1] Feller, W. *An Introduction to Probability Theory and its Applications, Vol. 1.* John Wiley and Sons, New York, New York, 1950.

[2] Graham, R. L., Knuth, D. E., and Patashnik, O. *Concrete Mathematics, A Foundation for Computer Science.* Addison-Wesley, Reading, Massachusetts, 1989.

[3] Knuth, D. E. *The Art of Computer Programming, vol. 1: Fundamental Algorithms*, second ed. Addison-Wesley, Reading, Massachusetts, 1973. tortoise and hare credited to floyd vol II, exer 6,7 p. 7.

[4] Scheinerman, E. R. *Mathematics: A Discrete Introduction.* BrooksCole, 2000.

[5] Skiena, S. *Implementing Discrete Mathematics.* Addison-Wesley Publishing Co., 1990.

Index