CSCI 189 Assignment 8 Problem Set

This assignment provides an opportunity to learn about counting, combinatorics, probability, and an interesting application of these ideas in computerized analysis of English sentences.

Summary of Concepts in Section 3.2 - 3.5 and Haskell

In many problem solving situations, we need to count the number of possible outcomes that the problem can have. This number is often computed by breaking the problem down into separate events and using the *multiplication principle.* That is, if there are n_1 different outcomes for one event and n_2 outcomes for a second event, then there are $n_1 \cdot n_2$ different ways that the first *and* second events can occur together.

For example, a 7-digit phone number has $10^7 = 10,000,000$ different values, since each digit can have 10 different values. On the other hand, if each digit in the phone number were required to be different from the other digits, there would be only $10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 = 604,800$ different values (the second digit must be different from the first, the third digit must be different from each of the first two, and so on).

The *addition principle* applies to *either one* of two independent events occurring. It says that if one event has n_1 outcomes and another event has n_2 outcomes, then the total number of ways that either one of these two events occurring is $n_1 + n_2$.

For example, suppose you want to buy either a donut or a bagel from Dunkin' Donuts. If they have 12 different kinds of donuts to choose from and 5 different kinds of bagels, then you have $12 + 5 = 17$ different choices. Of course, if you want both a donut and a bagel, you have $12 \cdot 5 = 60$ different choices (multiplication principle).

Often, these two principles are combined to count the number of outcomes for a problem. For instance, every computer connecting to the Internet must have a unique *IP address* that distinguishes it from all the rest. IP addresses are formed using four 8-bit numbers, translating them to decimal, and expressing them as a single value with a dot between each of the four parts.

For example, the computer known as "polar.bowdoin.edu" has IP address 139.140.1.1, whose four 8-bit numbers are `10001011 10001100 00000001 00000001`. To find out the IP address of your own computer, open a Web browser and link to `http://www.ed-phys.fr/htbin/ipaddress`.

Since a "bit" (short for "binary digit") can be either 0 or 1, we see that there are potentially $2^{32} = 4,294,967,296$ different IP addresses to assign (multiplication principle).

However, not quite that many are really available. That is, the first part of the IP address, called the *netid*, identifies the local network class to which the the computer belongs. The rest of the IP address is called the *hostid*, uniquely defines the machine itself. So the IP address is partitioned in the following way:

Class	Leading bits	Netid	Hostid
A	0	next 7 bits	remaining 24 bits
B	10	next 14 bits	remaining 16 bits
C	110	next 21 bits	remaining 8 bits

So the actual number of unique IP addresses available can be computed by combining the addition and the multiplication principle. That number is the sum of the class A, class B, and class C addresses, which is only $2^7 \cdot 2^{24} + 2^{14} \cdot 2^{16} + \cdot 2^{21} \cdot 2^8 = 3,758,096,384$

By the way, if you want to compute any of these large numbers using Haskell, just type the expression at the `Asst8>` prompt. For instance:

```
Asst8> 2^7*2^24+2^14*2^16+2^21*2^8
3758096384
```

As it turns out, this IP addressing scheme is now running out of unique addresses, due to the dramatic growth of the Internet. A new addressing scheme that uses 128 bits is now being adopted. For more information on Internet addressing, see `http://www.webopedia.com/TERM/I/IP_address.html`.

Decision trees can often be used to graphically display and count the alternatives in a complex problem. There is a good example on page 196 of your text.

Inclusion, Exclusion, and the Pigeonhole Principle

Often the alternatives in a problem are not mutually disjoint. For example, consider a group of students who are ordering pizza and some will eat sausage, some will eat pepperoni, some will extra cheese, and some will eat two or three of these extra toppings. These groups are not mutually exclusive, since a person who "will eat sausage" may or may not eat pepperoni. How do we determine how many students are in the group?

Such problems use the *principle of inclusion and exclusion*, and can be solved using "Venn diagrams" if there are 3 or fewer variables (e.g., pizza toppings in the above example). The pizza problem is solved with Venn diagrams on pages 203-204 of your text.

If there are more than three variables, a more general expression for counting the number of alternatives is presented on page 205, Equation (4):

$$\begin{aligned} \mid A_1 \cup A_2 \cup \ldots \cup A_n \mid &= \sum_{1 \le i \le n} \mid A_i \mid - \sum_{1 \le i < j \le n} \mid A_i \cap A_j \mid \\ &+ \sum_{1 \le i < j < k \le n} \mid A_i \cap A_j \cap A_k \mid - \ldots \pm \mid A_1 \cap A_2 \cap \ldots \cap A_n \mid \end{aligned}$$

The *pigeonhole principle* states that if there are more than k items competing for k places, then one of the places must contain two or more items. For instance, if our classroom had 11 chairs and there are 12 students, then someone will be sitting on someone else's lap!

Permutations and Combinations

The number of ways to arrange r items taken from n different possibilities is $P(n,r) = n!/(n-r)!$. Such an arrangement is called a "permutation." For example, the number of distinct 7-digit phone numbers is $P(10,7) = 10!/(10-7)! = 10!/3! = 604,800$, as we intuitively discovered above.

The number of ways to select r items from a set of n objects is $C(n,r) = P(n,r)/r!$. This is called the number of "combinations of r objects chosen from n" and it disregards how the objects are arranged. For instance, the two telephone numbers 123-4567 and 123-4576 represent the same combination of digits. Thus, the number of combinations of 7 digits taken from 10 is $P(10,7)/7! = 604,800/7! = 120$.

Probability

The set of all possible outcomes for an action called its "sample space" S. A subset of these outcomes is called an "event" E. The *probability* of an event E in a sample space S is defined by $P(E) = |E|/|S|$.

For example, the probability of a 7-digit phone number having all distinct digits = the number of ways of having a phone number with all distinct digits / the number of ways of forming a 7-digit phone number. In this example, S = all the different 7-digit phone numbers and E = all the phone numbers with all distinct digits. To compute the probability that an arbitrary phone number has all distinct digits, we just compute the quotient of the sizes of these two sets, which we can count using techniques discussed above. Since $|S| = 10^7$ and $|E| = 10!/3! = 604,800$, then $P(E) = 604,800/10^7 = 0.06048$, or roughly 6 in 100.

The "conditional probability" of a second event E_2 occurring, given that event E_1 has already occurred, is $P(E_2|E_1) = P(E_1 \cap E_2)/P(E_1)$. If events E_1 and E_2 are independent of each other, then $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$, in which case $P(E_2|E_1) = P(E_2)$.

Problems to be handed in

Section 3.2 (p 196) Exercises 8, 19, 23, 36, 40, 52, 57, 68.

Section 3.3 (p 207) Exercises 4, 10, 11, 13, 16.

Section 3.4 (p 217) Exercises 2, 12, 14bc, 17, 39, 64ab (use Haskell for some of these calculations).

Section 3.5 (p 232) Exercises 3, 10, 19, 33, 37, 54 (use Haskell for some of these calculations).

Complete Exercise 2 at the end of the Haskell tutorial Asst8.lhs