CS 70 Spring 2005	Discrete Mathematics for CS Clancy/Wagner	HW 13

Due Tuesday, May 10

Coverage: This assignment involves topics from the lectures of April 26 and May 3, and from Rosen section 3.2 (pages 233-236).

Administrative reminders: We will accept only unformatted text files or PDF files for homework submission. Include your name, login name, section number, and partner list in your submission. Give the command `submit hw13` to submit your solution to this assignment.

Homework exercises:

1. **(10 pts.) Using standard units**
Let Z be a random variable that has a normal distribution with mean 0 and variance 1. Given a real number z, there are tables that allow us to compute $\Pr[Z \geq z]$ as a function of z. The value z is sometimes called a z-score. For example:

   - The "right tail": $\Pr[Z \geq 1] \approx 0.1587$. $\Pr[Z \geq 2] \approx 0.0228$. The right tail is the area under the normal curve represented by all values greater than or equal to z.
   - The "left tail": $\Pr[Z \leq -1] \approx 0.1587$. The left tail represents the values less than or equal to $-z$.

   You can find resources for calculating these values—e.g., normal tables, normal calculators—at `http://www.cs.berkeley.edu/~daw/teaching/cs70-s05/tables.html`. These allow you to go back and forth between a z-score z and the area under the standard normal curve represented by all values greater than z (the "right tail"), or the corresponding area represented by all values less than z (the "left tail").

   (a) Let Z be normally distributed with mean 0 and variance 1. Use one of the tables or calculator mentioned above to find the approximate value of $\Pr[Z \geq 3]$.

   z-scores have many applications. For instance, if the random variable X is normally distributed with mean μ and variance σ^2, then the random variable X can be *normalized* to get a random variable X_{norm} defined by $X_{\text{norm}} = (X - \mu)/\sigma$. A useful fact is that, with these assumptions, X_{norm} will be normally distributed with mean 0 and variance 1. Note that the value of X_{norm} can be viewed as a z-score.

   (b) Suppose X is normally distributed with mean 100 and standard deviation 10.
   Calculate $\Pr[X \geq 125]$. You may wish to use the resources listed above.

   Here is another application of z-scores. Let B be the number of successes in n binomial trials each having success probability p. We have shown that $\mathbb{E}[B] = np$ and $\mathrm{Var}[B] = np(1-p)$. It turns out that, for large n, the binomial distribution B approximates the normal distribution with the same mean and variance. Let's normalize B, to get a random variable B_{norm} defined as follows:

$$B_{\text{norm}} = \frac{B - np}{\sqrt{np(1-p)}}$$

Given the assumption that B is approximately normally distributed with mean np and variance $np(1-p)$, then B_{norm} is approximately normally distributed with mean 0 and variance 1. Thus, the value of B_{norm} can be viewed as a z-score.

(c) Find a value k for which, when you flip a fair coin 10000 times, the probability of k or more heads is approximately 0.20.

2. (10 pts.) Countable or uncountable?

Determine whether the following sets are countable or uncountable. For each countable set, display a one-to-one correspondence between the set of natural numbers and that set, or an enumeration of the set. For each uncountable set, explain why it is uncountable.

(a) The set of binary strings which are palindromes. A string s is a palindrome if it can be written as the concatenation of some string t followed by the reversal of t.

(b) The set of real numbers in the interval [0,1] whose decimal representation contains a single "1" digit, and all other digits are "0".

(c) The set of real numbers in the interval [0,1] whose decimal representation contains only "0" and "1" digits (mixed in any order or combination).

(d) The set of rooted, finite binary trees, in which trees are distinguished only by their shape (that is, you should ignore node values).

3. (10 pts.) The Cantor set

The Cantor set is an amazing object. It is defined iteratively as follows. Start with the interval $[0,1]$ of real numbers between 0 and 1 (inclusive). In the first iteration, delete the middle third of this interval, namely, $(\frac{1}{3}, \frac{2}{3})$. You're left with the union of two smaller line segments, namely, $[0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. In the second iteration, delete the middle third of each of these two segments. Continue forever. In the nth iteration, you remove the middle third of each of the 2^{n-1} segments left over from the previous iteration. Let S denote the set of points that are left over after iterating forever, i.e., $x \in S$ if x is not removed in any iteration.

One can show that the Cantor set is, in some sense, very small; if we choose a real number X uniformly at random from the interval $[0,1]$, then $\Pr[X \in S] = 0$. However, the Cantor set is, in another sense, very large.

Show that the Cantor set is uncountably infinite.

HINT: There are two standard ways to prove that something is uncountable: find a bijection between it and some other uncountable set, or use diagonalization.

Also, you might find it useful to know the following alternative definition of the Cantor set: S is the set of real numbers $x \in [0,1]$ that can be represented in base 3 (ternary) using only 0's and 2's (i.e., no 1's). (Be warned that there is some ambiguity in ternary representations: $1/3$ could be represented as either $0.10000\ldots$ or $0.02222\ldots$. For this definition, we require that the ambiguity be resolved by always using representations that end in $02222\ldots$ rather than $10000\ldots$, whenever you have a choice.)

4. (10 pts.) Too many functions?

Let S denote the set of functions from $\mathbf{N}$ to the set $\{0,1,2,3,4,5,6,7,8,9\}$. Show that S is uncountable.

HINT: Establish a one-to-one correspondence between the set $[0,1]$ (of real numbers between 0 and 1) and a subset of S.

5. (20 pts.) QuickSelect

One way to find the kth smallest value in a set S of n integers ($n > k$) is to sort S into an array; the kth smallest element of the array will be the desired value. As we have seen, this requires $n \log n$ comparisons.

With an algorithm named QuickSelect, we can find the kth smallest value in a set of n (distinct) integers in linear time. Here's the algorithm.

$QuickSelect(S,k)$
    choose $p \in S$ at random;
    $lowVals = \{x \in S : x < p\}$; $n-1$ comparisons are required
    $highVals = \{x \in S : x > p\}$;   to compute *lowVals* and *highVals*
    if $k = size(lowVals) + 1$ then return p
    else if $k \le size(lowVals)$ then return $QuickSelect(lowVals,k)$
    else return $QuickSelect(highVals, k - size(lowVals) - 1)$;

The best case, when the first value chosen for p is the kth element, requires $n-1$ comparisons. The worst case, when p is always chosen as the maximum element in S and $k = 1$, requires $\frac{1}{2}n(n-1)$ comparisons. The expected number of comparisons, however—call it $T(n)$—is less than $4n$. Prove it.