Problem 0:

(a) X is a B(n.l/2) random variable with n >= 1, show that the probability that X is even is 1/2.

(b) Suppose that we independently roll two standard six-sided dice. Let X1 be the number that shows on the first die, X2 the number on the second die, and X the sum of the numbers on the two dice.

  i. What is E[X I X1 is even]?
  ii. What is E[X I X1 = X2]?
  iii. What is E[XI I X = 9]?
  iv. What is E[X1 - X2 I X = k] for k in the range [2, 12]?.

(c) Let X and Y be independent geometric random variables, where X has parameter p and Y has parameter q. You may find it helpful to keep in mind the memoryless property of geometric random variables.

  i. What is the probability that X = Y?
  ii. What is E[max(X, Y)]?
  iii. What is Pr(min(X, Y) = k)?
  iv. What is E[X I X <= Y]?

(d) Alice and Bob decide to have children until either they have their first girl or they have k children. Assume that each child is a boy or girl independently with probability 1/2 and that there are no multiple births.

  i. What is the expected number of female children that they have? What is the expected number of male children that they have?
  ii. Suppose Alice and Bob simply decide to keep having children until they have their first girl. Assuming that this is possible, what is the expected number of boys that they have?

(e) The geometric distribution arises as the distribution of the number of times we flip a coin until it comes up heads. Consider now the distribution of the number of flips X until the kth head appears, where each coin flip comes up heads independently with probability p. Prove that this distribution is given by Pr(X = n) = $\binom{n-1}{k-1}p^k(1-p)^{n-k}$ for n >= k. (This is known as the negative binomial distribution.)

(f) We roll a standard fair die over and over. What is the expected number of rolls until sixes occur twice?

(g) We roll a standard fair die over and over. What is the expected number of rolls until the first pair of consecutive sixes appears? (Hint: The answer is not 36.)

Problem 1:

A blood test is being performed on n individuals. Each person can be tested separately. but this is expensive. Pooling can decrease the cost. The blood samples of k people can be pooled and analyzed together. If the test is negative, this one test suffices for the group of k individuals. If the test is positive, then each of the k persons must be tested separately and thus k + 1 total tests are required for the k people. Suppose that we create n/k disjoint groups of k people (where k divides n) and use the pooling method. Assume that each person has a positive result on the test independently with probability p.

(a) What is the probability that the test for a pooled sample of k people will be positive?

(b) What is the expected number of tests necessary?

(c) Describe how to find the best value of k.

(d) Give an inequality that shows for what values of p pooling is better than just testing every individual.

Problem 2: (Please refer to Wikipedia for reservoir sampling)

The following approach is often called ***reservoir sampling.*** Suppose we have a sequence of items passing by one at a time. We want to maintain a sample of one item with the property that it is uniformly distributed over all the items that we have seen at each step. Moreover, we want to accomplish this without knowing the total number of items in advance or storing all of the items that we see.

(a) Consider the following algorithm, which always stores just one item in memory. When the first item appears, it is stored in the memory. When the kth item appears, it replaces the item in memory with probability 1/ k. Explain why this algorithm solves the problem.

(b) Suppose that we modify the above reservoir sampling algorithm of so that. when the kth item appears, it replaces the item in memory with probability 1/2. Describe the distribution of the item in memory.

Problem 3:

Let $a_1, a_2, \ldots, a_n$ be a random permutation of {1, 2, ... ,n}, equally likely to be any of the n! possible permutations. When sorting the list $a_1, a_2, \ldots, a_n$, the element a_i must move a distance of $|a_i$ - i | places from its current position to reach its position in the sorted order. Find the expected total distance that elements will have to be moved.

Problem 4:

Let $a_1, a_2, \ldots, a_n$ be a list of n distinct numbers. We say that a_i and a_j are inverted if i < j but $a_i > a_j$. The Buhblesort sorting algorithm swaps pairwise adjacent inverted numbers in the list until there are no more inversions, so the list is in sorted order. Suppose that the input to Bubblesort is a random permutation, equally likely to be any of the n! permutations of n distinct numbers. Determine the expected number of inversions that need to be corrected by Bubblesort.

Problem 5:

Linear insertion sort can sort an array of numbers in place. The first and second numbers are compared; if they are out of order, they are swapped so that they are in sorted order. The third number is then placed in the appropriate place in the sorted order. It is first compared with the second: if it is not in the proper order. it is swapped and compared with the first. Iteratively, the kth number is handled by swapping it downward until the first k numbers are in sorted order. Determine the expected number of swaps that need to be made with a linear insertion sort when the input is a random permutation of n distinct numbers.