CS 70 Discrete Mathematics and Probability Theory
Spring 2011 Demmel HW9

Due Apr 1, 8:00am

You may work in groups of up to 3 people (no larger!). Please read the group collaboration policies on bSpace or `www.cs.berkeley.edu/~demmel/cs70_Spr11` before beginning group work. You *must* write up the solution set entirely on your own. You must never look at any other students' solutions (from any semester, not even a draft), nor share your own solutions (not even a draft).

Please begin your answer to each question on a new sheet of paper, and make sure that each sheet is labeled with your name, GSI name, the assignment number, the question number, and "CS70-Spring 2011".

Turn in each question in a different box in 283 Soda Hall: Question 1 in the box labeled "CS70 - 1", Question 2 in the box labeled "CS70 - 2", etc. Reason: Different problems will be graded in parallel by different readers.

Warning: You risk receiving no credit, or losing points, for any homework that does not conform to the above regulations! Please take the time to write clear and concise solutions; we will not grade messy or unreadable submissions. No late homeworks will be accepted. We will drop the lowest two homework scores.

1. **(20 pts.) A Very Small Example of Hashing**
   Suppose we hash three distinguishable objects randomly into a table with three (labelled) entries. We are interested in the lengths of the linked lists at the three table entries.

   (a) List all the outcomes in the sample space of the experiment. How many of them are there?

   (b) Let X be the length of the linked list at entry 1 of the table. Write down X explicitly as a function on the sample space mapping to the real line (either in a figure as in class or as a list). Compute and plot the distribution and expectation of X.

   (c) Let Y be the length of the *longest* linked list among all three. Write down Y explicitly as a function on the sample space mapping to the real line. Compute and plot the distribution and expectation of Y.

   (d) Is the expectation of X larger than, equal to or smaller than that of Y?

   (e) Compute the distribution of X for the general case when m objects are hashed randomly into a table of size n, i.e. give an expression for the probability that X takes on each value in its range. (Computing the distribution of Y for the general case is not so easy, so we won't ask you to do it!)

2. **(15 pts.) Games**
   Here's a game. Alice and Bob will each roll a fair, six-sided die. If Alice's die comes up with a number higher than Bob's, Alice wins \$3 from Bob. If Bob's number comes up higher, or if they tie, Bob wins \$2 from Alice. Is this game a good deal for Alice? Explain.

   (Hint: Define a random variable and compute an expectation.)

3. **(20 pts.) Packets Over the Internet**
n packets are sent over the Internet (n is even). Consider the following methods for sending the packets, and models of how a packet could be lost:

(a) Each packet is routed over a different path and is lost independently with probability p.

(b) All n packets are routed along the same path, and with probability p, one of the links along the path fails and all n packets are lost. Otherwise all packets are received.

(c) The n packets are divided into 2 groups of $n/2$ packets, and each group is routed along a different path and lost with probability p. Losses of different groups are independent events.

Part 1: In each of the three models, compute the distribution of the number of packet losses, i.e. if the random variable $f =$ the number of packet losses, say what values $r = f(x)$ can occur, and their probability $P(f = r)$.

Part 2: Compute $E(f)$ for each of the three models.

Part 3: Which method for sending packets over the Internet do you think is more reliable, and why?

4. **(20 pts.)**
Twenty distinct numbers are arranged in a random order, so that all 20! orderings are equally likely. Going down the list, one marks every number that is larger than all earlier numbers on the list. (In particular, the first number of the list is marked.) What is the expected number of marked numbers on the list? **Hint:** Show that the probability that the i^{th} number marked is $1/i$.

5. **(20 pts.)**

Part 1: Let S be a sample space and P a probability function on S. Let $f : S \to \mathbb{R}$ be a random variable. Let $g : \mathbb{R} \to \mathbb{R}$ be another function, so that $h(x) = g(f(x))$ is also a random variable. Show that $E(h) = \sum_{r \in f(S)} g(r) \cdot P(f = r)$, where $r \in f(S)$ indicates the set of real numbers $r = f(x)$ generated by plugging in every $x \in S$. This may also be called the range of f, or sometimes the image of f (under S) - see the Wikipedia article "Range (mathematics)" for more details.

Part 2: You repeatedly throw a dart at a target until you hit the target. Each time you throw, the probability of hitting the target is $0 < p < 1$. Let the random variable f be the number of throws it takes to hit the target. What is $E(f^2)$?

6. **(20 pts.) Expectations**

Solve each of the following problems using linearity of expectation. Clearly explain the steps. (Hint: for each problem, think about what the appropriate random variables should be and define them explicitly.)

(a) A monkey types at a 27-character keyboard with one key corresponding to each of the lower-case English letters, plus the space key. Each keystroke is chosen independently and uniformly at random from the 27 possibilities. If the monkey types 1 billion letters, what is the expected number of times the sequence "banana" appears?

(b) A coin with Heads probability p is flipped n times. A "run" is a maximal sequence of consecutive flips that are all the same. (Thus, for example, the sequence HTHHHTTH with $n = 8$ has five runs.) Show that the expected number of runs is $1 + 2(n-1)p(1-p)$. Justify your answer carefully.