15-359: Probability and Computing

Assignment 3 | Due: February 10, 2012

Problem 1: Geometric digits (15 pts.)

Let X be a random variable defined similarly to the geometric distribution: in a sequence of independent Bernoulli trials with probability p of success, X is the number of failures before the first success (so that $X + 1$ has the Geometric(p) distribution defined in class).

Let D_i be the i-th digit of X (in particular, $D_i = 0$ when $X < 10^{i-1}$). Show that the random variables D_i and D_j are independent when $i \neq j$.

Problem 2: The randomized quicksort jumped over the lazy bubblesort (20 pts.)

We have seen in class that a randomized algorithm for median selection works in $O(n)$ time. A similar argument could be used to prove that a randomized version of quicksort (pivot is chosen at random from the given list) has expected running time $O(n \log n)$. Here is a different approach that focuses on the comparisons in quicksort.

Assume we are sorting some permutation $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$ of n distinct integers. Define the random variable C to be number of comparisons made during the execution of the algorithm on $\boldsymbol{a}$. Use the decomposition of C into a sum of indicator variables to determine $\mathrm{E}(C)$.

Problem 3: This is how graduate students spend their free time (20 pts.)

In a fit of boredom, you take out your matching set of n balls and n bins, and then throw each ball into a bin uniformly and independently at random. The bins are cleverly designed in such a way that each bin can hold exactly $\frac{10 \log n}{\log \log n}$ balls. If more balls than that are thrown into a bin, it overflows.

Let X be the number of bins that end up overflowing. Show that $\mathrm{E}(X) \leq 1/n$. Why does this imply that with probability $1 - 1/n$ (which approaches 1 as you invest in more balls and bins), no bin overflows?

Problem 4: The entropy cat is out of the bag now (20 pts.)

We define the *entropy function* of a real number $p \in (0, 1)$ as

$$H_2(p) = p \log_2 \left(\frac{1}{p} \right) + (1 - p) \log_2 \left(\frac{1}{1 - p} \right).$$

This function is distinct from, but related to, the *(information) entropy* $H(X)$ of a random variable X, defined in Homework 2 as the “excitement” $C(X)$. You can easily check that if X can take on only two values, its entropy is given by the function H_2.

Use bounds on the entropy of appropriately defined random variables to prove the inequality

$$\log_2 \binom{n}{k} \leq nH_2\left(\frac{k}{n}\right).$$

Problem 5: Unsorting algorithm (25 pts.)

Consider the following algorithm for shuffling a deck of n cards, initially ordered $1, \ldots, n$:

1. Pick up the top card of the deck.
2. Choose a position in the deck uniformly at random (including top and bottom) and re-insert the card at that position.
3. Repeat steps 1 and 2 until all cards have been picked up at least once.

Prove that after this algorithm terminates, all permutations of the cards are equally likely.
What is the expected running time of the algorithm?

Problem 6: Break up the monotony (extra credit) (15 pts.)

Recall that a function $f(x_1, \ldots, x_n) : \{0,1\}^n \to \mathbb{R}$ is *monotonic* if

$$f(x_1, \ldots, 0, \ldots, x_n) \leq f(x_1, \ldots, 1, \ldots, x_n)$$

(that is, switching a variable from 0 to 1 can only increase the function). Let f and g be monotonic functions, and choose x_1 through x_n by flipping n independent fair coins: $\mathrm{P}(x_i = 0) = \mathrm{P}(x_i = 1) = 1/2$. Prove that

$$\mathrm{E}(f(x_1, \ldots, x_n)g(x_1, \ldots, x_n)) \geq \mathrm{E}(f(x_1, \ldots, x_n)) \cdot \mathrm{E}(g(x_1, \ldots, x_n)).$$

(Hint: induct on n)