Homework 1

Due: Feb 7, 2019 at 2:30pm

Please follow these guidelines when writing up your CS 155 homeworks.

A major goal of this class is that you learn how to mathematically analyze probabilistic processes and events. Hence, make sure your proofs are rigorous, i.e., that every step is adequately justified. Try to keep your proofs as simple and as concise as possible, keeping in mind that you will potentially have to examine multiple proof strategies in order to achieve this. Make sure to CLEARLY STATE YOUR ASSUMPTIONS at the beginning of any solution, when necessary. As we are doing probabilistic analysis, we shall require that you clearly define the probability space you are working with and identify the events within it that you will analyze.

You are allowed to discuss these problems with other students, but you have to leave any meeting with other students without any written material about the problems, and the writeups must be done by yourself without help from others. If you have any questions regarding how these guidelines apply to a particular problem or what they mean in general, please email the TAs.

Problem 1

Exercise 2.25 in the textbook

Note: for part c, just describe the procedure you would use to find the best value of $k$ for a given value of $p$. There is no need to find a closed-form solution.

Problem text:

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

Exercise 3.7 in the textbook

Problem text:
A simple model of the stock market suggests that, each day, a stock with price q will increase by a factor \( r > 1 \) to \( qr \) with probability \( p \) and will fall to \( q/r \) with probability \( 1 - p \). Assuming we start with a stock with price 1, find a formula for the expected value and the variance of the price of the stock after \( d \) days.

Problem 3

With a \( b \)-bit counter, we can ordinarily only count up to \( 2^b - 1 \). With the following probabilistic counting algorithm, we can count up to a much larger value at the expense of some loss of precision.

We fix a priori an increasing sequence of integers \( n_0, n_1, n_2, \ldots \). When the counter stores a value of \( i \), it represents a count of \( n_i \). The initial value stored in the counter is 0, representing a count of \( n_0 = 0 \).

For example, if \( n_i = i \) for all \( i > 0 \), then the counter is an ordinary counter. More interesting situations arise if we select, say, \( n_i = 2^{i-1} \) for all \( i > 0 \). In this case, a counter storing a value of 10 would represent a count of \( 2^9 = 512 \). This counter can count up to \( 2^{2^b-2} \), which is much larger than the maximum count of \( 2^b - 1 \) that can be represented by an ordinary \( b \)-bit counter.

We now introduce a randomized increment operation on the counter, with the property that if we perform \( m \) increment operations, then the expectation of the count represented by the counter is exactly \( m \). Let \( i \) be the integer stored in the \( b \)-bits memory, and let \( C = n_i \) be the count represented by the counter.
We increment the counter in a probabilistic manner. The increment operation increases the stored value from $i$ to $i + 1$ with probability $1/(n_{i+1} - n_i)$, and it leaves the counter unchanged with probability $1 - 1/(n_{i+1} - n_i)$. (An exception occurs if $i = 2^b - 1$; then the counter overflows, because we can’t increase the stored value anymore.)

Note that performing an increment operation changes the value of $i$ (and the value of $C$) with some probability, and otherwise the values don’t change.

Throughout this problem, assume that $n_{2^{b-1}}$ is large enough that the probability of an overflow is negligible.

a. Show that, no matter which values are chosen for the $n_i$, the expected count represented by the counter after $m$ increment operations is exactly $m$. [Hint: show that the expected change in $C$ following an increment operation is always 1.]

b. The analysis of the variance of the count represented by the counter depends on the sequence of the $n_i$. Consider the case $n_i = 100i$ for all $i > 0$. Compute the variance in the count represented by the counter after $m$ increment operations.

c. Again consider the case $n_i = 100i$ for all $i > 0$. What can Markov’s inequality and Chebyshev’s inequality tell us about the probability that, after $m$ increment operations, the count represented by the counter is noticeably more than $m$, say at least $2m$?