Homework 9

Due Saturday, November 14

Implementation Part

1. Implement the discrete log collision algorithm `dLogCollide(g,h,p,n)` which takes as input a prime p, a primitive root $g \in \mathbb{F}_p^*$, an element $h \in \mathbb{F}_p^*$, and returns the discrete log $\log_g h$. It should also take an integer n which should default to $\lfloor \sqrt{p-1} \rfloor$ if no n is given, this is the length of your lists in the collision algorithm. It should do the following:

   (1) Compute a list $\{g^i\}$ for n random integers i.

   (2) Compute a list $\{hg^j\}$ for n random integers j.

   (3) Find an overlap $g^i = hg^j$ between the two lists and use this to return the discrete log.

   In order to optimize the search for an overlap, you can use a `set` or hash table like in HW5 Problem 2.

2. Test out your algorithm to compute the following discrete logs:

   (a) $\log_2 390$ in $\mathbb{F}_{659}$

   (b) $\log_{10} 106$ in $\mathbb{F}_{811}$

Written Part

3. Let $X : \Omega \to \mathbb{R}$ be a random variable, taking values in the set $\{x_1, x_2, \cdots x_r\}$.

   (a) Show that if $a \neq b$ then the events $(X = a)$ and $(X = b)$ are disjoint.

   (b) Show that
   $$\Omega = \bigcup_i (X = x_i)$$

   (c) Show that
   $$\sum_{i=1}^{r} f_X(x_i) = 1.$$

   (d) Recall that the expected value of X was defined to be:
   $$E(X) = \sum_{i=1}^{r} x_i f_X(i).$$

   Prove that it this is equal to the folloing value:
   $$\sum_{\omega \in \Omega} X(\omega) Pr(\omega).$$

4. In the following cases compute the expected value of the variable X

   (a) X is uniformly distributed on $\{0, 1, \cdots, N-1\}$.

   (b) X is uniformly distributed on $\{1, 2, \cdots, N\}$.

(c) X is uniformly distributed on the first 7 prime numbers.

(d) X is a random variable with a binomial density function. (Hint: use the binomial theorem and a differentiation to get a closed form for the sum).

5. In this problem we will use probability and expected values to study why the `findPrime` algorithm from problem 1 was so successful.

   (a) Let $L < U$ be a positive integers. Use the prime number theorem to estimate

   $$\rho = \rho(L, U) = (\text{the probability that a randomly chosen number } n \text{ with } L < n \leq U \text{ is prime})$$

   in terms of L and U.

   (b) Let Ω the set of outcomes consisting of infinite sequences of numbers between L and U:

   $$\Omega = \{n_1, n_2, n_3, \cdots | L \leq n_i < U \text{ for all } i\}.$$

   Let $X : \Omega \to \mathbb{Z}$ be the random variable whose value is number of guesses until the first prime. That is:

   $$X(n_1 n_2 n_3 ...) = i \iff n_i \text{ is prime and } n_j \text{ is not prime for any } j < i.$$

   Let a be a positive integer. Compute the probability density $f_X(i)$ in terms of i and the probability ρ from part (a). (That is, what is the probability that the ith number is the first prime?)

   (c) Compute the expected value $E(X)$. Interpret in words what this number means. (This computation should look a lot like the expected value of the coin flipping example in the 11/5 lecture).

   (d) Use part (c) to estimate the following:

   i. If I randomly guess 2 digit numbers how many guesses will it take to find a prime?

   ii. If I randomly guess 100 digit numbers how many guesses will it take to find a prime?

   iii. If I randomly guess 500 digit numbers how many guesses will it take to find a prime?

   (e) Use the evidence you've gathered to explain why `findPrime` from the first project was successful.

6. Suppose 23 random people are in a room. Compute the probability that 2 of them share a birthday. (This is the most well known statement of the *birthday paradox*).

The next problem concerns the following theorem from class, for which we did prove part (i).

Theorem 1. *Suppose there is an urn with N balls, of which n are red and $N-n$ are blue. Suppose further that you randomly choose m balls, replacing after each selection.*

(i) $Pr(\text{at least one red}) = 1 - \left(1 - \frac{n}{N}\right)^m$

(ii) $Pr(\text{at least one red}) \geq 1 - e^{\frac{-mn}{N}}$

(iii) If N is large and m, n are not much larger than $\sqrt{N}$ then the estimate from (ii) is quite accurate

7. Lets prove parts (ii) and (iii) of Theorem 1. We may use part (i) in our proofs since it was established in class.

   (a)
   $$e^{-x} \geq 1 - x \text{ for all } x.$$
   (Hint: use calculus to find the global minimum of $e^{-x} - (1 - x)$)).

   (b) Use part (a) and Theorem 1(i) to prove Theorem 1(ii).

   (c) Prove that for all $a > 1$ and $0 \leq x \leq 1$ the following inequality holds.
   $$e^{-ax} \leq (1 - x)^a + \frac{1}{2}ax^2.$$

   (d) Use part (c) and Theorem 1(i) to prove the following identity:
   $$Pr(\text{at least one red}) \leq 1 - e^{\frac{-mn}{N}} + \frac{mn^2}{2N^2}.$$
   Use this to deduce Theorem 1(iii).

8. In the Miller-Rabin problem I suggested that we interpret the Prime Number Theorem as saying the probability of a number n being prime is $\frac{\ln n}{n}$, but of course this way off the mark. The prime number theorem says there are $\frac{n}{\ln n}$ primes less than n. Thus the probability of one being prime in particular is approximately
   $$\frac{n/\ln n}{n} = \frac{1}{\ln n}.$$
   Notice that $\frac{1}{\ln n}$ is MUCH LARGER DENSITY than $\frac{\ln n}{n}$.

   (a) To really feel the difference between these two densities, use each to compute the probability that a random number less than 10^{100} is prime. This should illustrate the gravity of the mistake.

   The beauty of the prime number theorem is it says primes are rather dense, and the value I gave said they are extremely sparse. This likely affected the expected correctness of your Miller-Rabin computation.

   (b) Redo the computations from HW8 Problem 8 with this correct probability, so that we have computed the correct values. (Don't worry about re-deriving everything, just show the formula and plug in the correct values.) In particular you should show that:
   $$Pr\left(\text{n is prime} \mid \text{Miller-Rabin Fails } N \text{ times}\right) \geq 1 - \frac{\ln n}{4^N}.$$
   In particular, how confident are we in are primes when implemented RSA, where $N = 20$ and $2^{511} \leq n < 2^{512}$?