CS174 **Final Exam Solutions** **Spring 2001**
J. Canny **May 16**

1. Give a short answer for each of the following questions:

   (a) (4 points) Suppose two fair coins are tossed. Let X be 1 if the first coin is heads, 0 if the first coin is a tail. Let Y be 1 if the outcomes on the two coins are *the same*, and 0 if they are different. Are X and Y independent? Explain briefly.

   Answer: Each pair of outcomes (X, Y) corresponds to a unique combination of heads and tails for the two coins. There are four of these, and all have probability 1/4 for fair coins. So

   $$\Pr[X = u, Y = v] = \Pr[X = u]\Pr[Y = v] = 1/4$$

   for all pairs $u, v \in \{Heads, Tails\}$. Therefore X and Y are independent.

   (b) (4 points) Let X have the geometric distribution with parameter $p = 0.5$. What is $\Pr[X \geq 3]$?

   Answer: Notice that

   $$\Pr[X \geq 3] = 1 - \Pr[X < 3] = 1 - \Pr[X \in \{1, 2\}] = 1 - \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

   or you can compute the sum of the infinite geometric series $1/8 + 1/16 + \cdots\cdots$ which is another route to $\frac{1}{4}$.

   (c) (4 points) Let X be the number of fixed points in a random permutation of n items (assume n large). Which bounding method out of Markov, Chebyshev or Chernoff would you use to get the best bound for $\Pr[X > 5]$?

   Answer: The number of fixed points in a permutation is well-approximated as a poisson distribution with parameter $\lambda = 1$. The poisson distribution is a good approximation to the binomial distribution for large n, and we can apply Chernoff to binomial distributions. Chernoff gives the best bounds when it is applicable, so its the right choice here.

   (d) (4 points) Suppose $2n$ balls are randomly distributed into n bins, for large n. What is the expected number of empty bins?

   Answer: The probability of a bin being empty is well-approximated by a poisson distribution with parameter $m/n = 2$ here. The distribution is

   $$\Pr[X = k] = \frac{e^{-\lambda}\lambda^k}{k!}$$

   where $\lambda = 2$ and $k = 0$ (empty bins), which gives e^{-2}. This is also the expected value of the number of empty bins in this bin (think about it). The total expected number of empty bins is n times this, which is ne^{-2}.

(e) (4 points) Suppose we create a random graph on n vertices by adding one edge $\{x, y\}$ at a time, selecting x and y independently and uniformly at random from $\{1, \ldots, n\}$, discarding self-loops. Let m be the number of edges added until the graph is connected. What is $\mathrm{E}[m]$ in terms of n (you can give a big-O bound)?

Answer: From class, using epochs, the expected number of edges is $O(n \log n)$.

2. Let π be a random permutation of $\{1, \ldots, n\}$. Suppose the elements $\pi(1), \pi(2), \ldots, \pi(n)$ are inserted in that order into an initially-empty binary search tree.

(a) (6 points) What is the expected height of the final tree in terms of n?

Answer: From the analysis in class, we saw that the expected height of the tree is $O(\log n)$.

(b) (6 points) Let $a_1, \ldots, a_k$ be a subset of k elements of $\{1, \ldots, n\}$. What is the probability that a_1 comes before the other a_i's in the random permutation π?

Answer: By the arguments made in class, this is the same as if the k elements were the only ones under consideration. The number of permutations of k elements beginning with a_1 is $(k-1)!$, and dividing this by the total number of permutations of k elements which is $k!$, we get the answer of $1/k$.

(c) (8 points) What is the probability p_{ij} that elements $\pi(i)$ and $\pi(j)$ are compared together during the algorithm?

Answer: Remember that two elements will be compared iff one is an ancestor of the other. That happens iff there is no value in between them which occurs earlier in the permutation (because that element would become a common ancestor of both of them and prevent them from being compared). Let S be the set of elements between $\pi(i)$ and $\pi(j)$ inclusive, and notice that $|S| = |\pi(j) - \pi(i)| + 1$. These two elements will be compared iff either $\pi(i)$ or $\pi(j)$ occurs first in the permutation of elements of S. The probability of that is $2/|S|$, or in other words:

$$p_{ij} = \frac{2}{|\pi(j) - \pi(i)| + 1}$$

Comment: There were a lot of incorrect answers to this question with results like $(\ln n)/n$. That would be the *expected* probability if i and j were sampled uniformly at random. But the only random process here is the choice of π. Make sure you understand what the experiment and sample space are when thinking about a question.

3. Recall that a family of hash functions $\mathcal{H}$ is *two-universal* if for every x, y in U, with $x \neq y$, and for a hash function $h : U \to T$ chosen uniformly at random from $\mathcal{H}$,that $\Pr[h(x) = h(y)] \leq 1/n$, where $n = |T|$. Answer the following questions and give a brief justification for each.

(a) (5 points) Let $\mathcal{H}$ be the family of *all* functions from U to T. Is $\mathcal{H}$ two-universal?

Answer: Yes. Since h is drawn uniformly from the set of all functions, h is equally likely to assign every value to x and *independently* to assign every value in T to y. So the probability that $\Pr[h(x) = h(y)] = 1/n$.

(b) (5 points) Assume $|U| = |T|$, and let $\mathcal{H}$ be the set of all one-to-one and onto functions from U to T (you can think of this as the set of permutations). Is $\mathcal{H}$ two-universal?

Answer: Yes. Once we fix a value for $h(x)$, the one-to-one property forces $h(y) \neq h(x)$, so $\Pr[h(x) = h(y)] = 0$.

(c) (10 points) Let $U = T = \mathbb{Z}_p$ where p is a prime. Let $\mathcal{H}$ be the family of functions of the form

$$h_a(x) = ax (\mathrm{mod}\ p)$$

where a ranges over $\mathbb{Z}_p$. Is $\mathcal{H}$ two-universal?

Answer: Surprisingly, yes. Pick any distinct x and y. Then $h(x) = h(y)$ is the condition that $a(x - y) = 0 (\mathrm{mod}\ p)$, and since $\mathbb{Z}_p$ is a field, this is true iff either $x = y$ or $a = 0$. But x and y are distinct mod p, so it must be that $a = 0$. Since a is chosen uniformly at random from $\mathbb{Z}_p$, the probability $\Pr[a = 0] = 1/p = 1/n$, because $p = |T| = n$.

Comment: There were many incorrect answers to this question, all with the same issue that came up in question 2. The random choice here is the hash function h, not of x or y. Defining probabilities for a particular h doesnt make sense here. Check the sample space first.

4. Let p be a prime and q be another prime such that $q|p - 1$. Let a be an element chosen randomly from $\mathbb{Z}_p^*$.

(a) (4 points) Define the order of a.

Answer: The order of an element a is the smallest positive integer k such that $a^k = 1 (\mathrm{mod}\ p)$. This is the same as the order of the subgroup generated by a.

(b) (4 points) How would you test if the order of a is q?

Answer: Check first that $a \neq 1$ and then check whether

$$a^q = 1 (\mathrm{mod}\ p)$$

if both are true, then the element has order q. The test $a^q = 1 (\mathrm{mod}\ p)$ establishes that a has an order which divides q. Since the only value that divides q (other than q) is one, we eliminate the possibility that the order of a is 1 by checking if $a = 1$ itself. If not, then it must be q.

(c) (6 points) Suppose $p - 1 = 2q$, so that p is a strong prime. What is the probability that a random element of $\mathbb{Z}_p$ has order q?

Answer: First of all, we know that there is a unique subgroup G_q of $\mathbb{Z}_p$ of order q, and it has q elements which are the q different powers of a generator. Notice that all but one of these elements has order q, because the identity has order 1. So there are $q - 1$ elements in $\mathbb{Z}_p^*$ of order q. So the probability that a random element has order q is $(q - 1)/(2q)$.

Another proof is that we know that $\mathbb{Z}_p^*$ is cyclic since p is prime. So assume there is a generator g, which has order $2q$. Every even power of this generator has order q, while

every odd power has order $2q$. The number of even powers between 1 and $2q-1$ is $q-1$. Note that we dont count the zeroth or $2q$th powers, because those give the identity. So once again the probability of selecting one of those $q-1$ values is $(q-1)/(2q)$.

(d) (6 points) If $p-1=2q$, what are the possible values of $a^q \pmod p$ for $a \in \mathbb{Z}_p^*$?

Answer: We know that $\mathbb{Z}_p^*$ has order $2q$. So $a^{2q} = 1 \pmod p$ for every element in $\mathbb{Z}_p^*$. Let $b = a^q \pmod p$. We know that $b^2 = 1 \pmod p$ from the above. Since $\mathbb{Z}_p$ is a field, the only possible values for b are $+1$ and -1 (or $+1$ and $p-1$ which equals $-1 \pmod p$).

5. Here is the discrete-log ZKP from class. We assume as usual that p and q are primes and that $q|p-1$. Let g be an element of $\mathbb{Z}_p$ of order q. The prover chooses $s \in \mathbb{Z}_q$ at random, and computes $h = g^s \pmod p$. Prover publishes p, q, g, h. Now the prover wants to prove that she knows s. The proof is:

   (a) Prover picks an $r \in \mathbb{Z}_q$ at random, computes $a = g^r \pmod p$ and sends a to the verifier.

   (b) Verifier sends back a random challenge $c \in \mathbb{Z}_q$ to prover.

   (c) Prover computes $w = cs + r \pmod q$ and sends w to verifier.

   (d) Verifier checks that $g^w = h^c a \pmod p$.

   The conversation between prover and verifier consists of the messages (a, c, w).

   (a) (10 points) Give a simulation of the protocol, that is a program which generates tuples (a, c, w) with the same probability distribution, without knowledge of s.

   Answer: Prover picks c and w uniformly at random from $\mathbb{Z}_q$. Then prover computes

   $$a = g^w h^{-c} \pmod p$$

   and has a tuple (a, c, w). The probabilities in the real and simulated protocol are the same, because in both cases c and w are independent, uniformly distributed random variables. In the real protocol, w depends on the hidden value r, but you can check that it still has the uniform distribution for all c, hence is independent of c. The value a is always uniquely determined from the values of c and w.

   (b) (10 points) Show that from two conversations (a, c, w) and (a, c', w'), its possible to recover the secret s.

   Answer: Use part (c) of the protocol, and subtract the two values of w:

   $$w - w' = s(c - c') \pmod p$$

   there is no r term because the two r values are implicitly the same, because the value of a is the same in both cases. So we can compute s via:

   $$s = (w - w')(c - c')^{-1} \pmod p$$

6. Suppose you want to create a traceable, anonymous note for \$20 with the property that: there is a $k > 1$ such that the note can be spent $k - 1$ times without revealing the identity of the owner, but spending it k times will probably reveal the owners identity. Let m be the number of copies of the owner's identity in each note.

   (a) (10 points) How would you modify the anonymous cash protocol from class to do this? Dont give the whole protocol, just explain what needs to be changed.

   Answer: A bad way to do this is to have m copies of the owner's identity, where each copy is naively secret shared k ways, such that all k pieces of one identity are needed to reconstruct their identity. The problem is that if the spender takes the note to k different vendors, the probability that the vendors pick k different shares of one identity is $k!/k^k$ which is extremely small. Even with m copies of this identity, the probability of detecting a cheater is extremely small.

   A better approach is to have m copies of the identity, but where each copy is Shamir secret-shared k^2 ways, such that any k suffice to reconstruct the owner's identity. Now the probability that k vendors request k distinct shares from each identity is quite high (birthday paradox). See the next part.

   (b) (10 points) What is the probability that a user would be caught if they spend the note k times as a function of k and m? Your method should have good probability of detecting a cheater, say $1 - 2^{-O(m)}$ for full credit. If not, modify the protocol from part (a).

   Answer: Suppose as above that each identity copy is Shamir secret-shared into k^2 shares, where k suffice to reconstruct the identity. The probability that each vendor picks a share different from all the others is $\geq (k^2 - k)/k^2 = (k - 1)/k$. So the probability that all k vendors pick different shares of one identity is

$$\geq \left(\frac{k-1}{k}\right)^k \approx e^{-1}$$

   Since the process is repeated for each of the m identities, the probability that we succeed on some identity is:

$$\geq 1 - (1 - e^{-1})^m = 1 - 2^{-O(m)}$$