CS 261 HW8

Prof. Jared Saia, University of New Mexico

Due April 25th

1. Exercise 7.2.8 ("What is the probability of these events when we randomly select a permutation of $\{1, 2, \ldots n\}$")

2. Exercise 7.2.38 ("A pair of dice is rolled in a remote location")

3. Exercise 7.3.12 ("A space probe near Neptune")

4. If there are n people in a room and m days in a year, what is the expected number of groups of 3 people that all share the same birthday?

5. The game of Match is played with a special deck of 27 cards. Each card has three attributes: color, shape and number. The possible color values are {red, blue, green}, the possible shape values are {square, circle, heart}, and the possible number values are {1, 2, 3}. Each of the $3 * 3 * 3 = 27$ possible combinations is represented by a card in the deck. A match is a set of 3 cards with the property that for every one of the three attributes, either all the cards have the same value for that attribute or they all have different values for that attribute. For example, the following three cards are a match: (3, red, square), (2, blue, square), (1, green, square).

   - If we shuffle the deck and turn over three cards, what is the probability that they form a match? Hint: Fix outcomes for the first two cards, then what is the probability that when the third card is revealed that it forms a match?
   - If we shuffle the deck and turn over n cards where $n \leq 27$, what is the expected number of matches, where we count each match separately even if they overlap? Note: The cards in a match do not need to be adjacent! Is your expression correct for $n = 27$? Hint: Use Linearity of Expectation!

6. Imagine n points are distributed independently and uniformly at random on the circumference of a circle that has circumference of length 1. Let the distance between a pair of points on the circumference be the length of the arc between them. Show that the expected number of pairs of points that are within distance $\theta(1/n^2)$ of each other is greater than 1.[1]

   Hint: Partition the circumference of the circle into n^2/k arcs of length k/n^2 for some constant k; then use linearity of expectation (as in the Birthday paradox example from class) to solve for the necessary k.

7. In this problem, you will use the following facts: (1) any integer can be uniquely factored into primes; (2) the number of primes less than any number m is $\theta(m/\log m)$ (Prime number theorem).

   (a) Show that for any integer x, x factors into at most $\log x$ primes. Hint: 2 is the smallest prime.

   (b) Let x be some positive integer and let p be a prime chosen uniformly at random from all primes less than or equal to m. Use the prime number theorem to show that the probability that $p|x$ is $O((\log x)(\log m)/m)$.

   (c) Now let x and y both be positive integers less than n, such that $x \neq y$, and let p be a prime chosen uniformly at random from all primes less than or equal to m. Using the previous result, show that the probability that $x \equiv y \pmod p$ is $O((\log n)(\log m)/m))$.

   (d) If $m = \log^2 n$ in the previous problem, then what is the probability that $x \equiv y \pmod p$. Hint: If you're on the right track, you should be able to show that this probability is "small", i.e. it goes to 0 as n gets large.

   (e) Finally, show how to apply this result to the following problem. Alice and Bob both have databases x and y where x and y have value no more than n, for n a very large number (think terabytes). They want to check to see if their databases are consistent (i.e. they want to check if they are the same) but Alice does not want to have to send her entire database to Bob. What is an algorithm Alice and Bob can use to check consistency with small probability of error, by sending a lot fewer bits? How many bits does Alice

[1] This problem has applications in efficient routing in networks. The circle is the address space, and the points represent keys of stored data items.

need to send to Bob as a function of n, and what is the probability of failure, where failure means that this algorithm says the databases are the same but in fact they are different?

8. There are n cats and n cubbies, both labelled 1 through n, where cubby i belongs to cat i. The cats come into the room of cubbies in order from 1 to n. Cat 1 is a bit of a rebel, and so enters a cubby selected uniformly at random. Then for $i = 2, 3, \ldots n$, cat i first checks cubby i. If it's empty, cat i enters cubby i. If it's occupied, cat i enters a cubby chosen uniformly at random from all empty cubbies.

   In this problem, you will calculate the value $f(n)$, the probability that the last cat enters its own cubby when there are n cats and cubbies. Note that $f(1) = 1$. The remaining problems will consider the case $n \geq 2$.

   (a) What is $f(2)$?

   (b) If cat 1 enters cubby 1, what is $f(n)$? If cat 1 enters cubby n, what is $f(n)$?

   (c) If cat 1 enters cubby 2, write $f(n)$ in terms of $f(x)$ for values of x less than n. Think in terms of smaller subproblems!

   (d) If cat 1 enters cubby i for *any* i, $1 < i < n$, write $f(n)$ in terms of values of f that depend on n and i.

   (e) Now write a recurrence relation for $f(n)$. Hint: Remember cat 1 enters cubby i with probability $1/n$. The right hand side of your recurrence should contain many $f(j)$ values weighted by their probabilities of occurrence based on your answers above.

   (f) Compute $f(3)$ and $f(4)$ using your recurrence to spot a trend, and to get a good guess of $f(n)$ for $n \geq 2$. Now use strong induction to prove your guess is correct. Don't forget to include BC, IH, and IS.