Massachusetts Institute of Technology
6.042J/18.062J: Mathematics for Computer Science
Professors Nancy Lynch and Srinivas Devadas

Final Exam Solutions
December 15, 1998

Final Exam Solutions

Problem 1. **(10 points)** Suppose that A and B are non-empty sets such that A is strictly contained in B. Mark each statement below **T**rue or **F**alse.

_____ Set A is finite iff set B is finite.
(False.)

_____ If B is finite, then there must exist a surjection mapping B to A.
(True.)

_____ If B is finite, then there must exist an injection mapping A to B.
(True.)

_____ If B is finite and there is a bijection mapping A to $B \setminus A$, then $|B| = 2|A|$.
(True.)

_____ There can not exist a bijection from A to B.
(False.)

Problem 2. **(15 points)** Consider the following procedure for computing the greatest common divisor (gcd) of a non-empty list of positive integers $(x_1, x_2, \ldots, x_n)$.

If all entries in the list are equal, then the procedure outputs the common value and terminates. Otherwise, the procedure selects two entries with different values and replaces the larger one by the difference.

We can model this procedure with a state machine as follows.

- The set of states Q consists of all lists of n positive integers.
- The set of start states Q_0 consists of the single state $(x_1, x_2, \ldots, x_n)$.
- The set δ contains all transitions of the form

$$(y_1, y_2, \ldots, y_i, \ldots, y_j, \ldots, y_n) \quad \rightarrow \quad (y_1, y_2, \ldots, y_i - y_j, \ldots, y_j, \ldots, y_n)$$

  where i and j are distinct numbers in the range 1 to n and $y_i > y_j$. (Note that the transition is written with y_i preceding y_j, but that the reverse order is also possible.)

(a) **(5 points)** Prove that the procedure terminates.

We construct a non-negative, decreasing derived variable. This implies that the machine reaches at state with no outgoing transitions. All such states are terminating states of the procedure.

Define the derived variable S to take on the value $y_1 + y_2 + \ldots + y_n$ in the state corresponding to the list $(y_1, y_2, \ldots, y_n)$. This derived variable is non-negative, since all y_i are positive. The derived variable is decreased by each transition, since the list entry y_i is decrease by y_j (a positive quantity) and all other list entries are unchanged.

(b) (10 points) Prove that the procedure computes the gcd of $x_1, x_2, \ldots, x_n$. State explicitly any number theoretic facts that you use, but do not prove them.

The proof uses the Invariant Theorem. We will prove the invariant that in state $(y_1, y_2, \ldots, y_n)$,

$$\gcd(y_1, y_2, \ldots, y_n) = \gcd(x_1, x_2, \ldots, x_n).$$

The invariant hold trivially in the start state. Futhermore, the invariant holds after a transition if it holds prior to a transition:

$$\begin{aligned}\gcd(y_1, y_2, \ldots, y_i - y_j, \ldots, y_j, \ldots, y_n) &= \gcd(y_1, y_2, \ldots, y_i, \ldots, y_j, \ldots, y_n)\\ &= \gcd(x_1, x_2, \ldots, x_n)\end{aligned}$$

The first step is justified by the following number-theoretic fact. For positive integers $z_1, z_2, \ldots, z_n$,

$$\gcd(z_1, z_2, \ldots, z_i, \ldots, z_j, \ldots, z_n) = \gcd(z_1, z_2, \ldots, z_i - z_j, \ldots, z_j, \ldots, z_n)$$

provided $z_i > z_j$. The second step in the argument above is justified by the assumption that the invariant holds prior to the transition. Therefore, the invariant holds by the Invariant Theorem.

This invariant implies that the output of the procedure is correct. If $(y_1, y_2, \ldots, y_n)$ is a terminating state, then

$$\begin{aligned}y_1 = y_2 = \ldots = y_n &= \gcd(y_1, y_2, \ldots, y_n)\\ &= \gcd(x_1, x_2, \ldots, n_n).\end{aligned}$$

The second step uses the invariant.

Problem 3. (12 points) Find a closed-form expression for $T(N)$ in each problem part. Your answer may make use of Θ notation. No proofs are required.

(a) (6 points)

$$T(n) = \sum_{i=1}^{n} \frac{i^2 + 6\sqrt{i}}{i\sqrt{i}}$$

$$\begin{aligned}
T(n) &= \sum_{i=1}^{n} \frac{i^2 + 6\sqrt{i}}{i\sqrt{i}} \\
&= \sum_{i=1}^{n} \frac{i^2}{i\sqrt{i}} + \sum_{i=1}^{n} \frac{6\sqrt{i}}{i\sqrt{i}} \\
&= \sum_{i=1}^{n} \sqrt{i} + \sum_{i=1}^{n} \frac{6}{i} \\
&= \Theta\left(n^{3/2} + \ln n\right) \\
&= \Theta\left(n^{3/2}\right)
\end{aligned}$$

(b) (6 points)

$$T(n) = \begin{cases} 1 & \text{for} \quad n \leq 5 \\ \frac{1}{2}T(\lfloor \frac{n}{2} + \log n \rfloor) + \frac{2}{n} & \text{for} \quad n > 5 \end{cases}$$

We apply Akra-Bazzi. First, we find the parameter p by solving the equation

$$\frac{1}{2} \cdot \left(\frac{1}{2}\right)^p = 1.$$

The solution is $p = -1$. Therefore, we have

$$\begin{aligned}
T(n) &= \Theta\left(n^{-1} \cdot \left(1 + \int_1^n \frac{\left(\frac{2}{u}\right)}{u^{-1+1}}\, du\right)\right) \\
&= \Theta\left(n^{-1} \cdot (1 + 2\ln n)\right) \\
&= \Theta\left(\frac{\ln n}{n}\right)
\end{aligned}$$

Problem 4. **(10 points)** Suppose that five points are placed inside a square with sides of length 1. Prove that the distance between some pair of points is at most $\sqrt{2}/2$.

Divide the square into four subsquares with edge-length $\frac{1}{2}$. By the pigeonhole principle, some two of the five points must be placed inside the same subsquare. The distance between these two points can be at most $\sqrt{2}/2$.

Problem 5. **(20 points)** Solve the following counting problems. Simplified answers are not required.

(a) **(4 points)** In how many ways can the letters in BOOKKEEPER be arranged?

$$\frac{10!}{2!\ 2!\ 3!}$$

(b) **(4 points)** In how many ways can 160 distinguishable students be assigned to 8 indistinguishable recitations so that every recitation has 20 students?

If the recitations were distinguishable, then the answer would be $\frac{160!}{20!^8}$ by the multinomial formula. Since the recitations are identical, the answer is $\frac{160!}{20!^8\ 8!}$ by the quotient rule.

(c) **(4 points)** In how many ways can 6.042 TA John Dunagan assign the following objects to three desk drawers,

- 7 tasty treats,
- 27 thumb tacks, and
- 1 can of theoretically-enriched Spam,

so that the Spam shares a drawer with no tasty treats and at least 19 thumb tacks? (Desk drawers are distinguishable, but tasty treats and thumbtacks are not.)

Theory Pig (R. I. P.) and the tacks can be assigned to a drawer in 3 ways. The treats can be divided between the other two drawers in $\binom{7+1}{1} = 8$ ways. The 8 remaining tacks can be divided among the three drawers in $\binom{8+2}{2} = 45$ ways. Overall, there are $3 \cdot 8 \cdot 45$ assignments.

(d) **(4 points)** How many ways can 100 identical balls be placed in 10 distinguishable bins so that every bin contains an odd number of balls?

First, we put 1 ball into each bin. Then we pair up the 90 remaining balls into 45 pairs. Then we assign the 45 pairs to bins. This can be done in $\binom{45+9}{9}$ ways.

(e) (4 points) How many paths are there from $(0,0)$ to $(4,4)$ consisting of 10 steps, each of which is either north, south, east, or west?

Either there are 5 north steps, 1 south step, and 4 east steps, or else there are 5 east steps, 1 west step, and 4 north steps. Thus, there are

$$2 \cdot \frac{10!}{5!\ 1!\ 4!}$$

paths in total.

Problem 6. **(8 points)** Some MIT students go for dinner at the Border Cafe in Harvard Square. They want to order enough appetizers to feed everyone. Chili con queso will feed one person, Tejas chicken will feed two people, and quesadillas will feed three people. Write the generating function for the number of different appetizer orders that will feed exactly n people.

$$\frac{1}{1-x} \cdot \frac{1}{1-x^2} \cdot \frac{1}{1-x^3}$$

Problem 7. **(10 points)** The following identity holds for non-negative integers n, j, and k such that $n \geq j+k$.

$$\binom{n}{j}\binom{n-j}{k} = \binom{n}{j+k}\binom{j+k}{k}$$

(a) (5 points) Prove the identity algebraically.

$$\begin{aligned}
\binom{n}{j}\binom{n-j}{k} &= \frac{n!}{j!\ (n-j)!} \cdot \frac{(n-j)!}{k!\ (n-j-k)!} \\
&= \frac{n!}{j!\ k!\ (n-j-k!)} \\
&= \frac{n!}{(j+k)!\ (n-j-k)!} \cdot \frac{(j+k)!}{k!\ j!} \\
&= \binom{n}{j+k}\binom{j+k}{k}
\end{aligned}$$

(b) (5 points) Prove the identity combinatorially.

Both sides of the equation count the number of ways to choose distinguishable, disjoint subsets of sizes j and k from a set of n elements. On the left side, we first choose the j-element subset from the whole n-element set, and then choose the k-element subset from the remaining $n - j$ elements. On the right side, we first choose $j + k$ elements from the whole set. From these $j + k$ elements, we then choose the k-element set; the remainders form the j-element set.

Problem 8. **(5 points)** Suppose that I roll a die, but conceal the result from you. You then repeatedly guess at the number that I rolled, and I will tell you whether each guess is right or wrong. No proofs are required for this problem.

(a) (2 points) Suppose that you guess 1, 2, 3, 4, 5, and then 6 in that order. What is the expected number of guesses that you need to get the right number?

From the definition of expected value, the answer is

$$\frac{1}{6} \cdot (1 + 2 + 3 + 4 + 5 + 6) \quad = \quad \frac{7}{2}.$$

(b) (3 points) Suppose that you repeatedly guess numbers in the range 1 to 6 uniformly and mutually independently at random. What is the expected number of guesses that you need to get the right number?

Let the expected number of guesses be X. After a wrong guess, the expected number of guesses is the same as before. The rules of the game imply the following equation.

$$X \quad = \quad \frac{1}{6} \cdot 1 + \frac{5}{6} \cdot (1 + X)$$

By solving the equation, we find that the expected number of guesses is $X = 6$.

Problem 9. **(5 points)** A spider eats three flies a day. Each fly attempting to pass through his web has an even chance of becoming entangled. The spider eats the flies that become entangled in his web until he fills his quota. After his quota is filled, the spider allows flies that become entangled to work themselves free and escape.

Wilbur the Fly is about to attempt a flight through the spider's web. If five flies have already made the attempt today, then what is the probability that Wilbur survives?

(Assume that the flies become entangled mutually independently.)

Among the first five flies, the number that become entangled is either less than half (≤ 2) or more than half (≥ 3). These two events are equally likely, since each fly has an even chance of becoming entangled. Therefore, the spider has an even chance of filling his quota with the first five flies.

The probability that Wilbur survives is computed below.

$$\begin{aligned}
\Pr(\text{fly lives}) &= 1 - \Pr(\text{Wilbur dies}) \\
&= 1 - \Pr(\text{quota unfilled} \cap \text{Wilbur entangled}) \\
&= 1 - \Pr(\text{quota unfilled}) \cdot \Pr(\text{Wilbur entangled}) \\
&= 1 - \frac{1}{2} \cdot \frac{1}{2} \\
&= \frac{3}{4}
\end{aligned}$$

In the first step, we switch to computing the probability of the complementary event. The second step uses the fact that Wilbur the Fly dies only if the spider's quota is unfilled and the fly becomes entangled. The third step follows from the fact that flies become entangled mutually independently. The fourth step is substitution, and the last step is simplification.

Problem 10. (10 points) Smith and Wesson are shooting at a target. Smith's chance of hitting the target is double that of his friend Wesson. If they each take a shot, then the probability that at least one of them hits the target is $\frac{1}{2}$. What is the probability that Wesson hits the target with a single shot?

(The event that one hits the target is independent of the event that the other hits the target.)

Let S be the event that Smith hits, and let W be the even that Wesson hits. We are given two facts:

$$\begin{aligned}
\Pr(S) &= 2\Pr(W) \\
\Pr(S \cup W) &= \frac{1}{2}
\end{aligned}$$

Expanding the second equation using inclusion-exclusion gives:

$$\begin{aligned}
\Pr(S) + \Pr(W) - \Pr(S \cap W) &= \frac{1}{2} \\
\Pr(S) + \Pr(W) - \Pr(S) \cdot \Pr(W) &= \frac{1}{2} \\
2\Pr(W) + \Pr(W) - 2\Pr(W) \cdot \Pr(W) &= \frac{1}{2}
\end{aligned}$$

The second equation follows from independence, and the third equation follows by substituting in the first fact listed above. The root of this quadratic that makes sense as a probabilty is $\frac{3}{4} - \frac{1}{4}\sqrt{5} = 0.19\ldots$.

Problem 11. **(10 points)** A box initially contains n black balls. The following procedure is repeated. A ball is drawn from the box at random. If the ball is white, then it is returned to the box. If the ball is black, then with probability p a white ball is returned to the box and with probability $1-p$ a black ball is returned to the box. What is the expected number draws needed to fill the box with white balls?

Suppose that the box contains k white balls. The probability that a black ball is drawn and replaced by a white ball is $\frac{n-k}{n} \cdot p$. Therefore, the expected number of draws to introduce one more white ball is $\frac{n}{p(n-k)}$. By linearity of expectation, the expected number of draws required to fill the box with white balls is:

$$\begin{aligned}
\sum_{k=0}^{n-1} \frac{n}{p(n-k)} &= \frac{n}{p} \cdot \sum_{k=0}^{n-1} \frac{1}{n-k} \\
&= \frac{n}{p} \cdot \sum_{j=1}^{n} \frac{1}{j} \\
&= \frac{nH_n}{p}
\end{aligned}$$

The second step uses the substituion $j = n - k$.

Problem 12. **(5 points)** The number of people that enter Filene's Basement in the last hour before closing on Christmas Eve is a random variable taking on an integer value in the range 0 to 100 uniformly.

Each shopper picks at least one present to buy. After selecting each present, a shopper is equally likely to continue shopping and to go pay at the checkout register. Shoppers' decisions to continue buying are mutually independent.

The checkout clerk throws two fair, independent dice for each item and then charges a number of dollars equal to the sum.

What is the expected total amount of money spent by these last-minute shoppers at Filene's Basement?

Applying Wald's Theorem twice, the expected amount of money spent is the product of the expectations of the three random variables described in the problem.

$$\begin{aligned}\text{Ex}(\text{money spent}) &= \text{Ex}(\text{num people}) \cdot \text{Ex}(\text{num gifts}) \cdot \text{Ex}(\text{price}) \\ &= 50 \cdot 2 \cdot 7 \\ &= 700\end{aligned}$$

Problem 13. **(10 points)** Solve the following problems concerning grades at MIT.

(a) **(5 points)** Suppose that the average GPA at MIT is 4.0. Give a tight upper bound on the probability that a random student's GPA is 4.5 or higher.

From Markov's Inequality,

$$\begin{aligned}\Pr(\text{GPA} \geq 4.5) &\leq \frac{4.0}{4.5} \\ &= \frac{8}{9}.\end{aligned}$$

(b) **(5 points)** Suppose in addition that the standard deviation of GPA's at MIT is 0.2. Give a tight upper bound on the probability that a random student's GPA is 4.5 or higher, taking this new information into account.

From the one-sided Chebyshev inequality,

$$\begin{aligned}\Pr(\text{GPA} - 4.0 \geq 0.5) &\leq \frac{0.2^2}{0.2^2 + 0.5^2} \\ &= \frac{4}{29}\end{aligned}$$

Problem 14. **(10 points)** A primality testing algorithm takes as input an integer $n \geq 2$ and outputs either PRIME or COMPOSITE. The algorithm uses random numbers in a clever way to give a much better run-time. An unfortunate consequence is that the algorithm only ouputs a correct answer with probability $\frac{2}{3}$. The answers produced by multiple runs of the algorithm are mutually independent.

(a) (7 points) Suppose that we run the algorithm Q times. Use the Chernoff inequality to lower bound the probability that the most common answer is correct. You may treat the approximation $\ln\left(\frac{3}{2}\right) \approx \frac{2}{5}$ as an equality.

The expected number of wrong answers is $\frac{1}{3}Q$. Now we apply the Chernoff bound to upper bound the probability that the wrong answer is actually the most common answer.

$$\begin{aligned}
\Pr\left(\text{num wrong} \geq \frac{1}{2} \cdot Q\right) &= \Pr\left(\text{num wrong} \geq \frac{3}{2} \cdot \text{Ex(num wrong)}\right) \\
&\leq e^{-\left(\frac{3}{2}\right)\cdot\left(\ln\frac{3}{2}-1+\frac{2}{3}\right)\cdot\text{Ex(num wrong)}} \\
&\approx e^{-\frac{1}{10}\cdot\text{Ex(num wrong)}} \\
&= e^{-\frac{1}{30}Q}
\end{aligned}$$

Consequently, the most common answer is correct with probabiliy at least

$$1 - e^{-\frac{1}{30}Q}.$$

(b) (3 points) How many times must the algorithm be run to ensure that the most common answer is correct with probability at least $1 - \frac{1}{n}$? Give the smallest answer that you can justify.

Suppose that we run the algorithm at least $30 \ln n$ times. Then the probability that the most common answer is incorrect is at most:

$$\begin{aligned}
e^{-\frac{1}{30}\cdot 30\ln n} &= e^{-\ln n} \\
&= \frac{1}{n}
\end{aligned}$$

Therefore, the most common answer is correct with probability at least $1 - \frac{1}{n}$.

Problem 15. (10 points) The superhero Bozonmann is in hot pursuit of the archfiend Doctor Phermion. During each minute one of three events occurs.

- With probability p, Bozonmann pauses to recharge his powerpack, allowing the Doctor to get one step farther ahead.
- With probability q, the evil Doctor Phermion zaps his pursuer with a gluon gun, forcing him to fall two steps back.

- With probability $1 - p - q$ our resolute hero, Bozonmann, advances a step closer to the notorious villain.

If Bozonman comes within two steps of Phermion, then our hero can reach out and capture his foe immediately.

(a) **(5 points)** Write a recurrence and base cases describing P_n, the probability that Bozonmann catches Doctor Phermion, given that they start n steps apart. Do not solve the recurrence.

$$\begin{aligned} P_0 &= 1 \\ P_1 &= 1 \\ P_2 &= 1 \\ P_n &= pP_{n+1} + qP_{n+2} + (1-p-q)P_{n-1} \qquad (n > 2) \end{aligned}$$

(b) **(5 points)** Write a recurrence and base cases describing E_n, the expected number of minutes until Bozonmann catches up to the evil Doctor, given that they start n steps apart. Do not solve the recurrence.

$$\begin{aligned} E_0 &= 0 \\ E_1 &= 0 \\ E_2 &= 0 \\ E_n &= 1 + pE_{n+1} + qE_{n+2} + (1-p-q)E_{n-1} \qquad (n > 2) \end{aligned}$$