Massachusetts Institute of Technology
6.042J/18.062J: Mathematics for Computer Science
Professors Tom Leighton and Santosh Vempala

Handout 37
December 9, 1997

Practice Final Solutions

Problem 1. **(a)** A vampire arrives in Boston at day 0 and starts biting people at day 1. People bitten become vampires themselves. New vampires bite two persons on the next day after they were bitten and four persons every day afterwards. Write an expression for the number of vampires in n days for $n \geq 2$, assuming the first vampire is a newly made vampire. *Do not solve the recurrence relation.*

Answer:

Let $V(n)$ denote the number of vampires on the nth day. We have the initial conditions $V(0) = 1$ and $V(1) = 3$.

The number of vampires on day $n+1$ is the number of vampires on day n plus all the people that were bitten on day n.

$$V(n+1) = V(n) + 2(V(n) - V(n-1)) + 4V(n-1) = 3V(n) + 2V(n-1)$$

(b) Consider the recurrence $T(n) = T(n/2) + T(n/4) + n$ and $T(1) = 1$. Prove that $T(n) \leq 4n$, where n is a power of 2.

Answer:

The base cases are $T(1) = 1 < 4$, and $T(2) = T(1) + 2 = 3 < 8$.

The inductive step is

$$T(2n) = T(n) + T(n/2) + 2n \leq 4n + 2n + 2n = 8n \leq 4 \cdot 2n.$$

The induction is complete.

(c) Prove that $T(n) = \Theta(n)$ in the recurrence above.

Answer:

It suffices to show that $T(n) \geq n$. Indeed, the base cases are $T(1) = 1 \geq 1$, $T(2) = 3 \geq 2$.

$$T(2n) = T(n) + T(n/2) + n \geq 2n + n/2 \geq 2n.$$

The induction is complete.

Parts (b) and (c) establish that $T(n) = \Theta(n)$.

Problem 2. **(a)** Six people communicate via radio frequency with one another. Each pair of people choose one of **two** available frequencies with which to communicate. Show that there is a subset of 3 people who are each using the identical frequency with which to communicate with others in the subset.

Answer:

Each person communicates with five other persons. By the pigeonhole principle, this person (say a) communicates with another three persons (say b,c,and d) using the same frequency. If either of the pairs $(b,c),(b,d),(c,d)$ uses this frequency then we have three persons using it to communicate with each other and we are done. If not, then the pairs $(b,c),(b,d),(c,d)$ all use the other frequency, so we are done again.

(b) Now suppose there are seventeen people communicating via radio frequency with one another. Each pair of people choose one of **three** available frequencies with which to communicate. Show that there is a subset of 3 people who are each using the identical frequency with which to communicate with others in the subset.

Answer:

Now each person communicates with sixteen other people, and by pigeonhole uses the same frequency with at least six of them. If any pair among these six persons uses that same frequency then we are done. If not, then these six persons communicate with one another using two different frequencies so we are back in part (a) of the problem.

Problem 3. In this problem, A, B and C denote arbitrary sets. f and g are functions as specified in each part and I_A denotes the identity function on the set A. For each statement indicate whether it is **T**rue or **F**alse. No explanation is required.

True: There exists an injection (one-to-one) $f : A \to B$ if and only if there exists a surjection (onto) $g : B \to A$.

False: Let $f : A \to B$ and $g : B \to A$. $f \circ g = I_B$ implies $g \circ f = I_A$.

True: Let $f : A \to B$ and $g : B \to C$. If both f and g are injections, then $g \circ f$ is an injection.

False: Let $f : A \to B$ and $g : B \to C$. $g \circ f$ is a surjection implies both f and g are surjections.

False: Let $f : A \to B$ be an injection. If $|B| \leq |A|$, then f is a bijection.

Problem 4. Answer **T** or **F**:

True: Adding k edges to an undirected graph increases its chromatic number (the number of colors needed to color the vertices) by at most k.

False: The number of shortest paths in an n-vertex undirected graph is $O(n^2)$. (Note: there may be more than one shortest path between two vertices.)

False: The number of length-3 paths in an n-vertex undirected graph is $O(n^2)$.

False: The number of length-3 paths in an n vertex tree is $O(n)$

True: Reversing the directions of all edges in a directed graph leaves the number of simple directed cycles unchanged.

False: If an undirected graph has the degree of every vertex divisible by a prime p, then p must evenly divide the number of edges.

Problem 5. Find the number of permutations of the letters $\{a, b, c, d, e, f, g\}$ that do not contain "bag" or "def".

Let ALL be the set of all permutations, BAG the set of permutations containing "bag" and DEF the permutations containing "def". We want to determine the size of $ALL - (BAG \cup DEF)$.

We have $|ALL| = 7!$. To count the elements of BAG, notice that the set of permutations containing "bag" is in bijection with the set of permutations of $\{a, c, d, f, bag\}$. Therefore $|BAG| = 5!$. Analogously $|DEF| = 5!$ and $|DEF \cap BAG| = 3!$.

By inclusion exclusion:

$$|ALL - (BAG \cup DEF)| = |ALL| - |BAG| - |DEF| + |BAG \cap DEF| = 7! - 2 \cdot 5! + 3! = 4806$$

Problem 6. Consider the following recursive algorithm for coloring a graph G with natural numbers. For a vertex v, $G - v$ denotes the graph you get by removing v and all its incident edges from G.

```
Color(G)
    if G has 1 vertex v
        set the color of v to be 0
    else
        choose a vertex v of G
        Color(G − v)
        set the color of v to be the minimum n ∈ ℕ for which no neighbor of G has color n
```

Prove that this algorithm legally colors G with a number of colors no more than 1 plus the maximum degree of a vertex in G.

Answer:

We perform induction on the number of vertices of G.

$P(n) = \{$ If G has n vertices, Color(G) legally colors G with no more than one plus the maximum degree of a vertex in G $\}$

BASE CASE: $n = 1$, in which case the graph is colored with one color and the statement is true.

INDUCTIVE STEP: Consider now a graph with $n + 1$ vertices. Then $G - v$ has n vertices and the algorithm colors it legally with at most $1 + d$ colors, where d is the maximum degree of $G - v$.

If v is connected to some maximum degree vertex, then the maximum degree of G is $d + 1$, and since v is colored with at most one new color, the total number of colors for G is at most $d + 2 = 1$+maximum degree of G.

If v is not connected to some maximum degree vertex, then we can distinguish the following cases:

a) Either v has degree at least $d + 1$, in which case the maximum degree of G is at least $d + 1$, and at most $d + 2$ colors are required, or

b) v has degree at most d, which means that it can be colored with one of $d + 1$ colors. Therefore, the maximum number of colors remains $d + 1$.

Problem 7. For a certain test for a newly discovered virus, called JABOLA, it is known that 99% of the people carrying the virus will test positive, whereas only 1% of the people not carrying the virus will test positive. Suppose 0.5% of the population carries the JABOLA virus. Given that a person picked at random tests positive, what is the conditional probability that this person carries the JABOLA virus?

Answer: Let $T = \{\text{test is positive}\}$, $V = \{\text{person has the JABOLA virus}\}$. We are given that $P(T|V) = 0.99$, $P(T|\bar{V}) = 0.01$ and $P(V) = 0.005$.

$$P(V|T) = \frac{P(T|V)P(V)}{P(T)} = \frac{P(T|V)P(V)}{P(T|V)P(V) + P(T|\bar{V})P(\bar{V})}$$

The above expression evaluates to 0.3322.

Problem 8. A lake contains n trout. 100 of them are caught, tagged and returned to the lake. Later another 100 trout are caught, selected independently from the first 100.

(a) Assume that the second 100 trout are selected without replacement. Write an expression (in n) for the exact probability that these trout are found to contain exactly 7 tagged ones.

Answer: There are $\binom{100}{7}$ ways to pick tagged trout, $\binom{n-100}{93}$ to pick untagged trout, and $\binom{n}{100}$ ways to pick 100 trout out of the lake. Since all trout are equally likely to be caught, the probability of catching 7 tagged is

$$\frac{\binom{100}{7}\binom{n-100}{93}}{\binom{n}{100}}.$$

(b) Now consider selecting the second 100 trout with replacement. That is, you repeat 100 times the following steps: select at random a trout in the lake, check if the trout is tagged and return it to the lake before selecting a new trout.

What is the probability that exactly 7 of the selected trout are tagged?

Answer: The probability of catching a tagged trout is $p = 100/n$. The probability of catching 7 tagged and 93 untagged in a given sequence is $p^7(1-p)^{93}$. Thus the probability of catching 7 tagged trout (in any sequence) is

$$\binom{100}{7}\left(\frac{100}{n}\right)^7\left(1-\frac{100}{n}\right)^{93}.$$

Problem 9. A bridge hand consists of 13 cards. The *suit-distribution* of a hand is the number of cards in each suit. For example, one possible suit-distribution is 4 spades, 3 hearts, 3 diamonds, and 3 clubs.

(a) How many different suit-distributions are there?

Answer: This is equivalent to placing 13 balls (corresponding to each card in a hand) to 4 bins (corresponding to each suit). Thus the number of distributions is

$$\mathrm{BP}(13,4) = \binom{13+4-1}{4-1} = 560.$$

(b) How many different suit-distributions are there assuming that you get at least one card in each suit?

Answer: By using the same correspondence as in part (a), the number of distributions is

$$\mathrm{BP1}(13,4) = \binom{13-1}{4-1} = 220.$$

Problem 10. Suppose the entries a, b, c, d of a 2×2 matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

are chosen independently at random according to the uniform distribution on $\{0, 1, 2, 3, 4\}$. Let $X = \det(M) = ad - bc$.

(a) What is $E(X)$?

Using linearity of expectation

$$E(X) = E(ad + bc) = E(ad) - E(bc) = 0$$

because ad and bc are identically distributed.

(b) What is $E(X \mid b = 1)$?

Using linearity of expectation and the rule $E(XY) = E(X)E(Y)$ for independent random variables we have

$$\begin{aligned} E(X|b=1) &= E(ad|b=1) - E(bc|b=1) \\ &= E(ad) - E(c) \\ &= E(a)E(d) - E(c) \\ &= 2 \cdot 2 - 2 = 2 \end{aligned}$$

(c) What is $Var(X)$?

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= E(X^2) \\ &= E((ad-bc)^2) \\ &= E(a^2d^2 + b^2c^2 - 2abcd) \\ &= E(a^2)E(d^2) + E(b^2)E(c^2) - 2E(a)E(b)E(c)E(d) \\ &= 6 \cdot 6 + 6 \cdot 6 - 2 \cdot 2^4 = 40 \end{aligned}$$

Problem 11. We consider an algorithm based on a given rooted, undirected tree, oriented with the root on the top. Every node in the tree has an associated *value* variable. Initially, the *value* variables of the leaves contain natural number input values, and the *value* variables of the root and other internal nodes contain a special "undefined" value $\perp$. The algorithm is supposed to end up with the root's *value* variable containing the sum of all the *values* of all the leaves. This is accomplished by propagating summarized *value* information up the tree structure.

If $value(v) = \perp$

and for every $w \in children(v)$, $value(w) \in \mathbb{N}$

Then $value(v) := \sum_{w \in children(v)} value(w)$

The property we are interested in is that $value(r)$ for the root r, if it ever gets defined, is equal to the sum of the *values* of all the leaves. Formally:

Inv1: If $value(r) \in \mathbb{N}$, where r is the root, then $value(r) = \sum_{l \in leaves} value(l)$.

(a) Describe a state machine for this algorithm.

Answer:

This algorithm is described precisely by the following state machine. This machine assumes a fixed undirected tree $T = (V, E)$, rooted at node r. The state components are:

for each $v \in V$:
  $value(v) \in \mathbb{N} \cup \{\perp\}$;

Initially $value(v) \in \mathbb{N}$ if v is a leaf, and $value(v) = \perp$ if v is an internal node.

The state transitions are described by the algorithm.

(b) State and prove a strengthened version of Inv1, which can be proved by induction.

Answer:

Inv2:

If v is any node of T and $value(v) \in \mathbb{N}$, then $value(v)$ is the sum of the *value* of all the leaves that are descendants of v.

BASE CASE: For internal nodes, $value(v) \notin \mathbb{N}$, so Inv2 is vacuously true. For leaves, the only leaf below v is v itself, so $value(v)$ alone gives the correct sum.

INDUCTIVE STEP: When the algorithm is performed, the sums of all the *values* of v's children are added. The inductive hypothesis says that each of these *values* is the sum of the leaves below the corresponding child. Adding them all up gives the sum of all the leaves below v.

Problem 12. **(a)** At Marilyn's Casino, a new gambling game has been introduced. The game features 3 fair mutually independent 6-sided dice. To play, the contestant picks a number in the interval $[1, 6]$. The contestant wins \$1 if his/her number comes up on one or more of the dice and loses \$1 otherwise. This game has proved to be very popular with the gamblers such as Bobo who is a regular. Your job is to figure out the probability of winning the game.

Answer:

$\Pr(\text{WIN}) = 1 - (5/6)^3 = \frac{91}{216}$

(b) In an effort to attract even more gamblers to her casino, Marilyn is thinking about a variation of the game in which the bettor also picks a different second chance number and the bettor wins if the first pick comes up at least once or the second pick comes up at least twice. Compute the probability of winning this game.

Answer:

The probability that the second chance number comes up twice is

$\Pr(\text{2nd chance}) = \binom{3}{2}(1/6)^2(5/6) = \frac{15}{216}$

The probability that the first number comes up once *and* the second chance number comes up twice is $\frac{3}{216}$.

By inclusion/exclusion the probability of winning is

$\Pr(\text{WIN}) = \frac{91}{216} + \frac{15}{216} - \frac{3}{216} = \frac{103}{216}$

Problem 13. What is your expected return in each of the following games? You may assume the events are independent in each part.

(a) You get to roll six ordinary six-sided dice. You win the number of dollars equal to the total number of spots that appear on all the dice.

Answer:

The expected return from a single dice is 7/2. By linearity of expectation the total expected return is $6\frac{7}{2} = 21$.

(b) You roll one ordinary die. You win a number of dollars that is equal to six times the number of spots.

Answer:

Again by linearity of expectation, the expected return is 21.

(c) You toss a fair coin six times. For each time it comes up heads, you toss an ordinary six-sided die. You win the number of dollars equal to the total number of spots that appear on all the dice.

Answer:

Let N be the number of heads in the six coin tosses, X_i be the number of spots in the ith die toss. The expected return is $\sum_{i=1}^{N} X_i$

N is binomial and it's mean is 3, while each X_i has mean 7/2, so by Wald's identity, the expected return is $\frac{21}{2}$.

(d) You toss a fair coin repeatedly, until it comes up "tails" for the first time. (This could involve infinitely many tosses, if "tails" never comes up.) For each time the coin comes up heads, you toss an ordinary six-sided die. You win the number of dollars equal to the total number of spots that appear on all the dice.

Answer:

Again we have $\sum_{i=1}^{N} X_i$, but now N is geometric and has mean 2. Therefore, by Wald's identity the expected return is $2\frac{7}{2} = 7$.

(e) You roll two six-sided dice. You win the product of the number of spots that appear on the two dice.

Answer:

$E[X]^2 = \frac{49}{4}$

(f) You roll one ordinary die. You win a number of dollars that is the square of the number of spots.

Answer:

$E[X^2] = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$

Problem 14. There is a newly developed test called EQ (*Emotional Quotient*), that is used to decide whether or not you will succeed in life. EQ is expressed as a nonnegative integer. The average EQ over all humans is 1000.

(a) Knowing only the average EQ, find an upper bound for the fraction of humans that have EQ$\geq$ 1500?

Answer:

We use Markov's inequality $\Pr(X \geq cE[X]) \leq 1/c$, with $c = 1.5$, to obtain the upper bound 2/3.

(b) Assuming that the standard deviation of EQ is 100, what fraction of humans can have EQ$\geq$ 1500?

Answer:

We use Chebychev's inequality $\Pr(X \geq E[X] + c\sigma_X) \leq \frac{1}{1+c^2}$ with $c = 5$ to get an upper bound of $\frac{1}{26}$