Massachusetts Institute of Technology | Handout 54
6.042J/18.062J: Mathematics for Computer Science | May 24, 1996
Prof. David Karger

Final Exam

This is an OPEN BOOK Final: you may refer to books, handouts, or your own course notes. (Calculators are not allowed.) You may use in your problem solutions any of the mathematical results given in lectures, recitations, or the course text. There are twelve problems of roughly equal weight to complete within three hours.

Write all solutions in the exam book. You may use the back of the pages if you need more space.

Good luck!

Name: ______________________________

Problem	Grade
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
Total	

Problem 1. Let A be a set and R be a binary relation on A. For each of the statements in the box on the left below, indicate the number of the formula in the box on the right expressing that property. (Recall that the *composite* of two binary relations R and S, written $S \circ R$, is the relation $\{(a,c) \mid \text{there exists some } b \text{ such that } (a,b) \in R \text{ and } (b,c) \in S\}$. R^T is the *transpose* of R, i.e., $R^T = \{(a,b) \mid (b,a) \in R\}$, and I denotes the *identity* relation on A, i.e., $I = \{(a,a) \mid a \in A\}$.)

a.	R is transitive	2
b.	R is antisymmetric	3
c.	R is reflexive	5
d.	R is symmetric	1
e.	R is a bijection	6
f.	R is an equivalence relation	7
g.	R is a partial order	4

1.	$R = R^T$
2.	$R \circ R \subseteq R$
3.	$R \cap R^T \subseteq I$
4.	$R \cap R^T \subseteq I \subseteq R \circ R \subseteq R$
5.	$I \subseteq R$
6.	$R \circ R^T = R^T \circ R = I$
7.	$I \subseteq R \circ R \subseteq R \subseteq R^T \cap R$

Problem 2. Recurrences.

(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 $T(n) \leq 2n$.

Answer:

This part is wrong. However, we can show that $T(n) \leq 4n$. Indeed 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 3. **(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 4. In this problem, A, B and C denote arbitrary sets. f and g are functions as specified in each part and I denotes the identity function. For each statement indicate whether it is **T**rue or **F**alse. No explanation is required.

- **T**rue: There exists an injection (one-to-one) $f : A \to B$ if and only if there exists a surjection (onto) $g : B \to A$.
- **F**alse: Let $f : A \to B$ and $g : B \to A$. $f \circ g = I$ implies $g \circ f = I$.
- **T**rue: Let $f : A \to B$ and $g : B \to C$. If both f and g are injections, then $g \circ f$ is an injection.
- **F**alse: Let $f : A \to B$ and $g : B \to C$. $g \circ f$ is a surjection implies both f and g are surjections.
- **F**alse: Let $f : A \to B$ be an injection. If $|B| \leq |A|$, then f is a bijection.

Problem 5. Answer **T** or **F**:

- **T**rue: The mapping $f_a : Z_p \to Z_p$ defined by

$$f_a(x) = axp$$

  is a bijection for any nonzero $a \in Z_p$, when p is prime.
- **F**alse: For all a, b in $\mathbb{N}$, $\gcd(a^2b, ab^2) = ab$.
- **T**rue: If $a^{-1}b$ exists, then so does $b^{-1}a$.
- **F**alse: Let $\lambda(n)$ denote the number of distinct prime factors of n. Then $\lambda(n) = \Omega(\lg \lg n)$.

Problem 6. Answer **T** or **F**:

- **T**rue: Adding k edges to an undirected graph increases its chromatic number (the number of colors needed to color the vertices) by at most k.
- **F**alse: 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.)
- **F**alse: The number of length-3 paths in an n-vertex undirected graph is $O(n^2)$.
- **F**alse: The number of length-3 paths in an n vertex tree is $O(n)$
- **T**rue: Reversing the directions of all edges in a directed graph leaves the number of simple directed cycles unchanged.
- **F**alse: 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 7. 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 8. 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 $n \in \mathbb{N}$ 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 9. 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 10. 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 11. 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 12. Suppose the entries a, b, c, d of a 2×2 matrix

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

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 13. 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 *value*s of all the leaves. This is accomplished by propagating summarized *value* information up the tree structure.

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 transitions are:

$summarize(v)$, v an internal node
  Precondition:
    $value(v) = \perp$
    for every $w \in children(v)$
      $value(w) \in \mathbb{N}$
  Effect:
    $value(v) := \sum_{w \in children(v)} value(w)$

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 *value*s of all the leaves. Formally:

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

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 $summarize(v)$ is performed, the sums of all the *value*s of v's children are added. The inductive hypothesis says that each of these *value*s is the sum of the leaves below the corresponding child. Adding them all up gives the sum of all the leaves below v.