Massachusetts Institute of Technology
6.042J/18.062J, Fall '05: Mathematics for Computer Science — December 21
Prof. Albert R. Meyer and Prof. Ronitt Rubinfeld — revised December 22, 2005, 1118 minutes

Final Examination

Your name: ______________________________

Circle the name of your Tutorial Instructor: David Hanson Jelani Sayan

This exam is **closed book**, but you may have a one page, two-sided personal crib sheet.

There are 12 problems totaling 200 points. Write your solutions in the space provided, if need be running over to the back of the page. Total time is 180 minutes. GOOD LUCK!

DO NOT WRITE BELOW THIS LINE

Problem	Points	Grade	Grader
1	15	9	
2	20	18	
3	15	10	
4	15	15	
5	15	15	
6	15	15	
7	20	15	
8	20	7	
9	15	15	
10	15	15	
11	15	15	
12	20	14	
Total	200	81.5	

Problem 1 (15 points). Induction

Suppose $S(n)$ is a predicate on natural numbers, n, and suppose

$$\forall k \in \mathbb{N}\; S(k) \longrightarrow S(k+2). \tag{1}$$

If (1) holds, some of the assertions below must *always* (**A**) hold, some *can* (**C**) hold but not always, and some can *never* (**N**) hold. Indicate which case applies for each of the assertions by **circling** the correct letter.

(a) (3 points) **A** **N** **C** $(\forall n \leq 100\; S(n)) \wedge (\forall n > 100\; \neg S(n))$ ✓

(b) (3 points) **A** **N** **C** $S(1) \longrightarrow \forall n\; S(2n+1)$ ✓

(c) (3 points) **A** **N** **C** $[\exists n\; S(2n)] \longrightarrow \forall n\; S(2n+2)$ ✓

(d) (3 points) **A** **N** **C** $\exists n\, \exists m > n\, [S(2n) \wedge \neg S(2m)]$ ✗

(e) (3 points) **A** **N** **C** $[\exists n\; S(n)] \longrightarrow \forall n\, \exists m > n\; S(m)$ ✗

There exists an n, S(n) that implies for all n there exists an m greater than n. S(m)

9/15

Problem 2 (20 points). State Machines

We will describe a process that operates on sequences of numbers. The process will start with a sequence that is some *permutation* of the length $6n$ sequence

$$(1, 2, \ldots, n, 1, 2, \ldots, 2n, 1, 2, \ldots, 3n).$$

$3n$ $\quad 1 \overset{3}{\underset{3n}{-}} \overset{2}{\underset{2n}{-}} \overset{1}{-} 3n$

(a) (5 points) Write a simple formula for the number of possible starting sequences.

$2/5$ $\quad \dfrac{(6n)!}{n^2 (3!)(2!)}$

$(6n)! \leftarrow$ # of permutations of a unique 6n-length string

$n^2(3!)(2!) \leftarrow$ # of n-pairs = 3n of range 1-n which are the same & 2n of range n+1-2n which are the same

If $(s_1, \ldots, s_k)$ is a sequence of numbers, then the i and jth elements of the sequence are *out of order* if the number on the left is strictly larger the number on the right, that is, if $i < j$ and $s_i > s_j$. Otherwise, the ith and jth elements are *in order*. Define $p(S)$::= the number of "out-of-order" pairs of elements in a sequence, S.

From the starting sequence, we carry out the following process:

(*) Pick two consecutive elements in the current sequence, say the ith and $(i+1)$st.

I. If the elements are not in order, then **switch them** in the sequence and repeat step (*).

II. If the elements are in order, **remove both**, resulting in a sequence that is shorter by two. Then pick another element and remove it as well. If the length of the resulting sequence is less than three, the process is over. Otherwise, **reverse the sequence** and repeat step (*).

This process can be modelled as a state machine where the states are the sequences that appear at step (*).

(b) (5 points) Describe a simple state invariant predicate that immediately implies that if this process halts, then the final state is the sequence of length zero. (Just define the invariant; you need not prove it has the requisite properties.)

Preserved Invariant

The size of the sequence is always a multiple of 3.

✓

(c) (10 points) Prove that this process always terminates by defining a nonnegative integer valued derived variable that is strictly decreasing. (Just define the variable, you need not prove it has these properties.)

Variable = (~~length of sequence~~) + ~~p(S)~~

$$= |S| + (1 - p(S))$$ ✓

this is weakly decreasing

this can increase, but only when length decreases by three

∴ this sum is strictly decreasing

Problem 3 (15 points). Equivalence Relations and Random Variables

A random variable, X is said to *match* a random variable, Y, iff X and Y are on the same sample space and $\Pr\{X \neq Y\} = 0$. Prove that "matches" is an equivalence relation. *Hint:* Note that $\Pr\{X \neq Z\} = \Pr\{[X \neq Z] \cap [X \neq Y]\} + \Pr\{[X \neq Z] \cap [X = Y]\}$.

Proving equivalence means proving:

1) ~~Transitivity~~ Transitivity:

xRy & $yRz \longrightarrow xRz$

$\Pr\{X \neq Y\} = 0$ & $\Pr\{Y \neq Z\} = 0$

then this is always true.

This is true because

$$\underbrace{\Pr\{X \neq Z\}}_{\text{therefore this identity is always true}} = \underbrace{\Pr\{[X \neq Z] \cap [X \neq Y]\}}_{\text{evaluates to 0}} + \Pr\{\underbrace{[X \neq Z]} \cap \underbrace{[X = Y]}_{\text{this } = 1}\}$$

this is $= 0$ ∴

holds

ON BACK

X

2) Symmetry:

if $\Pr\{X \neq Y\} = 0 \rightarrow \Pr\{Y \neq X\} = 0$ holds

3) Reflexivity:

$\underbrace{\Pr\{X \neq X\}}_{\text{this is always true}} = 0$ ✓

holds

$\frac{10}{15}$

$$\Pr\{X \neq Z\} = \Pr\{\underbrace{\cancel{[X \neq Z] \cap [X \neq Y]}}_{0}\} + \Pr\{[X \neq Z] \cap [X = Y]\}$$

$$\Pr\{X \neq Y\} = 0 \qquad \Pr\{X = Y\} = 1$$

$$\Pr\{Y \neq Z\} = 0$$

$$\Pr\{Y \neq Z\} = \Pr\{\underbrace{\cancel{[Y \neq Z] \cap [X \neq Z]}}_{0}\} + \Pr\{[Y \neq Z] \cap [X = Z]\}$$

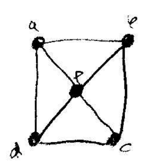
$$\underbrace{\Pr\{Y \neq Z\}}_{0} = \Pr\{[Y \neq Z] \cap [X = Z]\}$$

$$\therefore \Pr\{[Y \neq Z] \cap [X = Z]\} = 0$$

$$\Pr\{X \neq Z\} = \Pr\{[X \neq Z] \cap [X = Y]\} =$$

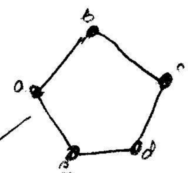
Problem 4 (15 points). Planarity.

(a) (8 points) Exhibit two planar drawings of the same 5-vertex graph in which a face in one drawing has more edges than any face in the other drawing.



max face is outer face abcda

4 edges



max face is outer face abcdea

5-edges

(b) (7 points) Prove that all planar drawings of the same graph have the same number of faces.

Proof by Contradiction:

Suppose a graph with vertices V and E did have a different number of faces, then $V-e+f=2$ would not be true for one of those. But we know Euler's formula is true for all planar embeddings, therefore the number of faces is always the same for a given graph

Problem 5 (15 points). Inclusion-exclusion

A certain company wants to have security for their computer systems. So they have given everyone a name and password. A length 10 word containing each of the characters:

a, d, e, f, i, l, o, p, r, s,

is called a *cword*. A password will be a cword which does not contain any of the subwords "fails", "failed", or "drop".

Use the Inclusion-exclusion Principle to find a simple formula for the number of passwords.

Number of passwords =

total # of len 10 permutations − words containing 'fails' − words containing 'failed' − words containing drop + words containing both 'fails' & 'failed' + words containing both 'fails' and 'drop' + words containing both 'failed' and 'drop' − words containing 'fails', 'failed' and 'drop'

$$= 10! - 6! - 5! - 7! + 0 + 3! + 2! - 0$$

total 'fails' 'failed' 'drop' no fails + failed fails + drop only possible in 'failedrop'

Problem 6 (15 points). Number Theory and Induction

(a) (5 points) Seashells are used for currency on a remote island. However, there are only *huge* shells worth 2^{10} dollars and *gigantic* shells worth 3^{12} dollars. Suppose islander A owes $m > 0$ dollars to islander B. Explain why the debt can be repaid through an exchange of shells provided A and B both have enough of each kind.

The debt can be repaid if $\gcd(2^{10}, 3^{12}) = 1$

since all of the first are even and the second are odd, this is true for this case.
(The two numbers are relatively prime)

(b) (10 points) Give an inductive proof that the Fibonacci numbers F_n and F_{n+1} are relatively prime for all $n \geq 0$. The Fibonacci numbers are defined as follows:

$$F_0 = 0, \qquad F_1 = 1, \qquad F_n = F_{n-1} + F_{n-2} \quad (\text{for } n \geq 2).$$

Base Case $n = 0$

$F_0 = 0$, $F_1 = 1$, true

Induction Hypothesis:

Assume that $F_0 \to F_n$ are all relatively prime with their predecessors,
we demonstrate F_{n+1} is also relatively prime

$$F_{n+1} = F_n + F_{n-1}$$

these two are relatively prime by induction hypothesis

F_{n+1} is relatively prime to F_n since the gcd of $(a+b)$ & a is 1 if ~~many~~ a & b are relatively prime

Problem 7 (20 points). Combinatorial Identities

(a) (10 points) Give a combinatorial proof that

$$\sum_{i=1}^{n} i\binom{n}{i} = n2^{n-1} \tag{2}$$

for all positive integers, n.

$$\sum_{i}^{n} i \frac{n!}{i!(n-i!)}$$

$$= \sum_{i}^{n} \frac{n!}{(i-1)!(n-i)!}$$

is equal to the total value of all bitstrings of n-length where each 1 is worth \$1

2^{n-1} is the total permutations of $(n-1)$ length bit strings, since the average value of a string is n, the total value is the same as the first part.

P

(b) (10 points) Now use the fact that the expected number of heads in n tosses of a fair coin is $n/2$ to give a different proof of equation (2).

if every head were worth \$1, then

$$\frac{\sum_{i}^{n} i\binom{n}{i}}{n}$$ would be the expected value of the wager

dividing $n2^{n-1}$ by n yields 2^{n-1} which is also that expectation.

5

Problem 8 (20 points). Generating Functions

Let a_n be the number of ways to fill a box with n doughnuts subject to the following constraints:

- The number of glazed doughnuts must be odd.
- The number of chocolate doughtnuts must be a multiple of 4.
- The number of plain doughnuts is 0 or 2.
- The number of sugar doughnuts is at most 1.

(a) (8 points) Write a generating function for each of the four doughnut types:

$\langle 1,0,1,0,1,\ldots\rangle$ $\langle 0,0,0,1,0,0,0,1,\ldots\rangle$

$G(x) = \frac{1}{1-x^2}$ X $\frac{x}{1-x^2}$ $C(x) = \frac{1}{1-x^5}$ X $\frac{1}{1-x^4}$

$\langle 1,0,1,0,0,0\rangle$

$P(x) = 1+x^2$ ✓ $S(x) = 1+x$ ✓

4

(b) (12 points) Derive a closed formula for a_n.

Convolution Rule

$$\left(\frac{1}{1-x^2}\right)\left(\frac{1}{1-x^5}\right)(1+x^2)(1+x) ✓$$

3

$$= \frac{(1+x^2)(1+x)}{(1-x^2)(1-x^5)}$$

$A =$

$$= \frac{A}{(1-x^2)} ✓ \frac{B}{(1-x^5)}$$

@ $x=0$ $A+B=1$ $B = A-1$ X

@ $x=2$ $\frac{A}{-3} + \frac{B}{-31} = \frac{(5)(3)}{(-3)(-31)}$

$\frac{A}{-3} + \frac{A-1}{-31} = \frac{15}{(-3)(-31)}$

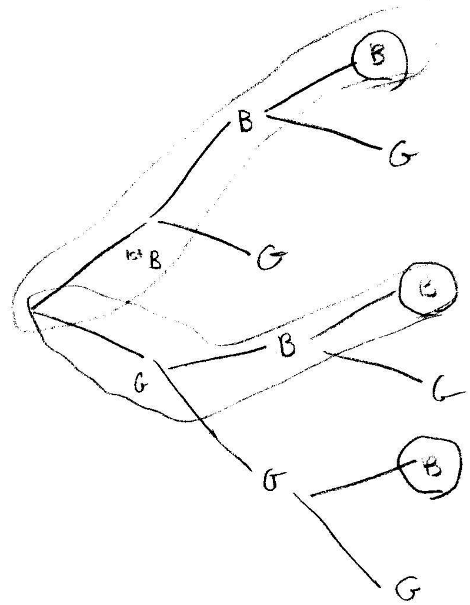
$-31(A) - 3(A-1) = 15$

$-34A = 12$ $A = \frac{-34}{12}$ $B = \frac{-34}{12} - 1$

Problem 9 (15 points). Conditional Probability

There are 3 children of different ages. What is the probability that at least two are boys, given that at least one of the two youngest children is a boy?

Assume that each child is equally likely to be a boy or a girl and that their genders are mutually independent. A correct answer alone is sufficient. However, to be eligible for partial credit, you must include a clearly-labeled tree diagram.



P(outcome) = 1/8

P(two boys | youngest boy)

$\boxed{P = 2/3}$

Problem 10 (15 points). Probability and Expectation

A box initially contains n balls, all colored black. A ball is drawn from the box at random.

- If the drawn ball is black, then a biased coin with probability, $p > 0$, of coming up heads is flipped. If the coin comes up heads, a white ball is put into the box; otherwise the black ball is returned to the box.
- If the drawn ball is white, then it is returned to the box.

This process is repeated until the box contains n white balls.

Let D be the number of balls drawn until the process ends with the box full of white balls. Prove that $\mathrm{E}[D] = nH_n/p$, where H_n is the nth Harmonic number.

Hint: Let D_i be the number of draws after the ith white ball until the draw when the $(i+1)$st white ball is put into the box.

First draw, $P(\text{drawing a black ball}) +$
$$\underset{1}{\shortparallel}$$

Let D_i = # of draws after ith white ball

$$E[D_i] = \frac{1}{\underbrace{\left(\frac{(n-i)}{n}\right)}_{\text{chance of drawing a black ball}}\underbrace{p}_{\text{chance of adding a white ball}}}$$

$$E[D_i] = \frac{1}{\left(\frac{n-i}{n}\right)p}$$

$$E[D] = \sum_i^n \frac{1}{\left(\frac{n-i}{n}\right)p}$$

$$= \frac{1}{p}\sum_i^n \frac{1}{\left(\frac{n-i}{n}\right)} = \frac{1}{p}\sum_i^n \frac{n}{n-i} = \frac{n}{p}\sum \frac{1}{i} = \boxed{\frac{nH_n}{p}}$$

Problem 11 (15 points). Deviation from the Mean

I have a randomized algorithm for calculating 6.042 grades that seems to have very strange behavior. For example, if I run it more than once on the same data, it has different running times. However, one thing I know for sure is that its *expected* running time is 10 seconds.

(a) (5 points) What does Markov's bound tell us about the probablity that my algorithm takes longer than 1 minute (= 60 seconds)?

$$\Pr\{R \geq x\} \leq \frac{E[R]}{x}$$

$$\Pr\{T \geq 60\} \leq \frac{10}{60}$$

The chance of it taking more than one minute is at most 1/6.

(b) (5 points) Suppose I decide to run the algorithm for 1 minute and if I don't get an answer by that time, I stop what I am doing, and completely restart from scratch. Each time that I stop and restart the algorithm gives me an independent run of the algorithm. So, what is an upper bound on the probability that my algorithm takes longer than 5 minutes to get an answer?

To do 5 minutes, the algorithm must have taken > 60 seconds each time or

$$\left(\frac{1}{6}\right)^5 = \frac{1}{6^5}$$ is the max probability of this occuring.

(c) (5 points) Suppose some 6.042 student tells me that they determined the *variance* of the running time of my algorithm, and it is 25. What is an upper bound on the probability that my algorithm takes longer than 1 minute?

$$\Pr\{|R - E[R]| \geq x\} \leq \frac{Var[R]}{x^2}$$

$$\Pr\{|T - 10| \geq 50\} \leq \frac{25}{50^2}$$

$$= \frac{25}{2500} = \boxed{1\%}$$

Problem 12 (20 points). Estimation and Confidence

On December 20, 2005, the MIT fabrication facility produced a long run of silicon wafers. To estimate the fraction, d, of defective wafers in this run, we will take a sample of n independent random choices of wafers from the run, test them for defects, and estimate that d is approximately the same as the fraction of defective wafers in the sample.

A calculation based on the Binomial Sampling Theorem (given below) will yield a near-minimal number, n_0, and such that with a sample of size $n = n_0$, the estimated fraction will be within 0.006 of the actual fraction, d, with 97% confidence.

Theorem (Binomial Sampling). *Let $K_1, K_2, \ldots,$ be a sequence of mutually independent 0-1-valued random variables with the same expectation, p, and let*

$$S_n ::= \sum_{i=1}^{n} K_i.$$

Then, for $1/2 > \epsilon > 0$, for an error less than one half

$$\Pr\left\{\left|\frac{S_n}{n} - p\right| \geq \epsilon\right\} \leq \frac{1+2\epsilon}{2\epsilon} \cdot \frac{2^{-n(1-H((1/2)-\epsilon))}}{\sqrt{2\pi(1/4-\epsilon^2)n}} \tag{3}$$

deviation from expectation $\geq$ error less than

where

$$H(\alpha) ::= -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha).$$

(a) (10 points) Explain how to use the Binomial Sampling Theorem to find n_0. You are not expected to calculate any actual values, but be sure to indicate which values should be plugged into what formulas.

$$Pr\left\{\left|\frac{\#\text{total defective}}{\#\text{total sampled}} - p\right| \geq 0.006\right\}$$

$$Pr\{|d-p| \geq \cancel{0.006}\} \leq \frac{1.012}{.012} \cdot \underbrace{\frac{2^{-n(1-H((1/2)-.006))}}{\sqrt{2\pi(1/4-(0.006)^2 n}}}$$

observed fraction

population fraction

forgot to set this to less than .03

5