6.042/18.062J Mathematics for Computer Science — December 20, 2006

Tom Leighton and Ronitt Rubinfeld

Final

$\frac{78}{84} = 92.9\%$

- This final is **closed book**, but you may have three 8.5 × 11″ sheet with notes in your own handwriting on both sides.
- Calculators are not allowed.
- You may assume all of the results presented in class.
- Write your solutions in the space provided. If you need more space, write on the back of the sheet containing the problem. Please keep your entire answer to a problem on that problem's page.
- Be neat and write legibly. You will be graded not only on the correctness of your answers, but also on the clarity with which you express them.
- If you get stuck on a problem, move on to others. The problems are not arranged in order of difficulty.
- For this final, $\mathbb{N}$ is the set of nonnegative integers (including 0): $\mathbb{N} = \{0, 1, \ldots, \}$.
- GOOD LUCK!
- **Important:** If you show your reasoning, even if your answer is wrong, you could earn partial credit.

TIME: 1:29

Problem	Points	Grade	Grader
1	8	8	
2	20	16	
3	8	8	
4	10	8	
5	10	10	
6	28	28	
7	~~16~~	—	
Total	~~100~~ 84	78	

17:53
17:58
0:05

Problem 1. [8 points] Prove that for all $n \in \mathbb{N}$, the following identity holds

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

8/8

By induction.

Base case, $n = 0$: $\sum_{1}^{0} i^2 = \frac{0 \cdot 1 \cdot 1}{6} = 0$ ✓

$n = 1$: $\sum_{1}^{1} i^2 = 1 = \frac{1 \cdot 2 \cdot 3}{6} = 1$ ✓

Inductive step :

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

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

$$= \frac{n(2n^2 - n - 2n + 1)}{6} + \frac{6n^2}{6}.$$

$$= \frac{2n^3 - 3n^2 + n + 6n^2}{6} = \frac{2n^3 + 3n^2 + n}{6}$$

$$= \frac{n(2n^2 + 3n + 1)}{6} = \frac{n(n+1)(2n+1)}{6}$$ ✓

$\square$

17:59
18:18

21:09
21:24

0:34

Problem 2. [20 points] Coin-Flip is a 2 player game. Each player wins with probability exactly 0.5. There are no ties.

16/20

n people are playing a Coin-Flip tournament. Every person plays a Coin-Flip game with every other person exactly once. Thus everybody plays $n-1$ games. The outcomes of all the games are mutually independent of one another.

We say that the tournament is a *success* if for every $i \in \{0, 1, \ldots, n-1\}$, there is exactly one player, which we will refer to as p_i, with exactly i wins.

(a) [10 points] Prove that if the tournament is a success, then for any integers j, k with $0 \le k < j \le n-1$, p_j defeats p_k.

(b) [6 points] What is the probability that the tournament will be a success?

(c) [4 points] Show that your answer to part (b) is $o(1)$.

(a) By induction. p_0 beats no one. p_1 beats only one person. Since everyone beat p_0, it must be true that p_1 beat p_0.
p_2 beats only two people. p_2 beat p_0, obviously. Since everyone beat p_1 except p_0, p_2 beat p_1.
Assume p_{j-1} beat p_k, $\forall\, j-1 > k$. Since everyone beat p_{j-1} except $\underbrace{p_{j-2}, \cdots, p_0}_{j-1}$, and p_j beat j people, $p_j > p_{j-1} > p_k$

(b) Player 0 can have 2^{n-1} outcomes for $n-1$ games.
Player 1 can have 2^{n-2} outcomes for $n-2$ games independent of 1 game w/ Player 0
" 2 " 2^{n-3} " " $n-3$ " " 2 " " Players 0, 1
" $n-2$ " 2^1 1 $n-2$ $0 \ldots n-3$

Total outcomes: ~~$(n-1)$~~ $2^{\sum_1^{n-1} i} = 2^{\frac{n(n-1)}{2}}$

Total ways to arrange the players in a "success" game: $(n-1)!$ $\xleftarrow{\times} n!$

$$P(\text{success}) = \frac{(n-1)! \xleftarrow{\times} n!}{2^{n(n-1)/2}}$$

(c) $P(\text{success}) = \frac{(n-1)!}{2^{n(n-1)/2}} \le \frac{n!}{2^{n(n-1)/2}} \sim \frac{n^n}{(\sqrt{2})^{n^2-n}} e^{-n} = \left(\frac{n}{(\sqrt{2})^{n-1}}\right)^n e^{-n} \to 0 \sim o(1)$

Stirling

YES, BUT... BAD FORM

18:18
18:24
0:06

$\frac{8}{8}$

Problem 3. [8 points] A person is passing time by advancing a token on the set of natural numbers. In the beginning, a token is placed on 0.

The person keeps playing *moves* forever. Each move proceeds as follows:

1. First the person tosses a fair coin (with heads/tails equally likely).
2. Suppose the token is currently placed on n. If heads came up, then the person moves the token to $n + 3$, otherwise he moves the token to $n + 4$.

For each $n \in \mathbb{N}$, let E_n be the event "There was a move on which the token landed on n". Let $p_n = \Pr[E_n]$.

Find a recurrence relation for p_n. *You do not need to solve the recurrence, but you should specify the boundary conditions that would be necessary to find a solution to the recurrence.*

$p_0 = 1 \quad p_1 = 0 \quad p_2 = 0 \quad p_3 = \frac{1}{2} \quad p_4 = \frac{1}{2} \quad p_5 = 0 \quad p_6 = \frac{1}{4} \quad p_7 = \frac{1}{4} \quad p_8 = \frac{1}{4} \quad p_9 = \frac{1}{8}$

$p_{10} = \frac{1}{4} \quad p_{11} = \frac{1}{4}$

$p_n = \frac{1}{2} p_{n-3} + \frac{1}{2} p_{n-4}$

B.C.s: $p_0 = 1, \ p_1 = 0, \ p_2 = 0,$
$p_3 = \frac{1}{2}$

18:25
18:32
0:07

Problem 4. [10 points] Exactly 1/5th of the people in a town have Beaver Fever©.

There are two tests for Beaver Fever, TEST1 and TEST2. When a person goes to a doctor to test for Beaver Fever, with probability 2/3 the doctor conducts TEST1 on him and with probability 1/3 the doctor conducts TEST2 on him.

8/10

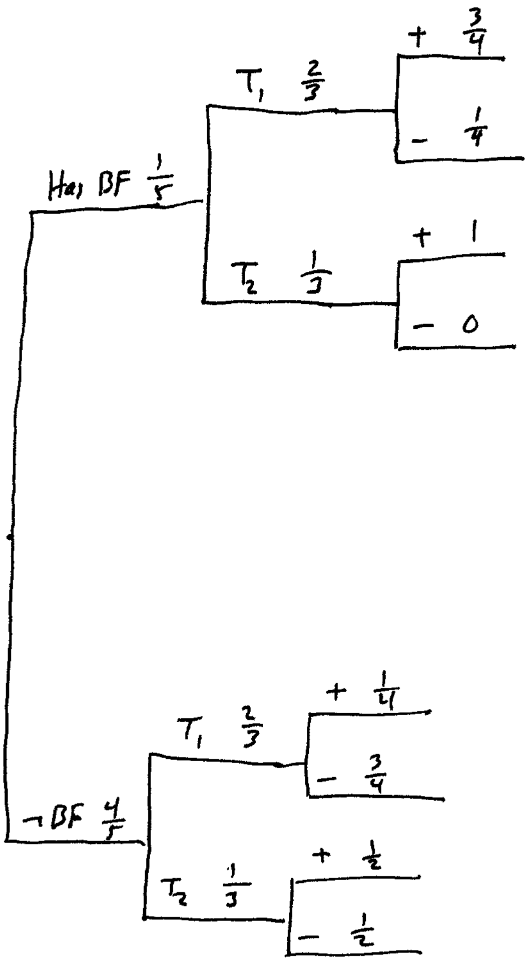
When TEST1 is done on a person, the outcome is as follows:

- If the person has the disease, the result is positive with probability 3/4.
- If the person does not have the disease, the result is positive with probability 1/4.

When TEST2 is done on a person, the outcome is as follows:

- If the person has the disease, the result is positive with probability 1.
- If the person does not have the disease, the result is positive with probability 1/2.

A person is picked uniformly at random from the town and is sent to a doctor to test for Beaver Fever. The result comes out positive. What is the probability that the person has the disease?



$+ \wedge BF$

$\frac{1}{5}\cdot\frac{2}{3}\cdot\frac{3}{4} = 6/60$

$+$

$\frac{1}{5}\cdot\frac{2}{3}\cdot\frac{3}{4} = 6/60$

$\frac{1}{5}\cdot\cancel{\frac{2}{3}}\,1/3\cdot 1 = \cancel{2/15}\ 1/15$

$\cancel{2}\,1/15$

$\frac{8}{60}$

$\frac{4}{30}$

$$P(BF|+) = \frac{\frac{6}{60}+\frac{\cancel{8}\,4}{60}}{\frac{6+8+8+\cancel{8}\,4}{60}} = \frac{6+\cancel{8}\,4}{\cancel{30}\,26} = \frac{14}{30} = \boxed{\frac{7}{15}}\ \cancel{4}\ 5/13$$

18:32
18:36
0:04

Problem 5. [10 points] Two identical complete decks of cards, each with 52 cards, have been mixed together. A hand of 5 cards is picked uniformly at random from amongst all subsets of exactly 5 cards.

10/10

(a) [5 points] What is the probability that the hand has no identical cards (i.e., cards with the same suit and value. For example, the hand $\langle Q\heartsuit, 5\spadesuit, 6\spadesuit, 8\clubsuit, Q\heartsuit\rangle$ *has* identical cards.)?

(b) [5 points] What is the probability that the hand has exactly one pair of identical cards?

(a) Total : $\binom{104}{5}$ No identical : $104 \cdot 102 \cdot 100 \cdot 98 \cdot 96$

$$P(\text{no identical}) = \frac{104 \cdot 102 \cdot 100 \cdot 98 \cdot 96}{\binom{104}{5}}$$

(b) 52 ways to choose 2 identical cards $102 \cdot 100 \cdot 98$ ways to choose other 3

$$P(\text{one pair}) = \frac{52 \cdot 102 \cdot 100 \cdot 98}{\binom{104}{5}}.$$

18:36
18:45
18:51
19:04
20:56
21:07
0:33

Problem 6. [28 points] Scores for a final exam are given by picking an integer uniformly at random from the set $\{50, 51, \ldots, 97, 98\}$. The scores of all 128 students in the class are assigned in this manner. For parts (a), (b), (c) and (d) you may NOT assume that these scores are assigned independently. For parts (e), (f), (g) and (h) you MAY assume that these scores are assigned independently.

28/28

Let $S_1, \ldots, S_{128}$ be their scores. Let $S = \frac{1}{128}(\sum_{i=1}^{128} S_i)$ be the average score of the class.

(a) [3 points] For $i \in \{1, \ldots, 128\}$, what is $\mathbb{E}[S_i]$? $98-50+1 = 49$ choices of score

$$E[S_i] = \sum_{50}^{98} p_k k = \frac{1}{49}\sum_{50}^{98} k = \frac{1}{49}\left[\frac{98(99)}{2} - \frac{49(50)}{2}\right] = \underline{74}$$

(b) [2 points] Show that $\mathbb{E}[S] = 74$. Make no independence assumptions.

$$E[S] = \frac{1}{128}E\left[\sum S_i\right] = \frac{1}{128}\sum E[S_i] = \frac{1}{128}\cdot 128\, E[S_i] = E[S_i] = \underline{74}$$

(c) [4 points] Prove that

$$\Pr[S \geq 88] \leq \frac{37}{44}.$$

Make no independence assumptions.

Markov, $P(S \geq 88) \leq \frac{74}{88} = \underline{\frac{37}{44}}$

(d) [5 points] Improve your previous bound by using the fact that the minimum possible score is 50. Prove that

$$\Pr[S \geq 88] \leq \frac{12}{19}.$$

Make no independence assumptions.

Let $z = S - 50 \Rightarrow S \geq 88 \Rightarrow z \geq 38$ $\quad E[z] = 74 - 50 = 24$

$$P(z \geq 38) \leq \frac{24}{38} = \frac{12}{19}$$

(e) [4 points] For the remaining problems, assume that all the scores are assigned mutually independently. Use Problem 1 of this final to find $Var[S_i]$.

$$Var(S_i) = \sum_{50}^{98} p_k (k - E[k])^2 = \frac{1}{49}\left(\cancel{\sum k^2 - 2E[k] \sum k} \right.$$

$$= E[S_i^2] - (E[S_i])^2 = \frac{1}{49} \sum_{50}^{98} k^2 - 74^2$$

$$= \frac{1}{49}\left(\frac{98 \cdot 99 \cdot 197}{6} - \frac{49 \cdot 50 \cdot 99}{6} \right) = 5{,}676 - 74^2 = \underline{200}$$

(f) [3 points] What is $Var[S]$?

$$Var(S) = \frac{128 \cdot Var(S_i)}{128^2} = \frac{25600}{\cancel{128}} = \cancel{200}\ 1.5625$$

(g) [2 points] What is the standard deviation of S?

~~14.14~~ $\cancel{\sqrt{200}}$

$$\sqrt{1.5625} = \underline{1.25}$$

(h) [5 points] Prove, using the Chebyshev Inequality, that

$$\Pr[S \leq 69] \leq \frac{1}{16}.$$

$$P(|S - E[S]| \geq a) \leq \frac{Var(S)}{a^2}$$

$$= P(S - E[S] \geq a \quad \text{or} \quad S - E[S] \leq -a)$$

$$= 2P(S - E[S] \leq -a) = 2P(S \leq -a + E[S])$$

$$= 2P(S \leq 74 - a) \leq \frac{Var(S)}{a^2}$$

for $a = 5$, $\quad P(S \leq 69) \leq \dfrac{Var(S)}{2 \cdot 5^2} = \dfrac{\overset{1.5625}{\cancel{200}}}{50}$

$$= \underline{0.03125}$$

← THIS IS AN EVEN BETTER BOUND

For some reason the solutions were included
on the test here. So I excluded this problem.

20:56

Problem 7. [16 points] 1000 files $F_1, F_2, \ldots, F_{1000}$ have just reached a disk manager for writing onto disk. Each file's size is between $0MB$ and $1MB$. The sum of all files' sizes is $400MB$.

The disk manager has 4 disks under its control. For each file F_i, the disk manager chooses a disk uniformly at random from amongst the 4 disks, and F_i is written to that disk. The choices of disk for the different files are mutually independent.

(a) [2 points] What is the expected number of files that will be written to the first disk?

We can use indicator variables. For each file, $P_i = 1$ if F_i is written to the first disk. The chance of an individual file being written to the first disk is $1/4$. By linearity of expectation, the expected number of files written to the first disk is the sum of the expected values of P_i's. The expected value of each indicator variable is $1/4$, and $\sum_{i=1}^{(} 1000)1/4 = 250$, so the expected number of files to be written to the first disk is 250.

(b) [2 points] What is the expected number of bytes written on the first disk?

We can say that each file F_i has bit size S_i. Each file has a $1/4$ chance of being written do the first disk. Therefore, by linearity of expectation, the expected number of bytes written to the first disk is the sum of the expected number of bytes per file written to the first disk, which is:

$$\sum_{i=1}^{1000} 1/4 \cdot S_i = 1/4 \sum_{i=1}^{1000} S_i = 1/4 \cdot 400 = 100$$

(c) [8 points] Find the best upper bound you can on the probability that $200MB$ or more are written on the first disk?

For this we can use the first Chernoff bound, which is:

$$\Pr(X \geq c\,\mathrm{Ex}(X)) \leq e^{-(c\ln c - c + 1)\,\mathrm{Ex}(X)}$$

The Chernoff bound only works if X is the sum of random variables that each take on a value between 0 and 1. The file size of each file in the first disk is between 0 and 1Mb . So we can define X to be the total number of bytes in disk 1. The expected value of X is 100, so we take c to be 2. We get:

$$\Pr(X \geq 2 \cdot 100) \leq e^{-(2\ln 2 - 2 + 1)100}$$

(d) [4 points] Find the best upper bound you can on the probability that there is some disk with $200MB$ or more written on it?

For this we can use the Union Bound along with our result from above. The probability of this event happening in one or more disks is upper bounded by the sum of the probabilities of the event happening in each disk. This gives us an upper bound of

$$4 \cdot e^{-(2\ln 2 - 1)100}$$