Massachusetts Institute of Technology | Handout 63
6.042J/18.062J: Mathematics for Computer Science | May 28, 1998
Professor Albert R. Meyer

Final Exam with Solutions

- This exam is closed book. Handout 60 with results and formulas is attached for reference.
- This exam contains twelve (12) problems totaling 250 points. Problems are labelled with point values.
- Put your name on the top of **every** page.
- Write your solutions in the space provided. If you need more space, write on the back of the sheet containing the problem.
- GOOD LUCK!

Your Name:

Circle the name of your TA:

Alex Duran   Stas Jarecki   Al Lee   Edmond Lee   Mark Roh   Tara Schenckel

Recitation Time:

DO NOT WRITE BELOW THIS LINE

Problem	Points	Grade	Grader
1	25		
2	15		
3	25		
4	15		
5	10		
6	20		
7	20		
8	25		
9	15		
10	25		
11	25		
12	30		
Total	250		

Problem 1. [25 points] Circle True/False.

T F $[\exists x : P(x) \implies \forall x : Q(x)]$ implies $[\forall x : [P(x) \implies \forall y : Q(y)]]$ **TRUE**

T F The number of distinct simple paths in a tree is bounded by $2^{(l+1)^2}$ where l is the number of leaves. **FALSE**: Consider a chain of length n: it's a tree with only two leaves, so n can't be bounded by any function of l.

T F If every vertex of an n vertex simple graph has degree at least $n/2$, then the expected distance between a randomly chosen pair of distinct vertices is strictly less than two. **TRUE**: If two nodes each are adjacent to $n/2$ nodes, then by Pigeonholing, their adjacency sets must overlap, and so distance is always ≤ 2. Since adjacent nodes are distance 1, the expectation must be strictly less than 2.

T F If there is a surjection from A to B and also a surjection from B to A, then there is a bijection from A to B. **TRUE**: This is the Schröder-Bernstein Theorem from Handout 34.

T F In the polynomial expansion of $(x+1)^{123}$, every coefficient occurs exactly twice. **TRUE**

T F Assume each Boolean variable is randomly assigned true or false with equal likelihood and the choices are independent. Then $\Pr\left[(p_1 \vee p_2 \vee \overline{p_3}) \wedge (\overline{q_1} \vee q_2 \vee q_3) \text{ is true}\right] \leq 3/4$. **FALSE**: $(7/8)^2 \approx .77 > 3/4$.

T F The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is a bijection. **TRUE**

T F If r is a real number and not an integer, then the set of infinite subsequences of the decimal representation of r is uncountable. **FALSE**: eg, $r = 1/10 = 0.100\ldots$

T F There is a surjection from the real line onto the plane. **TRUE**—this was highlighted in the cardinality Handout 34.

T F Every countable powerset is finite. **TRUE**: In order for $P(A)$ to be infinite, A must also be infinite. But then $\mathbb{N}\,\text{inj}\,A$, so $P(\mathbb{N})\,\text{inj}\,P(A)$, and since $P(\mathbb{N})$ is not countable, neither is $P(A)$.

T F All n-vertex simple graphs with chromatic number n (that is, that require exactly n-colors) are isomorphic. **TRUE**: Has to be the complete graph K_n.

T F $n! = 2^{f(n)}$ for some $f : \mathbb{N} \to \mathbb{R}$ where $f = \Theta(n \log n)$. **TRUE**: By Stirling's formula. The $\sqrt{2\pi n}$ and e^{a_n} can be absorbed into f.

Problem 2. [15 points] Let $T : \mathbb{N} \to \mathbb{N}$ be defined by the following recurrence. Prove that $T(n) = \Theta(n)$. *Hint*: Find a constant A such that $T(n) \le An$ for all $n > 0$.

$$T(n) = \begin{cases} 8 & \text{if } n < 5, \\ T\left(\left\lfloor \frac{n}{5} \right\rfloor\right) + T\left(\left\lfloor \frac{3n}{4} \right\rfloor\right) + 8n & \text{if } n \ge 5. \end{cases}$$

Solution: $T(n) \le (19/20)An + 8n = n(19/20A + 8) \le An$, providing $8 \le A/20$, so $A \ge 160$ will work. Since $T(n) \ge 8n$, we have $T = \Theta(n)$. ($T(n) \ge 8n$ follows assuming $T \ge 0$ which we'll take for granted; it's easy to verify by induction).

Problem 3. [25 points] Short answers. No need to be rigorous, though partial credit may be given based on the derivation.

(a) [8 points]
We'll say an integer $n \geq 2$ is "good" providing that in any set of 401 integers, there are three integers, j, k and l, such that the pairwise differences, $j-k, j-l, k-l$ are multiples of n. There is a bound $m \in \mathbb{N}$ such that n is good iff $2 \leq n \leq m$. What is m?
Solution: $m = 200$ by a simple Pigeonholing argument.

(b) [5 points] What is the coefficient of x^{11} in the polynomial $(3x+2)^{50}$? (Your answer may contain binomial coefficients and integer powers.)

Solution: $3^{11}2^{39}\binom{50}{11}$

(c) [7 points] The space shuttle runs 5 identical computers so that if two goof, the other three will out-vote them with the right answer. If each computer goofs independently with probability g, what is the probability of getting the right answer in a given computation?

Solution:

$$\Pr[5 \text{ correct}] + \Pr[4 \text{ correct}] + \Pr[3 \text{ correct}] = (1-g)^5 + \binom{5}{4}(1-g)^4 g + \binom{5}{3}(1-g)^3 g^2$$

(d) [5 points] Suppose two 6-sided dice are thrown, and then as many coins are flipped as the sum of the dice. The dice and coins are fair. What is the expected number of coins showing heads?
Solution: $\mathrm{E}[\text{sum of two dice}] = 2\mathrm{E}[\text{value of one die}] = 2[1+2+3+4+5+6]/6 = 7$; $\mathrm{E}[\text{number of heads from one toss}] = 1/2$. By Wald's Thm, expected number of coins showing heads is $\mathrm{E}[\text{sum of two dice}] \cdot \mathrm{E}[\text{number of heads from one toss}] = 7(1/2) = 7/2$.

Problem 4. [15 points]

(a) [5 points] Prove, by direct calculation, that for $1 \le r \le n$,

$$\binom{n}{r} = \frac{(n-r+1)}{r}\binom{n}{r-1}.$$

Solution:

$$\begin{aligned}
\binom{n}{r} \Big/ \binom{n}{r-1} &= \frac{n!}{r!(n-r)!} \times \frac{(r-1)!(n-(r-1))!}{n!} \\
&= \frac{1}{r\cdot(r-1)!(n-r)!} \times (r-1)!(n-r+1)! \\
&= \frac{1}{r(n-r)!} \times (n-r+1)\cdot(n-r)! \\
&= (n-r+1)/r.
\end{aligned}$$

(b) [10 points] Now prove the same identity combinatorially by comparing two procedures for choosing a committee of r members out of n people and also choosing a committee member to be chairman.

Solution: Straight from Recitation 8:
Consider the number of ways of choosing a committee of r members out of n people, and also choose a chairman for the committee. There are $\binom{n}{r}$ ways of choosing the committee, and for each such committee there are r choices for the chair. Alternatively, there are $\binom{n}{r-1}$ ways of choosing all the members but the chair, and then $n-r+1$ ways to choose the chair among the remaining people. It follows that

$$r\binom{n}{r} = (n-r+1)\binom{n}{r-1}$$

Problem 5. [10 points] Prove that the number of sequences of eleven letters which can be made from the letters in MISSISSIPPI is not divisible by four.

Solution: 11 letters: 4 S's, 4 I's, 2 P's, 1 M. After some cancelling,

$$\binom{11}{4, 4, 2, 1} = 11 * 10 * 9 * 7 * 5,$$

which is divisible by 2 but not 4.

Problem 6. [20 points]

Prove that there are 16,800 permutations of the letters $\{a, b, c, d, e, f, g, h\}$ such that neither the letters $\{a, b, c\}$ nor the letters $\{d, e\}$ occur in order. For example, $fbgadceh$ is not allowed because d occurs before e, and $fagbecdh$ is not allowed because a occurs before b and b occurs before c, but $fbgaedch$ is OK.

Solution:

One way is to find three positions for a, b, and c, then two other positions for d and e, and count the number of ways a, b, and c, can be out of order (namely, 5), the number of ways for de (namely, 1), and the permutations of remaining 3 elements (namely, 3!):

$$\begin{aligned}\binom{8}{3}\binom{5}{2}3!5 &= \frac{8*7*6}{3*2}\frac{5*4}{2}3*2*5 \\ &= (8*7)(5*2)30 = 56*300 = 16800.\end{aligned}$$

Inclusion-exclusion is another way: let ALL be the 8! permutations of all eight letters, let ABC be those permutations with a, b, c appearing in order, and DE those permutations with d, e in order. So $|\text{ABC}| = P(8,5) ::= 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4$, since a permutation in ABC is uniquely determined by the 5 out of 8 positions of the 5 letters besides a, b, c. Likewise $|\text{DE}| = P(8,6)$. Also, $|\text{ABC} \cap \text{DE}|$ is determined by the $P(8,3)$ positions of f, g, h and then the $\binom{5}{2}$ places to put d and e in order, that is $|\text{ABC} \cap \text{DE}| = P(8,3)\binom{5}{2}$. Then the number of allowed permutations is

$$\begin{aligned}|\text{ALL}| - |\text{ABC} \cup \text{DE}| &= 8! - |\text{ABC}| - |\text{DE}| + |\text{ABC} \cap \text{DE}| \\ &= 8! - P(8,5) - P(8,6) + P(8,3)\binom{5}{2} \\ &= 8*7*6[5! - 5*4 - 5*4*3 + 5*4/2] \\ &= 8*7*6[120 - 20 - 60 + 10] = 8*7*6*50 = 16,800.\end{aligned}$$

Problem 7. [20 points]

A high blood pressure test is known to be 90% accurate in detecting high blood pressure (that is, 10% false negatives), whereas only 5% of people without high blood pressure will falsely test positive. Suppose 30% of the population has high blood pressure. Given that a person picked at random tests positive, what is the conditional probability that this person has high blood pressure?

Solution: Let $T = \{\text{test is positive}\}$, $V = \{\text{person has HBP}\}$. We are given that $P(T|V) = 0.9$, $P(T|\bar{V}) = 0.05$ and $P(V) = 0.3$.

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

The above expression evaluates to 0.885.

The next two problems are based on the following scenario: an urn contains red, white and blue colored balls. The fraction of red, white, and blue balls is r, w, b, respectively (so $r + w + b = 1$). A gambler repeatedly pulls a ball at random from the urn. That is, at any point during succesive pulls, each ball in the urn is equally likely to be the next one pulled. The gambler puts the ball back into the urn before making the next pull. He keeps pulling balls until he pulls *two red balls in a row.*

Problem 8. [25 points]

(a) [5 points] Let N_n be the event that at least n balls have been pulled and the nth ball is *not* red; let $p_n = \Pr[N_n]$. Write an expression in terms of p_n for the probability that the Gambler stops pulling after exactly $n + 2$ balls.

Solution: $r^2 p_n$, because the Gambler stops at step $n + 2$ iff he pulls n balls with the nth ball not red (probability p_n), and then he pulls two reds in a row (probability r^2).

(b) [10 points] Write a linear homogeneous recurrence equation for p_{n+2} in terms of p_{n+1} and p_n.

Solution:

$$p_{n+2} = (1 - r)p_{n+1} + (1 - r)rp_n.$$

To see why, suppose N_{n+2} occurred. Consider whether the $n + 1$st pull was red. The probability the Gambler pulled non-red on the $n + 1$st turn is p_{n+1}, and the probability he pulled non-red again on the $n + 2$nd turn is therefore $(1 - r)p_{n+1}$. If the $n + 1$st pull was red, then the nth pull must have been non-red or the Gambler would have stopped at the $n + 1$st pull. So the probability the $n + 1$st pull was red and the $n + 2$nd pull was non-red is $r(1 - r)$ times the probability, p_n, that the nth pull was non-red.

(c) [5 points] Suppose $r = 2/3$. Write a closed form expression for p_n.

Solution: The characteristic equation $x^2 - (1 - r)x - (1 - r)r = x^2 - x/3 - 2/9 = 0$ has roots $r = 2/3, -1/3$, so $p_n = A(2/3)^n + B(-1/3)^n$. The initial conditions are that p_1 is the probability of throwing a non-red on the first throw, namely $p_1 = 1/3$. Likewise, p_2 is the probability of throwing a non-red on the second throw, namely $p_2 = 1/3$. So at $n = 1$, $2A/3 - B/3 = 1/3$ which implies $2A = B + 1$, and at $n = 2$, $3/9 = 4A/9 + B/9$ so $3 = 4A + B = 2(B + 1) + B = 3B + 2$ and $B = 1/3, A = 2/3$. Therefore,

$$p_n = (2/3)^{n+1} - (-1/3)^{n+1}.$$

GRADING NOTE. To avoid "double jeopardy," if a mistaken, nondegenerate recurrence from the previous part was correctly solved, we gave full credit on this part.

Problem 9. [15 points] The sequence of pulls by the Gambler can be regarded as a sequence of *tries*, where the first try begins with the first ball pulled. Each try consists of a sequence of pulled balls continuing until either a white ball, a blue ball, or two red balls in a row, are pulled. For example, if the colors of pulled balls are RBBWRWRR, then there are five tries which we indicate with slashes: /RB/B/W/RW/RR/. Note that each try consists of one or two consecutive balls.

(a) [15 points] What is the expected length of a try?

Solution: $1 \cdot (1 - r) + 2 \cdot r = 1 + r$

(b) [5 points] What is the expected number of tries?

Solution: A series of tries ends iff only the last try consists of two reds. We know from lecture that the "mean time to failure" is the reciprocal of the probability of failure, so $\mathrm{E}[\text{num tries}]$ is the $1/\Pr[2 \text{ reds}] = 1/r^2$.

(c) [5 points] Write a simple expression in terms of your answers to the previous parts of this problem for the expected number of balls pulled by the Gambler. Briefly justify your reasoning, citing appropriate results from the attached Handout 60.

Solution: By Wald's Thm

$$\mathrm{E}[\text{num pulls}] = \mathrm{E}[\text{pulls per try}] \cdot \mathrm{E}[\text{num tries}] = (1 + r)\frac{1}{r^2}.$$

Problem 10. [25 points] Let R_1, R_2 be random variables such that $\mathrm{E}[R_1 R_2] = \mathrm{E}[R_1]\,\mathrm{E}[R_2]$. Prove that

$$\mathrm{Var}[R_1 + R_2] = \mathrm{Var}[R_1] + \mathrm{Var}[R_2].$$

Solution: Let $m_i = \mathrm{E}[R_i]$ and $R = R_1 + R_2$. Then

$$\begin{aligned}
\mathrm{Var}[R_1 + R_2] &= \mathrm{E}\left[R^2\right] - \mathrm{E}^2[R] \quad \text{[by formula for } \mathrm{Var}[R]\text{]} \\
&= \mathrm{E}\left[R_1^2 + 2R_1R_2 + R_2^2\right] - (m_1 + m_2)^2 \quad \text{[by linearity of } \mathrm{E}[\,] \text{ and algebra]} \\
&= \mathrm{E}\left[R_1^2\right] + 2\mathrm{E}[R_1R_2] + \mathrm{E}\left[R_2^2\right] - (m_1 + m_2)^2 \quad \text{[by linearity of } \mathrm{E}[\,]\text{]} \\
&= \mathrm{E}\left[R_1^2\right] + 2m_1m_2 + \mathrm{E}\left[R_2^2\right] - (m_1 + m_2)^2 \quad \text{[by hypothesis]} \\
&= \mathrm{E}\left[R_1^2\right] - m_1^2 + \mathrm{E}\left[R_2^2\right] - m_2^2 \quad \text{[by algebra]} \\
&= \mathrm{Var}[R_1] + \mathrm{Var}[R_2] \quad \text{[by formula for } \mathrm{Var}[R_i]\text{]}.
\end{aligned}$$

Problem 11. [25 points] Psychological diagnosis of sociopaths is notoriously unreliable. For example, a random psychologist is as likely to diagnose a normal patient as sociopathic as he is to recognize that the patient is normal. A random psychologist is actually slightly *less* likely to correctly diagnose a real sociopath than not—sociopaths are really good actors. In these circumstances it might seem there was no reliable way to diagnose sociopaths, but in theory there is a way—if we have a large enough population of psychologists who reach their judgements independently. Explain how to obtain a reliable diagnosis in this case. Briefly explain why your method will be reliable, citing appropriate results from class. A qualitative answer is all that's required—you need not come up with any numerical estimates or growth rates.

Solution: Get a reliable test by repeatedly asking randomly chosen psychologists to do the diagnosis and computing the fraction, f, of "normal" diagnoses. Let $s > 1/2$ be the probability of a "normal" diagnosis on a sociopathic patient. Diagnose normal if f is closer to $1/2$ than to s.

By diagnosing independently enough times, there is small probability that f differs from the expected value of a diagnosis by very much (the expected value is either $1/2$ or s depending on whether or not the patient is normal—we don't know which). By the Law of Large Numbers, by Chebyshev's bound, or by the estimation of the binomial distribution as used in the Gallup Poll example from lecture, if we diagnose enough times, we can make this probability as small as desired, so the probability of a correct diagnosis can be as close to 1 as desired.

Problem 12. [30 points]

(a) [10 points] Knowing only that the average graduating MIT student's total number of credits is 200, find a tight upper bound for the fraction of MIT students graduating with at least 235 credits. (Ignore the fact that students who graduate do so after at most 700 years; that is, assume that there is no theoretical upper bound on the possible number of credits a student may earn.)

Solution: We use Markov's inequality $\Pr(X \geq c) \leq E[X]/c$, with $c = 235$, to obtain the upper bound $200/235 \approx .85$.

(b) [10 points] Assuming in addition only that the standard deviation of the total credits per graduating student is 7, give a tight bound on the fraction of students who can graduate with at least 235 credits.

Solution:
We use the one-way Chebychev inequality $\Pr(X \geq E[X] + c\sigma_X) \leq \frac{1}{1+c^2}$ with $c = 5$ to get an upper bound of $1/26$.

(c) [10 points] Different information about MIT credits is now provided by the Registrar. The standard deviation is not 7 after all. Instead, the registrar reports that 30% of past graduates are EECS majors and half of them had 235 or more credits. Only 1/7th of the non-EECS students graduated with that many credits. Use this information and the Chernoff bound to give a tight upper bound on the probability that among 1000 randomly chosen graduating students, at least 300 have 235 or more credits. (Since you don't have a calculator, you may leave ln and e in your answer).

Solution: We use the Chernoff bound. Among 1000 students, let N be the number with at least 235 credits. So $\mathrm{E}[N] = (1/7)700 + (1/2)300 = 250$ and

$$\begin{aligned} \Pr[N \geq 300] &= \Pr[X \geq (1.2)250] \leq e^{-(\ln 1.2 + 1/(1.2) - 1)(1.2)250} = e^{-(\ln 1.2 - .1666)300} \\ &= e^{-(.182 - .1666)300} \leq e^{-4.6} \approx .01. \end{aligned}$$