

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 | | |
| 2 | 20 | | |
| 3 | 15 | | |
| 4 | 15 | | |
| 5 | 15 | | |
| 6 | 15 | | |
| 7 | 20 | | |
| 8 | 20 | | |
| 9 | 15 | | |
| 10 | 15 | | |
| 11 | 15 | | |
| 12 | 20 | | |
| Total | 200 | | |

Problem 1 (15 points). Induction

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

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

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** $[\exists n S(n)] \longrightarrow \forall n \exists m > n S(m)$

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

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

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

If (s_1, \dots, s_k) is a sequence of numbers, then the i and j th elements of the sequence are *out of order* if the number on the left is strictly larger than the number on the right, that is, if $i < j$ and $s_i > s_j$. Otherwise, the i th and j th 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 i th 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.)

(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.)

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

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.

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

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.

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.

(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:

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

Problem 7 (20 points). Combinatorial Identities

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

$$\sum_{i=1}^n i \binom{n}{i} = n2^{n-1} \quad (2)$$

for all positive integers, n .

(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).

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 doughnuts 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:

$$G(x) = \underline{\hspace{4cm}} \quad C(x) = \underline{\hspace{4cm}}$$

$$P(x) = \underline{\hspace{4cm}} \quad S(x) = \underline{\hspace{4cm}}$$

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

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.

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 $E[D] = nH_n/p$, where H_n is the n th Harmonic number.

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

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 probability that my algorithm takes longer than 1 minute (= 60 seconds)?

(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?

(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?

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, \dots , 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 $1/2 > \epsilon > 0$,

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

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.

(b) (10 points) The calculations in part (a) depend on some facts about the run and how the n wafers in the sample are chosen. Write T or F next to each of the following statements to indicate whether it is True or False.

- _____ The probability that the ninth wafer in the *sample* is defective is d .
- _____ The probability that the ninth wafer in the *run* is defective is d .
- _____ All wafers in the run are equally likely to be selected as the third wafer in the *sample*.
- _____ The expectation of the indicator variable for the last wafer in the *sample* being defective is d .
- _____ Given that the first wafer in the *sample* is defective, the probability that the second wafer will also be defective is less than d .
- _____ Given that the last wafer in the *run* is defective, the probability that the next-to-last wafer in the run will also be defective is less than d .
- _____ It turns out that there are several different colors of wafer. Given that the first two wafers in the sample are the same color, the probability that the first wafer is defective may be $< d$.
- _____ The probability that all the wafers in the sample will be different is nonzero.
- _____ The probability that all choices of wafers in the sample will be different depends on the length of the run.
- _____ The probability that all choices of wafers in the sample will be different is $\Theta(e^{-an})$ for some constant $a > 0$.