CS 174 Fall 98
Final Examination

Read these instructions carefully

1. This is a **closed book** *exam. Calculators* **are** *permitted.*

*2. This midterm consists of **14** questions. The first ten questions are multiple choice; the remaining four require written answers.*

3. Answer the multiple choice questions by circling the correct answer (or the best answer if more than one is correct). You should be able to answer all of these from memory, by inspection, or with a very small calculation. Incorrect answers will attract a negative score, so if you do not know the answer **do not** *guess.*

4. Write you answers to the other questions in the spaces provided. None of these questions requires a long answer, so you should have enough space; if not, continue on the back of the page and state clearly that you have done so. **Show all your working.**

5. The questions vary in difficulty: if you get stuck on some part of a question, leave it and go on to the next one.

6. **Good Luck!**

Problem #1

A multiple choice exam has six possible answers for each question, only one of which is correct. A correct answer receives 4 points, while an incorrect answer incurs a penalty of *b* points. If we wish to ensure that the expected score for a student who randomly guesses on every question is zero, we should set *b* to be

1/6	4/5	5/6	1	6/5	5

Problem #2

Each of the 50 states has two senators. A committee is chosen by selecting a set of 50 senators uniformly at random.

(a) The probability that no senator from California is on the committee is

1/(2^50)	98!/100!	C(98, 50)/100!	(50!)/C(100, 50)	C(98, 50)/C(100, 50)	(98/100)^50

(b) The probability that the committe contains one senator from every state is

(2^50)/C(100, 50)	1/C(100, 50)	(50!)/C(100, 50)	(2^50)/100!	1/(100!)	50!/100!

Problem #3

(a) A coin with heads probability p is tossed n times. The expected number of heads is

$1/p$ $n/2$ np $np(1 - p)$ n/p

(b) A *run* in a sequence of coin tosses is a maximal subsequence of either heads or tails. (Thus, for example, the sequence HHHTTHTHH contains five runs.) The expected number of runs in n tosses of a coin with heads probability p is

$n/2$ np $np(1 - p)$ $2(n - 1)p(1 - p) + 1$ $(np)/(1 - p)$

Problem #4

(a) There are n bins, and balls are thrown into them in sequence, independently and uniformly at random. The process stops when *any three* bins are occupied. The expected number of balls thrown is

3 $n/(n - 3)$ $n[(1/n) + 1/(n - 1) + 1/(n - 2)]$ $C(n, 3)/(n^3)$ $3n$

(b) Consider the same process as in part (a), but now we stop when the *first three* bins (i.e., bins 1, 2 and 3) are all occupied. The expected number of balls thrown is now

3 $n/3$ $(11n)/6$ $3n$ $11n$

(c) Now let M be your answer to part (b). If M balls are thrown independently at random into n bins, then as n the expected number of the first three bins that are occupied is

3 M/n $3 - (M/n)$ $3(1 - n/M)$ $3[1 - e^(-M/n)]$

Problem #5

In a classroom, n students want to pick binary labels for themselves. If each student picks a random L-bit binary string uniformly and independently of all the others, then in order to ensure that all students have distinct labels with high probability, L should be on the order of

a constant $\ln \ln n$ sqrt($\ln n$) $\ln(n)$ sqrt(n) n

Problem #6

Let *G* be a random graph in the *G_np* model, and suppose that each vertex of *G* is colored u.a.r. with one of three colors, independently of the other vertices and of the edges of *G*. We call an edge of *G* *bad* if both its endpoints have the same color.

(a) The expected number of bad edges in *G* is

C(*n*, 2)**p*   (1/3)*np*   (1/3)*C(*n*, 2)**p*   (1/3)*(*n*^2)**p*   (2/3)*C(*n*, 2)**p*

(b) If there is a threshold value of *p* for the existence of bad edges in *G*, then this value must be on the order of

a constant   *n*^(-1/3)   *n*^(-2/3)   *n*^(-1)   *n*^(-2)

Problem #7

We have a group of five people. Each person picks, uniformly and independently at random, a number between 1 and 100. For each pair of distinct people, *i, j*, let *X_ij* be the indicator r.v. of the event "*i, j* both choose the same number", and for each triple of distinct people *i, j, k,* let *X_ijk* be the indicator r.v. of the event "*i, j, k* all choose the same number." Circle each of the following pairs of r.v.'s that are independent (assuming that the indices *i, j, k, l, m* are all distinct):

X_ij & *X_kl*   *X_ij* & *X_ik*   *X_ijk* & *X_ij*
X_ijk & *X_il*   *X_ijk* & *X_ijl*   *X_ijk* & *X_ilm*

Problem #8

Let *X* and *Y* be *independent* random variables, with E(*X*) = 1, Var(*X*) = 1, E(*Y*) = 3, Var(*Y*) = 3. Circle those <u>three</u> of the following statements that <u>must</u> be true of *X* and *Y*:

Pr[*X* < 0] = 0   Pr[(*Y* < 1) v (*Y* > 5)] <= 3/4   Pr[*X* + *Y* >= 0] >= 3/4
E(*X*^3) = 1   Pr[*X* = 1] < 1   Cov(*X*, *Y*) = 3

Problem #9

A certain casino game has a $5 stake and three possible outcomes: with probability 1/3 you lose your stake, with probability 1/3 the bank returns your stake plus $5, and with probability 1/3 the bank simply returns your stake.

(a) Let *X* denote your winnings in one play. The variance Var(*X*) is

0   10/3   5   50/3   25   125/3

(b) Now let ^2 denote your solution to part (a), and suppose you play the game 10000 times. You would expect the *magnitude* of your win or loss to be approximately (circle the closest answer)

^2   100*   100*^2   10000*   10000*^2

Problem #10

Consider the two-variable polynomial

$$Q(x_1, x_2) = (x_1 - 2*x_2)(x_1 - 3)x_2.$$

If the Schwartz-Zippel algorithm is applied to Q, with values for x_1, x_2 chosen from the set {0, 1, 2}, the probability that the algorithm fails is exactly

1/3 4/9 1/2 5/9 2/3 7/9

Problem #11 Take your partners

There are $2n$ people, consisting of n men and n women, in a dance class. A matching is a pairing-up of the people into n disjoint pairs, without regard to gender.

(a) When $n = 2$, there are exactly three possible matchings. List them.

(b) The dance instructor wishes to pair the people off randomly for dancing. He does this by constructing a matching uniformly at random (without regard to gender). What is the expected number of man/woman pairs he will create? [Hint: what is the probability that the ith man is paired with a woman?] **Check** your answer in the case $n = 2$.

(c) What is the probability that *all* pairs will be man/woman? [Hint: You may assume that the number of different matchings is precisely $(2n)!/(n!*(2^n))$. How many of these contain only man/woman pairs?] Again, **check** your answer in the case $n = 2$.

(d) If the matching the instructor creates consists only of man/woman pairs, he deems it a success and the dancing begins. Otherwise, he throws away the entire matching and starts again. Use your answer to part (c) to compute the asymptotic probability that the instructor succeeds in a single trail as n. [Hint: You will need Stirling's approximation, $n! \sim (n/e)^n*sqrt(2n)$.] What does this say about how reasonable the instructor's method is when n is large?

(e) Explain why the number of different matchings is $(2n)!/(n!*2^n)$, as we assumed in part (c). [Hint: What is the total number of permutations of the $2n$ people? Each permutation corresponds in a natural way to a matching; how many different permutations give rise to the same matching?]

Problem #12 The Usual Suspects

The local police are investigating a robbery, and have arrested *n* criminal suspects. They want to send these suspects to the scene of the crime so that witnesses can try to identify them.

Ideally the police would like to send all the suspects at once, but prior experience has shown that if some of the suspects know each other, the chance of them trying to escape is much higher. So the police want to select a group of suspects who do not know each other, and they would like to select as many as possible under this restriction.

We can view the *n* suspects as verticies V of a graph G = (V, E), where there is an edge *e* E between *i* and *j* if and only if suspects *i* and *j* know each other. Hence our aim is to choose a set S of vertices such that no two vertices in S have an edge between them. Such a set is called an *independent set*.

(a) For the following graph, show that there is an independent set of size 3 by circling the vertices in that set.

(b) Now let be a permutation of the vertex set V. Let S() denote the set of vertices *i* such that, for all neighbors *j* of *i*, we have (*i*) < (*j*). (In other words, *i* belongs to S() if, in the permutation , *i* comes before all its neighbors.) Argue that, for any permutation , S() is always an independent set.

(c) Now consider a *random* permutation . For a fixed vertex *i*, what is the probability that *i* is placed before all its neighbors by ? You should give your answer as a function of the degree d(*i*) of vertex *i*.

(d) Suppose now that every vertex has the same degree *d* (i.e., every suspect knows exactly *d* others). If we choose the permutation at random, what is the expected size of S(), as a function of *n* and *d*?

(e) Conclude that, if every vertex has degree *d*, then there exists an independent set in G of size at least *n*/(*d*+1).

(f) Devise a simple randomized algorithm that, with probability at least 1/2, finds an independent set of size at least *(*n*/(*d*+1)) in such a graph, where 0 < < 1 is a fixed constant. You should carefully justify the success probability of your algorithm, and also state its running time. [Hint: Use Markov's inequality to obtain an algorithm with a failure probability that depends on . Then use boosting. The running time will depend on .]

Problem #13 Walking Randomly on a Starfish

This problem concerns random walk on the *starfish* graph below, which consists of r "rays" each of length l edges, connected at a single center vertex c.

(a) What is the expected hitting time from one of the ray endpoints to vertex c, as a function of l? [Hint: use what you know about hitting times on the line graph.]

(b) Write down a formula for the expected commute time $H_{uv} + H_{vu}$ between any pair of vertices u, v in terms of the resistance R_{uv} between them. Hence compute the expected hitting time from the endpoint of one ray to the endpoint of another specific ray, as a function of r and l.

(c) Compute the expected hitting time from vertex c to the endpoint of a specific ray, as a function of r and l.

(d) Give a simple upper bound on the expected *cover time* starting from the center vertex c, as a function of r and l.

(e) Now we are going to get a better bound on the expected cover time. In preparation for this, argue that the expected time for a random walk starting at c to hit *any* of the ray endpoints (we don't care which one) and return to c is exactly $2*(l^2)$. [Note: this is a different question from that in part (c), where we are concerned with a *specific* endpoint. If you cannot see how to argue this part, assume the result and proceed to part (f).]

(f) Use coupon collecting to argue that the expected cover time starting at vertex c is $O[(l^2)*r\log(r)]$. [Hint: You may assume the following fact: if $T_1, T_2, \ldots$ are independent r.v.'s with common expectation μ, and if K is a r.v. that is independent of all the T_i, then $E(\text{sum } i=1 \text{ to } K \text{ of } T_i) = \mu E(K)$.]

(g) Prove the fact that was assumed in the hint for part (f). [Hint: Consider the conditional expectation $E\{(\text{sum } i=1 \text{ to } K \text{ of } T_i) \mid K=k\}$.]

Problem #14 Pairwise Independent Hash Functions

This problem concerns a property of hash functions that is closely related to the familiar 2-universal property. Let H be a family of hash functions from a universe $U = \{0\ldots m-1\}$ to a table $T = \{0\ldots n-1\}$. We say that H is *pairwise independent* if it satisfies both of the following conditions:

(i) For every *x* *U* and every *z* ∈ *T*,

Pr_*h*[h(*x*) = *z*] = 1/*n*.

(I.e., the hash functions distribute elements of *U* *uniformly* over *T*.)

(ii)For every *x_1, x_2* ∈ *U* with *x_1 ≠ x_2*, and every *z_1, z_2* ∈ *T*,

Pr_*h*[*h(x_1) = z_1 ∧ h(x_2) = z_2]* = Pr_*h*[*h(x_1) = z_1]* * Pr_*h*[*h(x_2) = z_2*].

(I.e., the destinations of any two elements in *U* are *independent*.)

In both of these conditions, the probability Pr_*h* means that the hash function *h* is chosen uniformly at random from *H*. *Study these two conditions carefully before proceeding.*

In this problem, we will explore the pairwise independence property and its relationship to 2-universality.

(a) Define what it means for the family *H* to be *2-universal*. Be sure that your definition is precise.

(b) Show that, if a family *H* is pairwise independent, then *H* is also 2-universal. [Note: In your argument, you should state clearly where you are using conditions (i) and (ii) above.]

(c) Let us now assume that *m* = 2^*l* and *n* = 2^*k*, so that we can represent each element *x* of the universe as an *l*-bit vector, and each table element *z* as a *k*-bit vector. For any *k* * *l* 0-1 matrix *A* and any *k*-bit vector *b*, we define the hash function *h_Ab : U → T* by

*h_Ab(x) = (Ax + b)*mod 2.

[Note: This family is the same as one we saw in a homework, except for the addition of the vector *b*.] How many functions are there in the family *H* = {*h_Ab*}?

(d) Show that the family {*h_Ab*} defined in part (c) satisfies condition (i) in the definition of pairwise independence.

(e) Show that the family {*h_Ab*} defined in part (c) satisfies condition (ii) in the definition of pairwise independence. [Hint: Consider first the simpler special case in which *k* = 1.]

(f) Recall the family *H* = {*h_A*} discussed in a homework, where for any *k* * *l* 0-1 matrix *A* we define

$$h\_A(x) = Ax \bmod 2.$$

In the homework, we showed that this family is 2-universal. How that it is *not* pairwise independent. [Hint: it is enough to show that just one of the conditions (i) or (ii) fails.]
