CS265/CME309, Fall 2017. Instructor: Gregory Valiant

Final Exam

Electronic submission via **Gradescope**, due **11:59pm Wednesday 12/13**.

You must work alone. You may consult the textbook/course notes, and other online resources (wikipedia, etc.). You may not discuss any aspect of this test with other students (whether or not they are enrolled in the class). You may NOT troll the internet looking for problem-specific solutions, or post course-related questions on online forums, with the exception of our class Piazza forum.

Please answer the questions thoroughly yet succinctly. Points may be deducted for needlessly elaborate solutions, or for vagaries or poor style. As always, if there is a hole in your proof, fewer points will be deducted if you acknowledge that you are aware of this hole, versus trying to hide it. Good luck!

1. For each of the following questions, provide a one to three sentence explanation.

   (a) (3 points) Suppose I just purchased gradescope's new "automatic grader" algorithm, which promises that it will grade each person's final exam, and comes with the guarantee that, no matter the distribution from which students' exams are drawn, the average number of grading errors that the program will make per exam is at most α. Assuming the guarantee is true, I would like to say something about what fraction of students will end up with lots of grading errors. What can I do?

   (b) (6 points) Describe three rather different approaches to showing that two distributions, A and B actually have the same distribution, or have very similar distributions, and indicate when each approach might be useful. Your three approaches could apply to different settings—for example, one approach could apply to the case where the distributions are continuous distributions over the reals, and are described to you via some complicated sums of other continuous random variables; another could apply to the case where the distributions A and B are described implicitly as the distributions resulting from applying some random process (e.g constructing a random graph via a strange process of drawing edges); perhaps you are given access to independent draws from A and B. As long as you clearly describe three approaches, and indicate when that approach might be useful, you will get full credit.

   (c) (3 points) Suppose you want to show the existence of some nice combinatorial object (such as a good encoding/decoding scheme, or a graph with special properties, etc.) What is one general approach to showing that such an object exists?

   (d) (3 points) Consider a dataset of n genomes, each represented as a length $m \gg n$ binary vector. You would like to compute all n^2 pairwise Hamming distances between the vectors, but you do not have time $O(n^2m)$. Instead, you decide to approximate each of these pairwise distances, and would be happy if most of the distances are accurate up to 10% error. What could you do, and what is the amount of runtime it would take to compute these n^2 approximate distances? [Justify your answer, thought feel free to ignore constant factors.]

(e) (6 points) Suppose you have n jobs, that you are allocating uniformly at random to m processors (i.e. for each job, you pick a number uniformly at random from $1, \ldots, m$ and assign that job to the corresponding processor). Let X denote the number of processors that end up with more than 2 jobs. Describe (in high-level terms) two different approaches for proving that X will be tightly concentrated about its expectation. [Here, tightly concentrated means with inverse exponentially decreasing tail bounds, as in a Chernoff bound.]

(f) (3 points) Suppose you have an enormous dataset of chord progressions found in pop music. You want to make a program that generates random (ideally novel) pop-sounding musical phrases. As an initial step, you first make a simple subroutine that takes a pair of phrases/sequences, and estimates their relative likelihoods of arising in pop music. (Such a subroutine could be based on the relative likelihoods in your dataset of all the 4-note subsequences, for example.) At a very high level, in one or two sentences, describe what techniques you might use to leverage this subroutine into a program that generates sequences according to some reasonable approximation of the distribution of pop-music phrases.

(g) (3 points) Suppose there are $n \geq 3$ houses on your street, arranged in a line. At time $t = 0$, a soccer ball appears in one of their yards. Suppose that each day, whichever house has the soccer ball, they wake up each day and leave it with probability p, or with probability $1 - p$ kick it into a random adjacent yard (the two end houses only have 1 choice of adjacent yard, and all the other houses have 2 choices of which yard to kick it into). Assuming all houses have the same probability of inaction on a given day, $p \in (0, 1)$, what do you expect is the distribution of the ball's location at day t for some extremely large number t?

(h) (3 points) In the above setting, if $p = 0$ (i.e. the ball gets kicked every day) describe the distribution of the ball's location at time t for some extremely large number t.

2. Suppose there are n different classes, and $m > 2n$ possible TAs. Each class needs at least 2 TAs, and no one should TA more than one class. Suppose each class submits a list of 20 "preferred" TAs, and it so happens that each TA appears in at most 3 lists.

(a) (6 points) Prove, via the Lovasz Local Lemma, that there exists an assignment of TAs to classes such that every class gets at least two of its preferred TAs. [Hint: Consider assigning each TA uniformly at random to one of the ≤ 3 classes that prefer that TA. Be sure to formally define the "bad" events, and argue about the degree of the dependency graph.]

(b) (2 points) Describe an algorithm, whose runtime is polynomial in n and m that could be used to find a satisfactory assignment. (No proof necessary, just clearly describe the algorithm.)

3. Suppose the class has A CS students, B Statistics students, and C Engineering students. Suppose a uniformly random subset of k students ends up submitting a course evaluation. Let X denote the number of CS students who submitted evaluations.

(a) (3 points) What is $\mathbf{E}[X]$ as a function of A, B, C, and k? [Prove your answer with at most one sentence.]

(b) (1 point) Suppose we iteratively select our random set of k students, by first choosing a random student to add to the set, then selecting a random student from the $A+B+C-1$ remaining ones, and so on until we have selected k students. Prove, in one sentence, why the resulting set consists of a uniformly random subset of k students.

(c) (3 points) Let X_i be the indicator random variable of whether the ith student chosen according to the above sequence is a CS student. Hence $X = \sum_{i=1}^{k} X_i$. Are the X_i's independent or dependent random variables? Prove your answer.

(d) (6 points) Prove that $\Pr[|X - \mathbf{E}[X]| > \lambda] \leq 2e^{-\frac{\lambda^2}{2k}}$.

4. Consider a fair betting game, where you bet i dollars at time i (and either win i or lose i dollars, independently, with probability $1/2$ each). Let $X_i \in \{\pm 1\}$ denote whether you win or lose in the ith round.

   (a) (1 point) Let $Y_i = \sum_{j=1}^{i} jX_j$ denote your net winnings through the first i rounds. Prove that $\{Y_i\}$ is a martingale with respect to $\{X_i\}$.

   (b) (5 points) Choose a function $f(i)$ such that if you define $Z_i = Y_i^2 + f(i)$, then the sequence $\{Z_i\}$ is a martingale with respect to $\{X_i\}$. [Hint: If you don't know where to start, first figure out what $f(1)$ should be so that $\mathbf{E}[Y_1^2] + f(1) = 0$, then figure out what $f(2)$ should be, then $f(3)$, etc. until you see the pattern, and then prove that it works.]

   (c) (5 points) Let T denote the (random variable) representing the first time when your net earnings are either more than a or less than $-a$. What is $\mathbf{E}[T]$ as a function of a? Its fine if your answer is only accurate up to a constant factor. [If you use the optional stopping theorem, be sure to argue why you can apply it to this setting.]

5. In this problem we will consider two different models of wealth distribution. Suppose we are in a world with n people and $100n$ dollars. At time $t = 0$ there is some initial assignment, X_0, of dollars to people.

   Model A: At each time $t = 1, 2, 3, \ldots$, the allocation changes according to the following protocol: choose a person uniformly at random from the n players, and call them player i; if i has no money, then nothing changes during that time-step; if i has at least one dollar, then select a player, j, with probability proportional to the amount of money they currently have, and transfer one dollar from i to j (note that it is possible that $i = j$). For example, if at time t, all players have \$100 then at time $t + 1$, with probability $1/n$, all players have \$100, and with probability $1 - 1/n$, one player has \$101 one has \$99 and the rest have \$100.

   Model B: At each time $t = 1, 2, 3, \ldots$, the allocation changes according to the following protocol: choose a person uniformly at random from the n players, and call them player i; then select a player, j, with probability proportional to the amount of money they currently have, and transfer one dollar from j to i (it is possible that $i = j$). [Note that in this model, j pays i, whereas in Model A, i pays j!!!!]

   (a) (2 points) Note that each model describes a Markov Chain over the "wealth allocation" of players. For each model, say whether the chain is periodic or aperiodic, and give one sentence of justification.

(b) (2 points) For each model, say whether the chain is irreducible or not, and give one sentence of justification.

(c) (2 points) For each model, say whether or not the fundamental theorem of Markov chains applies, and give a one sentence justification.

(d) (4 points) What is the stationary distribution of the Markov Chain described by Model B? (Prove your answer.) [Hint: Imagine that each of the dollars has a unique number, $1, \ldots, 100n$, and think about the state of the system at time t as being a list of which dollars each person has. Consider the update protocol that first picks a uniformly random dollar d, then picks a uniformly random person j, and takes that dollar from whoever has it and gives it to player j. This update scheme is one way of implementing Model B, no? Now, if you think of the states of this chain as nodes in a graph, the updates look kindof like a random walk on this graph....and you know all about the stationary distribution of random walks on a graph :)]

(e) (5 points) Prove that the mixing time of the chain corresponding to Model B is at most $O(n \log n)$. [Hint: define a coupling, and also use the hint of the previous part.] You can do this part even if you skip the previous part.

(f) (2 points bonus) As t gets large, in Model A one player will eventually end up with all the money. What is the expected time until this happens? Feel free to just give the answer up to constant factors, though give a rigorous proof.