Midterm

Name:

SUID Number:

[This is a closed-notes/closed-computer exam, though you may refer to 1 page (or 2 sides) of 8.5 x 11 notes that you have prepared. You must work alone on the exam.]

The following theorems might be helpful:

**Theorem 1** (Lovasz Local Lemma). Consider a set of 0/1 random variables, $A_1, \ldots, A_n$ defined over some probability space such that the maximum degree of the corresponding dependency graph is $s$. If, for all $i$, $\Pr[A_i = 1] \leq \frac{1}{e(s+1)}$, then there is a nonzero probability that $0 = A_1 = A_2 = \ldots = A_n$. (Namely, there is a positive probability that all the events $A_i$ are, simultaneously, avoided.)

**Theorem 2** (Chernoff Bounds). Let $X$ be the sum of independent random variables that take values in $[0, 1]$:

- If $E[X] \leq \mu$, then for $\delta \in (0, 1)$: $\Pr[X \geq (1 + \delta)\mu] \leq e^{-\delta^2 \mu/3}$.
- If $E[X] \geq \mu$, then for $\delta \in (0, 1)$: $\Pr[X \leq (1 - \delta)\mu] \leq e^{-\delta^2 \mu/2}$. 
1. For each of the following questions, provide 2-3 sentences of explanation. (2 points each)

   (a) Why is the Johnson-Lindenstrauss metric embedding a useful algorithmic primitive? Feel free to explain via a practical example.

   (b) For a large integer \( n \), roughly how many prime numbers are there less than \( n \)? Please give your answer as a function of \( n \).

   (c) Why is the above rough density of prime numbers, together with fast randomized primality-checking algorithms helpful for cryptography?

   (d) What is the relationship between the moment generating function of a random variable, and the moments of the random variable?
(e) Describe the high-level approach to proving a Chernoff-style bound. (One sentence is sufficient.)

(f) Why is the above approach often much easier to apply to a sum of independent random variables, rather than dependent ones?

(g) To prove that some object exists, via the probabilistic method, what are the high-level steps of the proof? [Hint: there should be 2 steps.]

(h) In what settings (i.e., for problems with what types of structure) is the Lovasz Local Lemma significantly more powerful than simply applying the naive union bound?
2. Consider the metric \((X, d)\) where the set \(X = \{a, b, c\}\) consists of three elements, and the distance function is defined by \(d(a, b) = 1\), \(d(a, c) = 1\), and \(d(b, c) = 1\).

(a) (2 points) Define an embedding of \((X, d)\) into \((\mathbb{R}^2, \ell_2)\) that does not distort distances at all (i.e. distortion 1). [The metric \((\mathbb{R}^2, \ell_2)\) simply denotes the 2-dimensional plane, under “Euclidean” distance.]

(b) (2 points) Define an embedding of \((X, d)\) into \((\mathbb{R}^2, \ell_1)\) that does not distort distances at all. [The metric \((\mathbb{R}^2, \ell_1)\) simply denotes the 2-dimensional plane, under “Manhattan” distance.]

(c) (2 points) What is the smallest distortion that can be achieved if you try to embed \((X, d)\) into \((\mathbb{R}, \ell_1)\)? (Justify your answer with a sentence or two.)

(d) (4 points) Is the following statement true or false: Any \(n\)-point metric can be embedded in \((\mathbb{R}^n, \ell_2)\) without distorting distances at all. Justify your answer with one to three sentences.
3. Suppose you go fishing, and catch $k$ fish, each one of which is an independent sample from a distribution of fish species in the ocean (i.e. $\Pr[\text{cod}] = .2$, $\Pr[\text{salmon}] = .1$, etc.)

(a) (2 points) Is the probability that you catch at least one “cod” dependent, or independent from the probability that you catch at least one “tuna”? (Explain with one sentence.)

(b) (3 points) Instead of catching exactly $k$ fish, suppose you choose $k'$ according to a Poisson distribution of expectation $k$, then catch $k'$ fish. In this “Poissonized” setting, is the number of “cod” you catch dependent or independent of the number of “tuna” you catch (answer from the perspective of someone who does not know $k'$, if you condition on the value of $k'$ then this is the same as part (a)....). (Explain with one sentence.)

(c) (6 points) Let $X$ denote the total number of species of fish that you caught exactly once in your $k$ fish. Prove that for sufficiently large $k$, $\Pr[|X - E[X]| > 10\sqrt{k}] \leq 0.1$. [Hint: If you use a Chernoff bound, clearly specify which bound you are using and why it applies. Also feel free to use the fact that for any sufficiently large integer $k$, the probability that a Poisson random variable of expectation $k$ deviates from its expectation by more than $5\sqrt{k}$ is at most 0.05.]
(d) (Bonus 2 points—come back to this part after finishing the rest of the exam)
Let $Y$ denote the total probability mass of all the species of fish that you did NOT catch. Equivalently, $Y$ denotes the probability that, if you were to catch a $k + 1$st fish, that fish is a new species that you did not see in your first $k$ catches. Prove that $\Pr[|Y - E[Y]| > 20 \log k \sqrt{k}] \leq 0.1$. For an extra bonus point, prove that $\Pr[|Y - E[Y]| > \frac{20}{\sqrt{k}}] \leq 0.1$. 
4. Suppose the class has \( n \) students, of which \( m \) pairs of students are friends.

   (a) (7 points) Prove that there exists a way to partition the students into 4 teams, such that at most \( m/4 \) friendships exists between team-members (i.e. at least \( 3m/4 \) friendships involve students split between different teams.)

   (b) (5 points) Give an efficient (running in time polynomial in \( n \)) DETERMINISTIC algorithm for finding such a partition, and justify its correctness with one to three sentences.
5. Suppose you are preparing a large dinner party for your dormitory, and plan to make many dishes. Each dish will contain exactly 5 spices. Also, to ensure that you don’t use too much of any one spice, each spice is in at most 2 dishes. Everything is planned, and you print out the recipes for all the dishes. In the middle of the night, an evil Grinch breaks into the kitchen and sees the recipes, and your bags of spices. The Grinch will add EITHER a special Lactose supplement, OR a special Gluten supplement to each spice jar.

(a) (7 points) Prove that the Grinch can add the supplements in such a way that EVERY dish will have both Gluten and Lactose. [Hint: if you use the Lovasz Local Lemma, be sure to define a probability space, and explicitly define the dependency graph, etc.]

(b) (4 points) Algorithmically, describe an efficient algorithm the Grinch can use to determine how to allocate the supplements.