Name: ______________________________

Instructions: You have 50 minutes to complete this exam. Show your work and justify all of your responses. No calculators, notes, or other external aids are allowed. You may use the following theorems (you may also use any version of the Chernoff bound you want):

Theorem 1 (Union Bound)**.** *Let $A_1, A_2, \cdots, A_n$ be events in a probability space.*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$$

Theorem 2 (Markov's Inequality)**.** *If X is a nonnegative random variable and $\lambda > 0$ is a real number, then*

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}.$$

Theorem 3 (Chebyshev's Inequality)**.** *Let X be a random variable with finite variance and $\lambda > 0$ a real number. Then*

$$\mathbb{P}(|X - \mathbb{E}(X)| > \lambda) \leq \frac{\mathrm{Var}(X)}{\lambda^2}.$$

Theorem 4 (Chernoff Bound)**.** *Let $X_1, \cdots, X_n$ be independent random variables with $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = 0) = 1 - p$. Let $S = X_1 + \cdots + X_n$. Then for any $0 \leq \epsilon \leq 1$,*

$$\mathbb{P}\left(S \leq (1-\epsilon)pn\right) \leq e^{-\epsilon^2 pn/2}$$

$$\mathbb{P}\left(S \geq (1+\epsilon)pn\right) \leq e^{-\epsilon^2 pn/3}$$

1. (10 points) Let A be a subset of integers in $[n]$. A is called a B_3 *set* if for $x_1, x_2, x_3, x_4, x_5, x_6 \in A$, if

$$x_1 + x_2 + x_3 = x_4 + x_5 + x_6,$$

then it implies that $\{x_1, x_2, x_3\} = \{x_4, x_5, x_6\}$. Show that there is a B_3 subset $A \subset [n]$ with $|A| = \Omega\left(n^{1/5}\right)$ (Hint: you may want to show that the total number of solutions to $x_1+x_2+x_3 = x_4+x_5+x_6$ with $x_1, x_2, x_3, x_4, x_5, x_6 \in [n]$ is $O(n^5)$).

Solution: For any fixed $x_1, x_2, x_3, x_4, x_5 \in [n]$, there is at most one $x_6 \in [n]$ such that $x_6 = x_1+x_2+x_3-x_4-x_5$. Therefore there are at most n^5 solutions to the equation $x_1 + x_2 + x_3 = x_4 + x_5 + x_6$.

Choose $S \subset [n]$ randomly, putting each integer in S independently with probability p. Let $X = |S|$ and Y count the number of solutions to the equation $x_1 + x_2 + x_3 = x_4 + x_5 + x_6$ with $x_1, x_2, x_3, x_4, x_5, x_6 \in S$. Given our set S, we may make it a B_3 set by removing at most one element of S for each solution to the equation. Therefore, there is a B_3 set of size at least $X - Y$ for every outcome of this random process. In particular, there is a B_3 set of size at least $\mathbb{E}(X - Y) = pn - p^6 \cdot \text{(the number of solutions)} \geq pn - p^6 n^5$. Choosing $p = \frac{1}{2} n^{-4/5}$ yields the result.

2. Throw m^2 distinct balls into m bins independently and uniformly at random.
   (a) Fix a bin. Give an upper bound on the probability that the bin has more than $m + \log m\sqrt{m}$ balls in it.
   (b) Show that no bin has more than $m + \log m\sqrt{m}$ balls in it with probability tending to 1.

Solution: Let X_i be the event that the i'th ball goes into the fixed bin and let $S = \sum X_i$. Since the X_is are independent, we may apply the Chernoff Bound with $p = \frac{1}{m}$ and $n = m^2$. Then we have that for any $\epsilon \in [0, 1]$

$$\mathbb{P}(S \geq (1+\epsilon)m) \leq e^{-\epsilon^2 m/3}.$$

Taking $\epsilon = m^{-1/2}\log m$ gives

$$\mathbb{P}(S \geq m + \sqrt{m}\log m) \leq e^{-\frac{1}{3}\log^2 m}.$$

For the second part, by the union bound the probability that any bin has more than $m + \sqrt{m}\log m$ balls in it is bounded above by

$$m\mathbb{P}(\text{a fixed bin has too many balls} \leq me^{-\frac{1}{3}\log^2 m} \to 0.$$

Taking complementary events gives the result.

3. (10 points) The k-color Ramsey number for C_4, denoted $r_k(C_4)$ is the minimum n such that any k-coloring of $E(K_n)$ contains a monochromatic C_4.

   (a) Show that $r_k(C_4) = O(k^2)$. (Hint: this is equivalent to showing that for $\varepsilon > 0$ small enough, any coloring of K_n with $\varepsilon n^{1/2}$ colors must contain a monochromatic C_4)

   (b) * Give the best lower bound on $r_k(C_4)$ that you can (equivalently, partition the edge set of K_n into C_4 free graphs using as few graphs as possible).

Solution: (a) If $E(K_n)$ is colored by $\epsilon n^{1/2}$ colors, then by the pigeonhole principle one color must have at least

$$\frac{\binom{n}{2}}{\epsilon n^{1/2}}$$

edges. If ϵ is a small enough positive constant, this is larger than $\text{ex}(n, C_4) \lesssim \frac{1}{2}n^{3/2}$, and therefore this color class must contain a C_4.

(b) Let q be a prime power and let $A_i \subset \mathbb{F}_q \times \mathbb{F}_q$ be defined by

$$A_i := \{(x, x^2 + i) : x \in \mathbb{F}_q.$$

We proved in class that A_0 is a Sidon set, and since any translate of a Sidon set is a Sidon set, we have that all of the A_i are Sidon sets. We also note that each has q elements. Next we claim that the A_i are disjoint. To see this, if for some $x, y, i, j \in \mathbb{F}_q$ we have

$$(x, x^2 + i) = (y, y^2 + j),$$

then $x = y$ which implies $i = j$. Since A_i and A_j are disjoint for distinct i, j, and since each set has size q, we have that (by counting)

$$\bigcup_{i \in \mathbb{F}_q} A_i = \mathbb{F}_q \times \mathbb{F}_q.$$

Now let K_n be a complete graph with $n = q^2$ and identify the vertex set with $\mathbb{F}_q \times \mathbb{F}_q$. We will color K_n with q colors such that no color class has no C_4. By the same density of primes argument in your homework, this implies that

$r_k(C_4) = \Omega(k^2)$. Let $(x_1, y_1), (x_2, y_2) \in \mathbb{F}_q \times \mathbb{F}_q$ be vertices in K_n. We color the edge between them with color i where

$$(x_1, y_1) + (x_2, y_2) \in A_i.$$

Since the A_i are disjoint and cover the whole $\mathbb{F}_q \times \mathbb{F}_q$ this coloring is well-defined. Further, since each color class is a Cayley sum graph with generating set a Sidon set, each color class does not contain a C_4.