1. (16pts) Consider a sequence of n independent coin tosses where $\Pr[\texttt{Heads}] = 1/4$ and $\Pr[\texttt{Tails}] = 3/4$. Let X be the number of heads.

   (a) What is the probability of getting $X = k$.

   Solution:

   $$\Pr[X = k] = \binom{n}{k}\left(\frac{1}{4}\right)^k\left(\frac{3}{4}\right)^{n-k}.$$

   (b) Apply Markov's bound to $\Pr[X \geq n/2]$.

   Solution:

   $$\Pr[X \geq n/2] \leq \frac{\mathbb{E}[X]}{n/2} = \frac{n/4}{n/2} = 1/2.$$

   (c) Apply Chebyshev's bound to $\Pr[X \geq n/2]$.

   Solution:

   $$\Pr[X \geq n/2] \leq \Pr[|X - E[X]| \geq n/4] \leq \frac{\mathbf{Var}[X]}{(n/4)^2} = \frac{3n/16}{n^2/16} = 3/n.$$

   (d) Compute the moment generating function for X.

   Solution: Let X_i be the i-th coin toss.

   $$\mathbb{E}[e^X] = \prod_{i=1}^{n} \mathbb{E}[e^{X_i}] = \frac{(e+3)^n}{4^n}.$$

2. (15pts) Suppose you throw $n/2$ balls into n bins, where n is even. What is the probability that exactly $n/2$ bins are empty?

   (a) Compute this exactly.

   Solution:
   $\frac{\binom{n}{n/2}(n/2)!}{n^{n/2}} = \frac{(n)_{n/2}}{n^{n/2}}$.

   (b) Use the Poisson approximation to give an upper bound on this value.

   Solution: Let $Y_i \sim \mathbf{Poisson}(1/2)$ for the i-th bin.

$$\begin{aligned}\Pr[\text{exactly } n/2 \text{ bins are empty}] &\le e\sqrt{n/2}\,\Pr[\text{exactly } n/2 \text{ of the } Y_i \text{ is } 0\,] \\ &= e\sqrt{n/2}\binom{n}{n/2}e^{-n/4}(1-e^{-1/2})^{n/2}.\end{aligned}$$

   Note that the event is monotone in the number of balls, hence one could get a marginally better bound.

   (c) Show that your bounds in part (a) is exponentially tighter than the bound in part (b) for large n.

   Solution: Recall that $n! < e\sqrt{n}\left(\frac{n}{e}\right)^n$, now let us consider the ratio of the two bounds:

$$\frac{\binom{n}{n/2}(n/2)!}{n^{n/2}e\sqrt{n/2}\binom{n}{n/2}e^{-n/4}(1-e^{-1/2})^{n/2}} < \frac{1}{(2(\sqrt{e}-1))^{n/2}}.$$

   Since $2(\sqrt{e}-1) > 1$, this is an exponentially small number.

3. (15pts) Consider the number of 3-cliques in a $G_{n,p}$ random graph, denoted by X.

(a) What is the expected number of 3-cliques in a $G_{n,p}$ random graph?
Solution: $\mathbb{E}[X] = \binom{n}{3}p^3$.

(b) Show that if $p = f(n) = o(1/n)$ then for any $\varepsilon > 0$ there exists some n sufficiently large such that the probability a clique will exist is less than ε.
Solution: By Markov's inequality, $\Pr[X \geq 1] \leq \mathbb{E}[X] = \binom{n}{3}p^3 = o_n(1)$. Recall that for any $\varepsilon > 0$, there exists n large enough such that $o_n(1) < \varepsilon$, which is exactly what we want to show.

(c) Show that $p = f(n) = \omega(1/n)$ then for any $\varepsilon > 0$ there exists some n sufficiently large such that the probability a clique will not exist is less than ε.
Solution: Let X_i be the indicator of whether the i-th clique is present. Recall that

$$\Pr[X > 0] \geq \sum_{i=1}^{\binom{n}{3}} \frac{\Pr[X_i = 1]}{\mathbb{E}[X \mid X_i = 1]},$$

where

$$\mathbb{E}[X \mid X_i = 1] = \sum_{j=1}^{\binom{n}{3}} \Pr[X_j = 1 \mid X_i = 1] = 1 + \binom{n-3}{3}p^3 + 3\binom{n-3}{2}p^3 + 3\binom{n-3}{1}p^2.$$

Thus we have,

$$\Pr[X > 0] \geq \frac{\binom{n}{3}p^3}{1 + \binom{n-3}{3}p^3 + 3\binom{n-3}{2}p^3 + 3\binom{n-3}{1}p^2} = 1 - o_n(1),$$

where the last equality follows from the fact that $\binom{n}{3} = (1 + o_n(1))\binom{n-3}{3}$.
Alternative second moment method is as follows:

$$\Pr[X = 0] \leq \frac{\mathbf{Var}[X]}{(\mathbb{E}[X])^2} \leq \frac{1}{\mathbb{E}[X]} + \frac{\sum_{i \neq j}(\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j])}{\mathbb{E}[X]^2},$$

where the covariance $\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \neq 0$ only if $|C_i \cap C_j| = 2$. If $|C_i \cap C_j| = 2$, $\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \leq \mathbb{E}[X_i X_j] = p^5$, and there are $12\binom{n}{4}$ such pairs.
So overall we have $\Pr[X = 0] \leq \Theta(1/p^3 n^3) + \Theta(\frac{n^4 p^5}{n^6 p^6}) = o_n(1)$.

4. (12pts) Consider a k-SAT formula with n literals and m clauses, where each literal is contained in at most 3 consecutive clauses, i.e. the clauses are ordered 1 to m and a specific literal can only be in 3 in a row, such as clauses $i, i+1, i+2$. In this problem you will apply the Lovasz local lemma.

   (a) Define the probability space and the set of bad events.

   Solution: The probability space will be the uniform distribution over $\{T, F\}^n$ assignments. For each clause of the form $(l_1 \vee l_2 \vee \cdots \vee l_k)$, we define a bad event: "this clause is violated" (i.e. $l_1 = l_2 = \cdots = l_k = F$).

   (b) Describe the dependency graph (construct the vertex set and edge set).

   Solution: The vertex set will be the set of bad events, i.e. the clauses. If two clause share a variable, then we add an edge to connect the two corresponding vertices. The maximum degree will be $\min\{4, 2k\}$, as a clause can only be connected to the previous two clauses, and the following two clauses, and every literal can connect at most two other clauses.

   (c) For what values of k must there exist a solution to the formula according to the Lovasz local lemma?

   Solution: We apply the symmetric Lovasz local lemma. The maximum degree of the dependency graph is $\min\{4, 2k\}$, the probability of a bad event is 2^{-k}, so setting $4\min\{4, 2k\}/2^k \leq 1$ we get $k \geq 4$.

5. (12pts) Consider a fair lazy random walk on $\{0, \cdots, n\}$ with reflecting boundaries, i.e. at every state i except 0 and n, the next step is chosen from $\{i, i+1, i-1\}$ with equal probability; while at state 0 it is 1 or 0 (unchanged) and at state n it is $n-1$ or n, both with equal probabilities.

   (a) For $n = 3$, draw the graph of this Markov Chain, including the transition probabilities.
   Solution: TBA.

   (b) For $n = 3$, explicitly write out the transition matrix of this Markov Chain.
   Solution: $P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}$

   (c) For arbitrary n compute the stationary distribution of this Markov Chain.
   Solution: Note: the best way is to use the detailed balance equation, see Theorem 7.10.
   Let $\pi = \begin{pmatrix} 2/(3n-2) & 3/(3n-2) & 3/(3n-2) & \cdots & 3/(3n-2) & 2/(3n-2) \end{pmatrix}$. In other words, only the first and last entry is $2/(3n-2)$, and every other entry is $3/(3n-2)$. It is easy to verify that $\pi P = \pi$.

   (d) If the Markov chain starts at state 0, show that the probability of not returning to state 0 within $\frac{3}{2}n^2$ steps is less than $1/n$.
   Solution: Note that the chain is ergodic. Let T be the random variable for the time to start from 0 and return to 0, then $\mathbb{E}T = 1/\pi(0) = (3n-2)/2$. Thus by Markov's inequality,

$$\Pr[T > 3n^2/2] < \frac{\mathbb{E}T}{3n^2/2} < 1/n.$$

6. (12pts) Consider a d-regular graph G with n vertices (d-regular means that every vertex has degree d). A dominating set S is a subset of the vertices such that every vertex in G is either in S or a neighbor of a vertex in S.

   (a) Show that $|S| \geq \frac{n}{1+d}$.

   Solution: Each vertex in S covers at most $d+1$ vertices, so in total S covers at most $(d+1)|S|$ vertices. Since S is a dominating set, it covers all the n vertices, thus $(d+1)|S| \geq n$.

   (b) Suppose $d \geq 4 \ln n$, and we choose a set S of vertices at random, where each vertex is chosen with probability $p = \frac{2 \ln n}{d+1}$. Show that the probability that S is a dominating set is at least $1 - 1/n$. (Hint: if $x > 0$, $(1 - 1/x)^x < 1/e$.)

   Solution: For each vertex, there is a constraint that says "at least one of its neighbors or itself has to be chosen". Let $N_G[v] = \{v\} \cup N_G(v)$ be the closed neighborhood of v, and B_v be such a event that $N_G[v] \cap S \neq \emptyset$.

$$\begin{aligned}
\Pr\left[\bigcap_{v \in V} B_v\right] &= 1 - \Pr\left[\bigcup_{v \in V} \overline{B_v}\right] \\
&\geq 1 - \sum_{v \in V} \Pr[\overline{B_v}] \\
&= 1 - n(1-p)^{d+1} \\
&\geq 1 - ne^{-2\ln n} \\
&\geq 1 - 1/n.
\end{aligned}$$

   (c) Construct an algorithm which improves on the approach in part (b) and produces a dominating set with $E[|S|] \leq n\frac{1+\ln(d+1)}{d+1}$. (Hint: sample and modify.)

   Solution: Recall that for every vertex, there's a constraint that says "at least one of its neighbors or itself has to be chosen". We will first run part (b) with $p = \frac{\ln(d+1)}{d+1}$. Let C_v be the indicator that $N_G[v] \cap S \neq \emptyset$, the expected number of violated constraints will be

$$\mathbb{E}\left[\sum_{v \in V}(1 - C_v)\right] = n(1-p)^{d+1} \leq n/(d+1).$$

   At this point $\mathbb{E}|S| = np = n\frac{\ln(d+1)}{d+1}$, but it may not be a dominating set yet. Then in order to fix the violations, for every violated constraint, we only need to add one vertex from the $(d+1)$ vertices to fix it. Since there are $n/(d+1)$ violated constraints in expectation, we are adding at most $n/(d+1)$ vertices to S in expectation. So overall we can find a set $\mathbb{E}|S| \leq n(1+\ln(d+1))/(d+1)$.

   Note: the greedy algorithm can also achieve the same approximation ratio, and the analysis is similar to that of the greedy algorithm for set cover (see CS170) but slightly more involved.

7. (12pts) Suppose that S and T are stopping times for the sequence $\{Z_n : n \geq 0\}$. Which of the following are necessarily stopping times for the sequence $\{Z_n : n \geq 0\}$? Justify your answers.

   (a) $S + T$.

   Solution: $S + T$ is a stopping time. Indeed,

   $$1_{S+T=n} = \sum_{i=1}^{n} 1_{S=i} 1_{T=n-1} = \sum_{i=0}^{n} f_i(Z_0, \ldots, Z_i) g_{n-1}(Z_0, \ldots, Z_{n-1}),$$

   which is a function of $Z_1, \ldots, Z_n$.

   (b) $\max(S, T) - \min(S, T)$.

   Solution: $\max(S, T) - \min(S, T) = |S - T|$ is not a stopping time. Here's a counterexample. Suppose that $Z_0, Z_1, \ldots$ is a simple random walk on the integers (that is, $Z_n = \sum_{i=1}^{n} X_i$ where $X_i = 1, -1$ with probability $\frac{1}{2}$.) Let S be the k such that $Z_k = 1$. Let $T = 1$ be constant. In the sequence $Z_0 = 0, Z_1 = 1$, we have $S - T = 0$. In the sequence $Z_0 = 0, Z_1 = -1$, we have $|S - T| \neq 0$. Since the term Z_0 is the same for these two sequences, this shows that the random variable $1_{|S-T|=n}$ is not determined by $Z_0, \ldots, Z_n$.

   (c) S^2.

   Solution: S^2 is a stopping time. Note that

   $$1_{S^2=n} = 1_{S=\sqrt{n}}$$

   which is a function of $Z_0, \ldots, Z_n$.